Purposive Variation in Recordkeeping in the Academic Molecular Biology Laboratory

David Francis Wilson

Submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Molecular, Cell, and Systems Biology
College of Medical, Veterinary, and Life Sciences
University of Glasgow

March 2011

Abstract

This thesis presents an investigation into the role played by laboratory records in the disciplinary discourse of academic molecular biology laboratories.

The motivation behind this study stems from two areas of concern. Firstly, the laboratory record has received comparatively little attention as a linguistic genre in spite of its central role in the daily work of laboratory scientists. Secondly, laboratory records have become a focus for technologically driven change through the advent of computing systems that aim to support a transition away from the traditional paper-based approach towards electronic recordkeeping. Electronic recordkeeping raises the potential for increased sharing of laboratory records across laboratory communities. However, the uptake of electronic laboratory notebooks has been, and remains, markedly low in academic laboratories.

The investigation employs a multi-perspective research framework combining ethnography, genre analysis, and reading protocol analysis in order to evaluate both the organizational practices and linguistic practices at work in laboratory recordkeeping, and to examine these practices from the viewpoints of both producers and consumers of laboratory records. Particular emphasis is placed on assessing variation in the practices used by different scientists when keeping laboratory records, and on assessing the types of articulation work used to achieve mutual intelligibility across laboratory members.

The findings of this investigation indicate that the dominant viewpoint held by laboratory staff other than principal investigators conceptualized laboratory records as a personal resource rather than a community archive. Readers other than the original author relied almost exclusively on the recontextualization of selected information from laboratory records into 'public genres' such as laboratory talks, research articles, and progress reports as the preferred means of accessing the information held in the records. The consistent use of summarized forms of recording experimental data rendered most laboratory records as both unreliable and of limited usability in the records management sense that they did not form full and accurate descriptions that could support future organizational activities.

These findings offer a counterpoint to other studies, notably a number of studies undertaken as part of technology developments for electronic recordkeeping, that report sharing of laboratory records or assume a 'cyberbolic' view of laboratory records as a shared resource.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Many people have helped me throughout the course of this studentship so it is a great pleasure that I have this opportunity to record my thanks.

To the researchers and technicians who participated in my studies, I offer my thanks for your generosity, commitment, and patience in accepting a stranger in your midst. I would have achieved nothing without your kindness, and I learned much from each of you. It is a great frustration that I can only offer this anonymous vote of thanks under the terms of the ethical approval for my studies.

To Mark and Phil who must by now both regret taking on the arduous role of supervisor, I offer my thanks for your advice and forbearance. To Barry, Matthew and Roger who each took on the role of assessor, I offer my thanks for struggling through various early drafts of my ramblings. Please don't let this experience deter any of you from taking on more refugees from industry in the future – they can't all be as difficult as me.

To Aislynn, Berit, Claudia, Colin, Darren, Graham, Jill, Richard, Tom and the many other members of the Anderson Level 5 club, I offer my thanks for your kindness, camaraderie, and humour. The fondest memories of my brief passage through academia will centre on you. My special thanks to Claudia and Jill for being there, to Graham for slowing down so that I could keep up whilst running, and to the legendary Richard for many entertaining discussions on history and a plethora of other topics.

To the Medical Research Council, I offer my gratitude for the financial support that enabled me to undertake this research.

Most importantly, to my beautiful Barbara, to our Velvet and to my parents, I offer my sincere thanks for the love, support, and encouragement that continue to carry me through the good and the bad times.

Author's Declaration

I certify that this thesis does not contain material previously published or written by any other person. The research reported in this thesis is my own work, except where otherwise stated, and has not been submitted for any other degree.

March 2011

List of Abbreviations

A°C:	annealing temperature
BPS:	British Psychological Society
BSC:	Biological Service Collaboratory
cAMP:	cyclic adenosine monophosphate
CAQDAS:	computer-aided qualitative data analysis
CSMD:	Core Scientific Metadata Model
DNA:	deoxyribonucleic acid
dNTP:	deoxynucleoside triphosphate
EAP:	English for Academic Purposes
EBV:	Epstein-Barr virus
EFL:	English as a foreign language
ELN:	electronic laboratory notebook
EMBL:	European Molecular Biology Laboratory
EMSA:	electrophoretic mobility shift assay
EOP:	English for Occupational Purposes
ESL:	English as a second language
ESP:	English for Specific Purposes
GOC:	Gene Ontology Consortium
ICC:	immunocytochemistry
IHC:	immunohistochemistry
IMRD:	introduction-method-results-discussion

ISO: International Organization for Standardization

IVT: *in vitro* transcription

JoVE: Journal of Visualized Experiments

L1: first language, as in L1 speakers of English

L2: second language, as in L2 speakers of English

LIMS: laboratory information management system

LSP: Language for Specific Purposes

MIAME: Minimum Information About a Microarray Experiment

NCBI: National Centre for Biotechnology Information

NNS: non-native speaker

NS: native speaker

OAIS: Open Archival Information System

PCR: polymerase chain reaction

POVMR: Problem-Oriented Veterinary Medical Record

qPCR: quantitative polymerase chain reaction

RNA: ribonucleic acid

RP-PCR: repeat-primed polymerase chain reaction

SFL: systemic functional linguistics

SNP: single-nucleotide polymorphism

T_A: annealing temperature

1 Introduction

1.1 Academic bioscience laboratory recordkeeping

Molecular biology is a scientific discipline that has delivered significant advances in a range of areas including medical therapies, forensic investigation, and plant and animal breeding for the agricultural industries. The focus of this branch of the biosciences is on "the study of the structure, functions, and molecular aspects (proteins, enzymes, nucleic acids) of the living cell" (Licker 2003:viii). Formed originally through a combination of physics, biochemistry and genetics (Muller 1936; Kellenberger 1989), molecular biology continues to develop new inter-disciplinary approaches both with other branches of the biosciences and with other disciplines such as mathematics and computer science. Research and development in molecular biology is undertaken across a range of market sectors encompassing academia, pharmaceutical companies, and health care services.

The primary remit of molecular biology and other bioscience laboratories is to advance and exploit our understanding of biological processes. However, bioscience laboratories have also contributed, albeit indirectly, to progress in fields of inquiry other than biology by participating as case studies in projects conducted by researchers working in those fields. This includes historical studies of the development of science and medicine (Abir-Am 2006; Judson 1996; Kay 1993; Magner 2002; Sturdy 1998), sociological studies of science and scientific knowledge (David 2005; Knorr Cetina 1999; Hine 2006; Jordan and Lynch 1993; Latour and Woolgar 1986; Lynch 1997; Mulkay 1995), technological studies of the design of computing systems to support laboratory work (Mackay *et al*. 2002; schraefel *et al*. 2004; Tabard *et al*. 2008; Yeh *et al*. 2006), and linguistic studies of spoken and written communication in professional and academic communities (Braine 1995; Dong 1998; Dubois 1987; Hyland 2000; Myers 1990; Samraj 2005; Swales 1990).

In common with most bioscience disciplines, research and development in molecular biology involves the planning, execution, and analysis of laboratory experiments in order to investigate hypotheses concerning the current understanding of the science (Reed *et al*. 2007; Wilson and Walker 2010). As an integral part of this process, bioscientists keep records of the details and outcomes of the experimental work that they conduct on a day-to-day basis (Barker 2005:89-100; Ebel *et al*. 2004:15-20; Kanare 1985). Benchwork laboratory scientists such as molecular biologists refer to the resulting documents variously

as laboratory records, laboratory notes, notebook entries, or experimental notes. In disciplines such as botany and ecology where experimental work centres on observations made in the wild, bioscientists may refer to the resulting documents as field notes. To ensure consistency, the term 'laboratory record' will be used throughout this thesis to denote this type of document.

This reliance on recordkeeping has a long and distinguished history in laboratory science. For example, the Codex Arundel in the British Library holds examples of scientific notebooks written by Leonardo da Vinci in the early sixteenth century. The National Library of France holds examples of the infamously radioactive notebooks written by Maria Skłodowska-Curie in the early twentieth century. Of particular significance to molecular biologists, James Watson's and Francis Crick's notebooks documenting the discovery in the 1950s of the double helix structure of DNA are now held respectively in the Cold Spring Harbor Laboratory Archive and the Wellcome Library Archive.

Given the importance of recordkeeping to laboratory science, an important question to consider is what makes a written record an effective means of communication? Both Reed (2005), and Shepherd and Yeo (2003) draw on the professional experience of archivists and records managers to highlight the dual importance of reliability and usability. In order to be reliable, a record must be a full and accurate description of the transaction that it embodies. In order to be usable, a record must be a tool that can support future organizational activities in multiple contexts of use.

Recordkeeping has previously been investigated as a central discursive practice in diverse professional and academic institutions. This includes studies of recordkeeping in the legal profession (Badger 2003; Bhatia 2008), accountancy practice (Coffey 1993; Flowerdew and Wan 2010), veterinary colleges (Schryer 1993), social work departments (Cicourel 1968; Paré 2004; Shemmings 1991), secondary schools (Cullingford and Swift 2002; Woods 1979), and a wide range of health care settings (Anderson *et al.* 2008; Berkenkotter 2008; Heath and Luff 1996; Pettinari 1988; Rooksby *et al.* 2007; Timmermans and Berg 2003). These studies have identified how professional and academic institutions construct their discourse according to a combination of situated organizational needs, external regulatory requirements, community expectations, and socio-cultural influences. The transactions to be embodied in, and the future contexts of use for, the records produced in each institution vary according to these needs, regulations, expectations, and influences. Recurrent issues highlighted by these and similar studies include, *inter alia*, the problem of

mismatched expectations between readers and writers of records, the role of contextualization in constructing meaning from written forms of communication, and the tensions involved in instituting implicit or explicit standardization of recordkeeping practices.

In the industrial and health service sectors, the recordkeeping practices at work in bioscience laboratories reflect hybrid membership of the corporate and scientific communities (Barabas 1990; Gaudillière and Löwy 1998; Rizova 2007). Both industrial laboratories and health service laboratories operate under rigorous regulatory regimes. This includes both internal regulation imposed by an organization's own quality assurance regime, and external regulation imposed by external supervisory agencies. The dominant organizational culture is typically orientated towards process rather than results, favours collectivism over individualism, formalizes the structure of activities in order to avoid uncertainty, and concentrates rather than distributes authority (Handy 1993; Hofstede 2001; Hofstede and Hofstede 2007). In these sectors, laboratory records acquire significant "evidential value" (Robek *et al.* 1995) in that they perform established communicative functions as proof of precedence for patent applications and proof of conformance for quality assurance procedures. Recordkeeping is embedded in the work of these sectors, and laboratory workflow controls both the records being kept and the tasks being performed in a tradition of recordkeeping known as "pre-action recordkeeping" (Reed 2005:114). In addition, laboratory records in these sectors are, in a direct sense, the products of collaborative authorship since supervisory staff members are required to verify and countersign any records kept by individual scientists during the course of their work.

In contrast to the large teams, high throughput, and preference for pre-defined repetitive tasks characteristic of work in many industrial and health sector laboratories, research laboratories in the academic sector still frequently operate as small groups of individuals applying relatively new techniques to small sample sets generating locally managed data (Borgman *et al.* 2007; Casper and Clarke 1998; Clarke and Fujimura 1992; Knorr Cetina 1999:81-87), in a setting characteristic of what Price (1963) has labelled "little science"[1]. Even the flagship bioscience project of recent times, the Human Genome Project (Watson

[1] Becher and Trowler (2001:105-108) draw on an analogy of population density concerning the "people-to-problem ratio" within research environments to identify a distinction between "rural" (*cf.* little science) and "urban" (*cf.* big science) research settings. Rural research settings exhibit reduced levels of collective activity, employ less heavily used information networks, and are organized according to a division of labour in which problems are locally scoped on the basis that "there is no point tackling one [research problem] on which someone else is already engaged."

1990), although heavily funded and responsible for processing large data sets, operated as a loose federation of individual laboratories through "coordinated encouragement of local initiative, by pluralist, decentralized efforts" (Kevles and Hood 1992:307). Given the distinct organizational culture of laboratories in the academic sector, it is possible that recordkeeping for academic laboratories may fulfil distinct communicative functions from recordkeeping in industrial and health service laboratories (*e.g*. Dias *et al.* 1999).

This thesis focuses specifically on molecular biology laboratories in the academic sector, and sets out to investigate recordkeeping as a discursive practice in these laboratories by examining the authentic records and practices of a range of scientists in multiple laboratories within a United Kingdom (UK) university.

1.2 Motivation for the study

Latour and Woolgar (1986:48) succinctly capture the importance of recordkeeping to laboratory scientists in their observation drawn from a study at the Salk Institute[2] that:

> "it strikes our observer that its members are compulsive and almost manic writers. Each bench has a large leather-bound book into which members meticulously record what they have just done".

This observation highlights a number of issues in laboratory recordkeeping that motivate the study presented in this thesis. Firstly, the product of laboratory recordkeeping is a written text. Secondly, these written texts have traditionally been captured using pen and paper. Thirdly, each scientist is responsible for constructing his or her own record concurrently with the experimental task that is being executed.

The first of the highlighted issues reinforces the concept of science as a "literate culture" (Smith 1993). Scientists devote a considerable amount of their working life to reading and writing (Flowerdew 2002b; Hyland 2000; Latour and Woolgar 1986) in the course of performing their research, promoting their research, and participating in their laboratory community, university community and the wider scientific communities. Consequently, becoming an effective molecular biologist requires not only competency in planning and

[2] The Salk Institute for Biological Studies is an independent research institute established in San Diego since the 1960s, which undertakes research in multiple branches of the biosciences including molecular biology. Further information is available at http://www.salk.edu [accessed 01 March 2011].

executing experiments in the laboratory, but also competency in communicating through reading and writing. One component of laboratory literacy is, of course, competency in reading and writing laboratory records.

Laboratory records do not, however, operate in isolation. Instead, laboratory records participate as one of a "constellation" (Swales 2004:12) of linguistic genres, both written and spoken, that co-constitute the discourse of an academic bioscience laboratory. Scientists read a range of texts from a range of authors, and read these texts for a variety of purposes. In academic bioscience laboratories, this includes, for example, reading research articles to acquire and maintain knowledge of current developments in the field, reading student reports to evaluate and guide student learning, and reading protocol manuals to learn and perform laboratory techniques. Similarly, scientists write a range of texts, and write these texts for a variety of audiences and purposes. In academic bioscience laboratories, this includes, for example, writing research articles to publish and promote one's work, writing grant applications to compete for funding, and writing course notes and handouts as teaching aids for students. Understanding the role of the laboratory record in laboratory discourse requires understanding the interactions between laboratory records and other genres that participate in the discourse of the laboratory community.

Previous studies of a range of written and spoken genres that participate in academic research settings have demonstrated that effective communication requires scientists to negotiate a complex mix of linguistic, social and rhetorical considerations. This includes studies of research articles (Bazerman 1988; Kanoksilapatham 2005; Myers 1990, 1991; Swales 2004; Tarone *et al*. 1998), conferences talks (Dubois 1987; Rowley-Jolivet 2002; Ventola *et al*. 2002), laboratory reports (Braine 1995; Dudley-Evans 1985), peer reviews (Gosden 2003), PhD theses (Bunton 2002; Dong 1998; Dudley-Evans 1991; Hyland 2004), PhD vivas (Grimshaw 1989; Maingueneau 2002), tenure track reports (Hyon 2008), research proposals (Cadman, 2002; Myers 1990), researcher websites (Cronin 2001), and textbooks (Hyland 2000; Love 2002).

The laboratory record has, however, received comparatively little attention in spite of its central role in the daily work of laboratory scientists such as molecular biologists (Shankar 2007; Wickman 2010). Consequently, part of the motivation for the study presented in this thesis is to add to the growing body of work on academic research genres by focusing on the laboratory record in bioscience laboratories.

The second of the highlighted issues concerns the technology used to capture and exchange laboratory records. For many years, the principal medium for recordkeeping in the biosciences has been the handwritten laboratory notebook. However, recent developments in computing technology for the bioscience laboratory have set out to change this aspect of laboratory work by providing tools for electronic recordkeeping together with tools for improved information exchange between laboratory staff. Both commercial ventures, for example CERF from Rescentris Inc[3] and eCAT from Axiope Ltd[4], and research projects, for example CombeChem (schraefel *et al.* 2004) and Labscape (Arnstein *et al.* 2002), have developed computing systems termed electronic laboratory notebooks (ELNs) that aim to support this transition from paper-based laboratory recordkeeping towards electronic laboratory recordkeeping.

Achieving this transition from paper to electronic recordkeeping has, however, proven problematic in various domains (Sellen and Harper 2001; Shepherd and Yeo 2003). In the context of bioscience laboratories, surveys such as those by Taylor (2006) and Nature (2005, 2007) report that computing systems such as ELNs have not been successfully deployed across all of the market sectors that engage in bioscience research and development. In particular, uptake has been markedly low in the academic research sector where the long-standing pen and paper approach to recordkeeping remains the preferred solution. Why might this be? Mirel (1993) emphasizes the need to understand how users interact with documents in actual work situations as a *sine qua non* for the effective design of computer systems that support document-mediated interaction.

Previous studies of medical recordkeeping have reported that the continued use of paper-based systems in conjunction with, or in preference to, computer-based systems often results from the failure of a computer-based system to recognize the entirety of the communicative transaction embodied in a record (Garfinkel 1967; Heath and Luff 1996; Nygren and Henriksson 1992). These studies have demonstrated that both the content and structure of useful records are highly dependent on, and limited by, the expectations and working practices of the intended readership. Studies such as those by Anderson *et al.* (2008), and Ellingsen and Monteiro (2003) have reported problems encountered during the introduction of computer systems for healthcare recordkeeping in situations where the

---

[3] Product information is available from Rescentris Inc at http://www.rescentris.com [accessed 01 March 2011]. All trademarks are acknowledged.

[4] Product information is available from Axiope Ltd at http://www.axiope.com [accessed 01 March 2011]. All trademarks are acknowledged.

working practices embodied in the computer systems attempt to enforce standardization over the existing diverse, local practices. Molecular biology laboratories warrant investigation in their own right on the basis that molecular biology differs from health care services in terms of the nature of its work practices, the diversity of its work practices, and the co-localization patterns of its staff.

Technology-based research projects have developed prototype ELN systems to investigate electronic recordkeeping in the context of chemistry, physiology, and biology laboratories. This has included the application of novel interaction techniques such as augmented reality systems to facilitate user interaction in ELNs (Borriello 2006; Mackay *et al.* 2002), the development of distributed systems incorporating handheld devices to support mobile laboratory recordkeeping (Arnstein *et al.* 2002; Yeh *et al.* 2006), the development of integrated experimental planning and recording tools using workflow models to structure the capture of experimental results (Arnstein *et al.* 2002; Frey *et al.* 2004), and the development of tools to encourage collaboration and creativity among laboratory communities (Chin *et al.* 2002; Chin and Lansing 2004; Farooq *et al.* 2005; Tabard *et al.* 2008). Technology adaptation has been shown to influence the role of written and spoken genres (Levy 2001; Myers 2000; Nickerson 1999; Warschauer 2002), and laboratory recordkeeping is currently a locus for technology-driven change. In particular, by promoting the sharing of records (Lysakowski 1997), tools such as ELNs may expand the readership for laboratory records placing new demands on authors of these records.

Many of the current crop of ELNs rely on the capture of free-form textual descriptions of experimental work, supplemented in some cases with a limited set of metadata to facilitate the subsequent indexing and retrieval of records. Comparatively little attention has been paid to the range of linguistic structures, content, and representations employed by molecular biologists to construct their laboratory records. Consequently, part of the motivation for the study presented in this thesis is to add to the growing body of work on understanding the potential for electronic recordkeeping in bioscience laboratories by focusing on the language used in constructing laboratory records.

The third of the highlighted issues concerns the role of the individual in academic laboratory recordkeeping. Each individual scientist is responsible for capturing his or her own records, making choices about which data to record, how to represent the data, and when to record the data. This agency can lead to diversity in the recordkeeping practices and products at work in a laboratory. In turn, this diversity can limit the reliability and

usability of the records that have been kept by individuals since adapting a recordkeeping system to the needs of one user or one group has been shown to render those records less intelligible to other users and groups (Bannon and Bødker 1997; Greenberg 1991). Poor recordkeeping on the part of an individual can translate to serious implications for an entire laboratory community as witnessed, for example, by the impact of "haphazard" recordkeeping on the infamous Baltimore case of alleged scientific fraud (Kevles 1998). Balancing the needs of the individual and the wider community remains a concern in any setting that seeks to achieve effective communication irrespective of whether paper-based or electronic approaches to recordkeeping are in use.

Individual scientists do not, however, operate in isolation. Instead, they participate as members of social communities ranging through project groups, the local laboratory community and the wider bioscience community. Each of these social communities exhibits homogeneity, albeit greater or lesser degrees of homogeneity, in their disciplinary practices and conventions (Bartholomae 1986; Becher and Trowler 2001; Bourdieu 1991; Hyland 2000; Killingsworth 1992; Prior 1998). Individual scientists must display an awareness of these disciplinary practices and conventions in order to "create successful texts which display one's disciplinarity, or tacit knowledge of its expectations, for the practical purposes of communicating with peers" (Hyland 2000:10). In this sense, the agency exhibited by individual scientists when constructing laboratory records is not "free agency" but "social agency" (Fairclough 2003:22) in that the actions of individual scientists are socially constrained but not socially determined by the disciplinary practices and conventions of the communities in which they participate.

A central concept in this community-orientated view of literacy is the concept of genre (Bakhtin 1986; Bhatia 2002; Devitt 1997; Martin 1992; Miller 1984; Swales 1990). Genres are socially recognized ways of using language, both written and spoken, to perform typified action in recurring situations. Mutual understanding of, and adherence to, genre conventions enables the producers and consumers of a text to interact in a socio-pragmatic manner. Theorists have proposed varying models of genre by focusing to varying degrees on text-internal factors and text-external factors, that is, by emphasizing either regularity of the linguistic and structural forms used in texts or regularity of the social context in which texts act (Fairclough 1992; Flowerdew 2002a; Hyon 1996). Each of these models recognizes the role of agency in genre, exemplified by Schryer's (1993:208) description of genres as "stabilized-for-now or stabilized-enough" and by

Devitt's (1997:54) conclusion that effective use of genres requires understanding them "as both constraint and choice, both regularity and chaos, both inhibiting and enabling".

Adoption of technologies such as ELNs could introduce *de facto* standardization into molecular biology laboratories. Restructuring and coordinating organizational practices through standardization is known to be a complex task (Grindley 1995; Mintzberg 1979; Timmermans and Berg 2003). Enabling the adoption of standardized practices for recordkeeping would not be straightforward in terms of its socio-political impact. Individuals within academic molecular biology laboratories may perceive a move towards standardization as either a restriction on the creativity and flexibility they require to achieve their work, or even a questioning of their professional competence. Such negative perception of standards would be fatal since successful standardization requires the active cooperation of the individuals in any organization. A more positive basis on which to introduce standardized approaches to recordkeeping would be one in which individual scientists accept standardization as a coordination mechanism that reifies and propagates best practice in recordkeeping throughout the laboratory.

In addition to refining our knowledge of laboratory discourse, understanding the dimensions of variation in recordkeeping could inform both the teaching of laboratory recordkeeping practice within academic molecular biology laboratories, and the potential role of customization in technologies such as ELNs for electronic recordkeeping. Consequently, part of the motivation for the study presented in this thesis is to assess the variability present in the structures, content, and representations used by members of molecular biology laboratories when keeping laboratory records, and to assess the types of articulation work used to achieve mutual intelligibility across laboratory members.

1.3 Research questions

The primary goal of the study presented in this thesis is to investigate the role played by laboratory records in the disciplinary discourse of academic molecular biology laboratories. This includes assessing how scientists write and interpret laboratory records, and assessing how the records kept by scientists are used to coordinate the work of, and mediate the interaction between, the members of academic molecular biology laboratories. By investigating these aspects of laboratory recordkeeping, the study aims to contribute to the understanding of discourse in academic molecular biology research settings.

As a secondary goal, insights gained into the role of laboratory records in academic molecular biology research settings may also be used to inform the design of tools such as ELNs to facilitate the capture and sharing of records in these settings.

Specific research questions to be addressed in the context of this study are:

1. *What roles do laboratory records play in the discourse of academic molecular biology laboratories?*

2. *What are the structures, content, and representations that characterize the genre of laboratory records in academic molecular biology laboratories, and to what extent do these vary across different contexts of use?*

3. *How do readers of laboratory records in academic molecular biology laboratories make sense of laboratory records in different contexts of use?*

1.4 Research method

1.4.1 Multi-perspective framework for genre analysis

In order to address the aforementioned research questions, the study presented in this thesis employs discourse analytic research methods to investigate both the role of recordkeeping in academic molecular biology laboratories, and the textual form of laboratory records.

Discourse analysis (Brown and Yule 1983; van Dijk 1997) is an inter-disciplinary field of inquiry that is concerned with the analysis of linguistic behaviour. Drawing on developments in sociology, psychology, organizational science and information science, discourse analysis has evolved a range of research methods to investigate linguistic behaviour. This includes both qualitative and quantitative approaches, approaches that focus solely on writing and/or speech, approaches that examine other semiotic modes such as graphics and gesture, and approaches that strike different balances between the interpretation of textual form and social context. A detailed discussion of the full range of approaches to discourse analysis is considered beyond the scope of this thesis. However, Bhatia *et al.* (2008) and Schiffrin *et al.* (2001) provide informative overviews of the range of approaches used in discourse studies, Hoey *et al.* (2007) discuss the role of corpora in discourse analysis, Wodak and Meyer (2009) discuss the application of critical theory to

discourse analysis, and O'Halloran (2004) discusses discourse analytic approaches that consider multiple semiotic modes.

Over its relatively short history, discourse analysis has expanded its focus from the "description to explanation" of linguistic behaviour (Bhatia 1993:3). The initial focus on description is concerned with the analysis of language as text by drawing, *inter alia*, on the work of structuralist schools of linguistics (*e.g.* Greimas 1966; Jakobson 1937). In this formalist viewpoint, discourse is typically defined in structural terms as a unit of language dimensionally larger than the sentence, clause, and morpheme leading to the classical definition of discourse analysis as the analysis of language *at levels beyond the sentence*. More recently, the shift in focus to explanation has emphasized the analysis of language in use in specific social settings, and has been driven by an increased recognition of the role of language in both shaping and being shaped by social practices and social functions (*e.g.* Fairclough 1992; Foucault 1972; Giddens 1984; Rafoth and Rubin 1988; Witte 1992). In this functionalist viewpoint, the definition of discourse is extended to include social function and context so that "discourse is viewed as a system (a socially and culturally organized way of speaking) through which particular functions are realized" (Schiffrin 1994:32). This has necessitated a revised definition of discourse analysis that takes function and context into account, for example, as "the analysis of linguistic behaviour, written and spoken, beyond the limits of individual sentences, focusing primarily on the meaning constructed and interpreted as language is used in particular social contexts" (Bhatia *et al.* 2008:1).

Hymes (1974:69-81) points out that formalist and functionalist approaches to linguistic behaviour place a different emphasis on the role of variation in language use. Whereas formalism focuses on the "replication of uniformity" across idealized speakers and listeners in a homogeneous community, functionalism focuses on the "organization of diversity" through functionally adaptive use of language in communities exhibiting heterogeneity. As Brown and Yule (1983:26) state, "the discourse analyst treats his data as the record (text) of a dynamic process in which language was used as an instrument of communication in context by a speaker/writer to express meanings and achieve intentions (discourse). Working from this data, the analyst seeks to describe regularities in the linguistic realisations used by people to communicate those meanings and intentions". Note that the terms 'text' and 'discourse' are differentiated in this regard to separate out respectively the product and process of linguistic behaviour. This use of the term 'text' is

subscribed to throughout this thesis in deference to Widdowson's (2004) clarion call for terminological clarity within discourse analysis.

The priority afforded to different participant viewpoints has been problematized[5] to some degree in discourse analysis. As Fairclough (2003:10) points out "there are three analytically separable elements in processes of meaning-making: the production of the text, the text itself, and the reception of the text". Analytical methods that focus on the production of a text may tend to privilege the viewpoints of the writers/speakers. Conversely, analytical methods that focus on the reception of a text may tend to privilege the viewpoint of the readers/listeners. The involvement of the discourse analyst adds a further dimension of complexity by questioning whether the primary responsibility for interpreting texts should rest with the insiders who actually participate in communicative events, or with the views of the analyst/linguist/rhetorician looking in from the outside. In the context of genre theory, this issue is dependent on the extent to which the conception of a genre is considered to be a purely pragmatic, analyst-defined tool for which the analyst's viewpoint must be primary (*e.g.* Rosmarin 1985), or a joint construction of the situated experience of the writers and readers of texts (*e.g.* Devitt 2000).

Increased recognition of the importance of social context and the situated nature of language use has led to increased methodological complexity in discourse studies. In particular, it has brought about a degree of methodological eclecticism through which discourse analysts combine multiple research methods, both textual and contextual, to investigate the complexity of linguistic behaviour in social settings (Askehave and Swales 2001; Bhatia *et al.* 2008; Flowerdew 2002b). Combining multiple methods in this way, particularly enriching textual analysis with contextual analysis, has proven to be an effective means of obtaining informative explanations of linguistic behaviour in discourse communities. The results obtained from each of the component methods 'in isolation' can be used to amplify, extend, or limit the 'net' findings of the combined framework. In this sense, each research method makes a dual contribution by adding its own findings, and by enabling evaluation of the findings from other methods in the framework.

[5] Given the broad theme of this research as academic discourse, it was interesting to note the critical reaction of some reviewers of this thesis to my use of the word 'problematize', a term which is more familiar in the discourse of social scientists than of bioscientists. Put mildly, these reviewers problematized the use of 'problematize'. A similar discussion can be found at http://userpages.umbc.edu/~korenman/wmst/problematize.html [accessed 01 March 2011].

Employing multiple research methods in a single framework raises the issue of how best to sequence these methods. Askehave and Swales (2001) propose the two procedural frameworks shown in Figure 1-1 overleaf as frameworks for sequencing the analysis of text and context in genre studies.

These frameworks are focused particularly on the "repurposing" (Askehave and Swales 2001:207) of genres, rendering them particularly suited to the goals of this study into laboratory recordkeeping. The term "repurposing" is used in this sense to indicate a re-evaluation of the communicative purpose of a particular set of texts, and is driven by a recognition that the social purpose served by any set of texts may expand or shrink over time and/or in line with changes in the social setting in which they act. As Swales (2004:73) highlights, these procedural frameworks have been designed to "support an orientation that acknowledges that sets of texts or transcripts may not be doing what they seem, or not doing what they have traditionally been assumed to have been doing". Each framework can be applied independently or the two frameworks can be combined in order to stage a combination of textual and contextual research methods in discourse studies.

The specific approach to discourse analysis employed in this study of laboratory recordkeeping in academic molecular biology laboratories is a relatively novel combination of ethnography (Garfinkel 1967; Hammersley and Atkinson 1995; Saville-Troike 1982), genre analysis (Bhatia 1993; Swales 1990), and reading protocol analysis (Ericsson and Simon 1993). This combined approach enables a study of 'language in use' that examines both the language use and the social practices at work in laboratory recordkeeping, and that examines multiple perspectives including, importantly, the viewpoints of both producers and consumers of laboratory records. Bhatia *et al*. (2008:14) employ the term "multi-perspective genre analysis" to describe frameworks that employ genre analysis in combination with other approaches such as ethnography, and this term is used throughout the thesis to describe the research method selected for this study of laboratory recordkeeping.

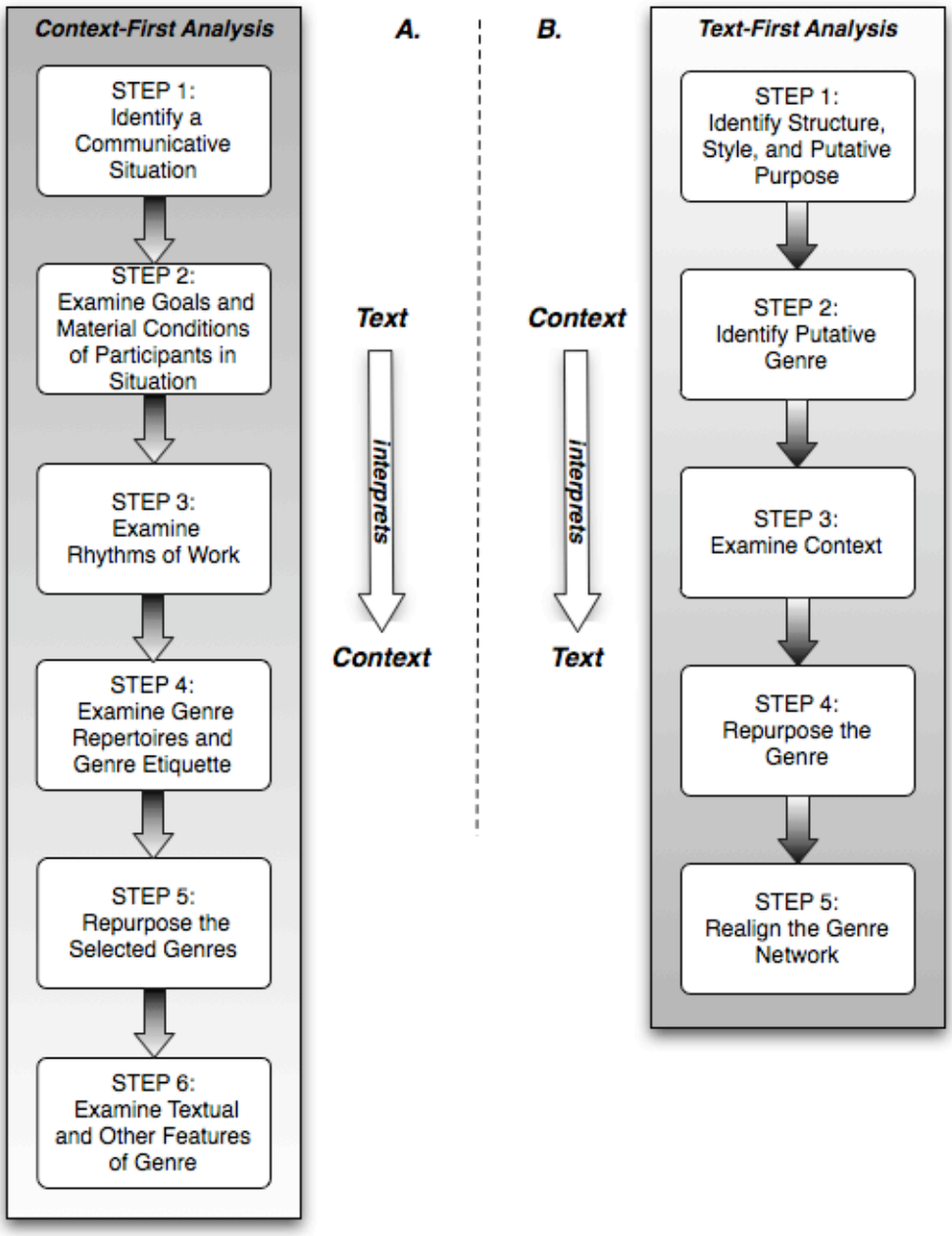


Figure 1-1: Procedural frameworks for genre analysis

Adapted from Askehave and Swales (2001). Showing two procedural frameworks for combining text-orientated research methods with context-orientated research methods in genre analysis. Both frameworks are designed to enable a re-evaluation of the communicative purpose served by a set of texts in order to potentially "repurpose" a genre. Each framework is defined as a sequence of procedural steps. Each step is defined in terms of the outcome to be achieved, and not in terms of the specific research method to be used to achieve that outcome. Note that these frameworks may be used independently or in combination with each other. (*A*) Framework for a *context-first* approach in which text-orientated methods are employed after context-orientated methods in order that knowledge of textual features can be used to refine the interpretation of the situational context in which a set of texts is written/read. (*B*) Framework for a *text-first* approach in which context-orientated methods are employed after text-orientated methods in order that knowledge of the situational context in which a set of texts is written/read can be used to refine the interpretation of the textual structure and style.

In line with the context-first procedural framework shown in Figure 1-1, an ethnographic study of laboratory recordkeeping in multiple academic molecular biology laboratories was followed by a genre analysis of a sample set of authentic laboratory records produced by multiple scientists during the course of their laboratory work. Following this, a study of the reading practices used by scientists from multiple laboratories to interpret a sample set of authentic laboratory records was performed using reading protocol analysis in line with the text-first procedural framework shown in Figure 1-1. In this way, the study investigated, in order, the social context in which laboratory records act, the textual features used by writers in constructing laboratory records, and the manner in which readers make sense of laboratory records.

1.4.2 Ethnography

Ethnography is a qualitative and interpretive approach to research that aims to develop an insider's view of socio-cultural practices through participant observation and the study of behaviour in naturally occurring settings. Drawing from its origins in anthropology (*e.g.* Malinowksi 1922) and sociology (*e.g.* Bulmer 1984), ethnography focuses on the collection of "naturally occurring data under normal conditions from numerous sources, typically over a period of time, without interfering with the context in any way" so that the data can be analysed "to convey the participants' subjective experiences" (Hyland 2006a:65). Data collection involves the ethnographer "participating, covertly or overtly, in people's daily lives for an extended period of time, watching what happens, listening to what is said, asking questions" (Hammersley and Atkinson 1995:1). In this way, ethnography privileges the systematic observation of the insider's own perspective and experiences over the use of pre-imposed conceptual frameworks to characterize the socio-cultural practices within a setting. This is not to say that ethnographers should, or even could, necessarily undertake studies of settings totally unencumbered of any theoretical constructs. Malinowski (1922:8), for example, advocates the role of "foreshadowed problems" rather than "preconceived ideas" as the starting point for ethnographic study, whilst Geertz (1983:57) constructs the ethnographer's role as identifying the insiders' own perspectives (*i.e.* "experience-near concepts") so that these can be merged with the concepts fashioned by theorists (*i.e.* "experience-far concepts") to produce an understanding of a community's behaviour.

Ethnography has a rich history in the investigation of literate cultures and oral cultures. Given this background, it is unsurprising that linguistic behaviour and texts continue to play significant roles in ethnographic studies as informative sources of insider data to be observed, collected, and analysed (Garfinkel 1967:186-207; Lynch 2009; Prior 1998; Swales 1998). This shared recognition of the analytic value imbued both in the language used in specific settings and in documents as the written product of discourse establishes a common foundational link between ethnography and genre analysis. Ethnography, however, operates from a wider analytical standpoint in which texts and linguistic behaviour constitute only one of the different types of social interaction that are significant and therefore to be observed and analysed by ethnographers.

Hammersley and Atkinson (1995), Lillis (2008), and Smart (2008) have pointed out that ethnography can be engaged at different levels within studies of linguistic behaviour. In this sense, ethnography as used is more accurately characterized as a range of approaches rather than a single approach since "there is considerable variety in prescription and practice, and along with this some dissensus about the proper nature of qualitative research" (Hammersley and Atkinson 1995:1). Some practitioners, for example, base their study on data collected through one collection method, whilst others prefer to use combinations of data collection methods. As Lillis (2008:355) points out, relying on only one data collection method such as insider interviews although widely used in "talk around text" can only be considered as a "minimal level" of ethnography that runs contrary to the holistic basis of data analysis and interpretation recommended in ethnography. Some practitioners rely on structured approaches to observation making use of pre-coded recording sheets to collate their observations (*cf.* time and motion studies), whilst others rely solely on event descriptions and quotations told from the insider's perspective. Practitioners choose to spend different periods of time in the setting under investigation. Some practitioners limit the report of their findings to the specific setting under investigation, whilst others attempt to suggest how their findings can be generalized to varying degrees. This last point emphasizes the fact that ethnographic studies in their characteristic form focus on discrete settings, and produce detailed, contextually-sensitive, "thick" descriptions (Geertz 1973) of these discrete settings.

Data analysis within ethnographic studies is essentially an iterative process that involves the detailed analysis of the relatively unstructured data collected through field notes, interviews, video recordings, and similar approaches in order to identify patterns of

behaviour, variations in the perspectives of different individuals or different groupings of individuals, differences between stated behaviour and actual performance, and other phenomena relevant to the focus of the study. These patterns of, and exceptions in, the behaviour and perspective of the study participants are reified into a set of thematic concepts that can be categorized, refined, related to each other, and grounded in the data collected during the course of the study in order to generate the findings of the ethnographic study.

There remain, however, contested schools of thought within ethnography over the representativeness of these thematic concepts. The key issue in this respect is to what extent is it valid to assume that the findings from discrete local settings will hold for other settings. Some theorists argue that it is unreliable to generalize the findings from an intense focus on such a discrete setting, so the thematic concepts should be used solely as the basis for a detailed description of the lifeworld[6] constructed by the specific community being observed. Other theorists argue that all generalizations are to some extent fuzzy, hence it is reasonable to generalize the findings from theoretical samples within discrete settings by developing thematic concepts into more systematic typologies to form the basis of a more generalized theory that may also inform other settings (Glaser and Strauss 1967; Lofland and Lofland 1995). Hammersley and Atkinson (1995:237) draw on earlier work to categorize the theories drawn from the findings of qualitative studies along two dimensions, *viz*. the scale of the setting under investigation (*i.e*. macro/micro theories), and the generality of the categories to which the cases belong (*i.e.* formal/substantive theories). Macro theories apply to large-scale social organizations such as nations or national societies, whilst micro theories are concerned with more local forms of social organization such as individual institutions. Formal theories are more general in nature and subsume substantive theories. The position subscribed to in this thesis is that advocated, *inter alia*, by Hammersley and Atkinson (1995:237) that "in many respects ethnography is better

[6] The term 'lifeworld' is used in the sense of Habermas (1987:131) as forming "the indirect context of what is said, discussed, addressed in a situation" by being "the intuitively present, in this sense familiar and transparent, and at the same time vast and incalculable web of presuppositions that have to be satisfied if an actual utterance is to be at all meaningful". Knorr Cetina (1999:217-220), *inter alia*, constructs "molecular biology laboratories in terms of small, lifeworldly arrangements focused upon single scientists and objectual relationships" where the context in these small lifeworlds is assembled from components including "besides a scientist, materials, instruments, bench space, and help from technicians or students".

suited to research on micro theory", and so is suited to the micro-substantive nature of this study into laboratory recordkeeping in academic molecular biology laboratories.[7]

The reasons for including ethnography in the framework of methods used in this study of laboratory recordkeeping are multiple. Firstly, ethnography enables observation of the interaction between different groupings of laboratory members. Secondly, ethnography privileges an emic (Headland *et al.* 1990) rather than an etic view of laboratory recordkeeping identifying the concepts and distinctions that are meaningful to the laboratory scientists themselves. Thirdly, ethnography enhances awareness of local exigencies in laboratory work by enabling direct and extended observation of the actual laboratory practices performed by scientists during the course of their work rather than the prescribed working practices. Fourthly, the longitudinal nature of an ethnographic study enables diachronic analysis of recordkeeping showing the potential for variation over time. Finally, ethnography has already proven to be an effective method in discourse analysis both in its own right and in combination with other approaches.

Specific methodological issues to be addressed when conducting ethnographic studies include the selection of which settings to investigate, managing the role of gatekeepers in gaining access to sites, ensuring the reliability and validity of the research findings, and selecting which process to follow in analysing the collected data. Considerations relevant to each of these issues for the design of this study into recordkeeping in academic molecular biology laboratories are discussed in chapter 3.

1.4.3 Genre analysis

Genre analysis is a discourse analytic research method that sets out with the dual aims, "first, to characterize typical or conventional textual features of any genre-specific text in an attempt to identify pedagogically utilizable form-function correlations; and second, to explain such a characterization in the context of the socio-cultural as well as the cognitive constraints operating in the relevant area of specialization, whether professional or academic" (Bhatia 1993:16). Although the emphasis in genre analysis is placed on "typical or conventional textual features", this does not preclude the potential for flexibility

[7] The issue of specificity has influenced, and continues to influence, the debate surrounding the appropriate use of ethnography in technology design (*e.g.* Crabtree *et al.* 2009).

but, instead, acknowledges that genre is a socially recognized way of using language in which authors "operate within a broad range of generic rules and conventions".

A central concern in genre analysis is the association between textual features and the socio-cultural context. Consequently, data collection in genre studies typically involves the collection of two distinct types of data, *viz.* data describing the social context in which the texts act, and data describing the linguistic features found in the texts in the corpus. Different levels of linguistic features may be analysed ranging from lexico-grammatical features such as the frequency of use of different tenses, to textualization features such as patterns of nominalization, up to structural features such as the move structures used in texts (*e.g.* Bhatia 1993). In this way, genre analysis can deliver quantitative analyses, qualitative analyses, or a combination of the two.

An increasingly important consideration when investigating the context of a specific setting is to understand the different types of relationship between the multiple genres at work in the setting. This includes understanding the relative value associated with each genre within the setting, understanding chronological dependencies between genres, understanding the specific subset of genres that are significant to individuals and/or categories of individual within the setting, and understanding the intertextual (Bakhtin 1986; Genette 1997) and other links between the full set of genres at work in the setting. Drawing on the work of others, Swales (2004:12-25) proposes the terms "genre hierarchy", "genre chain", "genre set", and "genre network" respectively for models of these four types of relationship between the genres in a setting.

The reference to "pedagogically utlitizable form-function correlations" in Bhatia's previously quoted definition identifies the traditional application domain for insights acquired through genre studies, *viz.* the provision of language support particularly in programmes such as the teaching of English for Specific Purposes (ESP) (Hutchison and Waters 1987) or its specialist branches in English for Academic Purposes (EAP) (*e.g.* Hyland 2006a) and English for Occupational Purposes (EOP) (*e.g.* Kim 2008). More recently, the potential for employing genre analysis in the design of computing systems[8] has also become the focus of attention (Antunes *et al.* 2006; Spinuzzi 2003; Swarts 2006).

[8] It is interesting to note that the use of genre analysis for technology design is at a much earlier stage compared to the use of some other social research methods for design, particularly the use of ethnography. Notwithstanding the body of experience already informing the use of ethnography in technology design, the issue of how best to deploy ethnography for design

The application of genre to pedagogy has, however, been problematic, and it is likely that its application to system design may also face similar challenges. One area of concern is specificity, questioning whether the focus of ESP should be on teaching the skills and language forms that are common across disciplines, or on teaching the skills and language forms that reflect the demands of specific disciplines such as molecular biology. A second area of concern relates to the power hierarchy at work within communities and is based on the understanding that "prestigious genres are often associated with precedent and proper procedure and this means that they represent an elite of expertise and power" (Hyland 2006a:31). This association of genre with power raises the ethical issue of whether promoting an understanding of genre forms is privileging the needs of the discipline over the needs of the individual student by reinforcing conformity to the existing social order in an uncritical manner (*cf*. enforced standardization) instead of promoting a more critical form of language awareness (Cadman 2002). Schryer (1994), in her study of literacy at a veterinary college, provides an informative example of the association between power and genre in the specific context of recordkeeping. In particular, Schryer (*ibid*.:122) distinguishes between the role of a researcher-orientated publication genre and that of a clinician-orientated recording genre, arguing that "research will continue to dictate to practice in disciplines such as medicine and engineering … as long as groups are socialized into different genres, especially where one genre is more highly valued".

A key task in performing genre analyses is the selection of an appropriate corpus, *i.e*. the sample set of texts to be analysed. The corpus appropriate for any given study can range from a single extended text to be analysed in detail up to a large set of texts to be investigated at a higher level of abstraction. In this sense, the selection of an appropriate corpus is dependent on the type of research questions posed for the genre study. It is important to note that the focus is on authentic texts that have been produced during the normal course of their everyday work by authors in real settings. This shared analytical focus on naturally occurring data establishes another common foundational link between ethnography and genre analysis.

It is important to recognize that genre analysis is not a unitary approach. In an influential mapping survey, Hyon (1996) differentiated three major schools of applied genre analysis,

---

remains open to discussion (*e.g.* Crabtree *et al.* 2009; Hartswood *et al.* 2008). Applying genre analysis for design is likely to face similar challenges. Specific areas to be addressed include which types of technology project may benefit from the use of these methods, how to represent and communicate the work products of these methods to technology project stakeholders, and how to coordinate the performance of these methods with other engineering tasks.

viz. the English for Specific Purposes (ESP) school (Bhatia 1993; Swales 1990), the Australian (or Sydney) school (Halliday and Hasan 1989; Martin *et al.* 1987), and the New Rhetoric school (Freedman and Medway 1994; Miller 1984) on the basis of the operational definition of genre and the intended application area for genre knowledge. The ESP school defines genre in terms of communicative events that are regularized in terms of "communicative purpose" and "patterns of structure, style, content, and intended audience" (Swales 1990), and is particularly concerned with tools for teaching spoken and written communication to non-native (L2) speakers for use in academic and professional settings. The Australian school draws on systemic functional linguistics (SFL) (Halliday and Hasan 1989) to define genre in terms of form-function relationships as "staged, goal-oriented, social processes" (Martin *et al.* 1987), and is particularly concerned with teaching at school rather university level and teaching of socially disadvantaged groups. The New Rhetoric school places greater emphasis on situational context than either the ESP or Australian schools by defining genre as regularity of social action and rhetorical function (Miller 1984), and is particularly concerned with tools for teaching rhetoric and composition to native (L1) speakers.

More recently, genre theorists (Devitt 2000; Fairclough 2003; Flowerdew 2000a; Swales 2004) have emphasized a simpler binary mapping based on the balance struck between text and context in the approaches to genre analysis. This is exemplified by Flowerdew's (2000a) designation of both the ESP school and Australian school as "linguistic approaches" that concentrate on lexico-grammatical and rhetorical features, whilst the New Rhetoric approach is designated as a "nonlinguistic approach" that concentrates more on situational context. Note that the investigation of textual and contextual features is not considered mutually exclusive in either of these approaches. As Flowerdew (2000a:91) states "the linguistic approach looks to the situational context to interpret the linguistic and discourse structures, where as the New Rhetoric may look to the text to interpret the situational context".

The specific approach to genre analysis subscribed to for this study of laboratory recordkeeping draws on the ESP school of applied genre analysis, and is essentially 'Swalesian' genre analysis (Askehave and Swales 2001; Swales 1981, 1990, 2004) following the analytic procedure described by Bhatia (1993:13-41). A key component of this approach is the analysis of individual texts in terms of move structures (Swales 1990), where moves are semantic and functional units of texts that can be identified on the basis

of their communicative purposes and linguistic boundaries to ‘‘describe the functions which particular portions of the text realize in the relationship to the overall task’’ (Connor *et al.* 1995:463). Bhatia (1993:30) emphasizes that a move is a cognitive structure that “serves a typical communicative intention that is always subservient to the overall communicative purpose of the genre”. Moves can vary in length from several paragraphs to parts of a sentence, and are realized through propositional and illocutionary content[9] of the text.

The reasons for including genre analysis in the framework of methods used in this study are multiple. Firstly, genre analysis emphasizes the role of conventionalized linguistic behaviour in institutional settings, and so directly addresses the issue of generic variation. Secondly, genre analysis involves the direct, detailed analysis of the authentic products of laboratory recordkeeping. Thirdly, genre analysis combines an evaluation of structural, linguistic, social, and cognitive aspects of text construction in laboratory recordkeeping. Finally, genre analysis has already proven to be an effective approach to analysing a range of written and spoken genres in both academic and professional settings.

Specific methodological issues to be addressed when conducting genre studies include the selection of an appropriate corpus or sample set of texts to be analysed, selecting the appropriate level of detail at which to analyse the texts, and selecting which process to follow in analysing the collected data. Considerations relevant to each of these issues for the design of this study into recordkeeping in academic molecular biology laboratories are discussed in chapter 4.

1.4.4 Reading protocol analysis

Initially, the design of this study into laboratory recordkeeping did not incorporate reading protocols on the basis that the ethnographic study would afford some opportunity to observe laboratory staff reading through other scientists’ laboratory notebooks. As will be discussed more fully in chapter 3, this did not prove to be the case. Consequently, the

---

[9] These terms are used in the sense of Austin (1975) to describe the types of function performed through linguistic utterances. In particular, utterances are characterized in terms of the locutionary act which designates the propositional content of an utterance, the illocutionary act (or force) which refers to the action a speaker performs in making the utterance (*e.g.* asserting, commanding, describing, promising, questioning, *etc.*), and the perlocutionary aspect which refers to the effect the speaker produces on a hearer (*e.g.* alarming, convincing, *etc.*).

framework of methods used in the study was extended to include reading protocols in response to the findings of the ethnographic study.

Protocol analysis is based on verbal reports made by individuals as they perform specific tasks and which are recorded for subsequent analysis. Reading protocols are verbal reports made by individuals as they perform reading tasks. Following work in psychology on the role of verbalization in problem solving (*e.g.* Gagné and Smith 1962; Marks 1951), verbal reports have become an established research method used in the study of the cognitive processes involved in tasks such as decision making, skill execution, writing, and reading. The core assumption underpinning the use of protocol analysis is that "the information that is heeded during the performance of a task, is the information that is reportable; and the information that is reported is information that is heeded" (Ericsson and Simon 1993:169). Recording the information verbalized by individuals when performing a task therefore provides "a trace of the cognitive processes that people attend to whilst doing a task" (Swain 2006:99).

Ericsson and Simon's (1993:12-24) influential work on verbalization categorizes verbal reports along two dimensions. The first of these dimensions concerns the time frame in which the verbal report is completed. Verbal reports completed whilst performing the task under investigation are termed concurrent reports; verbal reports made some time after the task under investigation has been completed are termed retrospective reports. The second dimension concerns the type of data to be verbalized by the subject. Verbal reports requiring subjects to verbalize their thoughts *per se* (thereby providing direct articulation of heeded information) are referred to as non-metacognitive reports (or Level 1 verbalization); verbal reports requiring subjects to verbalize additional information such as explanations and justifications are termed metacognitive reports (or Level 2/3 verbalization). Jourdenais (2001) similarly categorizes the verbal reports used in analysing language behaviour into retrospective reports, introspective reports, and think-aloud reports. Retrospective reports require the individual to think back and report upon the processes used and thoughts involved in completing a task at some point after the task was completed. Think-aloud reports and introspective reports are both made concurrently with the execution of a task. Think-aloud reports require individuals to verbalize directly whatever comes to mind during task execution, whilst introspective reports require individuals to interpret and explain their behaviour.

The use of verbal reports draws on a model of memory and information processing (*e.g.* Craik and Lockhart 1972; Newell and Simon 1972), which posits that the information accessible to individuals is held as the contents of short-term memory (or working memory) in which the information is initially stored and processed before potentially being added to long-term memory. Concurrent and retrospective reports operate differently under this model. In particular, concurrent reports are designed to tap directly into the short-term memory since they are collected concurrently with the execution of the task. Retrospective reports vary dependent on the time interval between the recording of the report and the completion of the task. If the retrospective report is given immediately after the completion of a task so that this interval is short, information may be retrieved partially from the individual's short-term memory and partially from the long-term memory. With longer intervals, the information for the retrospective protocol will be retrieved solely from the long-term memory.

The design and use of verbal reports for studies into linguistic behaviours must take account of the implications of this model of memory and information processing. Requesting individuals to explain or interpret their behaviour whilst performing a task may be reactive in the sense that it requires the individual to attend to an additional task and so may alter the cognitive processing reflected in any verbal report made by the individual (Bowles 2010; Nisbett and Wilson 1977). Requesting individuals to report retrospectively upon the processes used and thoughts involved in completing a task may limit veridicality in the sense that individuals may not accurately remember the processes used and may consequently report behaviours that they believe to be of interest to the study (*cf.* acquiescence bias) or that reflect the correct practice (*cf.* social desirability bias). As with other forms of qualitative interview, the type of prompting used during the collection of retrospective reports may also introduce bias into the verbal reports made by individuals by providing contextual cues that may trigger inaccurate responses (Ericsson and Simon 1993; Greene and Higgins 1994). In order to maximize the accuracy of verbal reports, practitioners of protocol analysis recommend the use of think-aloud concurrent protocols instead of introspective protocols, require immediate reporting for retrospective protocols, and emphasize the use of careful retrieval cues that typically begin with focused open-ended questions and follow through with questions on the specific threads raised by individual responses (Ericsson and Simon 1993; Greene and Higgins 1994; Jourdenais 2001; Sanz *et al.* 2009).

Analysing the verbal reports collected during concurrent and retrospective think-aloud protocols is an interpretative process that is driven by the research questions being addressed. In a similar manner to that described in section 1.4.2 for ethnographic studies, analysing the protocols involves the development and refinement of a thematic coding scheme that can be used to identify patterns of behaviour that are relevant to the research aims of the study. These coding schemes can either be driven in a bottom-up manner by the data found in the protocol reports, or can reflect the theoretical concerns on which the study is based (*e.g.* Jourdenais 2001:360). Each verbal report is systematically analysed and relevant segments of the protocol are coded appropriately. Ericsson and Simon (1993:272) emphasize that protocols can be analysed at different levels of abstraction, pointing out that "if the aim is to test theory more globally – the commonalities of behaviour, say, shared by a whole group of subjects – then it may be desired to code the protocol at a more aggregate level".

The reasons for including reading protocols in the framework of methods used in this study of laboratory recordkeeping are multiple. Firstly, reading protocols enable observation of the processes used by laboratory members to make sense of laboratory records produced by others, albeit in a less naturalistic setting. Secondly, reading protocols emphasize the cognitive processes involved in reading and so highlight the cognitive role of genre in linguistic behaviour. Thirdly, reading protocols used in conjunction with an appropriate aggregate level of thematic coding enable comparison between the different approaches used by laboratory staff to interpret different styles of laboratory recordkeeping. Finally, reading protocols have already proven to be an effective method in understanding linguistic behaviour in both reading and writing practices.

Specific methodological issues to be addressed when conducting reading protocol studies include the selection of an appropriate corpus or sample set of texts to be read by participants in the study, ensuring the veridicality of the data collected through the reading protocols, and selecting the appropriate coding scheme against which to analyse the reading processes. Considerations relevant to each of these issues for the design of this study into recordkeeping in academic molecular biology laboratories are discussed in chapter 5.

1.5 Reading guide

As befits a study of genre and variation in an academic setting, this thesis broadly conforms to an acknowledged genre variant, *viz.* the article-compilation PhD thesis genre variant described by Thompson (2001).

Chapter 2 of the thesis provides a survey of current literature relevant to recordkeeping and to the academic discourse of bioscience laboratories and related settings.

Chapter 3 presents the results of an ethnographic study into the recordkeeping practices at work in a sample set of academic molecular biology laboratories.

Chapter 4 presents the results of a genre analysis study of a corpus of authentic laboratory notebooks produced by a group of scientists during the course of their work in academic molecular biology laboratories.

Chapter 5 presents the results of a study using think-aloud protocols to investigate how a group of scientists read and interpreted a sample of authentic laboratory records that were written by other scientists during the course of their work in academic molecular biology laboratories.

Chapter 6 outlines the potential impact of the study findings on the design of computing technologies such as ELNs for use in academic molecular biology laboratories.

Chapter 7 summarizes the thesis conclusions, and outlines potential areas for future research.

A glossary containing definitions of a range of molecular biology and other terms is included after the appendices to this thesis. Many of the examples used to illustrate the results of the study will draw on molecular biology techniques, and it is recognized that these techniques may be unfamiliar to some readers of this thesis. In order to aid the reader, the first use of such terms will be defined in a footnote in addition to the definitions given in the glossary.

2 Review of Recordkeeping in the Discourse of Academic Bioscience Laboratories

This chapter of the thesis presents a review of the literature relating to recordkeeping and to the academic discourse of bioscience laboratories and related settings.

2.1 Models of recordkeeping

What is a record? Addressing this question is an appropriate starting point for any discussion of recordkeeping, and this thesis looks to the disciplines of records management and archival science to provide an answer. Both national records agencies such as the National Archives of Scotland[10], and professional societies such as the Archives and Records Association[11] have developed operational definitions of a record based on their work in recordkeeping for registration, governance, enterprise, and socio-cultural heritage. Representative of the majority of these is the definition provided by the International Organization for Standardization (ISO 2001), which states that a record is:

> "information created, received, and maintained as evidence and information, by an organization or person, in pursuance of legal obligations or in the transaction of business." [12]

This definition highlights three important issues in records management theory in respect of the role of containers, context, and value in recordkeeping.

2.1.1 Containers

The first of these issues, *viz*. containers, is concerned with the "objects we create in order to store records" (Upward 2000:124). As Upward (*ibid*.) stresses, containership can be viewed across multiple scales ranging from the container of an individual record, through

[10] NAS is an agency of the Scottish Government responsible for advising on, preserving, and providing access to national records encompassing state, church, and business activities. Further information can be found at http://www.nas.gov.uk [accessed 01 March 2011].

[11] ARA is a professional society formed by the recent amalgamation of the long-standing National Council on Archives, Association of Chief Archivists, and Society of Archivists (SoA). In the interests of full disclosure, the author of this thesis declares himself to be a member of ARA through his continued membership of the SoA.

[12] The ISO definition has its roots in a definition of record proposed by the International Council on Archives (*e.g.* Walne 1988), an international cooperative body including national archives agencies, special interest groups, and individual archivists.

the container of an archive holding multiple records, and up to the container of a group of archives. At the scale of the individual record as in the ISO definition, the container refers to the physical format or storage medium in which the recorded information is held. In this regard, the significant aspect of the ISO definition concerns what is not, rather than what is, stated in that there is specifically no restriction placed on the container of the recorded information. Constructing the concept of a record in this way both privileges the information contained in the record over the container, and acknowledges the evolution in recordkeeping that has been brought about by technological changes leading to diversity in containership including an increasingly common role for digital records (Ball 2010; Bearman 1994; Levy 2001; Research Information Network 2010).

Hartland *et al.* (2005:97), *inter alia*, emphasize that documents built up from multiple data in multiple semiotic modes (such as written text, photographs, speech recordings) and held in a range of physical formats and storage media have extended the concept of a record from that of relatively static, written text records to that of increasingly "virtual", "compound", and "fluid" records held in paper, digital, and hybrid forms. Although the container for laboratory records within academic bioscience settings has traditionally been paper notebooks, the advent of ELN systems (*e.g.* Arnstein *et al.* 2002; Tabard *et al.* 2008; Yeh *et al.* 2006) has raised the potential for a transition to digital or hybrid containership. However, as Ball (2010:4) points out in his recent review, "the curation of digital research data in a generalist context is still an immature discipline".

2.1.2 Content and context

The second issue highlighted by the ISO definition concerns the role of context in recordkeeping. By placing the producer of the record (in terms of the "organization or person") and the circumstances of its creation and use (in terms of the "legal obligations" or "transaction of business") on an equal basis with the information that is recorded, the ISO definition is emphasizing the increasing importance of context in recordkeeping (Day 2005; Eastwood 1992; Hartland *et al.* 2005). The importance attached to social context in records theory is attested by the increasingly important role assumed by metadata in archives in "recording the who, what, why, where of record creation … to provide the context and functionality of communicative activities" (Higgs 1998:108), and by a shift in focus to the function supported by records rather than the content of records as the preferred basis on which to appraise records (Cook 2005; Samuels 1991; Simpson and

Graham 2002). It is interesting to note that this turn to the social in records management mirrors similar turns to the social in discourse analysis[13] and in studies of human computer interaction.

Capturing context alongside the data content of a record enables the record to remain useful to a wider community over an extended period of time (Cumming 2007). The range of contextual knowledge that can be captured in metadata is extensive and encompasses knowledge about the agents involved in creating and using records, the social structures in which these agents interact, the social and business functions served by these agents, the types of transaction served by the records, and the rules and guidelines that regulate the social and organizational activities of these agents (Cook 2005; Levy 2001; McKemmish *et al.* 2005).

The influential Open Archival Information System (OAIS) reference model, defined originally for use with space science data but now adopted for wider use by the International Organization for Standardization (ISO 2003), embodies the importance of context via the notions of "representation information", "preservation information", and "designated community". Representation information, which includes structural, semantic and other types of information, is made available to potential users of a record in order to render the record understandable and therefore usable in different contexts of use. The scope of the information required to render a given record understandable to specific individuals will, of course, vary greatly dependent on the background knowledge possessed by each individual. The concept of designated community within the OAIS model provides a formal means through which to delimit the range of individuals for whom the representation information has been defined. Preservation information provides additional contextual colouring including information about the provenance of the record. The notion of designated community is particularly germane to this study of laboratory recordkeeping given the potential for tools such as ELNs to extend radically the scope of the readership for an individual scientist's laboratory records.

[13] Given the research focus of this thesis, perhaps this is better characterized as a turn to the discursive in sociology. Delanty and Strydom (2003:10), *inter alia*, characterize developments in the philosophy of social science as a series of epistemic shifts leading to an interpretative position in which "the very categories of science and of knowledge more generally came to be seen as shaped in and by language" (termed the 'linguistic turn') followed by "an extension of the linguistic turn into a full historical-cultural revolution which radically contextualized science" so that science is constructed "as a historically and socially shaped cultural artifact".

The Core Scientific Metadata Model (CSMD) (*e.g.* Matthews *et al.* 2010) provides a further example of development in the representation of metadata for scientific and laboratory settings to facilitate the sharing and reuse of scientific data in multiple contexts. This model has proven useful in experimental science, particularly laboratory science, and has been applied in e-Science related projects within the UK including projects in the bioscience fields of integrative biology (Gavaghan *et al.* 2005) and crystallography (Coles *et al.* 2006). e-Science is a UK national programme that seeks to promote the role of computing and information technology in enabling the work of scientific communities by supporting initiatives that seek "to develop advances in scientific data curation and analysis and to be a primary source of top quality systems and repositories that enable management, sharing and best use of research data."[14] Laboratory records form part of the product and process of laboratory research, and so constitute one type of research data. The range of metadata appropriate to the management, sharing, and best use of laboratory records within academic molecular biology laboratories remains open to investigation.

A common feature of both the CSMD and OAIS models is their support for collecting contextual information in the form of metadata as part of the scientific workflows in which records are originally created. This approach rejects the view of metadata capture as an epiphenomenon, and positions it as an essential part of the process to enable reuse of records such as laboratory records at a later date for diverse purposes by both the original creator and other users.

2.1.3 Value and purpose

The third issue highlighted by the ISO definition concerns the value, or more accurately the types of value, associated with a record. In particular, the ISO definition identifies the dual value attached to records in line with Schellenberg's (1956) taxonomy of values as both evidence of an activity, and a source of information about the activity. Evidential value is an essential characteristic that sets a record apart from other types of information source (Robek *et al.* 1995; Shepherd and Yeo 2003). It is, however, important to note that evidence in this sense does not necessarily imply legal evidence. In fact, current practitioner standards for records management recognize that the requirement for keeping

[14] This quotation forms part of the mission statement governing the work of the National e-Science Centre, and is taken from the National e-Science Centre website at http://www.nesc.ac.uk/nesc/mission.html [accessed 01 March 2011].

records may derive from multiple sources including external regulation, internal regulation, codes of practice, and community expectations (ISO 2001).

Shepherd and Yeo (2003:157), *inter alia*, implicitly extend the ISO definition of a record when they identify the additional value of a record as an "artefact or object" on the basis that users may be interested in the "aesthetic qualities, tangibility, physical form, associations or saleroom value" of the record. This extended definition is warranted in the domain of scientific recordkeeping. Examples of records valued, at least in part, as an artefact based on important associations might include the notebooks of Watson and Crick that document their work on the discovery of the structure of DNA, both of which have recently been appraised and retained in major laboratory archives.

The motivation for keeping records is, put simply, so that they can be used, and it is the evidential and informational value imbued in records that enables them to be used to satisfy different purposes. It is important to distinguish between two different contexts of use for records in terms of use within the organization that originally created the record, and use by other individuals or organizations. Couture (1996) proposes the terms "organic" and "inorganic" respectively to distinguish use within the organization that created the record from use by other organizations, whilst Schellenberg (1956) uses the terms "primary" and "secondary", and Shepherd and Yeo (2003:155-162) use the terms "internal" and "external". This thesis subscribes to the terms used by Shepherd and Yeo (*ibid.*). With regard to both internal and external use, records management theorists (*e.g.* Levy 2001; Reed 2005; Shepherd and Yeo 2003) distinguish three principal purposes for using records in terms of supporting accountability such as proving compliance with regulatory requirements, supporting organizational or business functions, and supporting cultural purposes such as historical studies or sociological studies of the role of an organization in the wider world. Understanding the role of the laboratory record in laboratory discourse requires understanding both the internal and external purposes served by these records.

2.1.4 Lifecycle of records

Lifecycle models for records, both the records continuum concept (Upward 1996, 1997) and earlier, linear lifecycle models (*e.g.* Couture and Rousseau 1987), emphasize that records may serve multiple purposes within an organization and that these purposes may vary over time. Earlier lifecycle models assume a linear progression in the purposes

served by a record over time from active records that are regularly used in the current business of an organization, to semi-active records, then to inactive records that are no longer needed for current business, culminating potentially in destruction if the records are appraised to be no longer of use. This progression may often signal an increased focus on external use of a record, particularly for cultural purposes, as records become semi-active or inactive in the setting in which they were created.

The records continuum concept, considered by many to be a more accurate lifecycle model and subscribed to by this thesis, draws on Gidden's (1984) structuration theory of space-time distanciation to construct a different perspective of "recordkeeping and archiving functions operating throughout the life of the record in four contemporaneous dimensions relating to their creation as documents, capture as records, organization as individual or corporate archive and pluralisation as collective archives" (McKemmish *et al.* 2005:165). The key distinction between the continuum lifecycle model and other lifecycle models for records lies in the fact that the records continuum adopts a more dynamic view of records as remaining in a continual state of development in which they "can even have multiple lives in spacetime as the contexts that surround their use and control alter and open up new threads of action, involving re-shaping and renewing the cycles of creation and disposition" (Upward 2000:120). In this viewpoint, even those records that are subsequently removed will leave visible memory traces in terms of their history of use within an organization and its archives. Variation in use of records is not assumed to progress through a linear sequence of stages but is held to be multi-threaded and considerably less predictable over space and time.

2.1.5 Allocative and authoritative recordkeeping

Different organizational settings may place different requirements on records, and may strike different balances between the need to support accountability, business functions, and cultural functions. In this sense, recordkeeping is known to be dependent on the organizational culture at work in any setting (Hofstede 2001; Hofstede and Hofstede 2007; Pheysey 1993). In particular, different organizational cultures will place a different emphasis on internal and external accountability on a basis which "will be partly predetermined by nationality, industry, task, and market, partly related to organizational variables such as structure and control systems, and partly unique products of idiosyncratic features such as the organization's history or the personality of its founder" (Hofstede

2001:401). This observation has important implications for recordkeeping in molecular biology laboratories. In particular, laboratory recordkeeping may vary between different molecular biology laboratories, between projects in the same laboratory, and over time.

Building on their experience as records management practitioners, Hartland *et al*. (2005) have also drawn on Giddens' (1984) theory of structuration to distinguish between the culturally dependent approaches of allocative recordkeeping and authoritative recordkeeping. Allocative recordkeeping is concerned with policies that enable the sharing of information within a community; authoritative recordkeeping, in contrast, is concerned with policies to control the production and dissemination of information.

Commercial ELN products have focused on support for authoritative recordkeeping by supporting accountability through automated date stamping of records, access control, and fixed record formats. In this way, these tools prioritize the role of laboratory record as authoritative evidence of work done over the expanded role of laboratory record as a coordination mechanism to support laboratory staff in achieving their work (Hurley 2005). This style of recordkeeping is typical of working cultures in large, hierarchical organizations with high power-distances, and which value collectivism over individualism (Hofstede and Hofstede 2007). In contrast to the large teams, high throughput, and pre-defined repetitive tasks characteristic of work in many industrial and health sector laboratories, research laboratories in the academic sector typically operate as small groups of individuals applying relatively new techniques to small sample sets (*e.g.* Knorr Cetina 1999:216-240). Arguably, the design criteria for ELN tools suited to academic molecular biology laboratories should support an allocative approach to recordkeeping, and should provide enhanced support for individualism over collectivism.

2.2 Recordkeeping in the laboratory

2.2.1 Role of laboratory records in science studies

Bioscience laboratories in academic settings have proven to be informative sites for studies across a range of disciplines, and the willingness of bioscientists to place themselves metaphorically 'in an Eppendorf tube'[15] is to be admired. Laboratory records have played

[15] This reference is made in honour of a laboratory worker who joked that he now knew what it felt like to be a sample in an Eppendorf tube whilst participating in the ethnographic study described in this thesis. An Eppendorf tube is a common brand of tube that is used within molecular

a central role in a number of these studies including studies in the history of science, studies in the sociology of science and scientific knowledge[16], studies undertaken in the development of ELN computing systems, and studies of laboratory discourse. In some cases, laboratory records have been used as a source of data to investigate aspects of laboratory science other than recordkeeping; in some cases, the role of recordkeeping in laboratory work has formed part of the research focus. The use of laboratory records in conducting these studies is warranted by the fact that "reconstructions of inventions and inventive processes are appropriate and valid, for the notebooks indeed record the state of the technology and provide evidence of the plans, thinking, trials, and solutions of the people who wrote the notebooks and built the technology" (Bazerman 1999:51).

It is important to note that these studies, whether historical, sociological, technological or linguistic in origin, have overlapped in their areas of concern under the broad agenda of science studies, and so have been able to contribute in a complementary manner to an understanding of laboratory recordkeeping. Particularly germane to this thesis, such studies have enabled researchers engaged in the field of science studies to reflect on the situated, social context in which laboratory recordkeeping is conducted across multiple settings. As Pickering (1995:2-3) points out, science studies encompass an interest in the "details of the day-to-day doing of science" which has expanded "our concept of the science object by documenting its sheer multiplicity and heterogeneity" across "all dimensions of science – the conceptual, the social, the material".

Surveying the use of written documents such as laboratory records in science studies does, however, require care. Latour (1987:63), for example, constrains the role of studies of technical literature in the laboratory on the basis that "no matter how interesting and necessary these studies are, they are not sufficient if we want to follow scientists and engineers at work; after all, they do not draft, read and write papers twenty-four hours a day." Collins (1975) privileges the analysis of the "contingent forum" of texts such as letters, conversations, and interview comments over the "constitutive forum" of texts such as published research articles on the basis that "the informal texts can show what goes on

biology experiments to hold DNA samples and other types of sample for both storage in freezers and processing in devices such as centrifuges. Product information is available at http://www.eppendorf.com [accessed 01 March 2011]. All trademarks are acknowledged.

[16] This includes studies grounded in different traditions including to a limited extent the Mertonian paradigm (sometimes characterized as the sociology of scientists), and to a much greater extent the strong programme (commonly referred to as the sociology of scientific knowledge). For comparative reviews of these and other schools in the sociology of science see, for example, David (2005) or Zuckerman (1988).

before discourse is fitted into the formalities of research articles" (Myers 1990:24). Gilbert and Mulkay (1984) distinguish an "empiricist repertoire" in scientific communication that emphasizes impersonality, objectivity, and experimental results from a "contingent repertoire" that acknowledges social factors, and warn that scientists switch back and forth between these repertoires even within the same document. The potential for subjectivity and partiality in written documents as in other data sources is also well known to historians, who seek out corroboration for their interpretation of the records held in notebooks in other sources including letters, interviews, and notebooks written by other authors (*e.g.* Edsall 1974; Holmes *et al.* 2003; Judson 1996). Similarly, written texts such as laboratory records represent only one source of data within sociological studies, which commonly employ a broader methodological framework in which records are combined with interviews, participant observation, and questionnaires (*e.g.* Hammersley and Atkinson 1995; Schryer 1993; Wickman 2010).

2.2.2 Containers of laboratory records

Holmes (2004), in his highly informative study, discusses the complexity of the investigative pathways traversed by different scientists throughout their careers. These pathways may resolve into multiple lines of investigation on the same or distinct research projects. Pathways that are regarded as sufficiently independent may, in some cases, be recorded in separate series of notebooks so that the organization of the containers used by a scientist to keep laboratory records in effect mirrors the cognitive partitioning of his/her research (Gruber 1989). Holmes (*ibid*.) emphasizes that the use of separate series of notebooks is dependent on the preferred strategy of the individual scientist by comparing the recordkeeping practices of Claude Bernard[17] who maintained two series of notebooks to separate his work on the chemical processes of nutrition from his work on the organization of the nervous system, and the recordkeeping of Hans Krebs[18] who maintained a single series of notebooks in which he shifted back and forth between multiple problems over the course of weeks and months. Tabard *et al.* (2008), in their recent analysis of work undertaken as part of the design of an ELN system, also reported

[17] Claude Bernard (1813-1878) was an eminent French physiologist and medical scientist. Michael Foster's recently reproduced (ISBN 0554852594) biography of Bernard is to be recommended. Details of his life and work in science are also available at http://www.claude-bernard.co.uk [accessed 01 March 2011].

[18] Hans A. Krebs (1900-1981) was an eminent German-born British physician and biochemist, who shared the Nobel Prize in Physiology for his discovery of the citric acid cycle. Biographical details are available at http://nobelprize.org/nobel_prizes/medicine/laureates/1953/krebs-bio.html [accessed 01 March 2011].

the use of both chronological and project-based organizational schemes for laboratory recordkeeping at the Pasteur Institute, an independent, not-for-profit research institute. Kaye *et al.* (2006), similarly, report variation in the organizational approaches used to structure a range of personal archives within academic settings including time-based schemes, project-based schemes, career-stage schemes, and random schemes.

In the large majority of cases in which laboratory notebooks have been consulted in the context of science studies, individual scientists have kept their records in personal notebooks. This includes studies of the records kept by scientists working in different disciplines and across different time periods ranging from the seventeenth century onwards up to current times (*e.g.* Barnes *et al.* 1996; Crick 2005; Eldredge 2005; Gross *et al.* 2002; Holmes 2004; Holmes *et al.* 2003; Judson 1996; Kay 1993; Mackay *et al.* 2002; Latour and Woolgar 1986; Yeh *et al.* 2006).

2.2.3 Personal and collaborative recordkeeping

Typical examples of these personal notebooks are the notebooks examined by Judson (1996) and Kay (1993) to construct their respective historical accounts of the race between researchers working at the California Institute of Technology (Caltech) and those working in the Cavendish Laboratory at Cambridge University that led to the discovery of the double helix structure of DNA. Judson (1996:125-173), in particular, examines the role of laboratory records in collaboration and sharing in laboratory research in his analysis of the events surrounding the privileged access afforded to James Watson of an X-ray diffraction photograph[19] taken by Rosalind Franklin and labelled in her records as "Structure B: Photograph 51". Gaining access to this photograph, by their own admission, helped Watson and Crick to reach their conclusions about the double helix structure, but remains controversial as the access was granted not by Franklin herself but by a third party.

These events highlight a number of issues relevant to the social context of laboratory research. Firstly, the controversy surrounding the access granted to Watson without the direct consent of Franklin reinforces the notion of unpublished experimental records as 'personal property', and laboratory etiquette requires other members of even the same

[19] X-ray diffraction photographs are the products of X-ray crystallography, a method of imaging that can be used to determine the arrangement of atoms within a crystal by observing the diffraction patterns formed by an X-ray beam directed at the crystal. This method has been used to investigate the structure of biological molecules such as proteins and DNA.

laboratory group to respect this notion. Secondly, detailed analysis of Franklin's laboratory records carried out by Judson (*ibid*.) in conjunction with Aaron Klug[20], a scientist and colleague of Franklin who inherited her notebooks, show that although Franklin was well on the way to discovering the structure of DNA in her own right, she was not able to make the final inductive steps. Judson (*ibid*.) cites Klug who contrasts the collaborative interaction of Watson and Crick against the isolated position of Franklin as a key reason for her inability to make the final inductive steps required at that time. Thirdly, Judson (*ibid*.:161) cites a letter sent to Franklin after she had moved away to a new position ordering her not to continue her work on the structure of DNA, a "prohibition that is unheard of in modern science" but is indicative of a degree of protectionism within laboratory research.[21]

Bazerman's (1999) study of work in Thomas Edison's laboratory, albeit not a bioscience laboratory, provides an informative counterpoint as a study of recordkeeping in a collaborative setting geared to industrial invention in which "the notebooks are the residuum of the communicative acts that brought the work of the many people in the laboratory together" (*ibid*.:48). Edison instituted a multi-perspective approach to recordkeeping based on three types of notebook, each of which served a distinct communicative function as reported by Bazerman (*ibid*.). Notebooks labelled "Experimental Researches" served as a container for laboratory records that constituted legal documentation of work, and were characterized by formality, explicitness, and structure in both the written and graphical entries used in these retrospective records of the experimental work that had already been completed. In contrast, laboratory notebooks were used to mediate the day-to-day operation of the laboratory and, interestingly, were conceived of as communal resources positioned at convenient locations throughout the laboratory for use by all staff. The laboratory records kept in these notebooks included less formal results and sketches in addition to administrative records such as equipment orders.

---

[20] Aaron Klug is an eminent biochemist, past-president of the Royal Society, and winner of the Nobel Prize in Chemistry for his development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes. Biographical details are available at http://nobelprize.org/nobel_prizes/chemistry /laureates/1982/klug.html [accessed 01 March 2011].

[21] Maddox (1996) provides more recent evidence for a trend towards protectionism based on his experience during two periods as editor of the journal Nature from 1966-1973 and from 1980-1995. Maddox (*ibid*.:xi) notes that "the most striking change [between these two periods as editor] was the huge increase in competitiveness" so that "by 1980, secretiveness had become commonplace" and "authors had taken to sending long lists of names to whom, please, manuscripts should not be sent for review, with the explanation, to authors always self-sufficient, that the listed names were those of people working on the same problem."

A third category of laboratory record was kept in scrapbooks, and included clippings drawn from the popular and technical press on inventions, theories, and discoveries by both Edison's group and other rival groups. The entries in these scrapbooks were concerned with popularizing the work of the laboratory and drew on linguistic forms aimed at promoting the work of the laboratory to the public at large, to potential customers, and to prospective investors.

2.2.4 Content of laboratory records

A notable aspect of the laboratory records consulted in studies such as Barnes *et al.* (1996), Bazerman (1999), Campbell (1990), Gross (2006), Gross *et al.* (2002), Holmes *et al.* (2003), Tweney and Gooding (1991), and Wickman (2010) is the range of information that was kept in laboratory records. This encompassed descriptions of experimental apparatus used, procedures followed, results obtained, observations made, and periods of reflective thinking. Particularly notable is the frequency and degree of reflective thinking that was present in the laboratory notebooks. In this sense, the laboratory records did not serve solely to record experimental data and processes in response to time spent at the laboratory bench, but also to memorialize periods of reflection on the implication of the results obtained for the refinement of scientific theory and experimental plans (Holmes 2004; Tweney and Gooding 1991). The importance of reflective thinking has recently begun to influence the design of tools such as ELN systems for use in laboratory settings. In particular, a limited set of these tools has sought to facilitate a degree of collaborative reflection on the results published in laboratory records (Chin and Lansing 2004; Farooq *et al.* 2005; Tabard *et al.* 2008).

The temporal sequencing of experimental work is also known to act upon the content and structures used to keep experimental records. Holmes (2004:150), *inter alia*, points out that "when experiments require extended periods for preparation and execution, or when other tasks intervene between experiments in a particular line, we would expect successive experiments more often to reflect, in addition to the experience of the preceding ones, some evolution in the thought of the investigator during the interval, or an influence impinging on the course of the investigation from some contemporary event in the field". The core issue in this respect is that laboratory records are not constructed in isolation but exist as part of a series of entries across one or more notebooks. Writing and reading laboratory records is therefore dependent, at least in part, on the sequence of records that

have gone before in the sense that this sequence constitutes a scaffold on which to build or interpret the current record.

Associating and interpreting related experimental work, particularly in the case of work separated by extended time intervals, can be a difficult process that Wickman (2010:262), in his study of records in a chemistry laboratory, identifies as dependent on "the writer or reader's ability to integrate context and text with his or her existing knowledge (intertextual connections, interpretants)". This analysis is driven by Witte's (1992) concept of writing and reading as a triadic constructivist semiotic based on text as "the object of production and interpretation", context as "the local sites and situations in which the texts are produced and interpreted", and intertextuality (Bakhtin 1986) as "the means by which individuals come to understand texts in relation to other texts and utterances" (Wickman 2010:262). A range of other texts exists within the genre system of the bioscience laboratory including other laboratory records, textbooks, research articles, and experimental protocols, so understanding the different types of intertextual dependency between records and these other texts may prove informative both for an understanding of laboratory discourse and, potentially, in the design of ELN systems.

Studies such as those by Campbell (1990), Crick (2005), and Gross (2006) focus on the rhetorical forms that influence the content of laboratory records, and employ diachronic analyses of laboratory records to investigate the dialogic nature of recordkeeping in laboratory notebooks including communal and personal notebooks. Each of these studies examines the evolution of the argument embodied in the laboratory records in Charles Darwin's celebrated red notebook[22] as it progresses from a form of self-persuasion during the initial development of his theory to a form of a public persuasion aimed at promoting and establishing the theory across a wider readership. Gross (*ibid*.) characterizes this shift as a transition from an inner mental debate to a public debate signalled by a corresponding shift in rhetorical device. It is interesting to note the correspondence between this transition in the communicative purpose of experimental recordkeeping and the transition in the style and structure of experimental articles over the history of scientific journals as reported by Bazerman (1988), Gross *et al.* (2002), and Valle (1999). A similar dynamic can be seen driving both these transitions in terms of the need to convince an external

[22] The red notebook, so-called because of its red leather binding, was used by Charles Darwin during his service on HMS Beagle and contains records of observation and reflective thinking that contributed to his theory of evolution. Copies of this and other texts by Darwin are available at http://darwin-online.org.uk [accessed 01 March 2011].

readership that the claims being made are acceptable and are a consistent and reasonable interpretation of the experimental work that has been carried out, characterized by Shapin (1994) as "virtual witnessing".

2.2.5 Role of inscription

Latour and Woolgar (1986) report on an ethnographic study of laboratory work in a bioscience setting within the Salk Institute, a leading independent, not-for-profit research institute. This study has been influential in identifying the central role played by documents in the practice of laboratory science on the basis that "every presentation and discussion of results entailed the manipulation either of slides, protocol sheets, papers, preprints, labels, or articles. Even the most informal exchanges constantly focussed either directly or indirectly on documents" (*ibid*.:53). The practice of laboratory science, following this model, is centred on the production of inscriptions produced by "inscription devices which transform pieces of matter into written documents. More exactly, an inscription device is any item of apparatus or particular configuration of such items which can transform a material substance into a figure or a diagram which is directly usable by one of the members of the office space" (*ibid*.:51). Graphs or images produced by laboratory instruments constitute one form of written inscription. Laboratory records and research articles constitute other forms of written inscription that are composed in part through a combination of these graphs and images.

It is important to note that inscriptions, in this sense, are not necessarily direct representations of the natural order (*e.g.* Lynch and Woolgar 1988). As Knorr Cetina (1999:26) states "one rarely works in laboratories with objects as they occur in nature. Rather, one works with object images or with their visual, auditory, or electrical traces, and with their components, their extractions, and their 'purified' versions". These indirect representations of the material objects found in the laboratory will be used in the construction and communication of laboratory records, and are also characteristic of other genres such as the research article. An informative comparison can be drawn here between the linguistic behaviour at work in research articles aimed at the scientific community and that used in popularizations of the same work aimed at the general public (*e.g.* Corbett 1992, 2009; Myers 1990). These two genres employ two distinct styles to present two distinct views of the work performed by the scientists, which Myers (1990:141) refers to as the "narrative of science" found in research articles and the "narrative of nature" found in

popularizations. Whereas the "narrative of science" emphasizes the conceptual structures of the scientific discipline, the "narrative of nature" emphasizes the plants, animals or other organisms as the subject of interest and not the scientific activity. Corbett (2009:88) points out that the "research-orientation in particular directs the specialist reader towards the abstract agents found in the 'narratives of science' and away from the physical agents more often found in the 'narratives of nature'."

Multiple semiotic modes are at work in laboratory discourse, and visual representations of data are particularly common in the laboratory (Amann and Knorr Cetina 1988; Latour and Woolgar 1986; Lynch 1985b). Amman and Knorr Cetina (1988:160), *inter alia*, identify the role of "montages" in ordering the documentation and interpretation of visual representations within science, whereby the visual data is processed "to suggest a particular reading of the display" by foregrounding the evidence of interest, reducing any background noise and superimposing "pointers" to mark out some features as significant (*cf*. Lynch 1985b). Accordingly, visual data, both in raw and montage forms, are a common component of laboratory records (*e.g.* Shankar 2007; Tabard *et al.* 2008; Wickman 2010; Yeh *et al.* 2006). A further interesting development in this direction is the use of video in place of traditional print journals as a means to publish research articles as exemplified by JoVE, the Journal of Visualized Experiments[23].

2.2.6 Purpose of laboratory records

Given the personal nature of laboratory notebooks, it is perhaps unsurprising that individual scientists differ in the approach they take to recording their laboratory work. Holmes (2004:14-15), drawing on a range of historical analyses of laboratory notebooks including his own extensive work on the notebooks of Antoine-Laurent Lavoisier[24] and Hans Krebs, points out that "like other people, scientists differ greatly in the care with which they make and retain records of their daily activities" with the consequence that "in the most sparse cases, such research notebooks may contain little more than numerical results, together with minimal descriptions of experimental conditions". This observation

[23] JoVE is a peer-reviewed, PubMed-indexed journal that publishes biological research on-line in video format. Further information is available at http://www.jove.com [accessed 01 March 2011].

[24] Antoine-Laurent Lavoisier (1743-1794) was an eminent French chemist and biologist, and is widely held to be one of the founders of modern chemistry. Arthur Donovan's biography (ISBN 052156672X) provides an informative account of Lavoisier's life both in and out of science until his untimely death during the French Revolution.

raises the issue of what should be recorded in a laboratory record, and what is the potential impact of different recordkeeping approaches. As for all records, the issue of what data/metadata should be recorded is crucially dependent on the accountability and organizational purposes to be served by the record. In practical terms, what is actually recorded in a laboratory record together with any associated metadata acts to constrain the accountability and organizational purposes that can actually be served by a record.

In line with the Mertonian norm of universality (Merton 1973), a common proposal for the purpose of a laboratory record is that it should enable a body of experimental work to be replicated (*e.g*. Ebel *et al.* 2004; Macrina 1995). This is illustrated by Wickman (2010:273) in his recent study of laboratory recordkeeping in a materials science laboratory when he reports the view of the scientist he was observing that "the main purpose of a lab notebook is a record for other people to reproduce your work". The ability to replicate experimental work based on a written report such as laboratory record has, however, been problematized (Collins 1985; Lynch 1997; Polanyi 1967; Schmidt 1997) underpinned by the notion of scientific practice as craftwork (Ravetz 1971). By characterizing science as craftwork, Ravetz (*ibid*.) emphasizes the role of tacit knowledge (Polanyi 1967) and the transfer of skills through experience in the work of scientists. Polanyi's (*ibid*.) concept of tacit knowledge refers to knowledge that is invisible and taken for granted by those using it, and encapsulates the ability of individuals to perform a skill without being able to articulate how they are doing it. Knorr Cetina (1981) further distinguishes the role of local knowledge that is resident in individual scientists or specific laboratories. Both tacit knowledge and local knowledge give rise to issues of representational adequacy in laboratory records. Tacit knowledge cannot be articulated and so will not be visible in a laboratory record. Individuals tend to assume that any knowledge they possess is shared by others (*e.g.* Hayes and Bajzek 2008), and this 'knowledge effect' suggests that local knowledge possessed by a scientist will also not be visible in his/her laboratory records.

The issue of replicating experimental work is also taken up by Collins (1985:159) who contrasts two models for learning, communication, and practice in science in terms of an algorithmic model and an enculturation model. The algorithmic model is based on the premise that "formal communication can carry a complete recipe for experiment" in which a reader "has been a 'virtual witness' of scientists' activities and can see the validity of the procedures and findings". In contrast, the enculturation model, which Collins (*ibid*.) holds to be a more accurate picture of the context in which laboratory records would be used,

asserts that "the locus of knowledge is not the written word" so that an individual's knowledge of how to perform an experiment "must be acquired by contact with the relevant community rather than by transferring programmes of instruction". Cambrioso and Keating (1988:246) point out that scientists are themselves aware of the fact "the unsaid is indeed a part of conscious scientific practice". Lynch (1985a, 1997) highlights the difficulties involved in achieving adequate representations of experimental work, and the limitations of experimental procedures as algorithmic models in his examination of the manner in which laboratory records have been deconstructed to reveal their inherent contingencies and reliance on tacit knowledge when used as forensic evidence in legal cases.

Ebel *et al.* (2004:16) emphasize the role of the laboratory records kept in notebooks "as the 'germ cell' of the scientific literature", alluding to the role of records in documenting a series of experimental procedures, results, and partial analyses from which research articles or other reports are retrospectively produced. One factor acting on the selection of which data to record in laboratory records is therefore the range of data that might be required for publication. It is important to note, however, that this is not an isomorphic relationship. Laboratory records form only part of the documentation set that is present in bioscience laboratories (Latour and Woolgar 1986; Shankar 2007; Wickman 2010), and a range of other information sources such as textbooks, equipment manuals, and previously published articles may contribute to the production of a report for publication.

Research articles and other published reports are distilled versions of the body of experimental work that has been carried out, which are typically shaped through choice of language and rhetorical device to promote, justify, or align the scientist's work within the scientific and other communities (Bazerman 1988; Kanoksilapatham 2005, Myers 1990; Swales 2004). In this sense, published reports tend to present science in an idealized, teleological form, which is stripped of the practicalities and contingencies that characterize the day-to-day practice of science (Shankar 2007). Consequently, preparing a report for publication is a reductive process in the sense that "much of what one enters into a notebook is likely never to see the light of day in an official report or publication" (Ebel *et al*. 2004:16).

Bowker (2005), Latour and Woolgar (1986), and Knorr Cetina (1981), *inter alia*, extend the role of laboratory records beyond that of a data repository geared to supporting later publication by emphasizing the importance of inscription as a tool for ordering the

complexity of day-to-day scientific practice. This includes managing what Knorr Cetina (1999:84) terms the "material culture" of the laboratory in terms of the interactions with the material objects of laboratory work including instruments and biological materials, and the "object-oriented processing" of the laboratory in terms of temporal and spatial rhythms of experimental protocols.

Barnes *et al.* (1996) highlight the role of interpretation in laboratory records by comparing two previous studies of Robert Millikan's[25] laboratory notebooks recording his work on determining the elementary charge. Millikan's records document a series of repeated experiments that were performed using the same apparatus over an extended period of time, but only a subset of these experiments (58 out of 175) was eventually selected to contribute to the results eventually reported in the published research article. The two studies referenced by Barnes *et al*. (*ibid*.) have investigated the basis on which the selection of results was made and the manner in which the discarded experimental results were documented in Millikan's laboratory records. The somewhat controversial conclusion of these studies is that there was no "smooth, automatic, unproblematic path joining the readings entered into the laboratory notebook and the data in the published papers that were used by Rutherford, Bohr and the rest of the scientific community" (*ibid*.:22). In terms of recordkeeping practice, it is important to note two particular aspects of the controversy generated by these studies. Firstly, the problem arises, in part, due to the minimalistic approach used by Millikan to record discarded results as " 'something wrong', but what that something is was never specified or tracked down" (*ibid*.:24). Secondly, different elements of an experimental record may exhibit more or less sensitivity to incomplete recordkeeping on the basis that "the item that maintains its identity readily across time and space isn't the actual sequence of experimental actions, but the *proper result* of the experiment, and the associated standards for deciding if it has been properly performed" (*ibid*.:40, emphasis in original).

Deficiencies in recordkeeping may impact to varying degrees on the potential purposes that can subsequently be served by laboratory records. Steinle (2003), for example, highlights the less than systematic approach taken by André-Marie Ampère[26] in his recordkeeping

[25] The elementary charge is the electrical charge carried by a single proton. The experiment that was designed by Millikan and Fletcher to measure this charge essentially involved suspending an oil drop in the air by balancing the gravitational force acting upon it against an electromagnetic force generated between two electrodes.

[26] André-Marie Ampère (1775-1836) was an eminent French physicist and mathematician, who laid the foundations for the study of electrodynamics (electromagnetism). James Hofmann's

including undated entries and limited, sparsely documented experiments. In particular, Steinle (*ibid*.) encountered significant difficulties when attempting to reconstruct Ampère's work retrospectively from Ampère's original records. Nevertheless, it is clear that Ampère himself remained able to interpret his own records in order to publish his associated findings on electrodynamics. Tabard *et al.* (2008) identify problems encountered by laboratory staff when attempting to locate previously recorded experimental procedures to the extent that it sometimes proved more economical to redo work from scratch rather than continue to search for existing records.

Kevles (1998) discusses the impact of a more recent example of "haphazard" recordkeeping in relation to the infamous Baltimore case where poor recordkeeping practice on the part of an individual played a central role in an allegation of scientific fraud, a political inquiry into scientific practice, and personal and professional implications for multiple laboratory staff. It is interesting to note that the Baltimore case emphasizes the evidential value of laboratory records far beyond the more common purpose of evidence for intellectual property rights, involving as it did detailed forensic examination of laboratory records as part of a United States Congressional investigation "on the subject of fraud in federally funded biomedical research" driven by a concern that "scientific corruption appeared to be going unpunished and research institutions were covering up to protect themselves" (*ibid*.:136-138).

2.3 Recordkeeping in other settings

Records are, of course, not solely the concern of laboratory science. On the contrary, records play a central role in a wide range of settings including the bureaucracy of government, commercial enterprise, and public sector institutions. Recordkeeping has accordingly been investigated as a central discursive practice in a range of professional settings including the legal profession (Badger 2003; Bhatia 2008), accountancy practice (Coffey 1993, Flowerdew and Wan 2010), veterinary colleges (Schryer 1993), social work departments (Cicourel 1968; Paré 2004; Shemmings 1991), secondary schools (Cullingford and Swift 2002; Woods 1979), and a wide range of health care settings (Anderson *et al.* 2008; Berkenkotter 2008; Heath and Luff 1996; Pettinari 1988; Rooksby *et al.* 2007; Timmermans and Berg 2003). These studies offer insights into multiple aspects of

biography (ISBN 0521566703) provides an informative account of Ampère's life both in and out of science.

recordkeeping in the discourse of professional and academic settings, both for paper-based recordkeeping and electronic recordkeeping. Recurrent issues highlighted by these and similar studies include, *inter alia*, the problem of mismatched expectations between readers and writers of records, the role of contextualization in constructing meaning from written forms of communication, and the tensions involved in instituting implicit or explicit standardization of recordkeeping practices.

Heath and Luff (1996) report on an attempt to replace the paper-based medical record cards used by UK general practitioners (GPs) during patient consultations with an electronic recordkeeping system. Adoption of the electronic recordkeeping system was limited, and instead most GPs continued to use the paper-based system during consultations whilst restricting use of the electronic system to a final report of the outcome of the consultation. Heath and Luff (*ibid*.:363) argue that the continued use of the paper-based system in preference to the electronic system derived from the fact that "by attempting to improve the record and formalise its contents, the system disregards the competencies and skills that general practitioners rely upon in assembling and interpreting the contents of the record in actual consultations". Examples of these competencies and skills include the use of descriptive economies within the records, the ability to infer additional information based on context dependencies between entries, and a reliance on implicit inferences that can be drawn from properties of the written record such as the layout or handwriting. For example, recognizing which GP had authored a record based on the handwriting could lead the GP reading a record to draw inferences based on *a priori* knowledge of the terminology, reporting style, or attitudes adopted by that specific author. The core problem is the partial characterization of "the record as a disembodied, retrospective account of the consultation, rather than an integral feature of the accomplishment of diagnostic and prognostic activities" (*ibid*.:363) In short, the evidential value of the record has been privileged over the informational value of the record (*cf*. Garfinkel 1967:186-207; Nygren *et al.* 1992).

Pettinari (1988) reports on the records kept by surgeons during operations, and emphasizes the central role played by these records in organizing the work carried out by the surgeons. In this sense, the records are not only shaped by the work that is carried out, but also participate in shaping that work. An important consequence in this respect is that learning to write competent records is seen as a central component of the training of junior surgeons, and competency with the genre is acquired over time. Schryer (1993), in her

influential work on records as genre, reports similar findings with regard to the role played by the Problem-Oriented Veterinary Medical Record (POVMR) within the training and practice of veterinary students. The sections of the POVMR embody not only a workflow but also a reasoning strategy to be used by students and practising veterinarians in diagnosing clinical problems. At the time of her study, Schryer (*ibid*.) reports that the POVMR had only recently been adopted as the new policy for recordkeeping within the college that formed the focus for her study. Evidence of the mutual dependency between record and process can also be found in the fact that some members of the college opted to transfer away as they did not agree with the new process that was being instituted by adopting these new recordkeeping practices.

Berkenkotter (2008) reports on the changing role of case history records within psychiatry over an extended history from the eighteenth century onwards covering the transition from the "the asylum age" to "the biomedical age". This diachronic study also emphasizes the co-constructive nature of records and practice by clearly mapping the changes over time in the organization, rhetoric, and linguistic features of the case history genre in terms of the rise and fall of a narrative style of reporting in response to the rise and fall of psychoanalysis as the dominant paradigm within psychiatry.

The complexity involved in instituting standardized recordkeeping to support a "gold standard", evidence-based approach to medicine within the Netherlands is examined by Timmermans and Berg (2003). This informative study highlights the problem of aligning multiple purposes in a single record, illustrated in this case by the need to support the diverse communicative purposes of hospital administrators, clinicians, and medical insurance representatives. Each of these various stakeholders require different levels of accountability and support for different types of business functions, which require in turn different forms of organization and linguistic features within the record. As Timmermans and Berg (*ibid*.) point out, these competing needs may prove impossible to reconcile. Anderson *et al*. (2008) report similar findings in their attempt to define a record to support the varying needs of different types of clinician within a hospital setting in terms of supporting multiple purposes, *viz*. diagnosis, treatment planning, and treatment provision.

It is, of course, important to recognize that recordkeeping within academic and professional settings is subject to different cultural and external influences. Shankar's (2007:1463) bold statement that "academic freedom, a certain amount of institutional autonomy, and science's privileged role in knowledge creation give scientists great leeway

and power to keep records in ways that suit them, without needing to address greater organizational mandates" may or may not hold true.

2.4 Technology for laboratory recordkeeping

Laboratory recordkeeping, including recordkeeping in academic bioscience laboratories, has recently become a focus for technology development. To date, the traditional approach to recordkeeping in academic science laboratories has centred on the use of pen and bound, paper notebooks. This approach to recordkeeping is reported, for example, in studies of laboratory work such as those by Knorr Cetina (1999), Latour and Woolgar (1986), Shankar (2007), Traweek (1988), and Wickman (2010). A perceived drawback of this pen and paper approach is that it potentially limits the ability of bioscientists to share information with colleagues in both the local and wider scientific communities. This limitation may impact both on individual scientists and on the laboratory as a whole by reducing opportunities for collaborative work, increasing the likelihood of 'reinventing the wheel', and minimizing the economic return on investment (Fry *et al.* 2008; Lysakowski 1997). Recent developments in computing technology for the bioscience laboratory have set out to address this aspect of laboratory work by providing tools for electronic recordkeeping together with tools for improved information exchange between laboratory staff.

2.4.1 Design features of ELNs

Taylor (2006) provides a recent survey of tools for electronic recordkeeping. These tools are referred to either as laboratory information management systems (LIMSs) or as electronic laboratory notebooks (ELNs). LIMSs are typically characterized as database-driven systems capturing highly structured data through rigid user interfaces; ELNs operate in a similar manner but offer more flexible user interfaces and more flexible data recording capabilities with support for a degree of mobile use. As Taylor (*ibid*.) points out, ELNs and LIMSs represent different points on a continuum of integrated tool support for the laboratory environment, and these two are likely to merge over time into a single, preferred form of integrated product. In light of this, no further distinction is made between LIMSs and ELNs within this thesis, and the term ELN is used to stand for this type of system.

A number of ELN developments have been commercial ventures aimed primarily, but not exclusively, at industrial bioscience laboratories. Examples of these include CERF from Rescentris Inc and eCAT from Axiope Ltd[27]. Other ELN developments have been influenced by national research programmes such as the UK e-Science[28] programme that aim to facilitate scientific research through the appropriate use of information technology. Examples of research-orientated projects that have investigated the design and use of prototype ELNs in bioscience settings include ButterfleyNet (Yeh *et al.* 2006), SmartTea/CombeChem (schraefel *et al.* 2004), Labscape (Arnstein *et al.* 2002), and Prism (Tabard *et al.* 2008).

schraefel *et al.* (2004) propose two dimensions along which to categorize ELN systems. The first dimension evaluates whether the data recording solution is personal to the scientist such as a notebook, or whether it is a communal resource using some form of social media such as wikis[29] or other Web-based approaches. The second dimension evaluates the degree to which paper is retained as part of the data recording solution. schraefel *et al.* (*ibid*.) classify the degree to which paper is retained as part of the data recording solution into four discrete levels, *viz*. replication, augmentation, supplementation, and replacement. Replication systems are simple repositories of scanned copies, perhaps with metadata to enable subsequent indexing, of the original records written by laboratory scientists in paper notebooks. Augmented systems use devices such as digitizing tablets[30] to enable laboratory scientists to interact with the paper notebook whilst simultaneously generating digital copies of the laboratory records that have been made. Supplementation systems enable some of the content of the laboratory record, commonly the sequence of steps in an experimental protocol, to be captured digitally but require paper notebooks to capture the remaining content. Replacement systems replace the use of paper notebooks with a computer system into which the laboratory records are directly entered.

---

[27] Further information on these ELNs including an indication of both the target markets and the product features is available from the vendors respectively at http://www.rescentris.com and http://www.axiope.com [both accessed 01 March 2011].

[28] Further details on the UK e-Science programme can be found at http://www.nesc.ac.uk [accessed 01 March 2011].

[29] A wiki is a collaborative form of website that enables multiple users not only to view but also to edit, delete, or modify the information content that is placed on the website. Ward Cunningham, the developer of the first wiki software describes it as “the simplest online database that could possibly work”. Further information is available at http://www.wiki.org [accessed 01 March 2011].

[30] A digitizing tablet, or graphics tablet, is a computer device that captures a digital image of the text or graphic sketched by a user on top of the pad using an attached pen.

State-of-the-art, commercial products such as CERF and eCAT would largely be categorized under schraefel *et al.*'s (*ibid*.) scheme as communal resource, replacement systems since they are designed around a Web-based solution in the form of a networked database that holds all laboratory records produced by the scientists working in a group. Laboratory records, which may typically contain both text and graphics, are entered and viewed via computer consoles including both desktop and portable computers. Form-based interfaces can be used to control the content that must be captured as part of the records. In some cases, a degree of automated data capture is possible due to interoperable interfaces to specific laboratory devices. Electronic signatures are in place to authenticate the resultant records. It is important to note, however, that not all commercial products operate at this level, and SCRIP-SAFE® from SCRIP-SAFE International Inc[31] is an example of a commercial product that would be classified under schraefel *et al.*'s (*ibid*.) scheme as a communal resource, replication system since it provides corporate security paper for use in retrospective scanning of handwritten laboratory records. In this sense, SCRIP-SAFE provides a basic type of records management system functionality by providing a "manuscript repository" (*e.g.* Robek *et al.* 1995).

Research-orientated projects have developed a range of prototype ELN systems to focus on diverse aspects of electronic recordkeeping in the context of chemistry, physiology, and biology laboratories. The majority of these projects are "demonstrator projects, which develop applications within a specific domain" in order to " 'explore the potential' of e-Science" and "to show what the potential pay-off of new large-scale computing facilities and data infrastructures might be." (Hine 2005:1). Given the range of research foci set for these projects, it is perhaps unsurprising that these research-orientated systems have been widely distributed across the dimensions of schrafel *et al.* (*ibid*.) classification.

Specific aspects of laboratory recordkeeping that have been investigated include the use of novel interaction techniques including augmented reality systems to facilitate user interaction in ELNs. This is exemplified by Borriello's (2006) use of a projection system to project information such as experimental protocols onto the laboratory bench, and by Mackay *et al.*'s (2002) use of a digitizer tablet to capture handwritten records. Some of the research-orientated ELNs (schrafel *et al.* 2004; Mackay *et al.* 2002; Tabard *et al.* 2008) have been designed as hybrid paper and electronic systems in order to retain the

[31] Product information is available from SCRIP-SAFE International Inc at http://www.scrip-safe.com [accessed 01 March 2011]. All trademarks are acknowledged.

affordances offered by paper in terms of its mobility, inexpensiveness, and ease of use (Sellen and Harper 2003). This hybrid approach combining paper and electronic documents has previously been used in non-laboratory systems as in the DigitalDesk (Wellner 1993). Systems such as Labscape (Arnstein *et al.* 2002) and ButterlyNet (Yeh *et al.* 2006) have investigated the use of distributed, mobile systems incorporating handheld devices such as personal data assistants[32], digital cameras and digital pens in order to support the mobile nature of laboratory work. Labscape is designed for work at the laboratory bench, whilst ButterflyNet is designed to support the work of field biologists. SmartTea/CombeChem (Frey *et al.* 2004) and Labscape (Arnstein *et al.* 2002) have both investigated the use of integrated experimental planning and recording tools using workflow models to structure the sequence and content of experimental records.

Prism (Tabard *et al.* 2008) and the BRIDGE workspace (Farooq *et al.* 2005) have begun to expand the role of ELN prototype beyond record capture to investigate the development of tools to encourage collaboration and creativity among laboratory communities including the collaborative evaluation of experimental results. OpenWetWare[33] is an international research community initiative that employs social media in an attempt to promote the sharing of expertise and information between research groups engaged in biology and biological engineering. Another important development in tools to facilitate and expand the role of collaboration between bioscientists is the Biological Services Collaboratory (BSC) (Chin *et al.* 2002; Chin and Lansing 2004). The BSC is an ongoing research project to promote the development of a collaboratory (Wulf 1993) for biology. A collaboratory in this sense is an extended environment offering shared tools and data sets to support collaboration between biology researchers irrespective of their location. Chin and Lansing's work is notable in that it highlights the need to understand diverse contexts of use of experimental data across heterogeneous laboratory environments in order to enable effective data-centric collaboration between laboratory scientists. Standardized representations for both the experimental data and any supporting metadata will be essential to enable a shift the focus away from tool-centric collaboratories towards data-centric collaboratories.

---

[32] A personal data assistant (PDA) is a handheld computer that is typically equipped with some form of Internet access, a mini keyboard or touch screen, and a small display screen.

[33] Further information on the OpenWetWare community is available from http://www.openwetware.org [accessed 01 March 2011].

2.4.2 Influence of standardization efforts

As with many domains, the drive towards standardization of the processes and data involved in molecular biology may encounter limits in the degree of determinism that can be achieved. This possibility is succinctly expressed by Searls (1998) who comments that "in biology, there are no rules without exception" made in reference to the grand challenges in computational biology. Nevertheless, collaboration and data sharing requires agreement to some degree between the parties involved as to the structure, content, and representations used for data, whether this be in the form of *de facto* or *de jure* standards (*e.g.* Bannon and Bødker 1997; Timmermans and Berg 2003).

Given the increased attention being paid to collaboration and data sharing, formal standardization efforts are continuing within both the molecular biology community and the wider bioscience community in order to facilitate collaboration and data exchange between scientists. This includes terminological standardization to establish a shared vocabulary for use across bioscience disciplines as in the ongoing work of the Gene Ontology Consortium (2000), information standardization to prescribe formats for use in exchanging specific types of data as in the MIAME format (Brazma *et al.* 2001) used to define the minimum information necessary to describe a microarray experiment so that it can be interpreted unambiguously[34], and procedural standardization to formalize experimental workflow (*e.g.* Conery *et al*. 2005). These types of standardization activity may impact on the laboratory recordkeeping practices at work in academic molecular biology laboratories.

It is important to note that these standardization efforts have originated from two camps. Some efforts have been driven by the need to achieve interoperability, construed in broad terms, as part of the development of computing systems. For example, under the auspices of the e-Science project for the Semantic Grid (De Roure *et al.* 2001, 2005), the CombeChem project has proposed an ontology that encompasses both protocols (referred to in their terminology as processes) and the materials involved in these protocols for use in the chemistry laboratory (Taylor *et al.* 2006). The development of the CombeChem

[34] A microarray is a relatively new technology used in molecular biology and related bioscience disciplines. Microarray technology enables multiple genetic tests to be performed in parallel so that a set of samples can be tested against multiple probes on a single array. For example, Affymetrix Inc's GeneChip® Human Mapping 500K Array, which is one of a number of commercially available microarrays, is capable of approximately 250,000 genotyping tests per microarray. Further information on MIAME is available at http://www.mged.org [accessed 01 March 2011]. All trademarks are acknowledged.

ontology has been motivated by the key requirement identified by chemists working on the e-Science project to support 'Publication at Source' of all aspects of their data including protocol definitions (Frey *et al.* 2004). This concept of 'Publication at Source' has been adopted in order to promote increased sharing of data through instant availability. Other efforts have originated from within the life sciences community itself, notably in the establishment of the Gene Ontology Consortium (GOC) that manages the definition of a number of integrated ontologies for representing biological knowledge. The primary focus of this consortium to date has been on the definition of biological terminology or vocabulary and not in the standardization of laboratory protocols. Nevertheless, it serves as a useful basis for the standardization of laboratory protocols within other projects as typified by the development of data integration and data mining techniques in (Smith *et al.* 2004).

Instituting standardized practices is a complex issue, and studies such as those by Anderson *et al.* (2008), Ellingsen and Monteiro (2003), and Timmermans and Berg (2003) have reported problems encountered during the introduction of computer systems for healthcare recordkeeping in situations where the working practices embodied in the computer systems attempt to enforce standardization over the existing diverse, local practices.

2.4.3 Influence of information sharing

Achieving the transition from paper to electronic recordkeeping has proven problematic in the context of academic bioscience laboratories, and surveys of electronic recordkeeping (Nature 2005, 2007; Taylor 2006) report that the uptake of ELNs remains markedly low in the academic research sector. Some of the benefits typically associated with automated recordkeeping using ELNs include controlled archiving of records, automated date stamping of records, and standardization of work processes across multiple staff. Whilst these features may accrue sufficient value to laboratories in the industrial sector to warrant adopting electronic recordkeeping, it is possible that the more individualistic nature of academic bioscience research may require other features from ELNs in order to convince academic researchers of the merits of this technology. However, the enhanced opportunity for collaboration and increased sharing of data appear to also be directly of benefit to those at work in academic laboratories.

Recent surveys of data sharing such as those by Borgman *et al.* (2007), David (2006), and the Research Information Network (2008, 2010) indicate, however, that the sharing of multiple types of research data between scientists in academia is highly constrained by a complex of "interrelated factors such as lack of demand, lack of standards, and concerns about publication, ownership, data quality, and ethics" (Borgman *et al.* 2007:17). These findings contrast with the more bullish projections about electronic recordkeeping reported by some technology-orientated research. The Research Information Network[35] (2010:9) summarizes the complexity of this situation in the following terms:

> "The focus on these new developments brings the risk that they may be taken as representative of life sciences as a whole, or as ideal types around which future life science research will converge. Our case studies seek to provide a broader evidence base about information practices across life science research; and they provide an important corrective to this kind of vision. In contrast to current discussions of transformation, they reveal a more uneven pattern, as life scientists individually and in their research groups grapple with the changing 'affordances' of emerging information tools and services available for their diverse activities."

Social studies of the sciences suggest that the adoption of new technologies in the laboratory environment may change the distinctive working culture of the laboratory. In light of the increased emphasis on standards and tools for sharing experimental data, studies of the biosciences such as those by Brown (2003), Cragin and Shankar (2006), Hine (2006), Lenoir (1998), and Nentwich (2004) have re-evaluated disciplines such as molecular biology from the viewpoint of an information science as opposed to the traditional viewpoint of a laboratory science, by reporting on the impact of community databases for the biosciences at different scales. Lenoir (1988) argues that the availability of large-scale genomic databases combined with the centrality of genome sequencing to biological research has re-shaped disciplines such as molecular biology in the form of an information science. Hine (2006:293), in her study of the working practices surrounding the setting up and use of a smaller scale database of mouse genomic data, offers a counterpoint to this "cyberbolic" viewpoint on the basis that "the database and the laboratory can therefore co-exist as different frameworks for organizing action, without one necessarily threatening the other." Much of the focus of these studies has been on the

---

[35] The Research Information Network (RIN) is a policy unit funded by the UK national higher education funding councils, the UK research councils including the Biotechnology and Biological Sciences Research Council (BBSRC) and the Medical Research Council (MRC), and the UK national libraries to investigate and support the use of information and knowledge resources within UK research. Further information is available at http://www.rin.ac.uk [accessed 01 March 2011].

practices surrounding community data pools such as those generated by the Human Genome Project (Watson 1990), and for which the roles of information producer and information consumer are relatively demarcated in space and time. Investigating the potentially different spatial and temporal patterns of use relating to the experimental records that are captured on a daily basis in laboratory notebooks will enable a comparison of these two facets of information exchange.

Bowker (2000) and Hine (2005) emphasize that the development and use of technology for data sharing in e-Science settings is influenced not only by technical considerations but also by political and policy considerations in terms of shaping both the community and its future actions. As Hine (*ibid*.:4) points out, an awareness of the policy situation and the institutional context must be present both " to promote design of infrastructures which accommodate the diversity of practice, and to occasion reflection on the prospects for success and potential consequences of the overall endeavour". Nentwich (2004) suggests that the typical factors affecting the adoption within scientific fields of information and communication technologies, of which ELNs are an example, include the size of the field, the pressures put on the members of that field to publish their work, the degree of collaboration that exists between members, and the existing communication conventions at work in the field. Hine (*ibid*.) reinforces the situated nature of scientific practice, and argues that it is also necessary to take into account the dynamic nature of work experienced by individuals in specific settings. Birnholtz and Bietz (2003), similarly, emphasize that sharing data has raised social problems for scientists in establishing communities of practice on the basis that data is seen as the prime economic resource for scientists resulting in an unwillingness to share, that tracking down the location of the required data is not a straightforward task, and that most data requires a significant degree of additional context to be able to make use of it but such contextual meta-data is typically unavailable at present.

2.5 Genres in academic science

Genre has proven to be an informative lens through which to investigate linguistic behaviour in diverse settings, confirming Candlin's (1993:ix) observation that it is "a concept that has found its time". Within academic settings, including the molecular biology laboratories that form the referent domain for this thesis, a wide range of genres has previously been studied encompassing research and administrative genres, written and

spoken genres, undergraduate and postgraduate genres, and occluded and public genres. Occluded genres, as defined by Swales (1996), are genres with a restricted profile that are not typically visible to apprentices in the community or to outsiders[36].

Specific genres that have been previously investigated include research articles (Bazerman 1988; Kanoksilapatham 2005; Myers 1990, 1991; Swales 2004; Tarone *et al*. 1998), popular press articles (Corbett 1992, 2009; Myers 1990), laboratory reports (Braine 1995; Dudley-Evans 1985), PhD theses (Bunton 2002; Dong 1998; Dudley-Evans 1991; Hyland 2004), PhD vivas (Grimshaw 1989; Maingueneau 2002), conference talks (Dubois 1987; Rowley-Jolivet 2002; Ventola *et al*. 2002), grant applications (Cadman, 2002; Myers 1990), and textbooks (Hyland 2000; Love 2002). Aspects of linguistic behaviour that have been highlighted by these studies of academic genres include forms of argumentation, patterns of citation, the cognitive organization of texts, lexical patterns used within texts, and approaches to audience engagement, where each of these aspects of linguistic behaviour is constructed not only as a textual feature but also as a response to the social context in which the texts are produced.

Ebel *et al.* (2004:16) stress the fact that laboratory records such as those captured by bioscientists in their laboratory notebooks act as "the 'germ cell' of the scientific literature". This metaphor, highly appropriate for a study of molecular biology settings, reinforces the fact that laboratory records interact with a number of other research and administrative genres in a complex of hierarchies, networks, chains, and sets within the discourse of academic molecular biology laboratories. In this sense, understanding issues relevant to the production and interpretation of other academic genres will prove informative for this study of laboratory recordkeeping. The two genres of the research article and the laboratory report are considered to be of particular relevance to the laboratory record on the basis that research articles and other published reports are distilled versions of the body of experimental work that has been carried out as discussed earlier in section 2.2.6.

[36] Swales uses the submission letter written by applicants to academic courses as an example of an occluded genre in the paper containing the original definition.

2.5.1 Research articles

Research articles play a central role in the discourse of academic disciplines both as the preeminent means of knowledge-making within academic disciplines, and as the primary capital used by individual researchers to advance their profile and career (Biagioli 2003; Latour and Woolgar 1986; Myers 1990; Swales 2004). In light of this position, it is perhaps unsurprising that the research article is the most widely researched of academic genres. This has included studies of the evolution of the genre, studies of the organization and linguistic features of articles as a whole, and studies of the structural units within an article such as the abstract (Lorés 2004; Samraj 2005), introduction (Swales 1990; Samraj 2005), and acknowledgements (Hyland 2004).

The genre of the research article has undergone significant evolution during the development of empirical scientific research from the seventeenth century onwards, leading to the now familiar and relatively stable introduction-methods-results-discussion (IMRD) structure of research articles that has become a pervasive model for reporting experimental work. The four moves in this cognitive structure address, in turn, what is the research question under study, how is the question to be studied, what were the findings of the study, and what do these findings mean. In addition to experimental research articles, the IMRD structure has also influenced other research-orientated genres such as the PhD thesis (Bunton 2002; Dudley-Evans 1991; Thompson 2001) and the laboratory report (Braine 1995; Dudley-Evans 1985; Lobban and Schefter 1992).

Bazerman's (1988) diachronic analysis of articles in the Philosophical Transactions of the Royal Society traces key phases in the socio-rhetorical shaping of experimental articles over time moving from the relatively uncontested reports of events prevalent in the mid to late seventeenth century, through stages of increased discussion over results, leading to articles that put forward theoretical claims on the basis of experimental proofs from the early ninteenth century article onwards. Bazerman (*ibid.*) contends that the need to communicate with an expanding body of scientists in the face of increasing levels of argumentation led to scientists using increasingly persuasive forms of language in experimental articles and drawing more heavily on the growing body of published literature in order to project precision and completeness of results. Gross *et al.* (2002), in their case studies of experimental articles written in English, French, and German between the seventeenth and twentieth centuries, report a similar shift away from an author-centred

discourse that is largely narrative in style towards a discourse representing science as an objective enterprise. Of particular relevance to this thesis, Gross *et al.* (*ibid.*) base their findings in part on the analysis of biology articles including a detailed comparative analysis of two articles written respectively in the seventeenth century and the twentieth century[37]. Valle (1999) reports corroborative findings based on a diachronic examination of experimental articles in the life sciences in publications made by the Royal Society. These studies are highly significant in that they identify the mutual dependency through which the laboratory record and the research article have been shaped.

Experimental reports are, however, not the only form of research article, nor are they the only form of research article at work in molecular biology settings. Tarone *et al.* (1998:115), in their work on astrophysics settings, point out that experimentation is replaced by reason of necessity in some situations by more theoretical expositions on the basis that " one cannot experiment on a star or galaxy in the way in which one can experiment on a chemical compound or bean plant", and consequently the IMRD structure of research articles is replaced by a top-down, logic-driven narrative form of argumentation (*cf.* earlier forms of experimental research articles). This form of argumentation is common in work based on mathematical modelling and simulation (Silver 2006; Swales 2004:207), and so is likely to influence research articles reporting the *in silico*[38] experiments that are playing an increasingly prominent role in molecular biology.

Another subgenre of research article that is already evident in molecular biology laboratories is that of the review article (Myers 1991; Noguchi 2006). In contrast to experimental research articles, however, these review articles do not present original research but set out with the purpose of presenting an overview of the literature pertaining to some aspect of the field of study. In this respect, they are substantially removed from the laboratory records, and it is questionable whether the information in the review article is distilled from the data held in the author's laboratory notebook.

---

[37] The seventeenth century article was published in the Philosophical Transactions of the Royal Society, and reports work carried out by Martin Lister on the "Nature and Differences of Juices, More Particularly, of our English Vegetables". The twentieth century article was published in the Proceedings of the National Academy of Sciences, and reports work carried out by Goodman and Rich on the "Formation of a DNA-Soluble RNA Hybrid and its Relation to the Origin, Evolution and Degeneracy of Soluble RNA".

[38] *in silico* experimentation refers to experiments conducted via computer models or computer simulations. This phrase is defined in opposition to *in vivo* experiments that are conducted in living organisms, and *in vitro* experiments that are conducted in a controlled environment outside living organisms.

The shift in the experimental article away from a largely narrative, author-centred discourse towards a discourse representing science as an objective enterprise is realized in part through the use of specific linguistic features. Gross *et al.* (2002:230) offer a useful summary of the findings from recent genre studies of experimental articles in their characterization of "scientific English" as "the international discourse of science, which involves not only a specific language but also a suite of stylistic features: relatively short, syntactically simple sentences containing complex noun phrase with multiple modification, verbs in the passive voice, noun strings, technical abbreviations, quantitative expressions and equations, and citational traces". The specialized nature of this use of language is driven by a concept of the intended readership being "almost exclusively other professionals engaged in similar research" (*ibid.*). Given the interrelationship between laboratory records and research articles, some of these stylistic features are also likely to be found in laboratory records. Whilst both genres are similarly intended for a technical readership, it may be the case that laboratory records are restricted to a readership drawn exclusively from the same research group, same laboratory, or close collaborators. The level of contextual knowledge that can be assumed in such a group differs from that which can be assumed of the wider scientific community.

Myers (1990) reports on an informative study describing the protracted series of interactions between the authors of two biology papers and the external reviewers/referees[39] as the authors attempt to have their papers accepted for publication in journals. This study highlights the social nature of the construction of research articles through an extended process involving multiple drafts and, in a real sense, multiple authors (*cf.* Berkenkotter and Huckin 1995). In each of the cases reported by Myers (*ibid.*), the authors submitted their articles to multiple journals over a period of time, and reworked aspects of their articles through a process of negotiation driven by the comments received from multiple referees. This process of negotiation involved modification of not only the form and style of the article but also modification of the content in terms of the claims that could be made based on the results obtained. This process reflects Knorr Cetina's (1981:106) observation that "the published paper is a multilayered hybrid co-produced by the authors and by members of the audience to which it is directed." The construction of laboratory records is, by contrast, instantaneous and the work of a single author. Laboratory guides (*e.g.* Barker 2005; Ebel *et al.* 2004; Kanare 1985) recommend, and in

[39] Myers expresses considerable empathy with journal referees on the basis that they are abused almost as much as football referees.

some cases mandate, that laboratory records be constructed concurrently with the work that is carried out at the bench, and that these records must not be altered although corrections can be made through addenda. The issue of how the intended audience of the laboratory record is engaged in its construction remains open to investigation.

2.5.2 Laboratory reports

The genre of the laboratory report offers, at least in one sense, a more direct comparison with laboratory records in that it is a direct report of a discrete package of experimental work. It is, however, essential to recognize that the laboratory report primarily serves a pedagogical function, and so is intended for a very different readership to that of laboratory records. As Dudley-Evans (1985) points out, a laboratory report is typically written by undergraduate students of the applied sciences such as the biosciences in order to enable supervisors to assess both the work that has been conducted and the form used to communicate that work. Accordingly, the laboratory report constitutes part of the training offered to scientists including bioscientists in how to maintain a laboratory notebook.

Lobban and Schefter (1992:4) identify the significant role played by the IMRD structure in laboratory reports by explicitly identifying the intertextual relationship to research articles. In particular, Lobban and Schefter (*ibid*.) advise the use of the IMRD structure on the basis that it is "the plan of a scientific paper and so of your report", whilst explicitly stating that the laboratory report remains a different genre by emphasizing that "while a lab report must have the structure of a scientific paper, it has a different audience and purpose" in that it is directed towards course supervisors rather than the wider scientific community. Dudley-Evans (1985:2) offers a similar, if slightly expanded, structure for the organization of laboratory reports in terms of "abstract, aim and objectives, introduction/theoretical background, equipment and materials, procedure, results, discussion of results, and conclusion."

Braine (1995), in his comparative study of laboratory reports across engineering disciplines, identifies considerable variation in the structure of the reports appropriate to each discipline. Specific disciplines adapted the broad organizational structure referred to by Dudley-Evans (*ibid*.) by not requiring abstracts, by requiring a description of experimental apparatus as part of the equipment and materials, or by requiring a consideration of potential hazards as part of the procedure. Indicative of the pedagogical purpose served by the laboratory report, some disciplines also required a description of the

theory on which the experiment was based. This is an important distinction to draw as laboratory records in the day-to-day work of molecular biologists are, of course, not pedagogical devices but are a core tool of scientific investigation.

2.6 Variation in genres

A number of earlier discourse studies focused on understanding the similarities between texts in order to identify common linguistic features that could be used to construct competent, persuasive texts across multiple academic settings, on the basis that "scientific writing was taken to be the prototypical exemplar of academic discourse" (Hyland 2006b:17). More recently, however, there has been an increased emphasis on contrastive studies (*e.g.* Braine 1995; Bunton 2002; Samraj 2005; Hyland and Bondi 2006; Hyland and Tse 2007) that compare the texts produced in different settings and characterize the factors influencing any perceived variation across these settings. The principal aim of these contrastive studies has been to further an understanding of the ways in which different disciplinary communities and their members construct knowledge, form persuasive arguments, engage their readerships, and project their identities through discourse. In short, these studies aim to improve our understanding of specialized discourses such as what it means to communicate as a molecular biologist as opposed to a computer scientist.

The community-focused orientation to literacy inherent in these studies is consistent with Swales' (1990) mutually dependent characterization of discourse community and genre. This concept of community encompasses a range of contextual features that impact on the production and interpretation of texts. Hyland (2006b:19), for example, identifies "knowledge of a cultural and interpersonal situation, knowledge of interlocutors, knowledge of the world, and knowledge of texts and conventions for saying things" as contextual aspects of community knowledge. Samraj (2002) identifies different levels of contextual variable that could influence the textual features used in academic writing in a hierarchy ranging from contextual variables at the level of the academic institution, to the level of the discipline, and down to the level of the individual reader/writer. Canagarajah (2005), Prior (1998) and others have, however, problematized the concept of community as too static, rightly pointing out that communities are fluid, composed of individuals with diverse backgrounds, different goals, varying levels of expertise and different levels of engagement with notional values, and that most individuals will be members of multiple communities at any time.

Driven by these insights, contrastive studies have investigated potential variation in writing resulting not only from disciplinary differences, but also from differences in the personal context of the individuals involved. In the case of the discipline or field of inquiry, studies have involved comparisons between specific disciplines (*e.g.* comparing molecular biology and ecology) and comparisons between broad categories of discipline (*e.g.* comparing natural sciences with the humanities). In the case of the individual, studies have involved comparisons between texts written by native and non-native speakers, between readers/writers of different nationalities, between male and female readers/writers, and between novices and experts.

2.6.1 Sources of variation between disciplines

Awareness of the epistemic conventions at work in a disciplinary community is required by academic writers on the basis that members of a disciplinary community are "more likely to persuade readers of our ideas if we frame our messages in ways which appeal to appropriate community recognised relationships" Hyland (2006b:21). Previous studies have examined the ways in which a range of linguistic features are used in multiple genres to frame messages across a range of disciplines. The intention in this thesis, however, is not to provide a survey of inter-disciplinary differences across a full range of academic genres. Instead, by focusing on two specific features, *viz*. the foregrounding of claims in research article abstracts and patterns of citation, the intention is to highlight social factors that could influence laboratory recordkeeping in academic molecular biology laboratories.

Hyland (2006b) examines the issue of foregrounding as a means of positioning the claims made by a researcher in a research article in relation to the existing body of knowledge within a community of practice, arguing that scientists such as bioscientists tend to foreground the novelty of their research in contrast to engineers who stress the utility of their research, and in contrast to sociologists and applied linguists who set out to establish unresolved lines of inquiry within their discipline[40]. Melander *et al.* (1997), in their study of article abstracts from the United States and Sweden, report that articles on biology and medicine tend not to employ rhetorical devices to foreground their work. Samraj (2005),

[40] The author of this thesis derives great comfort from this observation given the modest degree of resolution that he has been able to bring to the research questions posed for this thesis. At least, the problem has been identified!

in her comparison of abstracts and introductions from articles on conservation biology[41] and wildlife biology[42], reports that conservation biologists tend to promote the value of their work whilst wildlife biologists tend instead to present pragmatic statements of the goals, method and results. Hyland (2000) examines variation in the move structures used in a corpus of abstracts drawn from multiple disciplines, and concludes that social scientists/writers in the humanities preferred to situate their research using an introduction whilst physicists/engineers preferred to omit the introduction but provided a description of their experimental method. An interesting outcome of Hyland's (*ibid*.:70) study is that "biologists once again fell between the two groups" of soft and hard knowledge disciplines.

It is important to recognize that these example studies report findings based on comparisons of increasingly specific disciplines ranging from a comparison of science *v*. engineering *v*. sociology, to a comparison of two sub-fields within ecology. Findings at higher levels of abstraction may not persist in sub-fields, as illustrated by the differences observed by Samraj (2005) with respect to the two sub-fields of ecology. Samraj (*ibid*.) posits three alternative explanations for the use of rhetoric in the conservation biology abstracts in contrast to the approach taken by wildlife biology. Firstly, conservation biology is a relatively new, emerging field in which members may not yet have established agreed lines and methods of inquiry. Secondly, conservation biology is a cross-disciplinary field in which specialists from one discipline may not be fully aware of the concerns and practices appropriate to other disciplines. Thirdly, conservation biology is an applied discipline, which may necessitate explicit justification of practical applications to real world issues. Although it remains under investigation as to which of these explanations, if any, holds true for conservation biology, it is important to note these potential influences may prove relevant to the discourse of the relatively new, inter-disciplinary discipline that is molecular biology.

Citation of existing work offers a means of integrating the claims being made by a scientist into the existing body of accredited scientific knowledge within a discipline. In this sense, citations can be used to acknowledge a debt to previous work, to align one's own work with a particular group or viewpoint, or to establish the credibility of the claims being

---

[41] Conservation biology is concerned with the scientific study of phenomena affecting the loss and restoration of biological diversity.

[42] Wildlife biology is concerned with the scientific study of the conservation and management practices for wildlife species and their environments.

made. Hyland (2000), for example, reports on a study of the citations present in a corpus of research articles drawn from multiple disciplines including biology, concluding that biology articles tended to include more citations per paper than engineering, physics, and other 'hard' sciences. Moreover, the citations used in biology papers showed a greater use of integral reporting structures. Integral reporting structures in this sense indicate citations that give prominence to the cited author. Hyland (2000:35, emphasis in original) speculates that the influence for this approach to citation rests on "a disciplinary ethos which emphasises proprietary rights to claims" on the basis that "constructing knowledge in biology seems to involve rhetorical practices that give greater weight to *who* originally stated the prior work, rather than the traditional conventions of impersonalisation still observed in the other hard disciplines." Hyland (*ibid*.) clearly states that this remains, however, speculation.

2.6.2 Sources of variation between individuals

One potential source of variation in the language use exhibited by an individual stems from whether the individual is writing in his/her native language. The English language has achieved a dominant position within the world of research and scholarship (*e.g.* Mühleisen 2003; Pennycook 1994; Swales 2004), and estimates of worldwide English usage (*e.g.* Crystal 2003; Graddol 1999) suggest that the number of people who have learned English as a second language (L2) outweighs the number of people who speak English as their first language (L1). It is important to recognize that L2 speakers of English are a heterogeneous group encompassing multiple levels of proficiency up to bilingual competence. This group can be further divided into speakers of English as a second language (ESL) as in territories such as India, Pakistan, and Singapore where English has an institutionalized role in governance and education, and speakers of English as a foreign language (EFL) as in other territories where English has no such official role (Jenkins 2009; Kachru 1992).

Many researchers, whether L1 speakers of English or not, elect to publish their work in international journals that require submissions to be written in English on the basis that these journals have a wider readership and greater prestige. Interestingly, the choice of non-native speakers (NNS) to write and publish in English applies not only to research articles (Ammon 2001; Wood 2001) but also to other genres traditionally associated with a more local circulation such as PhD theses (Berg *et al.* 2001; Vandenbroucke 1989), prompting Swales' (2004:35) comment that "the 'English-only' bandwagon thus continues

to roll." Phillipson and Skuttnab-Kangas (1999), *inter alia*, emphasize that whilst researchers in some fields such as history and education maintain a preference for their native language, researchers in the natural sciences such as the biosciences consider English to be the natural choice for publication on the basis that it is the international language of science.

Against this background of increasing internationalization and 'Englishization'[43] of research, Wood (2001:81) has problematized the traditional privileged view of the native speaker (NS) as the authoritative norm of language use through his concept of "International Scientific English" as the "variety of scientific English used by scientists around the world of any linguistic background". As Woods (*ibid*.) argues persuasively "the fact that some, or even most, of the members [of the discourse community] are native speakers of English is irrelevant. What constitutes grounds for membership in the community is an acceptance by members of that community of a scientist." Swales (2004:54), similarly, problematizes the distinction between native speakers and non-native speakers in studies of discourse communities, contending that "if someone whose first language is other than English succeeds in getting published in an English-medium journal or gets invited to speak at an English-medium conference, then that itself, I would think, is sufficient ratification for inclusion in any analysis." In place of the NS/NNS dichotomy, Swales (*ibid*.) characterizes the linguistic proficiency of researchers against a continuum ranging from junior researchers to senior researchers in which English language proficiency is subsumed as one determinant of experience together with other factors such as rhetorical competency and genre awareness.

Another potential source of variation in linguistic behaviour between individuals is gender. Tse and Hyland (2006) point out that although gender has been shown to influence some types of social interaction, it has received little attention with regard to variation in academic discourse. However, the few studies conducted to date have reported somewhat contradictory findings. Flynn (1988), for example, advocates the need to recognize gender forms of argumentation and representation in the pedagogy of university students within composition classes based on her survey of feminist research on gender differences in social and psychological development. In contrast, Robson *et al.* (2002), report that that there was greater similarity than difference in the writing styles of men and women within

[43] With due deference to Phillipson and Skuttnab-Kangas who introduce the word in their 1999 paper, I feel obliged to apologize for using this particular neologism as it seems to me to be one "-ization" too far.

a corpus of university student essays, although there was some difference in form of argumentation. It is important to note that neither of these two studies directly investigated writing in scientific disciplines, but examined largely narrative forms of essay writing in the humanities. It remains unclear as to whether and in what way gender might influence laboratory recordkeeping.

Finally, it is important to recognize that writing is also subject to idiosyncratic influences. As Valero-Garcés (1996:281), *inter alia*, points out "some features of scientific discourse are provided by the genre of the text, others by the culture they belong to and also by the writer's own style." Whilst genre conventions can account for much of the conformity across writers, it is necessary to recognize that differences between texts may derive as much from differences between individual writes as from cultural differences.

3 An Ethnography of Recordkeeping in the Laboratory

This chapter of the thesis reports on an ethnographic study of academic molecular biology laboratories. The chapter is presented broadly in line with the IMRD structure (*cf.* Thompson 2001) to address the aims of the study, the method used to conduct the study, the results obtained, and a discussion of these results.

3.1 Aim of the study

The primary aim of this ethnographic study was to investigate the socio-cultural and organizational influences that shape the role of laboratory records in the discourse of academic molecular biology laboratory settings. To this end, the study set out to examine the situational context in which laboratory workers produce, maintain, interpret, and archive their laboratory records, use these records to coordinate their work, and use these records to collaborate with others. A secondary aim of the study was to understand current recordkeeping practices within the laboratory settings in order to evaluate the potential for electronic recordkeeping systems to support these practices within academic molecular biology laboratories.

Particular attention was paid throughout the study to the ways in which laboratory scientists employ information and procedural standards in laboratory recordkeeping as both these types of standardization have been shown to influence, and be influenced by, the role of records in knowledge production and knowledge dissemination within other domains such as medical recordkeeping (Anderson *et al.* 2008; Berkenkotter 2008; Ellingsen and Monteiro 2003; Heath and Luff 1996; Nygren and Henriksson 1992; Timmermans and Berg 2003). In this respect, the study was concerned not only with explicit forms of standardization but also with implicit standardization through enculturation. The analytic lens of standards also provided an additional basis for systematically comparing the recordkeeping practices between different laboratories and between different laboratory members. Whilst the principal motivation for the attention paid to information and procedural standards derived from a desire to understand the interplay between standards and sharing in laboratory recordkeeping, it also offered a means to evaluate the potential for customization in the design of tools such as ELNs to support laboratory recordkeeping.

3.2 Design of the study

The study was designed as an ethnographic study in order to enable detailed, first-hand observation of recordkeeping practices situated within laboratory settings. As discussed in section 1.4.2, the perceived strengths of this observational approach include the ability to observe recordkeeping over an extended period of time, and the ability to observe the actual working practices of the domain as opposed to the prescribed working practices.

For the purposes of the study, both procedural and information standards were interpreted in a broad, inclusive manner. Procedural standards were taken to include task descriptions, guidelines, scripts, or recipes for any element of laboratory work of varying degrees of formality. Typical examples include the descriptions of experimental tasks found in core molecular biology textbooks such as Berger and Kimmel (1987) and Sambrook and Russell (2001), and regulatory plans used within laboratory work such as the Control of Substances Hazardous to Health forms (Health and Safety Executive 2005). Information standards were taken to be any description, either formal or informal, of the structure, content, or encoding of experimental data. Typical examples include bioscience community formats such as MIAME for exchanging microarray experimental data (Brazma *et al*. 2001), device-specific formats such as the results formats generated by commercial laboratory equipment, and in-house formats defined within a specific laboratory setting.

3.2.1 Ethical approval procedure

Since the study involved observation of human participants, it was reviewed and approved by the Ethics Committee of the Faculty of Information and Mathematical Sciences at the University of Glasgow under application number FIMS00327. This approval ensured that the study conformed to the code of conduct set out by the British Psychological Society (BPS) for studies involving human participants (BPS Ethics Committee 1978). All participants were approached and recruited to the study only after ethical approval had been confirmed. Introductory and debriefing sessions were conducted with each participant in accordance with an interview script. The information sheet sent to prospective participants to describe the study is presented in Appendix 1 of this thesis.

3.2.2 Sample cases

3.2.2.1 Selection policy

The selection of sample cases for this study of laboratory recordkeeping was driven by both strategic and practical considerations.

The key strategic consideration was to adopt a form of stratified purposeful sampling (*e.g.* Patton 2001:240) in order to enable observation of multiple categories within the sampled cases, thus improving the representativeness of the sample set and the scope for observing potential sources of variation. Observing multiple types of participant across multiple laboratories also addressed, to some extent, a criticism that is often rightly levelled at ethnography. This criticism is concerned with the findings from such studies being too localized due to the focus on specific settings, rendering them difficult to generalize to other situations and therefore of limited utility in potential areas of exploitation such as the design of computing tools. For this study, multiple laboratories were observed rather than a single laboratory, and both research-orientated and service-orientated laboratories were observed within the university environment. This approach was chosen in order to consider potential variation in working practices both between workers in the same laboratory, and between workers in different laboratories. Laboratories of different sizes were chosen in order to allow scope for comparison of laboratory workers operating in relative isolation with those operating in collaborative teams.

Scientists at different stages of their academic career were observed within each laboratory. Academic career stages were ascribed on the basis of the scientist's function as technician, postgraduate researcher, postdoctoral researcher or principal investigator. This approach was chosen in order to consider potential variation in the working practices between different stakeholder responsibilities, and between novice and experienced users in line with Swales' (2004) gradation of junior and senior researchers (see section 2.6.2).

Practical considerations for the study derived both from the difficulty experienced in gaining access to laboratories, and the potential costs involved in travelling between multiple sites. Initial attempts via both shared contacts and 'cold-calling' to gain access to laboratories at multiple universities across Central Scotland did not meet with success, leading to the decision to abandon the observation of laboratories in multiple universities in favour of a sole focus on laboratories within a single UK university.

3.2.2.2 Sample laboratories

Table 3-1 lists the laboratories that participated in this ethnographic study of laboratory recordkeeping.

Table 3-1: Laboratories for the ethnographic study

Laboratory	*Type*	*Description*
ES-L1	Research	A small university research laboratory with approximately 7 members. Members are involved in human genetics research for projects in the field of sports and exercise science with a specific focus on the interaction between environmental factors and hereditary factors on human health and performance.
ES-L2	Research	A large university research laboratory with approximately 20 members formed as a close collaboration of two principal investigators. Members are involved in integrative physiology research using *Drosophila melanogaster*[44] as a model organism for a range of projects including some projects with commercial partners.
ES-L3	Service	A common services department housed within a university facility but offering laboratory services and consultancy in sequencing and data analysis to multiple client laboratories within the home university, in other universities, and in other research institutions.
ES-L4	Research	A small university research laboratory with approximately 6 members. Members are involved in human genetics research in projects investigating the genetics of human disease with a specific focus on one disorder[45].

Summarizing the laboratories that participated in the ethnographic study in terms of the identifier code assigned to the laboratory, the laboratory type, and a brief description of the laboratory setting.
The *identifier code* uniquely identifies each laboratory participating in the study whilst maintaining the anonymity required under the terms of the ethical approval for the study. The *laboratory type* is used to categorize laboratories into either *Research* laboratories that undertake research projects on their own initiative in order to investigate scientific questions of their own choosing, or *Service* laboratories that are commissioned to undertake specialist experimental work on behalf of client laboratories. The *description* outlines the size and broad research interests of the laboratory.

[44] *Drosophila melanogaster* is a species of fly that is also known as the common fruit fly. The common fruit fly is widely used as model organism in biological research, and humankind should be very grateful to them given the number that are sacrificed for research purposes.

[45] Many laboratories focus their research effort on the mechanisms underlying a specific disorder or related types of disorder. Particular examples of research foci, not specifically associated with the laboratories participating in this study, include diabetes mellitus, idiopathic epilepsy, psychiatric disease, vascular disease, myotonic dystrophy type 1 (DM1), and Epstein-Barr virus (EBV) associated cancers.

All four of the participating laboratories were housed within a single UK university. It is important to note, however, that the service laboratory *ES-L3* delivered its services to a number of external clients located throughout the UK. Only one of the research laboratories, *viz*. *ES-L2*, made use of the services offered by the service laboratory *ES-L3*. None of the three research laboratories were involved in any collaborative projects with each other. The number of laboratory members in each laboratory relates to the date of the study, and is presented as an estimate to reflect the fluid nature of laboratory staffing that results from temporary staff such as project students and visiting researchers.

3.2.2.3 Sample participants

Table 3-2 lists the range of laboratory members that participated in the ethnographic study. As far as possible under the access conditions granted within each laboratory, scientists at different stages in their academic careers were observed within each laboratory. A more detailed summary of each individual participant including a brief description of the participant's experience in laboratory work is presented in Appendix 2 of this thesis.

Table 3-2: Participants for the ethnographic study

Laboratory	***Participants (N = 13)***							
	Job Function				**Native Language**		**Gender**	
	PI	**PD**	**PG**	**T**	**L1**	**L2**	**M**	**F**
ES-L1	1	2	0	1	4	0	4	0
ES-L2	1	2	1	2	4	2	3	3
ES-L3	0	0	0	2	1	1	0	2
ES-L4	1	0	0	0	1	0	1	0
All Labs	**3** **23.1%**	**4** **30.8%**	**1** **7.7%**	**5** **38.5%**	**10** **76.9%**	**3** **23.1%**	**8** **61.5%**	**5** **38.5%**

Showing a numerical breakdown of all participants in the ethnographic study by laboratory of origin, by function in that laboratory, by native language, and by gender.
The *laboratory of origin* is identified using the code number assigned for the ethnographic study (see Table 3-1). The participant's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The participant's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks English as a second language. The participant's gender is indicated using F for a female participant, and M for a male participant.

Multiple participants were observed in the research laboratories *ES-L1* and *ES-L2*, and in the service laboratory *ES-L3* as shown in Table 3-2. Research laboratory *ES-L4* was included in the study at a later date in order to gain access to the perspective of an additional principal investigator as a means of improving the representativeness of the sample set used in the study.

All participants who spoke English as a second language participated regularly in laboratory meetings held in English, presented their own work at internal and external seminars in English, produced written publications in English, and had completed university-level courses taught in English (see section 2.6.2). Both male and female scientists were recruited to the study from each laboratory approximately in line with the gender balance represented within that laboratory (see section 2.6.2).

3.2.3 Experimental procedure

3.2.3.1 Participant contact

All contact with study participants before and during the study was made directly with the participant and independently of the principal investigator. This procedure was followed in order to minimize the potential for bias on the basis that access to laboratories had been gained initially through contact with the principal investigators in each laboratory. In this sense, the principal investigators acted as "gatekeepers" (*e.g*. Hammersley and Atkinson 1995:34). Given that the focus of this study, laboratory recordkeeping, is an activity that is often considered to be regulatory in nature, perception of the relationship between the ethnographer and the principal investigator was considered to be critical to the success of the study. In particular, it was important to ensure that participants did not perceive the ethnographer as acting in any way on behalf of the principal investigator, *i.e.* as 'a spy in their midst'. In addition, it was important to remain aware of the fact that gatekeepers typically have an interest in promoting a positive image of their organization and so may push or block the ethnographer in specific directions.

Prior to observing work in each laboratory, one-to-one meetings were arranged with each individual participant to explain the purpose of the study, to gain written consent, and to arrange convenient dates on which to observe the work of the participant.

3.2.3.2 Observation of laboratory practice

During the course of the study, the participating laboratory workers were accompanied and observed as they went about their daily work. This included observation of their work in the office areas, in the computing areas, and in the 'wet'[46] laboratories where bench work experimentation is carried out. Due to health and safety considerations, no work was observed in the 'hot' laboratories in which experimentation involving radioactivity is conducted.

Subject to the access granted for the study, observations were made in each of the participating laboratories except laboratory *ES-L4* over a period of approximately ten working days. The majority of this time was spent shadowing the individual laboratory members who participated in the study, but periods were also reserved during which to observe the general work of the laboratory and to attend the weekly laboratory meetings attended by all laboratory staff. In addition, attendance at a range of faculty-wide events such as the weekly guest speaker seminar series, the weekly principal investigator seminar series, and occasional postgraduate student research seminars offered scope for additional observations to complement the time spent observing work in each of the participating laboratories.

In addition to these initial periods of observation, it was possible to revisit two of the participating laboratories, *viz*. *ES-L1* and *ES-L2*, intermittently over a period of one year. Throughout this period, the investigator was also present on a full-time basis as a postgraduate research student in a neighbouring molecular biology research laboratory within the same university faculty as all four participating laboratories.

The actual time spent observing each participant varied, and was largely dictated by the work schedule of the participant. For some participants such as the technicians in the service laboratory *ES-L3*, observations were made over consecutive days in order to follow the full cycle of the extended experiments that are typical of that laboratory. For the majority of participants in the research laboratories, observations were made over non-consecutive days at the request of the participants. Some participants were not actively involved in performing laboratory work during the period of the study and so could not be

[46] Bioscientists use the term 'wet' laboratory to distinguish a laboratory in which bench work experiments are conducted using chemicals and biological materials. In contrast, 'dry' laboratory work typically involves computer modelling or other forms of bioinformatics.

observed performing work in the 'wet' laboratories; these participants were observed in the office environment and interviewed using unstructured interviews to provide insider accounts. Such interviews were always performed on a one-to-one basis. Observing laboratory work, on the other hand, took place in the communal laboratory areas and so enabled observation of spontaneous interactions between participants and other laboratory members.

When permitted by the participants, interviews were recorded using a digital voice recorder. A handheld video camera was also used to record specific aspects of the laboratory work performed by the participants, again with the express permission of the participants.

3.2.3.3 Data collection

Textual field notes were recorded by hand in each laboratory in the same type of notebooks and notepads used by the scientists at work in the laboratory settings in order to present a relatively familiar appearance whilst observing work in the laboratories. Table 3-3 overleaf identifies the range of observational data collected during the course of this study in addition to these written ethnographic field notes.

Not all participants were willing to be video recorded during the course of their work either due to personal preference or to concerns about causing disturbance to other laboratory members. For example, no video recording was made of the work carried out in research laboratory *ES-L1*. In this situation, textual field notes and photographs provided the sole record of observations of the work carried out by the participants.

All data collected during the study have been rendered anonymous in accordance with the terms of the ethical approval for the study. To facilitate subsequent analysis, the data was collated using version 2.8 of the HyperRESEARCH™[47] computer-aided qualitative data analysis (CAQDAS) software. This tool[48] was selected on the basis that it supports effective integration and the coding of multimedia data including audio/video recordings.

---

[47] Product information is available from Researchware Inc at http:/www.researchware.com [accessed 01 March 2011]. All trademarks are acknowledged.

[48] CAQDAS tools are software systems used by qualitative researchers working in multiple fields of inquiry to facilitate the component tasks of qualitative data analysis such as collating source data, transcribing source data, maintaining mappings between codes and source data, and

Table 3-3: Data collected for the ethnographic study

Laboratory	*Observational Data*			
	Audio Recordings	**Video Recordings**	**Photographs & Scans**	**Sample Documents**
ES-L1	5hrs 30 mins Interviews with laboratory members; Unstructured work chat.	0 hrs	11 Benches; Equipment; Protocols; Guides.	4 Primers spreadsheets; Laboratory device manual.
ES-L2	15hrs 52 mins Interviews with laboratory members; Unstructured work chat	3 hrs Execution of laboratory work covering multiple techniques; Bioinformatics tool.	27 Offices & Laboratories; Benches; Bioinformatics tools; Notebooks; Protocols.	5 Paper consulted for protocol; Spreadsheets; Kit manual; FlyAtlas[49] web pages.
ES-L3	0 hrs	9 hrs Execution of laboratory work covering analysis of client samples; Kit usage.	27 Offices & Laboratories; Benches; Bioinformatics tools; Notebooks; Protocols.	8 Web pages; Service brochure; Order forms; Example protocols; Kit manual.
ES-L4	1 hr Principal investigator interview.	0 hrs	0	0
All Labs	***22 hrs 22 mins***	***12 hrs***	***65***	***17***

Summarizing the observational data collected in each laboratory for the ethnographic study by type of data, and by amount of data.
The laboratory of origin is identified using the code number assigned for the ethnographic study (see Table 3-1). The type of data is categorized into audio recording of interviews, video recording of laboratory work, scans and photographs of the laboratory environment and artefacts, or copies of sample documents. The amount of audio/video recording is specified in hours. The amount of photographs and sample documents is specified as an item count. The range of data items collected in each laboratory is described in text.

relating coded source data to theory. Reviews of a range of CAQDAS tools are available at http://caqdas.soc.surrey.ac.uk/softwareoptions.html [accessed 01 March 2011].

[49] The FlyAtlas community database (Chintapalli *et al.* 2007) is available at http://www.flyatlas.org [accessed 01 March 2011]. This database provides an example of using social media within a specific research community to promote information sharing between laboratory groups.

3.2.3.4 Data analysis

The data collected for the study have been analysed, coded, and categorized in line with the data analysis procedure for developing grounded theory[50] (Charmaz 1983; Glaser and Strauss 1967; Strauss and Corbin 1998). This procedure involved an iterative process of open coding, axial coding, and selective coding in order to classify and interrelate the data obtained during the course of the study. These levels of coding mirror the item level, pattern level, and structural level of analysis recommended by LeCompte and Schensul (1999) for ethnographic data analysis.

Open coding involved the systematic examination of the field notes, sample documents, audio recording, and video recordings collected during the study in order to identify and describe categories of behaviours, events, actions, and other concepts in relation to laboratory recordkeeping. Axial coding was concerned with refining the understanding of laboratory recordkeeping by identifying generalization/specialization relationships[51] to capture semantic links between the categories and their associated subcategories. Finally, selective coding was concerned with integrating and refining the set of categories in order to build an understanding of laboratory recordkeeping that was derived in an inductive manner from the data collected during the course of the study. It is important to note that coding was not a static procedure, but instead proceeded in an iterative manner throughout the course of the study so that categories could be compared, modified, and refined in the light of new observations. This constant comparison approach proved particularly appropriate to the study in terms of facilitating a robust comparison of laboratory recordkeeping across different laboratory settings and members.

---

[50] It is recognized that grounded theory is something of a contested term as witnessed by the separate approaches subsequently developed by its originators, Glaser and Strauss. Titscher *et al.* (2000), *inter alia*, point out that whilst grounded theory is widely used in qualitative research, it has been adapted to varying degrees from Glaser and Strauss's original conception. As Charmaz (1983:125) points out, "each researcher who adopts the approach likely develops his or her own variations of technique".

[51] A generalization/specialization relationship, sometimes referred to as a supertype/subtype relationship, is an information and type modelling construct that identifies a specific form of relationship between two concepts in which one concept (the specialization or subtype) represents a more specialized form of the other concept (the generalization or supertype). For example, a 'car' is a specialized form of the more generic concept 'vehicle', and 'molecular biology' is a specialized form of the more generic concept 'biology'. Generalization/specialization relationships form part of many systems and data analysis techniques such as object-orientated modelling and entity-relationship modelling in addition to their use in grounded theory data analysis.

In order to preserve the rich information content inherent in audio and video data (Goodwin 2000; Heath and Luff 2000; Kress and van Leeuwen 2001), these types of data were not transcribed prior to analysis but were analysed and coded using the multimedia coding facilities in the HyperRESEARCH CAQDAS tool selected for the study.

3.2.3.5 Data validation

Ensuring reliable and valid research findings based on ethnographic data requires an awareness of the complexity of ethnography. This complexity stems from a number of areas including the fact that the ethnographer is observing a culture with which he/she is largely unfamiliar, participants may not act as they would normally do when in the presence of an observer, and only snapshots of the setting are being observed corresponding to the limited period in which the ethnographer is present in the setting. In addition to prolonged exposure to the setting, the principal strategies used to promote the reliability and validity of the findings based on the observational data collected during this study of laboratory recordkeeping were triangulation and participant validation (Hammersley and Atkinson 1995; Lillis 2008; Schryer 1993; Swales 1998).

Triangulation essentially involved cross-checking the data collected during the course of the study both from different participants and using different methods in order to enable a comparison of the data pertaining to a single phenomenon that has been captured from a range of sources. To enable triangulation within this study, multiple data collection methods were used to collect data on laboratory recordkeeping including interviews, field notes, video-recording, and sample documents so that both between-method triangulation (*i.e.* cross-checking data collected using during methods) and within-method triangulation (*i.e.* cross-checking data collected from different sources using the same method) could be employed.

Participant validation involved presenting and discussing the findings from the study both with individual participants and, to a much lesser extent, via group presentations in order to evince feedback on insiders' perspectives on the validity of the findings. The role of participant validation in ethnography can be problematic (*e.g.* Bloor 1978). Crucially, the feedback gained from these sessions, whether supportive or hostile, was not to be interpreted uncritically as direct validation (or invalidation) of the findings from the study. Instead, it was held to be an additional source of data for use in the shaping the findings of the study.

3.3 Results

The results of this ethnographic study of laboratory recordkeeping are presented in the following subsections, arranged around the categories identified during the data analysis and "illustrated by characteristic examples of data" (Glaser and Strauss 1967:5). References to individual participants in the study results make use of the participant identifiers listed in Appendix 2. Excerpts from audio and video recordings used in the study results have been transcribed using the conventions identified in Appendix 10.

3.3.1 Locally mobile recordkeeping

3.3.1.1 Distributed working environment

Laboratory records were produced, used, and archived within the specific physical working environment of the laboratory settings[52]. In each of the participating laboratories, this working environment had been partitioned into a number of distinct areas providing general office accommodation, computing facilities, 'wet' laboratory areas, specialized laboratory areas, and utility areas such as chemical storage and weighing rooms. Figure 3-1A illustrates this arrangement within the service laboratory *ES-L3*. The rationale behind this arrangement in each setting derived partly from a need to comply with health and safety considerations, partly from a need to observe best laboratory practice for avoiding contamination, and partly from a need to share expensive resources.

In particular, health and safety considerations dictated that office areas and other general accommodation should be separated from the laboratory areas. Office accommodation was at a premium within laboratories *ES-L2*, *ES-L3*, and *ES-L4*, and consequently personal office space was available only to the principal investigators in those research groups whilst the postdoctoral researchers, postgraduate researchers, and technicians were housed in a communal writing room. Research laboratory *ES-L1* differed in this respect with personal office space allocated to postdoctoral researchers and technicians in addition to the principal investigator.

---

[52] The influence of physical environment on aspects of laboratory work has been acknowledged previously in studies such as those by Knorr Cetina (1999), Livingstone (1995), and Lynch (1991). Knorr Cetina (1999:43, emphasis in original), for example, contends that "laboratories *recast* objects of investigation by inserting them into new temporal and territorial regimes". Lynch (1991:74) argues that "scientific praxis is bound up in locally organized topical contextures."

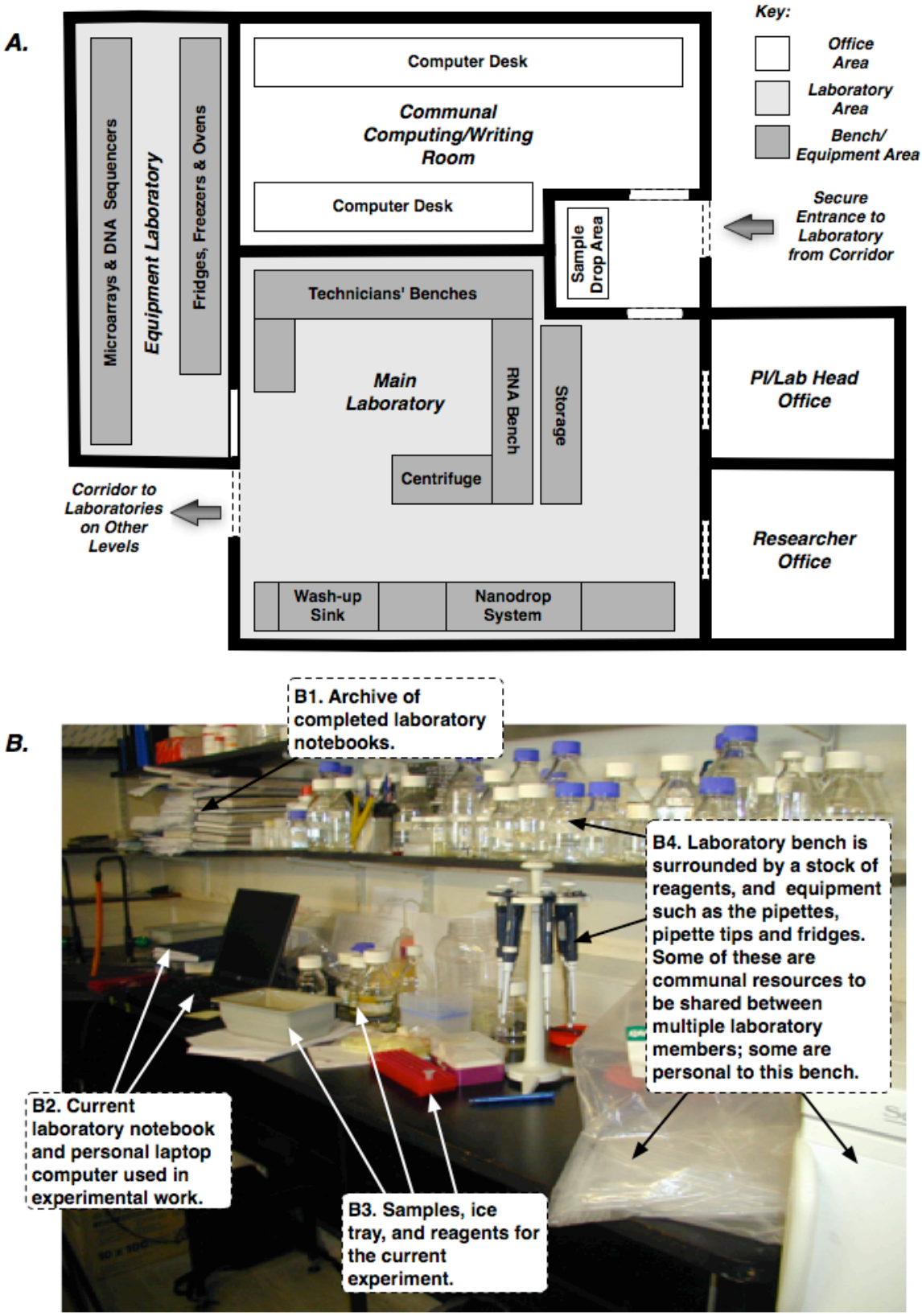


Figure 3-1: Physical environment of the laboratory settings

Depicting the working environment in two of the laboratories that participated in the ethnographic study. (*A.*) Schematic diagram illustrates the topography of the main area of service laboratory *ES-L3*. (*B.*) Photograph shows the laboratory bench used by one of the postdoctoral researchers in research laboratory *ES-L2*.

In contrast to the shared office accommodation, all postdoctoral researchers, postgraduate researchers, and technicians in each of the participating laboratories were allocated a personal bench in a 'wet' laboratory area. In a reversal of fortune, the principal investigators in each laboratory had no allocated bench space, which was indicative of the limited time they now spent directly engaged in bench work.

In addition to the main 'wet' laboratory area, specialized laboratories and designated regions within laboratories were reserved in each setting for specific categories of experiment. Reservation of specialized areas was driven by multiple criteria including a consideration of the biological properties of the sample types being investigated, and the need to manage access to resources with limited availability such as NanoDrop® instruments[53]. The specific configuration of specialized laboratories varied across each of the participating laboratories, due primarily to the differing nature of the experimental work conducted by the members of each laboratory. For example, the service laboratory *ES-L3* maintained a separate laboratory area to house the equipment used to provide its two core services of microarrays (*e.g.* Craighead 2006) and DNA sequencing[54], and a separate bench within the main laboratory for RNA work[55] as illustrated in Figure 3-1A. In contrast, the research laboratory *ES-L4* maintained a 'bug room' for growing bacterial cultures, a 'fly room' for maintaining multiple genetic strains of *Drosophila melanogaster*, a confocal microscope[56] area, and a bench reserved for gel electrophoresis[57] work.

Access to devices that formed a limited resource such as the NanoDrop instrument was also made available to researchers from other groups, and resource sharing formed a basic

---

[53] A NanoDrop® is a laboratory instrument that uses fluorospectrometry to provide micro-volume sample quantitation. Product information is available from Thermo Fisher Scientific Inc at http://www.nandrop.com [accessed 01 March 2011]. All trademarks are acknowledged.

[54] Automated systems now enable high-throughput DNA sequencing. DNA sequencing is the process of identifying the internal structure of a strand of DNA in terms of the ordered sequence of the possible nucleobases of which DNA can be composed. These bases are adenine (A), guanine (G), cytosine (C), and thymine (T) leading to the familiar alphabet soup used to encode DNA sequences as in $(GGC)_3G(CCG)_{20}(CCGCTG)_{14}(CTG)_{35}$. Thanks are due to a friend and colleague, Dr Claudia Braida, for providing this example sequence taken from her recent work characterizing DNA mutations associated with myotonic dystrophy.

[55] RNA work is particularly sensitive to contamination due to the presence of ribonucleases (RNases) that can rapidly degrade RNA samples. Consequently, additional precautions are required to maintain benches that are RNase-free, and to main stocks of suitable reagents such as RNase-free water.

[56] A confocal microscope is a laboratory device that provides improved microscopic imaging of fluorescently labelled specimens, and is often used within molecular biology laboratories to visualize *in vivo* cells or tissues that have been labelled with fluorescent probes.

[57] Gel electrophoresis is a core molecular biology laboratory technique that uses electrical charge to sort fragments of DNA/RNA based on their size and charge.

level of collaboration with partners for each of the three research laboratories. A range of more routine laboratory equipment was also available in each of the participating laboratories including freezers and refrigerators for sample and reagent storage, and devices for performing routine laboratory tasks such as gel electrophoresis tanks and PCR machines[58]. Such devices were distributed throughout the main laboratory areas in order to facilitate communal use, and were interspersed between the personal benches used by the scientists at work in those laboratories.

Figure 3-1B shows the laboratory bench used by the postdoctoral researcher *ES-L2R1* in research laboratory *ES-L2*. Scientists in each laboratory made use of these benches as a dedicated workspace at which to perform a range of experimental procedures. To this end, each bench was equipped with a dedicated set of basic devices such as pipettes[59] for use in performing the experimental procedures. Various chemicals and other reagents used in the experimental procedures were stored either at the bench in overhead shelves, in close proximity to the bench in freezers or refrigerators within the laboratory area, or in secure cupboards away from the benches in the case of particularly hazardous substances. During the execution of experimental work as shown in Figure 3-1B, the bench functioned as a locus for combining bioscience materials such as samples and reagents, technological artefacts such as pipettes and experimental kits, informational resources such as research articles and protocol manuals, and recordkeeping containers such as laboratory notebooks.

3.3.1.2 Patterns of mobility

Given the extended layout of the laboratory workspace, it is perhaps unsurprising that all scientists observed during the course of the study were highly mobile workers[60]. Each of the study participants and their colleagues in both the service laboratory and the research

---

[58] A PCR machine is a laboratory device used to run polymerase chain reactions (PCRs). PCR is a core laboratory technique that is used to copy and amplify strands of a DNA sequence of interest. Amplify in this sense refers to bulk replication, and PCR can generate a large number of copies of a target DNA sequence starting from a single or small number of copies.

[59] A pipette is a laboratory device that is used to transfer accurately a measured volume of a solution. Manual, graduated pipettes that are calibrated to handle different volumes were routinely used in each of the participating laboratories. Those with an interest in history of technology are directed to a website dedicated to this workhorse of laboratory devices at http://www.pipetteuk.com [accessed 01 March 2011].

[60] Arnstein *et al.* (2002) monitored a scientist working in a cell biology laboratory using video recording who was recorded as changing location 76 times during a work session lasting one hour. schraefel *et al.* (2004) report on the range of movement exhibited by a worker in a chemistry laboratory arising from the chemist's need to combine working at a bench, a fume cupboard, and various other locations housing specialized equipment.

laboratories were obliged to move around the physical geography of the laboratory whilst performing experiments. Typically, this was required in order to access communal devices such as PCR machines or centrifuges[61] that were distributed throughout the laboratory area.

For the most part, this involved the study participants moving around a single laboratory area. In some cases, this involved moving between rooms on different floors within a building. For example, both technicians in the service laboratory *ES-L3* would periodically require to take samples down two flights of stairs to a basement laboratory to make use of a vacuum concentrator[62]. Less frequently, this could also involve moving between different buildings. For example, the technician *ES-L2R2* in research laboratory *ES-L2* was observed preparing a sample in her own laboratory and then transporting[63] the sample together with various pieces of laboratory equipment across to a remote laboratory in another building in order to gain access to a specialized device for flame emission spectroscopy[64].

3.3.1.3 Patterns of mobile record production

In order to manage information in this locally mobile[65] environment, all study participants observed at work in the laboratory areas situated both their experimental work and associated recordkeeping around their dedicated bench as a central hub. The laboratory notebook together with any other information resources required for the planned experimental work were collated prior to beginning the experiment and kept at this hub. Examples of information resources used by the study participants included printed laboratory protocols often with handwritten annotations, copies of published articles

---

[61] A centrifuge is a laboratory device used to isolate and separate substances of different density within a tube by spinning the tubes at high speeds.

[62] A vacuum concentrator is an adapted form of centrifuge that is used to concentrate and dry samples by applying a vacuum, centrifuge, and heat to achieve controlled evaporation. This device is typically used within molecular biology laboratories to concentrate small samples of DNA, RNA, or protein.

[63] Experimenters working in ethnography can engage as participant, participant-observer, or observer. For the majority of this study, the experimenter resolutely remained an observer due to his lack of laboratory skills. At this point, however, he briefly took on the role of participant to help transport the equipment and samples up one of the many hills at the university site to the remote laboratory.

[64] Emission spectroscopy is a laboratory technique that is used to determine the elements in a compound by observing the electromagnetic radiation spectra obtained when the compound transitions between energy states. In the case of flame emission spectroscopy, this transition is caused by subjecting a solution of the compound to high temperatures over a flame burner.

[65] The term ‘locally mobile’ is used here in the sense of Bellotti and Bly (1996) to indicate mobility within a building or set of buildings relating to an organization.

providing useful background or additional protocol data, and manuals for off-the-shelf laboratory kits. Information resources that pertained to multiple experiments and so held value over an extended period were physically attached by the study participants to their personal bench as shown in Figure 3-2, which depicts the bench allocated to postdoctoral researcher *ES-L2R4* in research laboratory *ES-L2*.

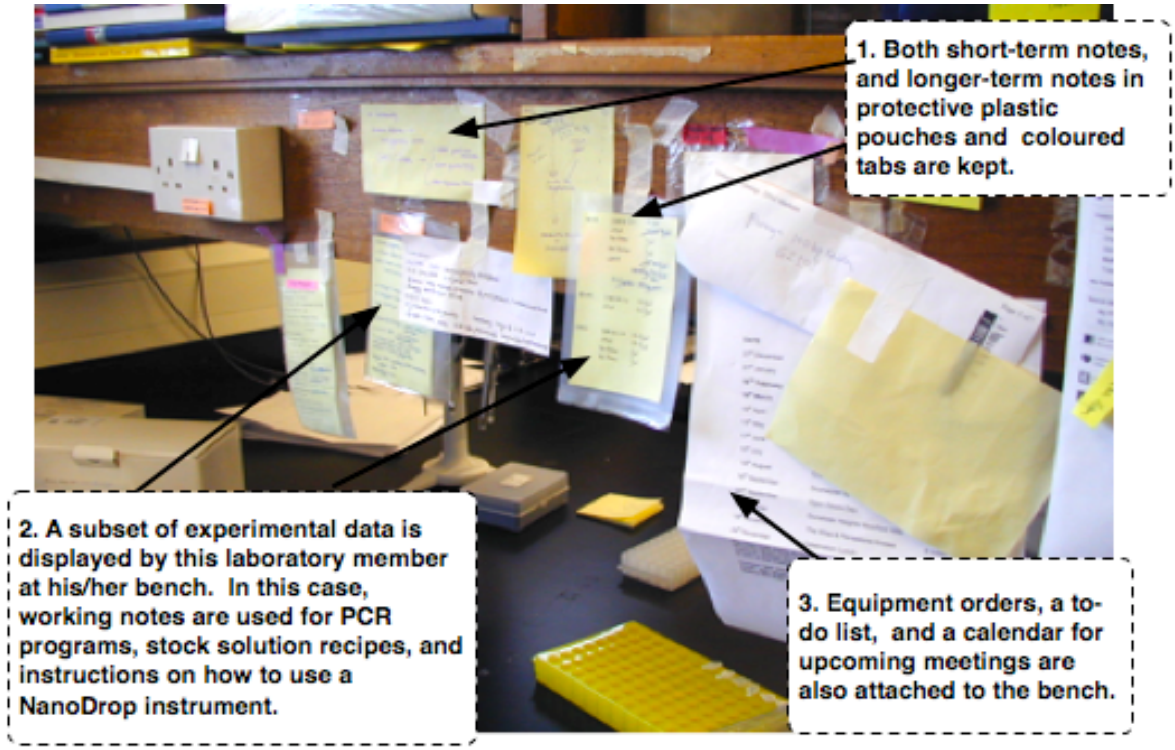


Figure 3-2: Bench information resources

Depicting the role of the laboratory bench as an information hub. Photograph shows the laboratory bench of a postdoctoral researcher within research laboratory *ES-L2*. Both short-term notes and longer term notes in protective plastic pouches are taped to the bench to record a range of experimental information including PCR thermal cycles and recipes for making up stock solutions. A subset of this information is repeated or referenced in notebook entries, forming one aspect of intertextuality in laboratory records.

Whenever necessary, study participants would transfer to the location of a shared device taking samples and other reagents with them, perform the necessary work, and then return to the bench. Data recording in the laboratory notebook in relation to this work took place at the central hub. This pattern of mobile recordkeeping situated at the personal laboratory bench was observed in each of the participating laboratories in the work of multiple study participants executing a range of experimental procedures including DNA sequencing experiments and quantitative PCR (qPCR) experiments[66] in research laboratory *ES-L1*, fly

[66] qPCR is a variant of the PCR laboratory technique that is used to both amplify and quantify strands of a DNA sequence of interest. This technique is often used within molecular biology laboratories to provide quantitative data on the expression of a gene over time, perhaps in response to specific stimuli.

secretion assays[67] and plasmid preparations[68] in research laboratory *ES-L2*, and microarray experiments in service laboratory *ES-L3*.

Whilst data recording centred on the 'bench as hub' was by far the most prevalent approach employed by the study participants to balance recordkeeping with local mobility, it was not the only approach observed. In particular, the study participants employed variant approaches to manage their interaction with remote devices, complex devices, and devices generating digital results.

When working with remote devices and some complex devices as in the previous example of technician *ES-L2R3* using the flame emission spectrometer, the technicians in each participating laboratory and most of the researchers used spiral-bound notepads or loose scraps of papers to record device settings and partial results away from the bench at the site of device. These temporary results were subsequently transcribed into the notebook or pasted into the notebook when the participant returned to the bench. Some participants such as postdoctoral researcher *ES-L2R1* preferred to rely on memorizing the settings and writing them into the laboratory notebook when back at the bench, an approach that was not without risk[69]. A range of current laboratory devices such as PCR machines was programmable. Depending on the anticipated level of use, such devices were typically configured by each of the study participants with the required settings in order to facilitate repeated use, thus changing the nature of data recorded for device use from individual settings to named programmes.

Laboratory devices geared to visualization or quantification were often capable of generating digital results. An example of one such device in routine use within the participating laboratories was the gel documentation system[70] that generated digital images

---

[67] A secretion assay is a laboratory technique used to identify cells that are secreting a specific protein of interest by using specific antibodies that bind to the protein of interest.

[68] Plasmid preparation is a laboratory technique used to extract and purify the DNA stored in a plasmid. This technique was routinely performed in the participating laboratories using a range of off-the-shelf laboratory kits.

[69] Ebel *et al.* (2004:18) sum up the view of many laboratory workers including all principal investigators in each participating laboratory when they counsel that "under no circumstances should one trust observations to memory, even for brief periods" on the basis that "things 'remembered', regardless of the time span, far too often turn out to be remembered only vaguely, or perhaps incorrectly!"

[70] Gel electrophoresis plays an important role in molecular biology experiments as a means of isolating and visualizing DNA fragments. A gel documentation system is a laboratory system that combines a camera, transilluminator, and image capture and analysis software to support the digital capture of gel electrophorerograms, commonly referred to as gel images.

of electrophoresis gels. In each of the three research settings, using this system involved shifting the locus of record production away from the bench over two stages. First of all, study participants would move to a specialized laboratory area or dark room in which to make use of the gel documentation system. Thereafter, the locus of recordkeeping would typically transfer out of the laboratory area to the office accommodation where the study participant would access the digital gel image via the laboratory's computing network in order to continue with recordkeeping.

3.3.2 Ownership and containership

3.3.2.1 Containers of current records

All study participants had direct access to personal computers within the laboratory settings. Indeed, two of the study participants in research laboratory *ES-L2, viz*. postdoctoral researcher *ES-L2R1* as illustrated in Figure 3-1B and postgraduate researcher *ES-L2R3,* kept laptop computers at their personal laboratory benches. Nevertheless, all study participants and their colleagues in each of the four participating laboratories relied on bound, hardback paper notebooks for their primary record of experimental work. Typically, the current laboratory notebook was kept to hand either on the individual's personal desk within the office accommodation or at their personal bench in the 'wet 'laboratory area. Completed laboratory notebooks relating to previous experimental work were similarly stored within easy reach at the individual's desk or laboratory bench. Clearly, the laboratory notebook was held to be a valuable resource.

All principal investigators and technicians in each participating laboratory demonstrated a clear vision of these laboratory notebooks as community property, held as a historical archive of experimental work for use by current and future generations of laboratory staff as illustrated by the comment:

> "we tell everybody that comes into this lab (.) that at the end of their period in this lab, they have to leave all their lab books and materials with us (.) the lab books are not theirs, they can take photocopies away."
>
> [Principal Investigator in Research Laboratory]

This view contrasted sharply with that of all but one (*ES-L2R3*) of the postdoctoral and postgraduate researchers in the research laboratories who considered their laboratory notebooks as personal property. For these subjects, the laboratory notebook as primary

record of experimental data served to maintain the subject's sense of identity (*e.g.* Biagioli 2003; Kaye *et al.* 2006; Riley 2007). Interestingly, this viewpoint was not only recognized by the principal investigators in each laboratory but was also validated by them to a degree as shown by the comment:

> "I mean I think the other thing is that people (.) feel quite rightly ownership of data, y'know your grad students and your post docs all feel very strong ownership of data … because it's the data that makes you special in science."
>
> [Principal Investigator in Research Laboratory]

Each of the four postdoctoral researchers that participated in the study kept a personal archive of laboratory notebooks produced during the course of their previous work as postgraduate students or in previous postdoctoral positions. In contrast to the first of these two quotations (see previous page), the archive for three of these four researchers did not contain photocopies but instead contained the original notebooks pertaining to the earlier work. It was not that the case these notebooks had been 'liberated' by their authors[71]. Instead, photocopies of the notebooks' content had been left where requested with previous laboratories, each of which was also an academic molecular biology laboratory based within the UK.

3.3.2.2 Containers of the archived records

The stack of laboratory notebooks visible in Figure 3-1B in the shelves above postdoctoral researcher *ES-L2R1*'s bench illustrates the fact that each study participant kept a separate archive of his/her own records whilst he/she was present in the laboratory. In laboratories *ES-L1 and ES-L4,* this archive of previous notebooks was typically kept at the participant's desk in the communal writing room or in the personal office allocated to each participant. In laboratories *ES-L1* and *ES-L3*, the archive of previous notebooks was typically kept at the participant's bench.

Career progression, particularly in light of the three-year postgraduate studentships and shorter postdoctoral positions typical of many UK academic molecular biology laboratories, meant that individual scientists would routinely leave the research laboratories. Given the intended role of the records contained in laboratory notebooks as a community resource, it was interesting to observe what happened to laboratory notebooks

[71] Under the motto *ignorantia juris neminem excusat*, I felt obliged to confirm that no crime had been committed.

when the original author had left each of the participating laboratories. In research laboratory *ES-L1*, no central archive was in place to hold notebooks from previous staff so that these notebooks were distributed across multiple offices and individuals. In research laboratory *ES-L2*, notebooks produced by previous staff were held centrally in the principal investigator's office, and some of these the notebooks had been loaned out by the principal investigator to those individuals now tasked with continuing the work of previous staff. In service laboratory *ES-L3*, notebooks produced by previous technicians were archived in the communal computer/writing room. In research laboratory *ES-L4*, the notebooks were again held in shelves in the communal office space where they were freely available to any member of the laboratory who may require them.

No written index, in any form, of previous notebooks existed in any of the participating laboratories. Instead, identifying and locating notebooks that may be of use to current staff was dependent in the first instance on the retained memory of the principal investigator and other staff of long-standing, followed by a process of trial-and-error. It was important to note that all of the study participants including principal investigators in each of the participating laboratories had rarely attempted to retrieve notebooks written by previous staff.

3.3.3 Audience for laboratory records

3.3.3.1 Personal expectations

The view of the laboratory notebook as personal property was manifested most clearly in the readership expectations voiced by the study participants. It is, of course, important to distinguish here between the intention of an author as to whether he/she is writing for others to be able to interpret the records, and the reality of whether other laboratory members could actually make sense of the recorded entries. With the exception of postgraduate researcher *ES-L2R3* and the principal investigators, all study participants in both the research and service laboratories recorded entries in their laboratory notebook on the basis that:

> "it's only for me to read; nobody else could make sense of it."
>
> [Postdoctoral Researcher in Research Laboratory]

Interestingly, the expectation of these postdoctoral researchers, postgraduate researchers, and technicians was that other laboratory members would follow the same personalized approach. Consequently, they did not expect to be able to interpret the entries recorded in laboratory notebooks produced by other laboratory staff as illustrated by the comments:

> "I'm now very guilty (.) I'm sure most people are (.) writing enough in the lab book that you can remember (.) what you're doing."
>
> [Postdoctoral Researcher in Research Laboratory]

> "I think if you were to take a straw poll of anybody in the lab and ask them if they found anybody else's lab book useful, I just don't (.) I think it depends what you're looking for."
>
> [Postdoctoral Researcher in Research Laboratory]

As indicated previously, the 'personal property' view of laboratory notebooks was not universal amongst the postgraduate and postdoctoral researchers. One of the study participants was the postgraduate researcher *ES-L2R3* who had previously worked in a diagnostics laboratory in the healthcare sector and so was accustomed to an administrative environment in which individual notebooks would be verified by other staff. This individual commented that he intended to transfer the disciplines learnt from that setting to his work in academic laboratories.

3.3.3.2 Supervisory expectations

The motivating factor behind the concept of the laboratory notebook as community property is that it provides a vehicle for communicating 'who did what, and when'. In the research laboratories *ES-L2* and *ES-L4* in particular, those in supervisory roles stressed the use of laboratory notebooks during the weekly one-to-one sessions that were held with each of the research staff to monitor progress as illustrated by the comment:

> "Yes (.) I mean we talk over stuff with them ((the students)) (.) and so, when people come to see us for their one-to-one, they bring their lab books (.) eh (.) if somebody doesn't bring their lab book (.) eh it's commented upon (.) you should always bring your lab book whenever you come to talk to us (.) because that's why you're here (.) it's to go over results, problems and everything else."
>
> [Principal Investigator in Research Laboratory]

However, during both these progress meetings and other meetings with colleagues, the laboratory notebook served less as a written inscription of work done to be read by others

and more as an *aide-memoire* for its author during verbal communication. The experimental data contained in the laboratory notebook was interpreted by the author of the laboratory notebook and communicated verbally. The laboratory notebook was typically only consulted directly by others to view specific images or printouts from specific devices. This approach was also followed in the service laboratory *ES-L3* as shown by the comment:

> "he'll ((the principal investigator)) just say what did you do ((imitating the voice)) (.) if something of the (.) things aren't (.) the results don't quite tie up the way they're meant to (.) he'll say (.) is there anything special that happened to this one (.) you can just try and look back and see if anything happened (.) no, he doesn't usually look."
>
> [Technician in Service Laboratory]

Study participants demonstrated a clear preference for verbal communication with colleagues rather than trying to read through the colleagues' laboratory notebooks. On one occasion, for example, a senior researcher in the laboratory *ES-L2* preferred to re-arrange his scheduled experimental work by a week whilst he waited for a colleague to return from holiday rather than try to read through the colleague's laboratory notebook. To some extent, this was due to the practical difficult of finding the required information. However, it was also considered that to do otherwise might be a breach of conduct as illustrated by:

> "it would be very unusual (.) you wouldn't ask someone to go and look at someone else's lab book (.) who is still here, OK (.) so, yeah, while that person is here, it is kinda their personal property."
>
> [Principal Investigator in Research Laboratory]

3.3.4 Structure and content of laboratory records

3.3.4.1 Learning to keep records

None of the four participating laboratories provided in-house standards or guidelines that could be referred to by laboratory staff to define the recommended structure or content for recording experimental data in laboratory notebooks. Instead, the only written guide to recordkeeping observed in any of the laboratory settings was present in the form of instructions provided by the notebook manufacturer in the preface of a specific brand of

notebook[72] as shown in Figure 3-3 overleaf. This brand of notebook was used by most, but not all, of the laboratory staff in research laboratory *ES-L4*. A straw poll of current and archived notebooks in the laboratory indicated that none of the laboratory staff had signed to confirm that the instructions had been read and understood.

Only two study participants in the shape of postgraduate researcher *ES-L2R3* and technician *ES-L3T2* reported that they had previous experience working to a defined form of recordkeeping. That previous experience had been in a health service laboratory in one case, and in an industrial laboratory in the other case.

Further, the majority of the study participants including the principal investigators and others now in supervisory roles reported that they had received no formal instruction in their current laboratory nor in any previous laboratories on how to structure laboratory notebooks. Instead, all study participants had "learned on the job" and developed a style to suit their needs over the course of their work in multiple laboratories[73]. To this end, some of the study participants in each of the participating laboratories expressed a comparative awareness of the structures and representations used by their colleagues in the same laboratory to keep laboratory records. Many study participants acknowledged that their recordkeeping approach had developed over a considerable period incorporating influences from their experience of laboratory report writing as undergraduates and before as illustrated by the comment that recordkeeping is:

> "one of (.) the fundamentals of science as we learned it at school (.) I'm sure you were the same (.) sometimes for the most obvious experiment you would have to take these very careful notes (.) what your rationale was, what you did, what you found, write up your conclusions."
>
> [Postdoctoral Researcher in Research Laboratory]

[72] This brand of bound hardback laboratory notebook was produced by NALGENE® Labware. The guide to recordkeeping was included on the inside front cover of the notebook before any of the pages used by research staff for records of their experimental work. Each page in this brand of notebook was numbered, duplicated, and included areas for both author and supervisor signatures. Further information is available from Thermo Fisher Scientific Inc at http://www.nalgenelabware.com [accessed 01 March 2011]. All trademarks are acknowledged.

[73] Assessment, both self-reflection and expert assessment, plays an important role in literacy learning (*e.g.* Douglas 2000; Johns 1997). The source, form, and extent of feedback received by laboratory staff on their recordkeeping remains open to investigation given the limited direct use of laboratory records in their written form by other staff.

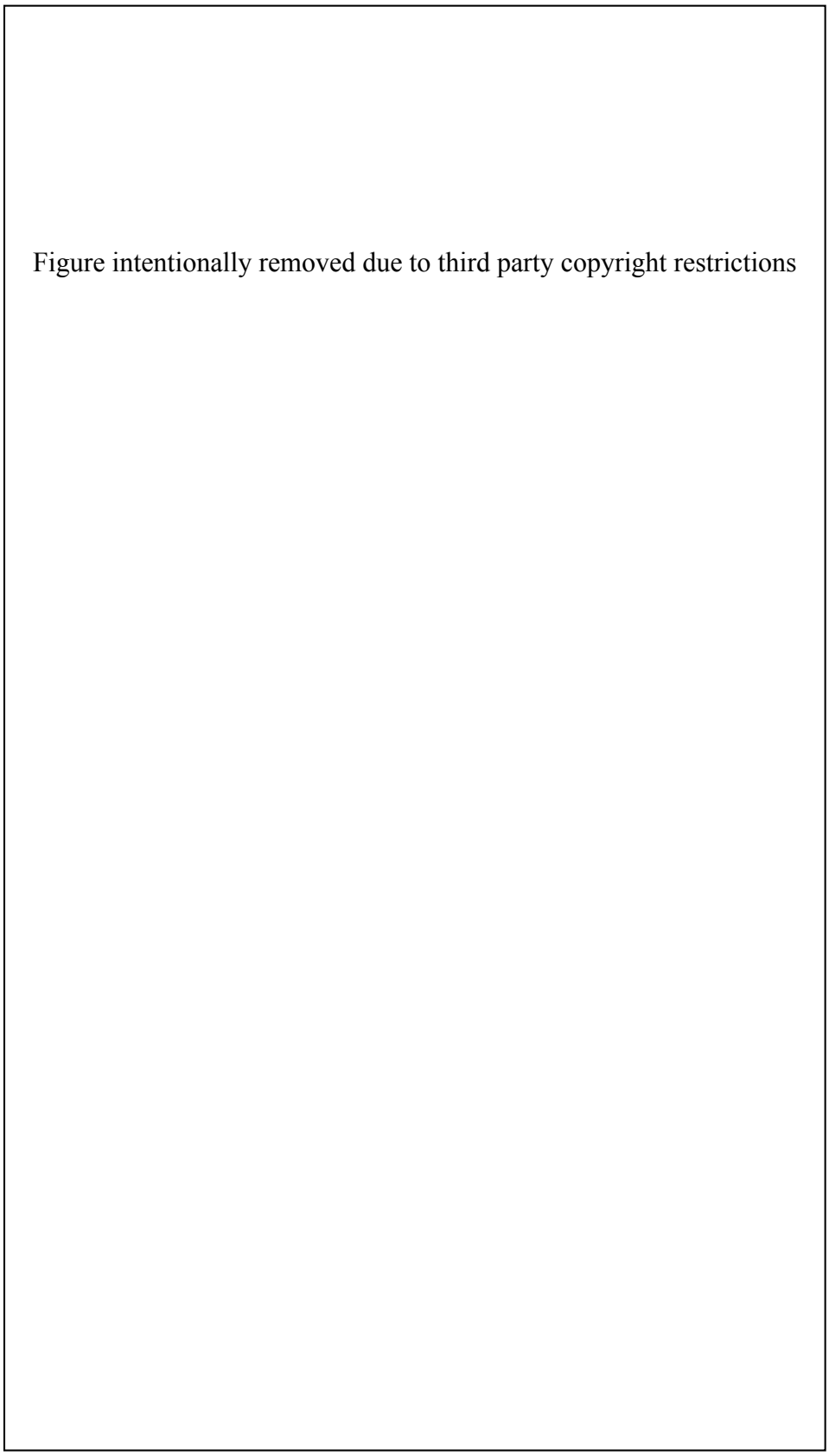


Figure 3-3: Notebook instructions for recordkeeping

Depicting the instructions provided by the manufacturer in the preface of a branded type of paper laboratory notebook used by the majority of the staff in research laboratory *ES-L4*. This brand of notebook is produced by NALGENE® Labware, part of Thermo Fisher Scientific Inc. The recordkeeping instructions are copyright of Thermo Fisher Scientific Inc.

3.3.4.2 Gold-standard laboratory record

Notwithstanding the lack of explicit guidelines for recordkeeping within each participating laboratory combined with the limited formal training in recordkeeping experienced by the study participants, each participant in both the research and service laboratories expressed a remarkably consistent concept of the ideal information required in a notebook as typified by:

> "date is important for a start (.) and (.) the (.) experimental samples (.) the experiment first, the experimental samples (.) the N number (.) the protocol if it's a standard one that's fine, if there's anything difficult it needs to be recorded (.) and then the results at the end of the experiment (.) and so (.) I would (.) and add a conclusion unless you know it in your head unless it's something unusual, write it down."
>
> [Principal Investigator in Research Laboratory]

An important feature of this construction of the laboratory record is that it highlights the contextual competency required of individual staff in determining when experimental data becomes sufficiently "difficult" or "unusual" to require it to be recorded in the notebook. It remains a challenge to position the ideal reader for the laboratory records given the wide range of knowledge and experience that exists in every academic molecular biology laboratory by dint of its dual role in conducting advanced research whilst educating a new generation of scientists. In the case of those taking on new supervisory responsibilities for junior research staff, recognizing this challenge has led to a change in the structures and content used for recordkeeping as illustrated by the comment:

> "and I'm now making more of an effort 'cause often if (.) if someone's having a problem (.) one of the student's having a problem (.) (obviously) we check in my lab book (.) to see what we did (.) so I'm now just over the last month making a conscious effort to make it legible to others rather than code words for me."
>
> [Postdoctoral Researcher in Research Laboratory]

3.3.4.3 Breakdowns in recordkeeping

Verifying the recordkeeping practices actually used by individual laboratory staff was pursued to a degree within research laboratory *ES-L2,* particularly in the context of projects with industrial funding. Within the other three participating laboratories, the quality of

recordkeeping was largely dependent on the behaviour of the individual scientist. The focus instead was on the progress of the research results as illustrated by the comment:

> "at the end of the day (.) I can't micro-manage (.) everyone (.) I'm not a very good micro-manager (.) I'm much more a show me the money person ((laughing))."
>
> [Principal Investigator in Research Laboratory]

Although an infrequent occurrence, the principal investigators in each of the research laboratories reported situations in which they had attempted to consult the laboratory notebooks produced by staff that had now moved away from the laboratory. In many cases, this approach had proved unsuccessful as typified by:

> "it's (.) can be tricky to interpret, yeah … and so (.) but that's important (.) it's invariably with people who've left already, OK."
>
> [Principal Investigator in Research Laboratory]

It is important to note that difficulty in interpreting the records kept in laboratory notebooks was not only encountered in notebooks produced by other writers. Most of the study participants appeared, at least on initial inspection, to maintain a largely consistent style in the records kept over time within their notebooks. However, a degree of temporal variation was evident in the approach used by some research staff, both postdoctoral researchers and technicians, when constructing laboratory records and this variation could lead to difficulties as illustrated by:

> "but I go through phases with it (.) sometimes I write really good notes and sometimes I don't (.) and it's mostly me writing really good notes after I've been stung by the fact that I can't remember something I did."
>
> [Postdoctoral Researcher in Research Laboratory]

3.3.4.4 Laboratory notebook as multi-purpose diary

All experimental records in the laboratory notebooks were date-stamped, and in this sense the laboratory notebook functioned as a diary of experimental work. One particular participant, postdoctoral researcher *ES-L2R5*, had extended the role of the laboratory notebook as a diary to incorporate day planning and week planning sections. A separate sheet of paper was taped to the front cover of the current laboratory notebook with dates of any upcoming meetings, deadlines and other activities; 'to do' lists were included in the laboratory notebook entries so that activities could be ticked off when completed.

Postdoctoral researcher *ES-L1R2* used his laboratory notebook not only for experimental work but also regularly made use of the notebook to include meeting reports for internal administrative meetings, summaries of budget meetings, and draft plans for papers to be written. These meeting reports and other types of record were interspersed among the experimental records.

3.3.4.5 'Optimizing' and 'results generating' experiments

Initial inspection[74] of the notebooks produced by the postdoctoral researchers, postgraduate researchers, and technicians in each of the participating laboratories highlighted patterns of both consistency and personal variation in the structures, content, and representations used in the laboratory records. One issue in this respect concerned the staging of experiment tasks into periods of 'optimization' followed by 'results generation'. Optimization work was concerned with a process of trial and error to determine the optimal experimental conditions including temperatures, concentrations, and reagents to use for a specific experimental purpose. Results generation work was concerned with applying these optimized conditions to batches of samples in order to generate results for interpretation.

Both technicians in the service laboratory *ES-L3* consistently recorded full statements of each experiment so that individual entries in the laboratory notebook could be read in relative isolation. Each of these experiments was undertaken as part of a separate client order, and involved the execution of one of a limited set of standard services with the client samples. In this sense, the technicians in the service laboratory were involved in repeat executions of well-defined experimental tasks to generate results for clients, with each discrete execution largely independent of the others due to the separation of client concerns. Interestingly, however, technician *ES-L3T2* also maintained a separate notebook for "custom arrays", in which the experimental process involved a period of optimization in order to test and evaluate different experimental conditions prior to running the optimized conditions with the client samples. It was interesting to observe that the style of recordkeeping in this "custom array" notebook varied from that used in the technician's other notebooks. In particular, the optimization entries displayed more frequent use of abbreviations and acronyms, use of implicit and explicit cross-references to previous

[74] More detailed inspection of the content of a range of laboratory records was undertaken within the framework of the genre analysis study in line with the multi-perspective approach employed for this study of laboratory recordkeeping.

notebook entries, and omission of experimental conditions that would normally be recorded.[75]

The sequence of optimizing experimental conditions followed by executing the optimized protocol for a batch of samples was highly typical of the experimental work recorded in the notebooks of those study participants at work in research-orientated laboratories. Here again, shorthand notations and reliance on multiple forms of intertextuality were also evident to some degree within the records kept by study participants in the research laboratories. It is important, however, not to underestimate the effort devoted by research staff to optimization. For example, one of the laboratory notebooks archived by postdoctoral researcher *ES-L1R1* covering his work as a postgraduate researcher consists solely of optimization experiments culminating in a penultimate entry containing the result interpretation "Weyhey. It's worked. Gallus Ya beauty … Best rx^n was Wynton's[76] reagents with my buffer". In contrast to the relative independence between successive entries that was characteristic of service laboratory notebooks excluding the "custom array" notebook, successive entries in research notebooks were often interdependent and formed part of a single larger block of work within a project. This interdependence established an extended contextual time frame over which to manage the structure and content of the laboratory records.

3.3.4.6 Records as compound documents

The concept of the laboratory notebook as a pen and paper solution for recording experimental results did not wholly reflect the situation observed in any of the participating laboratories. As mentioned previously, a range of the laboratory instruments such as NanoDrop instruments and gel documentation systems that were in regular use within the laboratory settings were able to generate digital and/or printed results. In the case of a NanoDrop instrument, the generated results consisted of a formatted table of samples with the measured concentrations. In the case of a gel documentation system, the generated result consisted of a digital image of an electrophoresis gel.

---

[75] Heath and Luff (1996) use the term "descriptive economies" to characterize these elements of recordkeeping style.

[76] All names have been changed to protect the privacy of the participants. Wynton is chosen in honour of the great Wynton Marsalis whose recordings with the Wynton Marsalis Septet at the Village Vanguard provided the soundtrack to the writing of much of this thesis.

This production of digital results gave rise to distinct issues. One concrete issue was how to combine the digitally formatted data with the handwritten sections of the experimental record. The approach preferred by the majority of the study participants was to generate hardcopy of the digital data, for example a printout of a gel image, and physically paste that hardcopy into the laboratory notebook as shown in Figure 3-4. In the case of electrophoresis gel images and other such images, all the study participants followed this approach. In the case of the NanoDrop, however, some study participants preferred to make handwritten records of the results displayed on screen by the instrument. An interesting example of this variation in approach that again highlights the individual choice involved in recordkeeping was observed in the service laboratory *ES-L3*. Of the two technicians who worked side-by-side at neighbouring benches, one routinely inserted hardcopy of the formatted NanoDrop results whilst her colleague routinely copied the results by hand.

Inspection of a small set of laboratory notebooks highlighted the occasional use of a hybrid system by a limited number of research staff and by the technicians in the service laboratory. Under this hybrid system, the file location of the results data was also recorded in the laboratory notebook alongside the pasted hardcopy of results.

Using digital result formats also seemed to raise a degree of concern with some of the study participants, particularly those in supervisory roles. The objection was that the digital tool represented another potential source of introducing error into the experimental analysis raising the issue of how you control the reliability of the digital tool as illustrated by the following remark:

> "rather if they've come to me with some results that don't look right (.) y'know you go back to the lab book to try and work out what's happening (.) then (.) although, you see, it might be that a lot of the time now you get something spewed out of a machine (.) it comes back at you as an Excel[77] spreadsheet so you have to make sure that the spreadsheet hasn't mutated during the processing."
>
> [Principal Investigator in Research Laboratory]

[77] Excel® is a trademark of Microsoft Corporation. All trademarks are acknowledged.

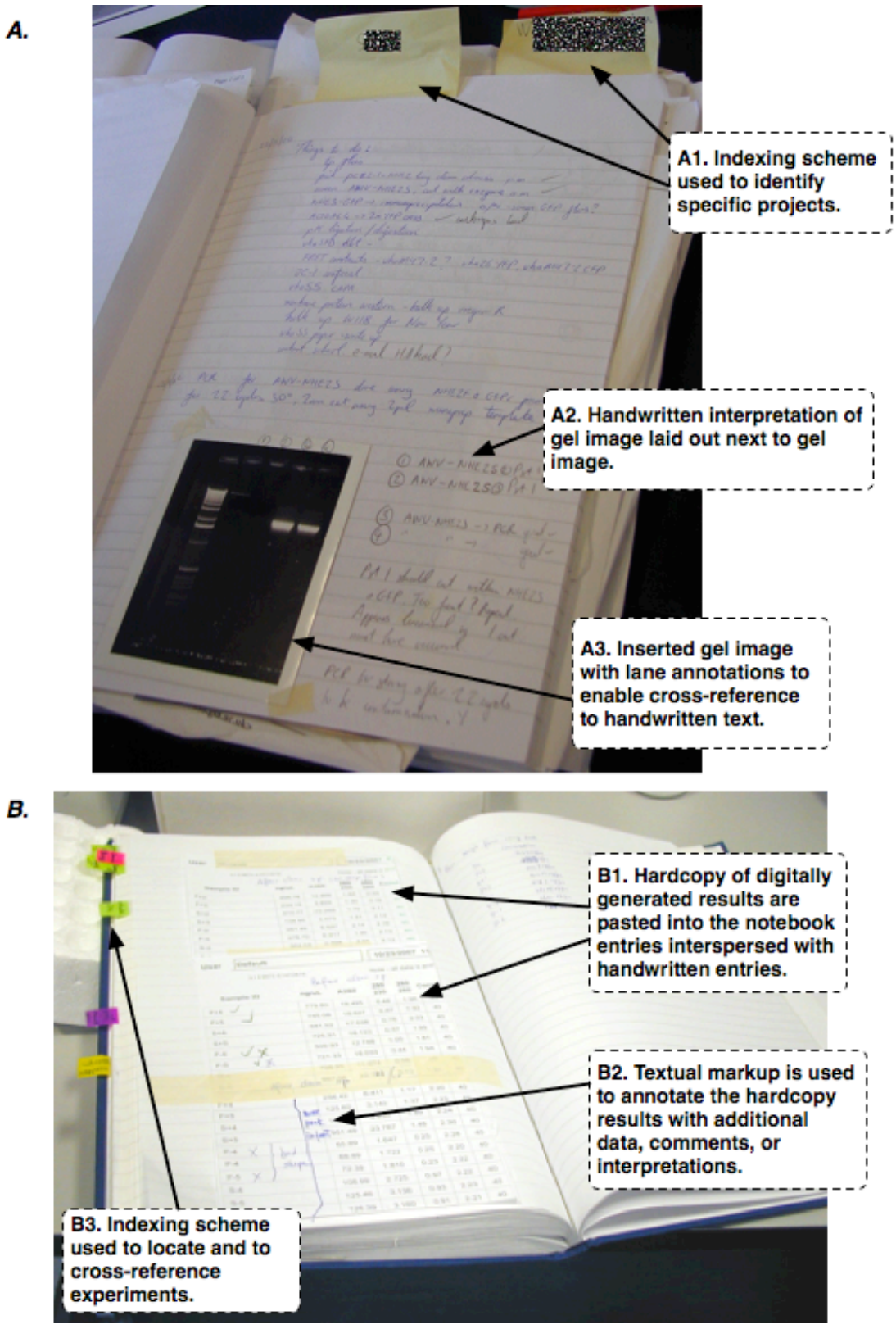


Figure 3-4: Composition of laboratory records

Depicting aspects of manifest intertextuality in the composition of laboratory records in notebooks used within the research- and service-orientated laboratories. (*A.*) Photograph shows an entry containing an electrophoresis gel image in the notebook of a postdoctoral researcher in research laboratory *ES-L2*. Note that names on the index tabs have been blocked out to preserve anonymity. (*B.*) Photograph shows an entry containing results generated by a NanoDrop® instrument in the notebook of a technician working in the service laboratory *ES-L3*. This hardcopy was generated by a proprietary software application that displays and prints the results obtained by the instrument.

It is essential to recognize that the study participants were not technophobes; instead there was a straightforward desire to ensure that the adoption of digital tools did not affect the quality or reliability of their data. The data are the primary artefact of the bioscientists' work, and any supporting technology must not interfere with the quality of the data.

3.3.5 Workflow for recordkeeping

3.3.5.1 Chronological and experiment-focused records

Much of the experimental work observed in all four participating laboratories involved natural intervals when a reaction had been started but the laboratory worker would have to wait an extended period of time for the reaction to complete. In order to manage their workload, many researchers and technicians routinely used these gaps as an opportunity to run multiple experiments simultaneously.

Interleaving experiments also engendered different approaches in the study participants in terms of the way they made use of their laboratory notebook. A majority of study participants, including some of the research staff and all of the technicians in both the research and service laboratories, preferred to start the record of each experiment on a new page. The amount of space needed to complete the record would be estimated in order to leave sufficient blank space between entries so that subsequent notes could be added for each experiment as necessary. This approach ensured that the record of each individual experiment was presented as a single block of text; discrete entries in this block of text were accrued over time. Some of the postdoctoral researchers and postgraduate researchers in each laboratory preferred a 'running sequence' approach and would order the entries in the laboratory notebook in the same sequence in which the associated task was performed. This approach led to disjoint records of each individual experiment that were interleaved through each other. The trade-off in this respect concerned the prioritization of the current use of the record over the retrospective use of the record by simplifying the production of experimental records at the expense of later interpretation.

3.3.5.2 Finished product recordkeeping

Variation in laboratory recordkeeping was not limited to the structure and content used to record experimental work, but was also evident in terms of the dynamic sequencing used

by study participants to construct the laboratory records held in the notebooks. Whilst all participants conceptualized the laboratory notebook as the primary record of their experimental work, the records contained in the notebook were not always constructed contemporaneously with the execution of the experimental work.

In the case of the majority of the study participants including all principal investigators and technicians, the laboratory notebooks were maintained in real-time with the experimental work recorded directly in the laboratory notebook as and when it is performed. These were truly authentic records in the sense of records management theorists, and represented the consistent approach advocated by supervisory staff in each of the participating laboratories who stressed that laboratory notebooks should be used:

> "to record (.) everything (.) at the time (.) things are done (.) so that there are no scrap bits of paper, there is no final version of the lab book (.) there is one lab book that has all the notes (.) on a continuous basis."
>
> [Principal Investigator in Research Laboratory]

However, some of the study participants interpreted the role of the laboratory notebook in a different manner by viewing it as a 'finished product' record that presents a polished presentation of the experiment. These participants made use of a temporary notepad to record the experimental work whilst it was performed. At some later point, which could range from later on during the same day to up to a week later, a fair copy of the experiment would be transcribed to the laboratory notebook often with shorthand notations from the notepad expanded to a fuller text. It was particularly important to note that some proponents of this two-tier approach based on a 'working copy' and a 'finished product' made use of the transcription activity as an opportunity to validate their work as well as a means of avoiding the overhead of detailed record keeping whilst performing experiments.

Keeping 'working copy' records in notepads was, however, not endorsed by the principal investigators as illustrated by the comments:

> "I get annoyed when people don't make a copy as you go along because you'll end up with somebody with a book full of loose pieces of paper then."
>
> [Principal Investigator in Research Laboratory]

3.3.6 Derived information products

The notebook served as the primary, if not necessarily contemporaneous, record of the experimental work conducted by all of the study participants. In each of the participating laboratories, a range of additional information artefacts, both written and verbal, were derived from this primary record including progress reports for industrial research partners, client report forms for the service laboratory, papers for publication in scientific journals, and presentations at laboratory meetings and conferences. In contrast to the notebook entries, the structure and content of these other information artefacts were more explicitly standardized and typically defined by an external agency such as the industrial partner or the publication committee of a scientific journal.

From the viewpoint of the supervisory staff, these derived work products served as a preferred means of communicating experimental data in a written format as illustrated by the comment:

> "and y'know we've rarely needed to go (.) to lab books because (.) eh (.) y'know we publish quite well so we're generally on top of things (.) not everything that somebody's ever done in their entire career in the lab (.) but the essential things."
>
> [Principal Investigator in Research Laboratory]

These derived information artefacts were both reductive and supplemental in comparison to the experimental records in the laboratory notebooks. On the one hand, the derived products were reductive in that they tended to summarize a set of experiments. Consequently, none of the study participants, and particularly not the principal investigators, considered these derived genres as replacements for the records in laboratory notebooks as illustrated by the viewpoint:

> "but it is important to have the raw data in case anything goes wrong to (.) to backtrack."
>
> [Principal Investigator in Research Laboratory]

Many of the derived products such as progress reports provided an essential supplement to the laboratory records in the sense that these reports would also include interpretation and reflective thinking. Initial inspection of a subset of the notebooks produced by the study participants indicated interpretation and reflective thinking was a component of experimental work that was missing in many of the records produced by the majority of the

study participants. It is important to note that this omission was expected by the principal investigators who emphasized the interpretation of experimental results as one component of an experimental record that they would not necessarily expect to find in a laboratory notebook as shown by:

> "and especially the interpretation (.) the interpretation is largely held in the (.) brain rather than formally written down (.) it only really gets written down in papers and grant application and things."
>
> [Principal Investigator in Research Laboratory]

In the absence of these derived information products, the expectation of the principal investigators was that the interpretation of results remained unwritten and instead:

> "just goes in to the (.) the lab (.) consciousness (.) which at some point is me ((the principal investigator himself/herself)) (.) it may or may not persist out there ((with other members of the research group))."
>
> [Principal Investigator in Research Laboratory]

3.3.7 Sharing data

3.3.7.1 Group data resources

Web-based Intranet facilities were available at each of the participating laboratories, and research laboratories *ES-L1* and *ES-L2* had both attempted with varying levels of success to make use of these facilities to establish a group-wide resource to promote sharing of experimental data, experimental protocols, and other information such as published papers. No such initiative was in place within research laboratory *ES-L4*. Service laboratory *ES-L3* operated to a different communication model relying on Web-based services to source externally defined protocols from commercial vendors, and direct communication with clients to exchange data.

Within research laboratory *ES-L1*, a senior technician *ES-L1R4* had taken the initiative to set up a group resource in order to promote the sharing of data between the various researchers active in the group. This resource had been announced to all members of the group, and each of the study participants from *ES-L1* was aware of the resource and its intended purpose. It was interesting to note that none of the study participants were considering publishing their experimental data via this group resource. This opinion was not unique to the study participants but was common to many of the staff in that

laboratory. Accordingly, use of the group resource as a means of sharing data at the time of this study had not been wholly successful as illustrated by the following comment.

> "I think the technology is there (.) I mean we've got (.) like a research drive that all the people in here can access it (.) I don't actually know if there's anything on it (.) I mean I got it set up (.) and the idea was that they could drop their data in (.) so that they could take a look at it."
>
> "This is where you start to despair (.) because (.) I'm the one that had to deal with the guy ((from computing/IT support))."
>
> [Technician in Research Laboratory]

A different emphasis was observed within the research laboratory *ES-L2*. In this case, a similar Intranet-based group resource had been set up on the initiative of one of the principal investigators to promote sharing of data, publications, and protocols. Many of the study participants working in this laboratory made use of this group resource to publish and to retrieve information within the laboratory group. However, it was not the case that all experimental data, either raw or processed, was routinely published to the laboratory group resource. Furthermore, relying on this group resource was not without the usual problems associated with computer-based solution as when the website went offline during the course of the study:

> "everyone was stumbling about looking for paper copies ((of experimental protocols))."
>
> [Principal Investigator in Research Laboratory]

In addition to this local group resource, one of the principal investigators in research laboratory *ES-L2* had been instrumental in setting up a database resource termed FlyAtlas (Chintapalli *et al.* 2007) for gene expression data that aimed to serving the wider *Drosophila melanogaster* research community. It is important to point out that the intention with FlyAtlas was not that research laboratory *ES-L2* would publish all or even much of its data to the wider community. Instead, the intention was that a specific set of what one of the principal investigators labelled as "background data" could be made available to the wider community, with the request that any use of FlyAtlas be referenced in published work and the request that other groups also submit their "background data" to this centralized resource.

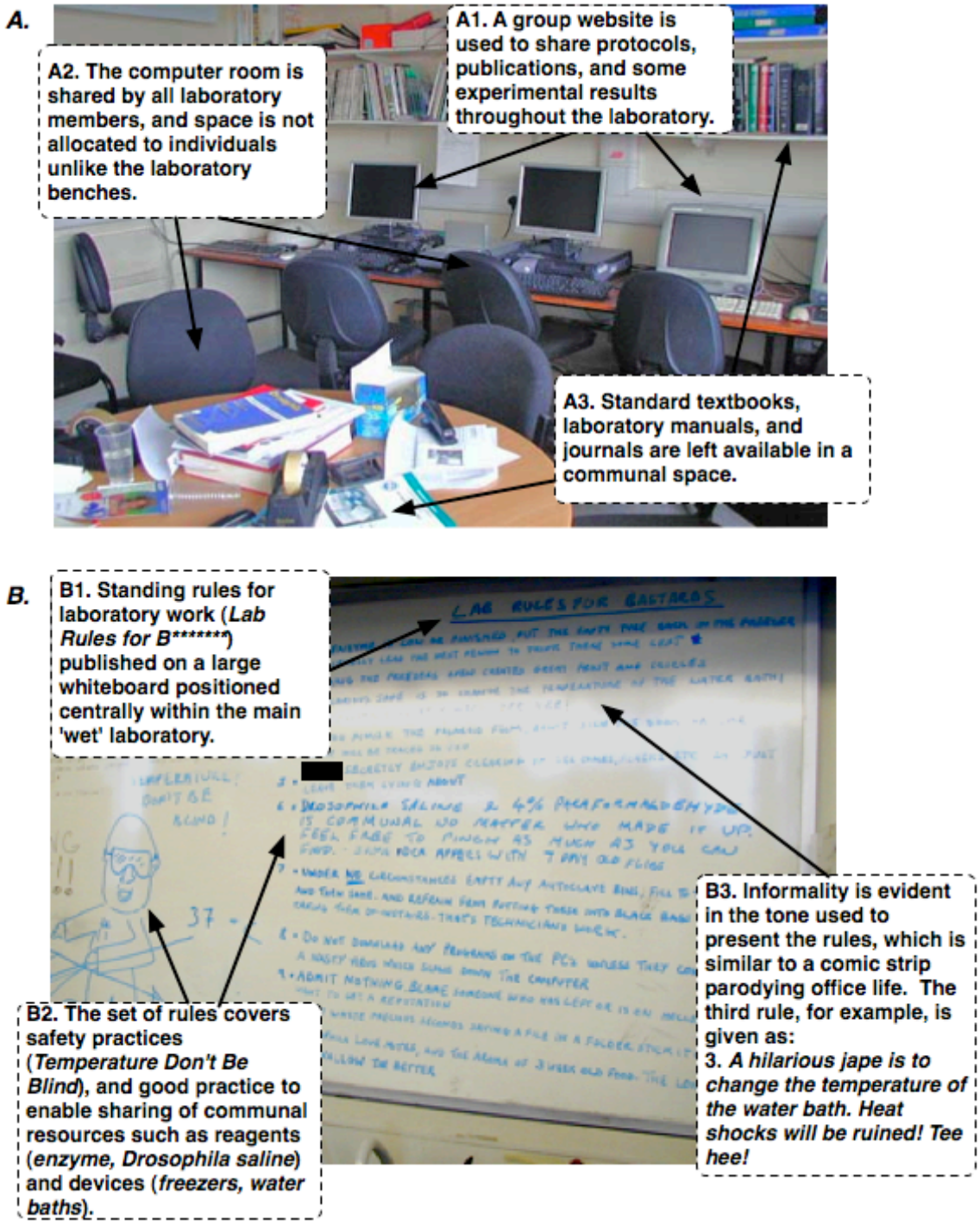


Figure 3-5: Group data sharing within the laboratory settings

Depicting some of the tools in use within the laboratory settings to promote sharing of information between laboratory members. (*A.*) Photograph shows the communal computer room/social room within research laboratory *ES-L2*. (*B.*) Photograph shows a large whiteboard that is located centrally within the main 'wet' laboratory area of research laboratory *ES-L2*. The writing on this whiteboard was a spontaneous reaction by an experienced laboratory staff member to codify correct behaviour within the laboratory in terms of resource sharing[78]. A participant name in the laboratory record has been blocked out to preserve anonymity.

[78] Apologies to those offended by the title shown on the whiteboard. A number of the bioscientists observed during the course of the study displayed an impressive ability to combine the sacred and the profane in their choice of language, often relating to the outcome of their experiments.

Computer-mediated information exchange was not the only approach in place with the participating laboratories as illustrated in Figure 3-5. Informal verbal exchanges were commonplace in the communal office space[79] of laboratories *ES-L2*, *ES-L3*, and *ES-L4*. Research laboratories *ES-L2* and *ES-L4* both held weekly group meetings, which all group members were required to attend. As part of these weekly group meetings, one of the postdoctoral researchers, postgraduate researchers or technicians would give a presentation typically lasting around forty-five minutes concerning their recent work[80]. Interestingly, no such group meeting took place in research laboratory *ES-L1* at the time of the study, although postdoctoral researcher *ES-L1R1* was attempting to arrange such a meeting in negotiation with the principal investigators in the laboratory.

3.3.7.2 Personal data resources

Study participants in all of the participating laboratories made extensive use of publicly available information tools such as the National Centre for Biotechnology Information[81] (NCBI) database to access genetic sequences and other data when planning their experimental work. For some of the researchers observed during the study, this activity took up the majority of their working day. For example, postdoctoral researcher *ES- L1R1* spent approximately 60% of his time engaged in searching through published databases to retrieve information necessary to design his experimental work. Other study participants were observed to be at a different stage in the lifecycle of their projects and so devoted considerably less time to this activity.

Retrieving the necessary information often involved negotiating multiple databases and Internet websites, and combining the potentially conflicting results obtained from searching these tools. Conflicts in this sense encompassed both contradictory information and incompatible data formats. In order to leverage the considerable investment made in retrieving experimental planning information from multiple external sources, all of the

---

[79] And in the nearby coffee shop in the case of research laboratory *ES-L2* where many of the members of that group would gather each morning after setting up their initial experiments.

[80] The range of presentations delivered during these meetings corresponded with Weissberg's (1993) categorization of graduate seminars into the four genres of project proposals, preliminary literature reviews, in-progress reports, and completed research reports.

[81] The NCBI was founded as a branch of the United States National Library of Medicine, and houses a number of important resources including the GenBank DNA sequencing database and the PubMed index of research articles. As part of its role, NCBI coordinates its DNA databases with those of other territories including the European Molecular Biology Laboratory (EMBL). Further information is available at http://www.ncbi.nlm.nih.gov [accessed 01 March 2011].

study participants maintained personal databases of this information, often as some form of spreadsheet. For example, all of the postdoctoral researchers and postgraduate researchers in the research laboratories maintained a database of primer[82] information for use in their experiments. These databases contained both information pertaining to the current experimental work, and historical information accrued from previous projects. Although most researchers maintained such information, each participant personalized the structure, content and representation used to capture this type of information as illustrated in the following excerpt:

> "If you've seen Thomas[83]'s spreadsheet (.) Thomas is very, very meticulous about the information he keeps in it. My view is different from Thomas (.) I'm sure I'll probably be proved wrong in the future is that he's got too much there (.) he's got information in that spreadsheet that I would never use (.) maybe because I don't (.) wouldn't know how to use it but (.) eh, so I'm nowhere near as meticulous as that. I've the primer sequence and any other basic info that I would want to know about the primer and that's about it (.) there's nothing (.) Thomas's is a huge big database, I don't have anything near that (.) and it's all self-defining the info that I can see myself using."
>
> [Postdoctoral Researcher in Research Laboratory]

3.3.7.3 Controlled data sharing

A common theme in the development of computing support for scientific communities such as the molecular biology laboratory is that of promoting data sharing. This is typified by the work of the UK e-Science programme. Interestingly, all of the research-orientated study participants expressed strong reservations about sharing their data with others since it is seen as their main economic resource. Data plays a key role in determining their publication prospects, grant funding prospects, and promotion prospects. This view prevailed across all roles within the research laboratories from technicians to principal investigators, and is illustrated by the following typical comment:

> "There's certainly no way I would release unpublished data to anyone (.) uh, even intentions for work that you plan to do I'd be guarded over because you don't want (.) I've seen it in our department and others (.) of course you don't

[82] A primer is a short nucleotide sequence that is complementary to a target sequence and enables a polymerase to begin synthesis of a DNA chain as in a PCR.

[83] Again, all names have been changed to protect the privacy of the participants. Thomas is chosen in honour of Sir Thomas Beecham, who conducted the Royal Philharmonic Orchestra in a performance of Handel's Messiah that also figured frequently in the soundtrack to the writing of this thesis.

> want people to nick your ideas if it's been you that developed them (.) like in F1[84], MacLaren nicking Ferrari's ideas."
>
> [Postdoctoral Researcher in Research Laboratory]

The economical attitude towards data sharing also extended to the use of publicly available databases such as the NCBI database for sequence information. Although extensive use is made of databases such as the NCBI database for retrieving information during the planning and design of experimental work, use of these resources was considered to be 'one way traffic' by the study participants. A common attitude towards reporting newly discovered sequence information back to the community resource is illustrated by the comment:

> "Why should I spend time writing into NCBI when I get no credit for it (.) so I'll keep it to myself and get on with things."
>
> [Postdoctoral Researcher in Research Laboratory]

3.3.8 Sharing protocols

3.3.8.1 Sharing protocol via written archive

Written definitions of common laboratory techniques were available in each of the participating laboratories both as locally defined protocols, and externally defined protocols such as those found in textbooks or those available on-line from bioscience community websites. These written definitions offered a means of sharing and standardizing the experimental techniques that were used within the laboratory.

However, all study participants commented that it was relatively unusual to learn from a written definition when setting out to learn a new technique. Instead, the routine approach to learning experimental techniques was to track down an individual in the same or another laboratory who is already experienced in performing the technique and observe that individual executing the technique. Whilst watching the demonstration, the novice would make his/her own written definition of protocol, possibly in a laboratory notebook or possibly in a separate 'methods book'. This approach enabled the novice to document the protocol definition using structures and representations that would be meaningful to the

[84] The quotation refers to a dispute between two Formula 1 motor-racing teams in which one team claimed that technical details of their design had been given without permission to another team. For those interested, more details are available from the Fédération Internationale de l'Automobile at http://www.fia.com [accessed 01 March 2011].

individual. All study participants preferred the use of demonstrations both as a means of learning new techniques and as a means of teaching new techniques.

The study participants in each of the four laboratories commonly referred to one particular textbook as the "Bible" for molecular biologists. Sambrook and Russell (2001) are the authors of the most recent edition of this 'charter document'[85], although many of the study participants still referred to the textbook as "Maniatis" after the first author of an earlier edition. As illustrated by the comment below, most workers in the field did not use the protocols in this textbook directly but had instead developed their own variations on the core protocol that was defined in the "Bible"[86]. In this sense, the written definition in the textbook served more as a detailed record of experience and a source of detailed guidelines.

> "And it's funny (.) the one kinda (.) book that we might go to if we're struggling is Maniatis, the textbooks in the lab, Maniatis they're known as ( ) I think Sambrook is the first author now but (.) again it's funny (.) we had a visitor over from China recently, a very high profile geneticist who was saying he had no-one teach him any molecular biology (.) his way of learning was he would read all of Maniatis until he learned about everything (.) and it's funny because everyone professes that this is the book that they'll go to but (.) equally things that Levi[87] or others have taught me in the past (.) I'll often will go to Maniatis now without having read it all thoroughly but if there's something that's troubling me, I'll maybe go there as a first point of reference and often it's different from the way that other people taught me to do things and then I wonder that it's just Chinese whispers (.) y'know where each person's changed something very slightly you end up with something very different from what the standard might be."
>
> [Postdoctoral Researcher in Research Laboratory]

An essential part of research articles published in the field of molecular biology is the description of the materials and methods used in the experimental work. In principle, this section of any paper also provides a written definition of a protocol that can be used to replicate the work that is being described in that paper. Hence, published papers also serve as a means of sharing and standardizing protocols for laboratory work. For example, two

---

[85] The term 'charter document' is used in the sense of McCarthy (1991) to indicate a document that has a dominant influence on its field. Typical examples include religious texts such as the Bible, Tanakh, or Qur'an.

[86] In molecular biology as in theology, personal interpretation of the "Bible" is not without consequences.

[87] All names have been changed to protect the privacy of the participants. Continuing with the musical theme, Levi is chosen in honour of Mr Levi Stubbs, singer with The Four Tops, who also contributed to the soundtrack for this thesis.

of the study participants *ES-L2R1* and *ES-L2R2* were observed using a research article published by another research group from an external laboratory as the protocol definition for one aspect of their experimental work before switching to the use of a kit manual. In addition, principal investigators within *ES-L2* also relied in some cases on their own published articles as a means of standardizing protocols that have been defined within their own organization as shown by the comment:

> "I mean in all our papers there'll be (.) em (.) some form of the ICC[88] protocol, it's usually shortened to reflect the one that was published in 2001(.) now that this is the one that everyone uses but they refine."
>
> [Principal Investigator in Research Laboratory]

All of the study participants maintained a personal archive of their own variations of the core laboratory protocols, typically in a separate 'methods book'. Initial inspection of a range of laboratory notebooks produced by the study participants highlighted the fact that there were very few, if any, standalone protocol definitions included in the primary laboratory notebooks. In many cases, these personal copies of the standard protocol had been annotated with handwritten notes to record personal variations on the standard or to highlight critical steps in the protocol when extra attention would be required during execution of the protocol.

3.3.8.2 Sharing protocols via demonstration

All study participants from each of the laboratories demonstrated a clear preference for visual demonstrations of how things are done over written or verbal explanations of how things are done[89]. Further, all study participants had considerable experience of learning experimental techniques by observing a more experienced worker. This learning process typically consisted of observing an experienced worker perform the technique, then performing the technique under the watchful eye of the tutor, and finally being let loose to apply the technique independently. It was interesting to note that this attitude to teaching

[88] ICC stands for immunocytochemistry, a laboratory technique that is used to assess and visualize the presence of specific proteins within a cell by introducing specific tagged antibodies that will bind to the protein of interest.

[89] In this way, teaching laboratory protocols may offer a solution to the 'say-do problem' (Goguen and Linde 1993).

techniques in laboratory science has recently motivated the inception of an on-line journal named JoVE[90] for publishing bioscience experiments and research using video.

Many core laboratory tasks such as using pipettes to load large numbers of samples into electrophoresis gels required a certain degree of manual dexterity. Demonstrating protocols to novices also seemed to promote the transfer of the practical skills or 'tricks of the trade' that facilitated these tasks but were typically not found in the written inscriptions of laboratory protocols. When demonstrating protocols, experienced laboratory workers routinely passed on such tricks that improved the efficiency of the task or reduced the chances of errors occurring in the execution of a laboratory protocol[91].

One of the benefits of this approach was perceived to be the standardization of a "lab way of working" by ensuring that the current generation of workers in any given laboratory are responsible for training the next generation of workers. Hence, preferred protocols for performing experimental work could be propagated throughout the laboratory. It seemed somewhat ironic, however, for biologists to rely on this as a means of standardization given their understanding of the mutations that can occur from one generation to the next. A complicating factor in this approach concerned the means of rationalizing the advice gained from multiple experts, even in the case of multiple tutors within the same laboratory. The typical experience of many of the study participants is captured in the comment:

> "Talking about standards there (.) I was just learning this the same as Renaldo [92] by listening to people, having people like Lawrence or Abdul tutoring me (.) and I could see straightaway that different people were telling me different things (.) how things should be done (.) and in terms of a standard way to do things, there didn't seem to be one … I still feel like I don't really know,

---

[90] Examples of these video research articles in JoVE are available at http://www.jove.com [accessed 01 March 2011].

[91] An example of this witnessed during the course of the study involved loading samples into multiple 96-well plates. These 96-well plates are laid out as a two-dimensional grid of 8 x 12 wells; quantities of the sample solution and other reagents have to be individually loaded into each well using a pipette. This can be a tedious process, but should you be interrupted whilst midway through the process of loading each well, it is essential that you remain aware which of the wells you have loaded and which remain to be loaded. A common approach used by experienced workers to handle possible interruptions is to align the 96-well plate with the box holding the 96 pipette tips that will be used to load the reagents; the tip holder box is laid out in the same 8x12 grid as the plate. At a glance, it is possible to tell whether a well has been loaded since the tip at the matching position in the tip holder box will be missing.

[92] All names have been changed to protect the privacy of the participants. Here are the other three members of The Four Tops, whose 1966 live recording from the Roostertail contributed to the soundtrack for this writing the discussion for this thesis.

> y'know, what's the justification for the rules (.) how (.) y'know, why do they do it one way (.) mostly I think it's just based on what works for one individual person."
>
> [Postdoctoral Researcher in Research Laboratory]

One particular issue with this approach to propagating experimental protocols within a laboratory concerned the ability to track down the local expert who is capable of demonstrating a protocol to a novice. Often, a crucial component of this approach was the role commonly played by the principal investigator and other long-serving members of a laboratory who served as the "lab consciousness" and could direct novice users to the local expert for a given technique.

> PI: "… so check with her what it is you're meant to be looking at (.) which antibody you're using and that will then (.) em (.) direct you to which particular refinement of the ICC protocol to use Having said that, Vivien[93] is the queen of ICC ((laughing)) for some things (.) for some of the proteins so y'know (.) maybe just -"
>
> PG: "Yeah, she's trying to find the protocol so (.) she says she's got it somewhere so I'll leave her to find it."
>
> PI: "Yeah, OK (.) most people in the group should have one stuck in their book somewhere, OK.
>
> [Principal Investigator and Postgraduate Researcher in Research Laboratory]

[93] All names have been changed to protect the privacy of the participants. Vivien is chosen in honour of Vivien Ellis, the vocalist with the Dufay Collective whose repertoire of early music also contributed to the soundtrack for this writing the discussion for this thesis.

3.4 Discussion

Three research questions were posed for this investigation. The following discussion evaluates the results of this ethnographic study of laboratory recordkeeping with respect to each of these questions in turn. Given the strength of ethnography as an approach to understanding socio-cultural practices, the findings of this particular study are directed principally towards the first research question. Greater focus is placed on the other two questions by the subsequent genre analysis and reading protocol studies.

1. *What roles do laboratory records play in the discourse of academic molecular biology laboratories?*

The study confirmed the central role assumed by laboratory notebooks as the primary written record of day-to-day experimental work kept by technicians, postgraduate researchers, postdoctoral researchers, and those principal investigators who were still directly involved in bench work within each of the participating laboratories. This applied within both the service-orientated laboratory and the research-orientated laboratories.

However, as identified in Figure 3-6 overleaf, the laboratory records in these notebooks formed only one part of a wider genre system encompassing a range of information artefacts in multiple semiotic modes to support diverse organizational functions. The genre systems at work in each of the research laboratory settings were broadly similar but not identical. In this sense, Figure 3-6A should be considered illustrative of the research laboratories. In particular, research laboratory *ES-L1* did not, at the time of this study, conduct regular laboratory group meetings that required research staff to present 'lab talks'. Figure 3-6B presents the system found within the service laboratory *ES-L3*.

The central role played by the laboratory records was manifested in their involvement in multiple patterns of interaction within these genre systems to support a range of communicative functions. In particular, the study identified specific patterns of interaction in which notebook records supported the planning and design of experimental work, the reporting and monitoring of experimental work, and the learning and execution of experimental work.

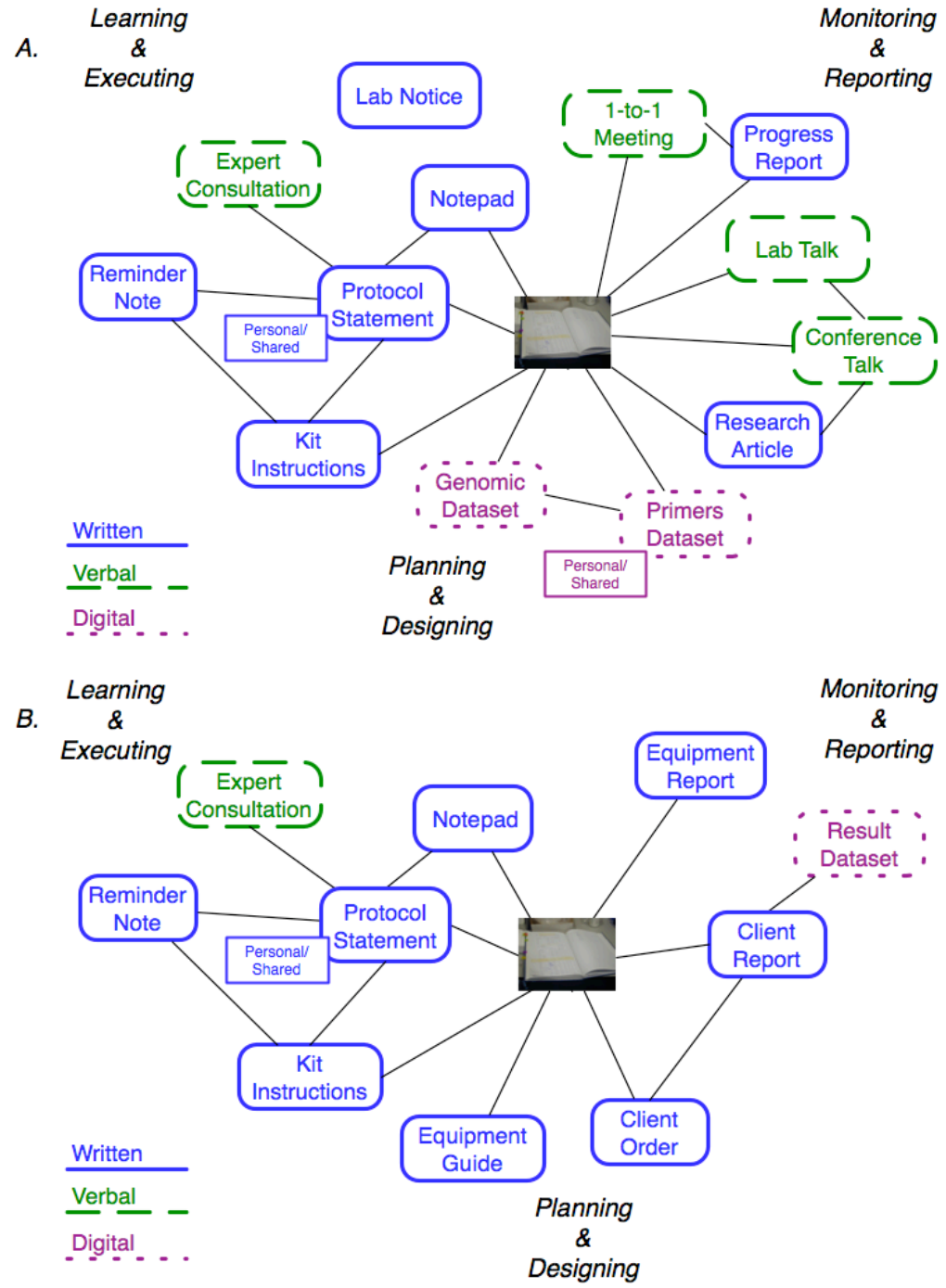


Figure 3-6: Genres in the laboratory settings

Summarizing the genres at work in the laboratory settings together with the mediatory relationships between these genres and the records held in laboratory notebooks. The diagram is presented as an informal genre ecology framework as in Spinuzzi (2003) but extended to represent semiotic mode, and shared/communal instantiation of genres. The semiotic mode (written, verbal, digital) of each genre is indicated by colour and line style. *Protocol statements* and *Primers datasets* are instantiated within the laboratories as both personal copies owned by individual scientists, and shared copies owned by groups of scientists. Connecting lines indicate patterns of mediation in which the connected genres are used jointly to perform some laboratory activity. The principal type of activity associated with subgroups of genres in these frameworks is indicated as either learning how to and performing experiments, planning and designing experiments, or monitoring and reporting experiments. (*A.*) Genre framework in the research laboratories *ES-L1* and *ES-L2*. (*B.*) Genre framework in the service laboratory *ES-L3*.

In some of these interactions, information from other genres was recontextualized[94] for inclusion within the laboratory records; most interactions involved recontextualizing information from the laboratory records towards other genres. In this way, the notebook records could act as a conduit for propagating certain types of information within the laboratory settings. In particular, the study highlighted the importance of these patterns of interaction as a means of transforming laboratory records via multiple genre chains into 'public genres' directed at wider audiences. This included transformation into both written and spoken modes, and transformation aimed at audiences both within and beyond laboratory boundaries.

These transformations were performed to support both the reporting and monitoring of experimental work, and the learning and execution of experimental work. For example, selected information held in the laboratory records was regularly used in the construction of 'lab talks' or oral presentations given to the local laboratory group. These talks were used to both report and affirm individual progress, as a means of propagating technical knowledge regarding variations on protocols and experimental kits, and as a means of discussing/solving problems associated with the experimental work on a local community basis (*cf*. Jacoby and Gonzalez 1991; Weissberg 1993). Via separate genre chains, selected information from laboratory records was also used in the construction of research articles for the wider scientific community and written progress reports for industrial sponsors. This observation is in line with recent studies of recordkeeping in academic laboratory settings such as Shankar (2007) and Wickman (2010) who both report the use of research articles and other publication-orientated genres for communicating parts of the data held at source in laboratory records to the wider scientific community.

It is important to recognize that these derived information products were simultaneously reductive and supplemental to the experimental data held in the laboratory notebooks. Reductive in the sense that the public genres typically presented a processed version of the data held in the notebook records; supplemental in the sense that the public genres often included data rarely recorded in the notebooks such as result interpretations and reflective thinking.

---

[94] Recontextualization is used in the sense of Linell (1998:144) as "the dynamic transfer-and-transformation of something from one discourse/text-in-context to another", *cf*. Goffman (1974)'s concept of reframing.

The study highlighted an almost exclusive reliance on these other genres as the means of communicating the information held in records in laboratory notebooks. In this sense, one of the most striking findings from this study was the minimal use of laboratory records in their direct written form by laboratory staff other than the original author of the record. This finding offers a counterpoint to the findings from other studies, notably a number of studies undertaken as part of technology developments for electronic recordkeeping that report sharing of laboratory records or assume laboratory records to be a shared resource (*e.g.* Arnstein *et al.* 2002; schraefel *et al.* 2004; Tabard *et al.* 2008). Within the service-orientated and research-orientated laboratory settings examined for this study, only those notebooks written by authors who had since left the laboratory were read as written records, typically only by principal investigators, and then only if the information sought was not available in other derived products such as research articles. It was not the case that research staff simply did not require to read other scientists' laboratory records. More fundamentally, it was that research staff did not expect to be able to make sense of other scientist's records, and that it would in some circumstances breach laboratory etiquette to read the notebook of another scientist.

Laboratory records are not entirely unique in this respect, and the sharing of various types of scientific data within academic settings has proven a source of contention (*cf*. Borgman *et al.* 2007; David 2006; Research Information Network 2008, 2010). It is important to recognize that the availability of technology to enable sharing is often not the prime concern in this respect. Instead, the concerns centre on more fundamental issues including questions of ownership, ethics, publication rights, and the definition of meaningful data sets to exchange[95]. This includes sharing internally within a laboratory, and external sharing. In contrast to the controlled approach to sharing experimental data, most laboratory members readily participated in the exchange of experimental methods, often in the role of experts training novices internally within a laboratory group or research division.

In principle, the laboratory records held in notebooks delivered significant evidential value to the research laboratories by performing an established communicative function as proof

[95] The author of this thesis has vivid memories of attending more than five presentations by e-Science representatives within university bioscience faculties. The meetings were attended primarily by principal investigators, and the e-Science representative discussed the available technology for centralized databases and Web-based communication. The series of questions following these talks adhered to same pattern by wholly ignoring the technology, and focusing on why would a scientist want to share data and how could you ever define a dataset that would be meaningful to others.

of precedence. In practice, principal investigators and other research staff relied on research articles as the primary basis for establishing the priority of their work and securing the rewards of authorship in the academy (*cf.* Biagioli 2003). Of course, it is essential to distinguish in this respect between different types of laboratory, and to acknowledge that some laboratories direct their research towards commercial exploitation. Commercialization was, however, not the focus of the "little science" research laboratories participating in this study, although research laboratory *ES-L2* was engaged in some projects with industrial sponsorship. Interestingly, a specialized genre for progress reports to industrial partners was at work for these projects, which also included recontextualized summaries of selected data from the notebook records.

In practice, the laboratory records kept in notebooks also delivered significant informational value to the authors themselves. This included supporting multiple organizational functions both during and after the execution of experimental work. In particular, the dynamic process of accumulating the entries in the record over time provided a progress monitor that was used by laboratory staff to manage their experimental work. Most importantly, the laboratory record offered a single text or "compound document" in which to collate information from multiple sources such as laboratory devices, reagents, digital results, written protocols, and colleagues' advice whilst managing the complexity of daily laboratory practice (*cf.* Latour and Woolgar 1986; Knorr Cetina 1981, 1999; Wickman 2010). In addition, the laboratory records functioned as a written record of completed work that was consulted retrospectively by the author. Typical examples of retrospective use included retrieving the experimental conditions applied in previous work, and identifying the results obtained by previous experiments.

In each of the four laboratories, the experimental records were kept in bound, paper notebooks. It was clear, however, that technology solutions had been adopted for some aspects of laboratory work. This was typified by the use of spreadsheets in each of the laboratory settings to maintain information on DNA primers, and by the use of computing systems for generating certain types of experimental result such as electrophoresis gel images. It was interesting to note, however, that the widespread adoption of information technology for data such as DNA primers within the laboratory settings had not translated to sharing of this data between laboratory members, even between those members working on related projects. Instead, each individual maintained his/her own personal spreadsheet

so that the use of information technology was focused on the organization and storage of this personal information.

2. *What are the structures, content, and representations that characterize the genre of laboratory records in academic molecular biology laboratories, and to what extent do these vary across different contexts of use?*

The study identified a central tension in which the laboratory records held in notebooks were conceptualized on the one hand by principal investigators as a 'community archive', and on the other hand by postdoctoral and postgraduate researchers as a 'personal resource'. Other studies have reported a similar tension. For example, Tabard *et al*. (2008:569) report in their study of research at the Pasteur Institute that "although usually viewed as the personal record of an individual scientists, lab notebooks are written to be read by others, imposing a corresponding discipline in the style and choice of what is recorded". In a study of laboratory recordkeeping from an information science perspective, Shankar (2007:1457) similarly reports that "these documentary products of daily activity represent tensions between standardization and flexibility, the collaborative nature of science and the practical and personal needs of the individual scientist". Whilst the legal ownership of the laboratory records is clear in that the laboratory has full or joint ownership, it is a question of a different order as to whether the records are actually capable of forming a reliable and usable community archive that could be read by others. As mentioned previously, from a records management viewpoint (*e.g.* Reed 2005; Shepherd and Yeo 2003), a reliable record is one that provides a full and accurate description of the transaction that it embodies, and a usable record is one that can support future organizational activities in multiple contexts of use.

In this sense, it is essential to recognize the impact of the 'personal resource' viewpoint on the structures, content, and representations used in laboratory records. The laboratory notebook contributed significantly to the sense of identity felt by individual laboratory members. Individual researchers employed personal approaches to multiple aspects of laboratory recordkeeping including the use of personal shorthand and other descriptive economies when writing the experimental records, the use of personal indexing schemes to store and retrieve the records, and the use of varying levels of abstraction to document the experimental work. In addition, individual researchers adopted different schemes to manage the temporal intervals that could arise between different steps in experimental

work. In one style of recording, authors produced disjointed records in which separate entries for a single experiment could be distributed over multiple pages.

No explicit standards or guidelines were employed in any of the participating laboratories to constrain the structure, content, or representations used in laboratory notebooks. In contrast, many of the public genres such as external progress reports that could be derived from the personal notebook records were subject to a greater degree of explicit standardization. In many cases, the structure, content, and representations in these derived information artefacts were typically imposed by an external agency such as an industrial partner, or by journal editors in the case of research articles. In addition, the production of many of these derived information artefacts often involved collaborative authorship with both internal colleagues and potentially external collaborators. The production of laboratory records did not involve collaborative authorship.

However, all researchers and technicians espoused a consistent understanding of the 'gold standard' genre structure for laboratory records, which was centred on the IMRD move structure typical of research articles in many of the biosciences (*cf*. Gross *et al*. 2002; Kanoksilapatham 2005; Myers 1990, 1991; Samraj 2005; Swales 2004; Valle 1999). Although all laboratory staff expressed a remarkably consistent concept of the information required in a laboratory record, writing records involved considerable contextual competency by requiring individual authors to judge when experimental data and/or method becomes sufficiently "difficult" or "unusual" to require it to be recorded in the notebook.

It is interesting to note that this 'gold standard' structure for a laboratory record differed from the pedagogical genre of laboratory reports used in undergraduate science (*cf*. Braine 1995; Dudley-Evans 1985) by omitting some of the structures such as abstract, theoretical background, and conclusion that established the wider context in which the experiment was undertaken.

3. *How do readers of laboratory records in academic molecular biology laboratories make sense of laboratory records in different contexts of use?*

As discussed in relation to the first research question, the majority of scientists relied on recontextualization of the data held in laboratory records into other public genres as the means of retrieving the information held in the records in laboratory notebooks. The

original author performed this recontextualization, and the reader made sense of the derived information artefact.

Direct interpretation of the laboratory records, in the limited circumstances when this occurred, relied on prior understanding of the context in which the experimental work was carried out together with domain-specific knowledge of the schemata appropriate for specific types of experimental work and a shared notion of the genre of laboratory records. Given the diversity in the structure, content, and representations used by individual laboratory staff, both these coordination mechanisms could and did break down. Breakdown of these coordination mechanisms severely restricted the retrospective use of the experimental records, limiting the reliability of the laboratory records as a community archive. As an important measure of the difficulty involved in reconstructing meaning from the laboratory records kept in notebooks, multiple researchers experienced breakdowns when interpreting their own laboratory records during the course of experimental work. In this sense, breakdowns in interpretation were not limited to the community archive but also impacted upon the personal archive.

4 A Genre Analysis of Laboratory Notebook Records

This chapter of the thesis reports on the analysis of a corpus of authentic laboratory records from notebooks written by scientists during the course of their work in academic molecular biology laboratories. The chapter is again presented broadly in line with the IMRD structure (*cf*. Thompson 2001) to address the aims of the study, the method used to conduct the study, the results obtained, and the conclusions drawn from these results.

4.1 Aim of the study

The primary aim of this study was to investigate the range of textual features that characterize the laboratory records written by scientists at work in academic molecular biology laboratories. To this end, the study set out to examine the structures, content, and representations used in a corpus of authentic laboratory records produced by a range of laboratory staff. A secondary aim of the study was to provide a qualitative survey of the content and structures used for recordkeeping in laboratory settings in order to evaluate the potential for formatted record capture and display in electronic recordkeeping systems within academic molecular biology laboratories.

By focusing on the textual features of laboratory records, this study aimed to build on the contextual insights gained through the ethnographic study of recordkeeping in academic molecular biology laboratory settings that was reported in chapter 3. Specific issues highlighted by the ethnographic study that were germane to this analysis of textual features included variation across laboratory members in the expected readership for laboratory records, variation in the expected range of purposes to be served by laboratory records, variation in the temporal rhythms of laboratory recordkeeping, variation in the attitude of laboratories and laboratory staff towards sharing both experimental data and protocols, and the production of laboratory records as compound documents.

4.2 Design of the study

The study was designed as a genre analysis of a corpus of laboratory records written by multiple scientists in their notebooks. This design was chosen in order to focus the

analysis on language use in authentic texts produced by scientists during their work in academic molecular biology laboratories, and to enable a comparative analysis of the language used in these laboratory records from the "pattern seeking" (Hart 1986:280) viewpoint of genre analysis. As discussed in section 1.4.3, the perceived strengths of the genre analysis approach include the ability to combine a multi-level evaluation of structural, linguistic, and cognitive aspects of the construction of text in laboratory records, and the ability to examine conventionalized linguistic behaviour in situated language use.

4.2.1 Ethical approval procedure

Since the study involved inspection of notebooks written by human participants, it was reviewed and approved by the Ethics Committee of the Faculty of Information and Mathematical Sciences at the University of Glasgow under application number FIMS00423. This approval ensured that the study conformed to the code of conduct set out by the British Psychological Society for studies involving human participants (BPS Ethics Committee 1978). All participants were approached and recruited to the study only after ethical approval had been confirmed. Introductory and debriefing sessions were conducted with each participant in accordance with an interview script. The information sheet sent to prospective participants to describe the study is presented in Appendix 3 of this thesis.

4.2.2 Sample cases

4.2.2.1 Selection policy

The selection of sample cases for this study of records kept in laboratory notebooks was driven by similar strategic and practical considerations to those that influenced the design of the previous ethnographic study of laboratory recordkeeping (see section 3.2.2.1).

The key strategic consideration was to adopt a form of stratified purposeful sampling (Patton 2001:240) to enable inspection of multiple notebooks drawn from multiple categories of author, thus improving the representativeness of the sample set of laboratory records and the scope for observing potential sources of variation. In particular, notebooks were selected from authors in multiple laboratories rather than a single laboratory, and notebooks were selected from both research-orientated and service-

orientated laboratories within the university environment. This approach was chosen in order to consider potential variation in the structure, content, and representations used in laboratory records both between workers in the same laboratory, and between workers in different laboratories. Laboratories of different sizes were chosen in order to allow scope for comparison of laboratory records kept by workers operating in relative isolation with those operating in collaborative teams.

Notebooks written by scientists at different stages of their academic career were examined within each laboratory. Academic career stages were again ascribed on the basis of the scientist's function as technician, postgraduate researcher, postdoctoral researcher, or principal investigator in a similar manner to that used for the previous ethnographic study of laboratory recordkeeping. This approach was chosen in order to consider potential variation in the recordkeeping practices between different stakeholder responsibilities, and between novice and experienced users in line with Swales' (2004) gradation of junior and senior researchers (see section 2.6.2).

An additional strategic consideration for this study was to focus primarily, but not exclusively, on notebooks produced by scientists at work in those laboratories that had been observed as part of the previous ethnographic study of laboratory recordkeeping. Insights gained from the ethnographic study enabled the structures, content, and representations used in the laboratory records to be interpreted in a contextualized manner. In this way, the findings of the ethnographic study and the findings of this genre analysis could be integrated directly within the overarching multi-perspective study framework.

Practical considerations for the study derived primarily from the difficulty experienced in gaining access to laboratory notebooks. Recruiting authors from the same laboratories that participated in the previous ethnographic study also delivered practical benefits. Gaining access to notebooks was facilitated by the existing professional and personal contacts that had been established with a range of scientists within these laboratories. This level of trust was essential given the importance attached by individual scientists to their laboratory notebooks as discussed in the findings of the previous ethnographic study.

4.2.2.2 Sample laboratories

Table 4-1 overleaf lists the laboratories from which notebook authors were recruited for this genre analysis study of laboratory records.

Table 4-1: Laboratories for the genre analysis study

Laboratory	*Type*	*Description*
GS-L1 *(ES-L1)*	Research	A small university research laboratory with approximately 7 members. Members are involved in human genetics research for projects in the field of sports and exercise science with a specific focus on the interaction between environmental factors and hereditary factors on human health and performance.
GS-L2 *(ES-L4)*	Research	A small university research laboratory with approximately 6 members. Members are involved in human genetics research in projects investigating the genetics of human disease with a specific focus on one disorder.
GS-L3 *(ES-L3)*	Service	A common services department housed within a university facility but offering laboratory services and consultancy in sequencing and data analysis to multiple client laboratories within the home university, in other universities, and in other research institutions.
GS-L4 *(ES-L2)*	Research	A large university research laboratory with approximately 20 members formed as a close collaboration of two principal investigators. Members are involved in integrative physiology research using *Drosophila melanogaster* as a model organism for a range of projects including some projects with commercial partners.
GS-L5 *(No previous participation)*	Research	A small university research laboratory with 2 members. Members are involved in human genetics research in projects investigating the genetics of human disease with a specific focus on identifying genes and pathways for one type of disease[96]. This is a new laboratory in its first year of operation under a recently promoted principal investigator.

Summarizing the laboratories that provided authors for the genre analysis study in terms of the identifier code assigned to the laboratory, the laboratory type, and a brief description of the laboratory setting.
The identifier code uniquely identifies each laboratory that participated in the study whilst maintaining the anonymity required under the terms of the ethical approval for the study. The laboratory type is used to categorize laboratories into either *Research* laboratories or *Service* laboratories. *Research* laboratories undertake research projects on their own initiative in order to investigate scientific questions of their own choosing, whilst *Service* laboratories provide support services to other laboratories and are commissioned to undertake specialist experimental work on behalf of these client laboratories. The description of the laboratory setting outlines the size of the laboratory, and the broad research interests of the laboratory.

[96] The research questions addressed by a number of molecular biology laboratories involve locating candidate genes associated with disorders, identifying the genetic mutations that predispose individuals to certain disorders, and describing the biological mechanisms and pathways that give rise to the disorders. These and other research questions are addressed by research staff such as those in the participating laboratories in the context of a wide range of disorders such as neurological disorders, developmental disorders, vascular disease, myotonic dystrophy, and types of cancer.

All five of the participating laboratories were housed within a single UK university. It is important to note, however, that the service laboratory *GS-L3* delivered its services to a number of external clients located throughout the UK. Only one of the research laboratories, *viz*. *GS-L2*, made use of the services offered by the service laboratory *GS-L3*. None of the five research laboratories were involved in any collaborative projects with each other. However, laboratory members in research laboratories *GS-L1* and *GS-L5* were allocated space in the same 'wet' laboratory area and communal office accommodation. In addition, members including the study participants from these two laboratories presented their work at the same weekly laboratory group meeting.

As noted in Table 4-1, all of these laboratories except research laboratory *GS-L5* were observed as part of the previous ethnographic study.

4.2.2.3 Sample authors

Table 4-2 overleaf lists the range of laboratory members that participated as authors in the genre analysis study. As far as possible under the access conditions granted within each laboratory, notebooks written by scientists at different stages in their academic careers were examined within each laboratory. A more detailed summary of each individual participant including a brief description of the participant's experience in laboratory work is presented in Appendix 4 of this thesis.

Gaining access to laboratory notebooks written by principal investigators proved particularly difficult. This was due principally to the fact that most principal investigators no longer work in the laboratory on a regular basis but are instead involved in administration and directing the work of others. To a much lesser extent, the difficulty in gaining access to principal investigator notebooks also derived from a greater resistance on the part of principal investigators to make their notebooks available[97]. The sole principal investigator who participated as an author in the study was a research fellow who was active in the laboratory on a daily basis at the time of the study.

All participants who spoke English as a second language had produced written publications in English, participated regularly in laboratory meetings held in English, presented their

[97] In personal communications at departmental seminars and social events, a number of principal investigators from laboratories other than those involved in this study suggested that they had not written their notebooks intending them to be read by others and so felt that the notebooks would offer little to the study.

own work at internal and external seminars in English, and had completed university-level courses taught in English (see section 2.6.2). Gender was not used to discriminate between those participating in the study as authors, and both male and female scientists were recruited from each laboratory wherever possible (see section 2.6.2).

Table 4-2: Authors for the genre analysis study

Laboratory	*Authors (N = 14)*							
	Job Function				**Native Language**		**Gender**	
	PI	**PD**	**PG**	**T**	**L1**	**L2**	**M**	**F**
GS-L1	0	2	0	0	2	0	2	0
GS-L2	0	1	3	1	3	2	1	4
GS-L3	0	0	0	2	1	1	0	2
GS-L4	0	0	2	1	1	2	2	1
GS-L5	1	0	0	1	1	1	1	1
All Labs	***1 7.1%***	***3 21.4%***	***5 35.7%***	***5 35.7%***	***8 57.1%***	***6 42.9%***	***6 42.9%***	***8 57.1%***

Showing a numerical breakdown of all participating authors in the genre analysis study by laboratory of origin, by function in that laboratory, by native language, and by gender.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The author's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks/writes English as a second language. The author's gender is indicated using F for a female participant, and M for a male participant.

4.2.2.4 Sample notebooks

Where possible, two or three notebooks were examined for each author in order to enable an evaluation of the structures, content and representations used in the written records over a period of time. The range of notebooks available from each author was dependent in part on the previous experience of the author. In particular, some authors were able to provide notebooks relating to work that they had carried out at different stages in their academic career in laboratories other than the one in which they were now employed. Authors who had only recently started out on their academic careers were able to provide only a single notebook relating to their current work.

Table 4-3 lists the range of notebooks that were examined in this genre analysis study of laboratory records.

Table 4-3: Notebooks for the genre analysis study

Author Function	*Notebook Set*					*Notebooks (N = 30)*
	Single	**Double, Same Function**	**Double, Different Functions**	**Triple, Same Function**	**Triple, Different Functions**	
PI	0	0	0	1	0	***3 10.0%***
PD	0	1	1	0	1	***7 23.3%***
PG	2	1	0	1	1	***10 33.3%***
T	1	3	0	1	0	***10 33.3%***

Showing a numerical breakdown of all notebooks examined in the genre analysis study by function of the notebook author, and by the type of notebook set provided by the author.
The author's function is indicated using *PI* for a principal investigator/head of laboratory, *PD* for a postdoctoral researcher, *PG* for a postgraduate research student, and *T* for a laboratory technician. The notebook set provided by each author is categorized along two dimensions in terms of the number of notebooks in the set (single, double, triple), and whether the notebooks in the set relate to the same or different stages in the author's academic career as indicated by the author's function at the time of writing the notebook.

Two postdoctoral researchers, now working respectively in research laboratories *GS-L1* and *GS-L2*, provided laboratory notebooks relating to previous work as postgraduate researchers in other laboratories. In both cases, this involved work in academic molecular biology laboratories outside the UK university selected for this study, one in a different UK university and the other in a North American university. One postgraduate researcher provided a laboratory notebook written in Spanish relating to previous work in a laboratory at a South American university in addition to two notebooks written in English relating to current work in research laboratory *GS-L2*. One recently appointed technician in research laboratory *GS-L5* provided a single notebook, as did two recently appointed postgraduate researchers in research laboratories *GS-L2* and *GS-L4* respectively. All other authors provided multiple laboratory notebooks relating to work undertaken as part of their current functions in the laboratories participating in this study.

4.2.3 Experimental procedure

4.2.3.1 Participant contact

All contact with study participants before and during the study was made directly with the participant and independently of the principal investigator in order to minimize the potential for bias on the basis that access to laboratories had been gained initially through contact with the principal investigators in each laboratory. As with the ethnographic study of laboratory recordkeeping (see section 3.2.3.1), this procedure was followed in order to ensure that participants did not perceive the experimenter as acting in any way on behalf of the principal investigator. In particular, it was important that participants felt able to offer notebooks for analysis that were representative of their actual recordkeeping practices rather than limited examples of what could be termed 'best practice'.

Prior to observing work in each laboratory, one-to-one meetings were arranged with each individual participant to explain the purpose of the study, to gain written consent, to agree a selection of laboratory notebooks for subsequent analysis, and to arrange convenient dates on which to gain access to the laboratory notebooks for content analysis.

4.2.3.2 Definition of content analysis framework

A content analysis framework (*e.g.* Neuendorf 2002) was used to direct the analysis of language use within each laboratory notebook in the corpus. This approach was adopted for two reasons. Firstly, the use of a framework ensured a systematic and consistent approach to the analysis of language use in each individual notebook within the corpus. This was achieved by providing a structured questionnaire to direct the range of issues to be considered with regard to the structures, content, and representations used in the laboratory records held in the notebooks. Secondly, the use of a framework enabled a relatively efficient analysis of each notebook by minimizing errors of omission during the course of the analysis. This second consideration was motivated by one of the central findings from the previous ethnographic study of laboratory recordkeeping concerning the importance attached by individual scientists to their laboratory notebooks. This was confirmed in the context of the genre analysis study by a subset of the study participants expressing reservations about losing control of their notebooks for any extended period of time.

Consequently, the initial stage in the conduct of this genre analysis study was concerned with the construction of an appropriate content analysis framework to use in characterizing the linguistic behaviour found in laboratory records. The findings from the previous ethnographic study of laboratory recordkeeping proved invaluable in this respect. The set of issues embodied in the content analysis framework used for this genre analysis study was defined by combining the findings of the ethnographic study both with insights gained from a survey of recordkeeping issues in other domains such as healthcare, and with the outcomes of a preliminary survey of a subset of the notebooks offered for the genre analysis study. The range of issues is summarized in Table 4-4.

Table 4-4: Content analysis framework issues

Scale	*Framework Issues*
Notebook	Description of context of production including author and laboratory; Description of the time period covered by the notebook; Summary of the range of records in the notebook including types of record (experimental/protocol/plan/meeting/other); Description of the notebook structure such as separate sections for record types; Description of the range of structures and textual features for experimental records; Description of the range of structures and textual features for protocol records; Description of the range of structures and textual features for other types of record; Description of indexing/cross-referencing schemes used in the notebook; Description of workflow sequencing schemes used in the notebook.
Record	Description of the molecular biology experiment being recorded; Is it a continuation entry?; Length of the entry in pages; Description of the sequence of semantic units used to compose the record of the experimental; Description of any inserts, cross-references, or indexes used in the record; Description of textual features such as abbreviations, omitted structural elements, use of summary levels of specification, and use of way-markers and other navigational aids to guide readers/writers through the text.

Summarizing the range of issue embodied in the content analysis framework used to assess laboratory notebooks for the genre analysis study. The issues embodied in the framework support two scales of analysis in terms of the individual records within the notebook, and the notebook as a whole.

Laboratory notebooks are containers of multiple laboratory records, so the content analysis framework used for the study accordingly operated at two scales by assessing both the properties of the individual laboratory records in the notebook, and the properties of the

notebook as a whole. As illustrated by Table 4-4, the records and notebooks were assessed across a range of contextual, structural, and linguistic features.

Appendix 5 of this thesis provides an example of a completed content analysis framework for one of the notebooks submitted to the genre analysis study.

4.2.3.3 Definition of laboratory record structure

An essential part of the analysis of individual records within the notebook was the description of the structure of the laboratory record in terms of the domain-specific semantic units used by scientists to compose the records. Examples of these domain-specific semantic units include units such as 'title', 'sample list', 'purpose statement', and 'gel loading conditions'. These semantic units are considered domain-specific in that they are units of language use that are used to encode specific elements of molecular biology laboratory work. Within laboratory notebooks, these units may be expressed in different modes including text and/or graphics. Since no such list of units was already available, it was also necessary to define a set of semantic units for use in characterizing the structure of laboratory records as part of the initial stage of this genre analysis study. In a similar manner to the content analysis framework, this list was defined by combining the findings of the previous ethnographic study with insights gained from a preliminary survey of a subset of the notebooks offered for the genre analysis study. Where necessary, the list was extended to include additional semantic units as they were discovered during the subsequent analysis of notebooks in the corpus.

Appendix 6 of this thesis provides a summary of the semantic units used to describe the composition of laboratory records for this genre analysis study.

4.2.3.4 Data collection

Analysis of each notebook in the corpus was conducted by reading through the notebook on a record-by-record basis in order to complete the associated sections of a content frame that described both the notebook and the laboratory records contained in each entry in the notebook. Separate content frames were completed for each individual notebook. Completing the frame for each notebook was essentially a two-pass operation. The first pass through the notebook focused on completing those sections pertaining to individual record entries in the notebook. Thereafter, a second pass through the notebook focused on

completing those sections pertaining to the notebook as a whole. Exemplar records demonstrating typical examples of linguistic and structural variation within the notebook were scanned to produce digital copies. Depending on the number and the complexity of the entries in the notebook, the content analysis of each notebook took from three up to five days to complete.

During the course of the study, notebooks were typically borrowed for an agreed period of time so that they could be evaluated away from the participants' laboratory sites. This approach was preferred partly in order to minimize any potential disruption to the participants' work, and partly in order to enable an initial evaluation of the notebook contents that was independent of the original author. All study participants were presented with the option to require that the analysis of notebooks be conducted at the participant's own laboratory site. However, this option was not exercised by any of the study participants.

Whenever necessary, short open interviews were conducted with the study participants after the main analysis of the notebooks had been completed in order to clarify any points that may have arisen during the completion of the content frame. Depending on the availability of the study participant, these interviews were either conducted in a single session recorded using a digital voice recorder, or conducted over multiple sessions with written notes made recording the results of the interview. Two of the study participants, postgraduate researchers *GS-L2R5* and *GS-L4R3* were unavailable for interview since they had transferred to new positions during the course of the study. In both these cases, interviews were conducted with colleagues of the study participants who worked in the same laboratory.

All data collected during the study have been rendered anonymous in accordance with the terms of the ethical approval for the study. To facilitate subsequent analysis, the data was collated using version 2.8 of the HyperRESEARCH™ CAQDAS tool. This tool was selected on the basis of its use during the previous ethnographic study of laboratory recordkeeping, and due to its integrated support for multimedia data including text and digital scans.

4.2.3.5 Data analysis

Content analysis frameworks may be used to support both quantitative analyses and qualitative, thematic analyses. For the purposes of this study, the intention of the content analysis framework has been to provide a systematic means of collecting primarily qualitative data characterizing the use of language in laboratory records. The qualitative data collected for the study have been analysed, coded, and categorized in line with the data analysis procedure for developing grounded theory (Charmaz 1983; Glaser and Strauss 1967; Strauss and Corbin 1998). This procedure involved an iterative process of open coding, axial coding, and selective coding in order to classify and interrelate the data obtained during the course of the study.

Open coding involved the systematic examination of the content analysis frames and interviews with the study participants collected during the study in order to identify and describe categories of linguistic behaviour and generic variation in relation to laboratory records. Axial coding was concerned with refining the understanding of linguistic behaviour by identifying generalization/specialization relationships to capture the semantic links between the categories and their associated subcategories. Finally, selective coding was concerned with integrating and refining the set of categories in order to build an understanding of linguistic behaviour in laboratory records that was derived in an inductive manner from the data collected during the course of the study. As for the ethnographic study, it is important to note that coding was not a static procedure, but instead proceeded in an iterative manner throughout the course of the study so that categories could be compared, modified, and refined in the light of new observations. This constant comparison approach proved particularly appropriate to the study in terms of facilitating a robust comparison of laboratory recordkeeping across different laboratory settings and members.

4.2.3.6 Data validation

Selinker (1979) and Huckin and Olsen (1984)[98], *inter alia*, highlight the important role played by domain specialists in validating the findings from studies of specialist texts such as academic molecular biology laboratory records. Huckin and Olsen (*ibid*.:129), in

[98] Both these studies are particularly relevant to the work of this thesis given their collaboration with a specialist in genetics to help understand the language in research articles.

particular, recommend that "perhaps the most useful specialist informant one can find for an LSP[99] text is the actual author of that text". The approach used to validate the findings of this study was driven by these observations, and relied on interacting with the notebook authors in order to present the findings concerning the notebooks produced by that author.

As with the previous ethnographic study of laboratory recordkeeping, participant validation not only involved presenting and discussing the findings from the study with individual participants but also, albeit to a much lesser extent, via group presentations in order to evince feedback on insiders' perspectives on the validity of the findings. The problems highlighted for participant validation in ethnographic studies (*e.g.* Bloor 1978) applied equally to this genre analysis study. Here again, it was important not to interpret the feedback gained from these sessions, whether supportive or hostile, uncritically as direct validation (or invalidation) of the findings from the study.

Of particular interest in the context of this thesis was the scope for an extended form of between-method triangulation afforded by the multi-perspective framework of ethnography and genre analysis chosen for this research project. This approach enabled a degree of cross-validation between the data generated across the two studies, supported by the overlap in the laboratories and individual scientists that participated in both studies and so established a localized basis for comparison.

4.3 Results

The results of this genre analysis study of laboratory records are presented in the following subsections, arranged around the categories identified during the data analysis and "illustrated by characteristic examples of data" (Glaser and Strauss 1967:5). In this case, the characteristic examples of data consist of annotated extracts taken from laboratory notebooks[100]. References to individual participants and notebooks in the study results make use of the identifiers listed in Appendix 4. The set of domain-specific semantic units used to represent the molecular biology tasks documented within laboratory records are described in Appendix 6.

---

[99] LSP stands for Language for Specific Purposes, a more cosmopolitan project than ESP or English for Specific Purposes involving other languages in addition to English.

[100] It appears that the anecdotal evidence for medical doctors and poor handwriting may also apply to some doctors of philosophy in the biosciences. It is hoped that the box comments included in figures will be useful in communicating the intent of the laboratory records.

The linguistic behaviour identified in the corpus of laboratory notebooks is surveyed, in particular, along three dimensions. These dimensions are concerned with generic variation in the structural units that are used to compose laboratory records, temporal variation in the construction of laboratory records, and variation in laboratory records driven by laboratory-wide policies. Although the study is primarily concerned with a qualitative survey of language use within laboratory records, an indication is also given of the distribution of specific behaviours across the study participants wherever possible.

4.3.1 Unit-level variation in records

The following examples survey the language features used to communicate the internal structural units from which laboratory records are composed. Detailed analysis of the laboratory records present in the notebook corpus identified a common set of core molecular biology tasks that were routinely used in the experimental work documented in these records[101]. These have been collated in the set of domain-specific language units listed in Appendix 5. Particularly common tasks that appeared in the majority of the laboratory records analysed during the course of this study included stating the purpose of experimental work, preparing reaction mixes, running PCRs, and running electrophoresis gels. Given that these tasks were routinely used by all notebook authors, the variations in linguistic behaviour associated with these tasks is indicative of the range of linguistic behaviour found across the laboratory staff.

The ethnographic study reported in chapter 3 highlighted the fact that although all laboratory staff expressed a remarkably consistent concept of the information required in a laboratory record, writing records involved considerable contextual competency by requiring individual authors to judge when experimental data becomes sufficiently "difficult" or "unusual" to require it to be recorded in the notebook. In particular, the following subsections examine such variation in the use of language to specify three common aspects of laboratory records, *viz*. the recording of experimental purpose, the recording of PCR thermal cycles, and the recording of gel electrophoresis runs.

[101] This is consistent with work carried out, for example, by Arnstein *et al.* (2002) in cell biology laboratories and Frey *et al.* (2004) in chemistry laboratories. Arnstein *et al.* (*ibid*.) identify a set of only six core task types encompassing combination tasks, incubation tasks, dispensing tasks, separation tasks, detection tasks, and storage tasks. Whilst the scope of the protocols defined in manuals such as Berger and Kimmel (1987) and Sambrook and Russell (2001) indicates that molecular biologists are required to perform a complex range of tasks, the complexity derives in part from the complexity of a workflow combining multiple tasks, and in part from the instantiations of these task types to suit the biological properties of specific cells or tissues.

4.3.1.1 Recording experimental purpose

Figure 4-1 overleaf illustrates the two conventional approaches identified within the corpus of laboratory records that characterize how scientists record the purpose of an experiment.

In both examples shown in Figure 4-1, the laboratory records document an optimization experiment to determine the optimal experimental conditions for a PCR. In both examples, this PCR is designed for use in DNA sequencing work aimed at determining allelic variation[102] within samples for a named gene. In both examples, the experiment is designed to optimize a combination of experimental conditions consisting of the choice of primers to be used in the PCR and the annealing temperature[103] to be used in the reaction. In short, the two example records document a similar type of experimental work, and this type of experiment was contained within each of the notebooks in the corpus.

Determining suitable primers for a PCR involves running multiple test reactions with known samples and different pairs of primers designed specifically for the alleles being studied. The primers are paired in order to form forward and reverse primer pairs that match respectively the start and end of the DNA fragment targeted for amplification. The results of this reaction can be visualized using gel electrophoresis to determine whether the primers have been successful in separating out the DNA fragments necessary to enable identification of the various alleles of interest. Determining the optimal annealing temperature involves running multiple test reactions across a temperature gradient. This temperature gradient represents a stepped range of temperatures, which is conventionally specified using the low temperature in the range together with the interval spanned by the gradient. Again, the results of this reaction can be visualized using gel electrophoresis to determine which of the annealing temperatures has produced the best amplification signal.

[102] Different individuals may have variant forms of the DNA sequence of a given gene. Each of these possible variant forms of the DNA sequence of a gene is termed an allele.

[103] The annealing temperature forms part of the thermal cycle used in PCRs. In particular, the annealing temperature is the temperature set for the annealing step in the PCR, during which the strands of DNA that were separated by denaturation are recombined in the presence of DNA primers to amplify the target DNA.

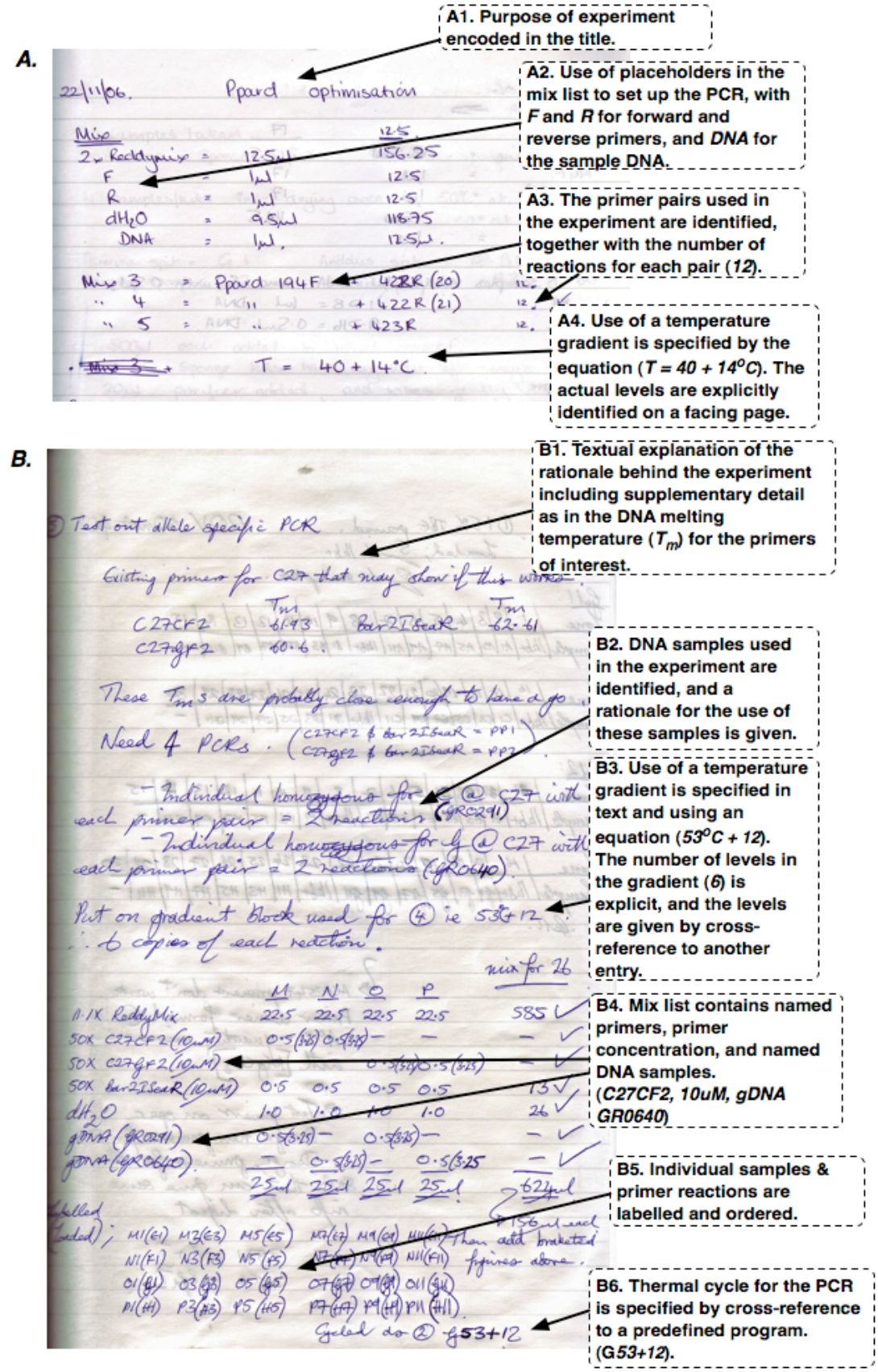


Figure 4-1: Example record entries for purpose statements

Showing variation in the representation and content used to specify the purpose of an experiment within experimental records. In both cases, the experimental purpose is to test out different primer pairs and annealing temperatures for use in a PCR. (*A) and (B).* Scanned excerpts from laboratory records written by two postdoctoral researchers in research laboratory *GS-L1*.

A postdoctoral researcher in research laboratory *GS-L1* recorded the example shown in Figure 4-1A. In this example record, the purpose of the experiment is partially encoded in the title of the record, and partially encoded by the range of domain-specific semantic units that have been included and excluded from the experimental method. For the purpose of this study, this style of recording[104] is termed '*title-focused*' recording of experimental purpose.

Decoding the title relies on a contextual understanding of the phrase "Ppard optimisation". The reference to "Ppard" in this title identifies the gene of interest for the experiment, which in this case is the *PPARD*[105] gene. The specific research interests within laboratories engaged in molecular biology are often reified into a set of target genes, proteins, or other biological constructs that encapsulate the work of the laboratory. In this sense, knowledge of the specific research interests within the laboratory setting may facilitate the interpretation of this title. The reference to "optimisation" in the title establishes the theme of an optimization experiment for the record. It is important to note, however, that multiple experimental conditions could require optimization for a PCR including volume of reagents, concentration of reagents, reagents in the reaction mix, annealing temperature, and primer choice. Accordingly, it is necessary to examine the semantic units that are listed in the experimental method further on within this entry in order to understand exactly what is being optimized. In this case, the list of primer pairs named as "Mix 3", "Mix 4", "Mix 5" indicate multiple forward and reverse primer pairs, whilst the equation "T = 40 + 14ºC" symbolically identifies a temperature gradient from a low temperature of 40ºC across an interval of 14ºC. The rationale behind the use of the three primer combinations is not explained, and the sample DNA used in the reaction is not identified. Interpreting the reaction mix list[106] in combination with the list of primers pairs requires an

[104] The term 'styles of recording' is used here to indicate variation in the patterns of language used by laboratory staff to realize laboratory records in whole or in part. It is recognized that the term 'style' has evolved in a number of ways within linguistics (*e.g.* Coupland 2007). This linkage between style and variation has remained key as illustrated, *inter alia*, by de Beaugrande and Dressler (1981:16, emphasis in original) who comment that "despite the diversity of approaches, nearly all work reflects the conviction that style results from the characteristic *selection of options* for producing a text or set of texts." Eggins (1994:25-48) identifies "realization patterns" as one of the dimensions in the realization of genres where "the boundaries between stages, and the function of each stage of the genre, are expressed through language choices (discourse-semantic, and lexico-grammatical) realized in a text".

[105] The *PPARD* or peroxisome proliferator-activated receptor delta gene encodes a type of nuclear receptor protein that has been shown to be involved in glucose and lipid metabolism amongst other functions. As a result, it has been investigated in studies relating to diabetes mellitus and obesity. Note that gene names are, by convention, written in italics.

[106] A reaction mix list identifies the list of the reagents to be combined for a reaction together with the volume of each reagent to be used in the reaction. The typical format used in all laboratory

understanding that the placeholders "F" and "R" in the mix list are intended as cataphoric substitutions for the forward and reverse primers named in the primer pairs.

A different postdoctoral researcher in the same research laboratory *GS-L1* recorded the example shown in Figure 4-1B. In this example record, the purpose of the experiment is encoded in the title of the record, and expounded in a narrative block of text to include a statement of the expected results of the test. For the purpose of this study, this style of recording is termed '*narrative*' recording of experimental purpose. As a measure of the influence of laboratory policy on recordkeeping, it is interesting to note that these two postdoctoral researchers worked side-by-side at neighbouring benches.

Perhaps the most obvious difference in this example record is the significant level of narrative text used in addition to the symbolic representations that formed the exclusive content of the previous style of recording. The theme of the first phrase in this narrative block is "existing primers", which establishes that the purpose of the test is primers. The difference between these examples is, however, not limited to issues of representation. This second example also includes additional content in the statement of purpose including the rationale behind the choice of primers. The title of this second example record contains an explicit reference to the type of experimental task in the form of "allele specific PCR", and identifies this experiment as an optimizing experiment through the word choice in the verbal phrase "test out". The rationale behind the choice of primers is explained in text adjacent to the list of primer pairs to be used for the experiment. Very importantly, what would be considered a positive result of the test with regard to the primer pairs is also stated together with the named DNA sample to be used in the experiment in the phrase "Individual homozygous[107] for G @ G27 with each primer pair". No placeholders/references are used in the specification of the reaction mix lists, but all primers are explicitly stated in the list together with the concentration of the primer as in "10μM". The specific DNA sample used in the reactions is also explicitly stated as in

notebooks for these mix lists is a table with three columns. Each row lists the volumes for a single reagent. The first column lists the names of the reagents used in the reaction mix. The second column identifies the volume of each reagent required for a single reaction whilst the third column identifies the scaled up volume required for the actual number of reactions to be run. Expanded, multi-column variants of this format may be used to specify combinations of reactions. Of the thousand or so reaction mix lists recorded in the notebook examined during the course of this study, only one did not correspond to this format; that mix was specified in narrative text in a format typical of research article methods.

[107] Humans and other diploid organisms have two chromosomes and therefore two copies of each gene, one per chromosome. Individuals that have the same allele for both copies of the gene are said to be homozygous for that gene. Individuals that have different alleles are said to be heterozygous for that gene.

"gDNA (GR0640)". Finally, each of the individual tubes containing the reaction mix for the different combinations of primer pairs and annealing temperature are individually listed in a two-dimensional grid that lists the labels used to identify the individual test reactions. This form of labelling establishes a basis for endophoric referencing at the granularity of individual samples within subsequent entries to the record, specifically the electrophoresis gel that will be used to visualize the results of the test reactions.

Whilst the 'title-focused' style provides implicit cues to allow the purpose of the experiment to be understood, it relies on symbolic representations and requires an understanding of the context of the laboratory and of the experiment type on the part of the reader to reconstruct the experimental purpose. In contrast, the 'narrative' style provides a detailed level of specification using a narrative statement of purpose that describes the rationale behind the design of the experimental and establishes a cohesive link between experimental purpose and design.

Figure 4-2 provides an indication of the patterns of use of these two approaches to recording experimental purpose in the notebooks produced by authors in different laboratories and by authors performing different laboratory functions.

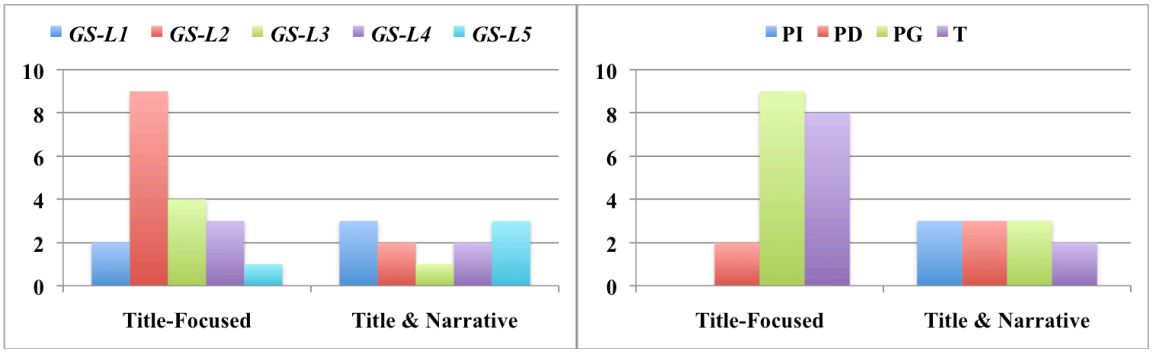


Figure 4-2: Distribution of styles for recording experimental purpose

Indicating the number of notebooks using different patterns of recording experimental purpose by laboratory of origin of the notebook author and by function of notebook author in that laboratory.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician.

It is important to note that the 'narrative' style of recording was not used as the sole means of recording experimental purpose in any of the notebooks analysed for this study, but was instead reserved for records that documented non-routine experimental work such as the

design of complex reactions or the execution of 'charter experiments'[108] that were critical in determining the direction of a project. For this reason, the two categories reported in the charts are concerned with the use of the 'title-focused' style only, and with the combined use of the 'title-focused' and 'narrative' styles of recording.

The 'title-focused' style of recording was used in nineteen (63.3%) of the thirty notebooks in the corpus, whilst the combined 'title and narrative' style was used in eleven (36.7%) of notebooks. The charts indicate a tendency towards the use of the 'title-focused' style of recording experimental purpose across authors who are postgraduate researchers and technicians. The data for principal investigators are unrepresentative since only one such author participated in the study.

4.3.1.2 Recording PCR thermal cycles

Figure 4-3 overleaf illustrates the three conventional approaches identified within the corpus of laboratory records that characterize how scientists record the thermal cycle for a PCR.

In each of the examples shown in Figure 4-3, the PCR thermal cycle forms a discrete part of the experimental data documented in the laboratory record to describe the reaction. In each example, the PCR thermal cycle was fixed for the purpose of the experiment in the sense that the focus of the experimental work was not on varying any aspect of the PCR thermal cycle. In short, the PCR thermal cycles were recorded in a similar context of use.

PCR employs repeated cycles of heating and cooling of DNA fragments in reaction with other chemicals in order to copy and amplify strands of a target DNA sequence of interest. The reaction proceeds through three important steps of denaturation, annealing, and extension, each requiring different temperature conditions in order to first melt the DNA and then enzymatically replicate the target DNA sequence. Defining a PCR thermal cycle therefore requires specifying the temperatures to be used for the denaturation, annealing, and extensions steps, how long to hold the reaction at these temperatures in each of these steps, and how many iterations of the thermal cycle to run. In addition to the denaturation

[108] The term 'charter experiment' has been paraphrased from the term 'charter document' (McCarthy 1991), and is used to differentiate experiments that were of particular significance within an overall body of experimental work. Examples include experiments where the findings would refute a central hypothesis within a project, or experiments that were generating results for direct inclusion with a research article.

and extension steps in the main cycle, temperature and duration are also required for an initial denaturation step and a final extension step.

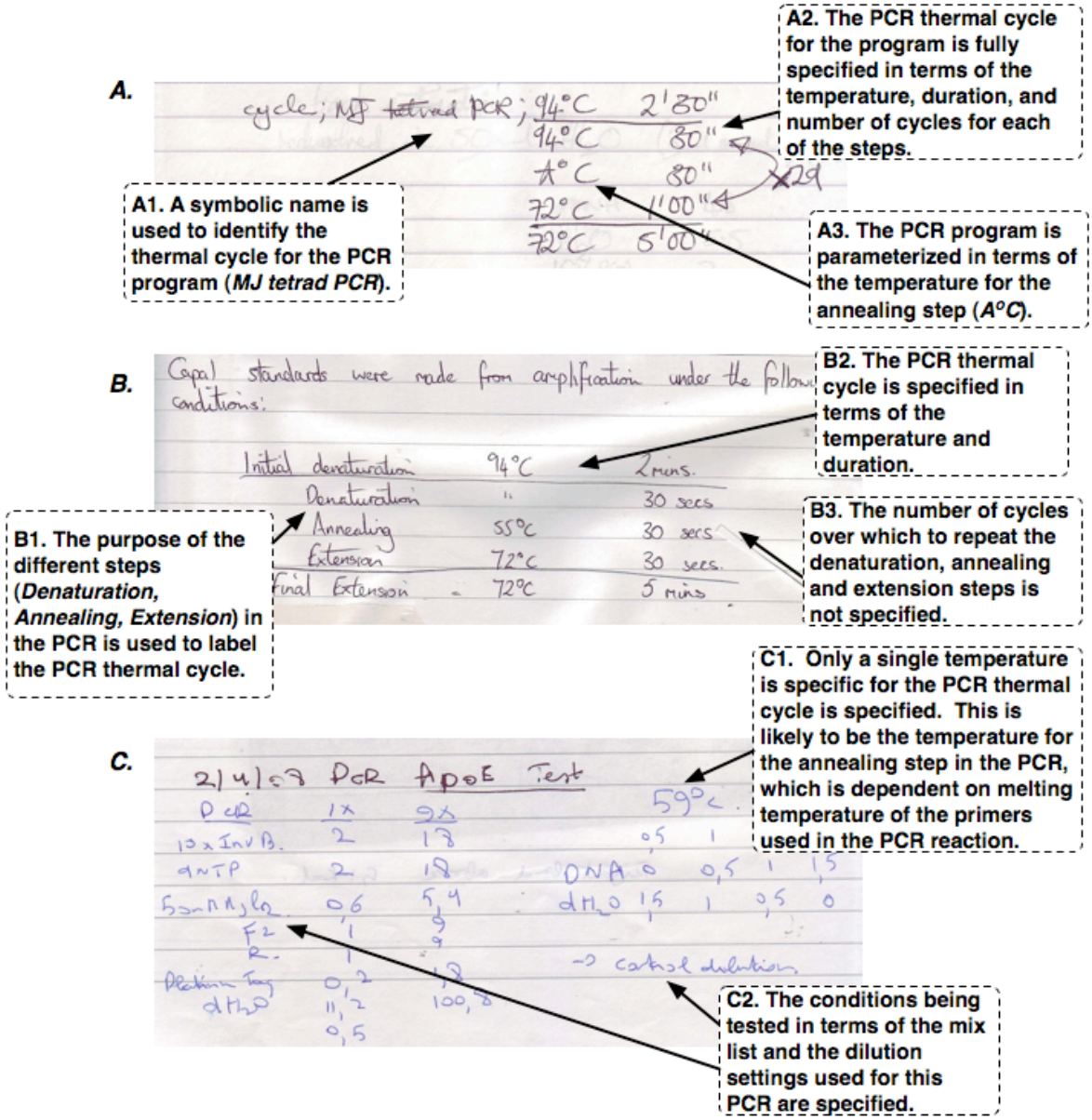


Figure 4-3: Example record entries for PCR thermal cycles.

Showing variation in the representation and content used to record PCR thermal cycles within experimental records. PCR is used to copy and amplify strands of DNA, and is a core molecular biology technique that is now a routine component of many experiments. The reaction relies on repeated cycles of heating and cooling in order to melt DNA fragments and then enable replication of the DNA. The thermal cycles in these excerpts describe the cycle of temperatures to be used within PCRs. (*A*). Scanned excerpt from a laboratory record written by a postdoctoral researcher in research laboratory *GS-L1*. (*B*). Scanned excerpt from a laboratory record written by a postgraduate researcher in research laboratory *GS-L4*. (*C*). Scanned excerpt from a laboratory record written by a principal investigator in research laboratory *GS-L5*.

PCR relies on the use of specific primer sequences that are designed to form forward and reverse primer pairs matching the specific DNA fragment targeted for amplification. The reaction also relies on the use of DNA polymerases, which are the enzymes that enable the synthesis of new strands of DNA. A single polymerase, named *Taq* polymerase[109], was widely used for PCRs in each of the laboratories that participated in this study. It is important to note that the temperatures used in each of the three steps in a PCR are sensitive to different properties of the reagents used in the reaction. In particular, the annealing temperature is sensitive to properties of the primers used in the reaction, and the extension temperature is sensitive to properties of the DNA polymerase used in the reaction. The PCR machines used within the participating laboratories were programmable devices that could memorize the temperatures and durations in PCR thermal cycles.

The example shown in Figure 4-3A was recorded by a postdoctoral researcher in research laboratory *GS-L1*. In this example record, the PCR thermal cycle is fully specified and presented in a tabular layout that signifies the internal steps within the cycle. For the purpose of this study, this style of recording is termed '*specified*' recording of a PCR thermal cycle.

A number of variations around the 'specified' style of recording were identified within the laboratory records in the notebook corpus, and some of these are also illustrated in Figure 4-3A. In addition to the individual temperatures and durations, an identifier is used to name the PCR thermal cycle as in "MJ tetrad PCR". This identifier establishes a further basis for linking the use of the cycle in this record with its use in other laboratory records. It is important to recognize that the distance[110] between records referring to a thermal cycle by name may span an entire notebook, which is typically in the order of 160 pages. The use of an identifier was in many cases driven by the assignment of programme names for use with programmable PCR machines, and so establishes an exophoric reference to the physical laboratory environment in the shape of the device used to perform the reaction. Control of the laboratory environment forms a crucial part of the positivist approach in

[109] *Taq* polymerase is named for the bacterium *Thermus aquaticus* from which it was isolated. This bacterium, which is found in thermal pools, is able to withstand high temperatures. It is this thermostable property of *Taq* polymerase that has led to its widespread use in PCR since it is able to withstand the high temperature used during the denaturation step.

[110] The concept of 'distance' is used here in the sense of Bunton (1999) to characterize the separation between a reference in a text and the referent segment of text. In Bunton's (*ibid*.) case, the distance is defined for PhD theses in terms of chapter (*i.e.* references that cross chapter boundaries), section in chapter, same section, and immediate sentence. In the case of laboratory notebooks, the range would include notebook, record in notebook, entry in record, and immediate entry.

molecular biology, and this has motivated some scientists to record the individual machines used during the performance of experimental work. This style of recording is termed '*specified and identified*' recording of a PCR thermal cycle.

Figure 4-3A also shows a less common variant of the 'specified and identified' style of recording a PCR thermal cycle in which all temperatures in the cycle are fixed apart from the annealing temperature. Instead, the annealing temperature is parameterized using a placeholder "AºC". This placeholder provides a cataphoric reference to subsequent laboratory records that may make use of the cycle with different settings for the annealing temperature. In this sense, the cycle is defined in the scope of an extended body of experimental work that spans multiple records. This variant style of recording is termed '*parameterized and identified*' recording of a PCR thermal cycle.

The example shown in Figure 4-3B was recorded by a postgraduate researcher in research laboratory *GS-L4*. In this example record, the PCR thermal cycle is also fully specified and presented in a tabular layout that reflects the internal steps within the cycle in a similar manner to the 'specified' style of recording. In addition, textual labels have been placed adjacent to each of the individual steps to identify its role in the PCR. These labels bring the internal functioning of the PCR technique into the foreground, and provide an explicit basis for referencing and describing the internal technical elements of the cycle in the remainder of the record. This foregrounding of the internal technical construction of a thermal cycle serves to explain the purpose of each internal step at a level that is appropriate for a novice to the PCR technique, as was the case with the author of this example record. For the purpose of this study, this style of recording is termed '*specified and labelled*' recording of a PCR thermal cycle.

The example shown in Figure 4-3C was recorded by a principal investigator in research laboratory *GS-L5*. In this example record, the PCR thermal cycle is represented solely by the annealing temperature, which is presented as a single number. In contrast to both the previous examples, this record employs a significant degree of ellipsis to achieve a highly economical style of recording. Interpreting the single number as a PCR thermal cycle, however, requires contextual interpretation on the part of the reader. This is cued in part by the use of temperature units "ºC", in part by the purpose of the experimental as expressed in the 'title-focused' style, and in part by the adjacency of layout where the temperature is placed next to other settings used for the PCR. The durations and other temperatures in the PCR thermal cycle are not specified, and no reference is provided to

help a reader navigate to where these may be defined. In practice, these other settings could be recorded externally to the notebook in a separate protocol folder, or could remain unrecorded but known to the author of the notebook as the routine settings used in all his/her experiments. In stark contrast to the use of the 'specified and labelled' style by novices to the PCR technique, this style of recording a PCR thermal cycle is characteristic of an expert user of PCR. For the purpose of this study, this style of recording is termed '*annealing only*' recording of a PCR thermal cycle.

Figure 4-4 provides an indication of the patterns of use of these approaches to recording PCR thermal cycles in the notebooks produced by authors in different laboratories and by authors performing different laboratory functions. It is important to note that the 'parameterized' recording style has been subsumed within the 'specified' recording style or the 'specified and identified' recording style as appropriate since it is consistently used in conjunction with one of these two approaches. An additional category 'None' is included in the charts to register notebooks in which PCR thermal cycles are routinely omitted from the laboratory records.

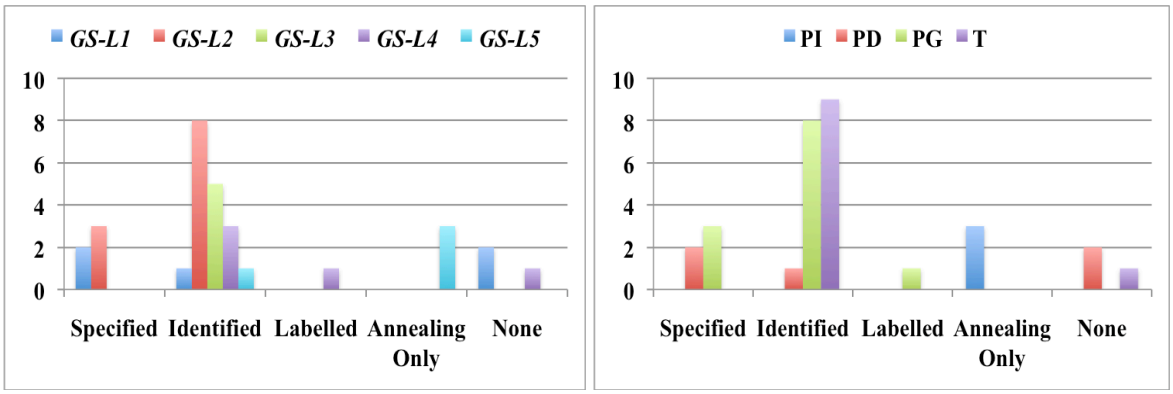


Figure 4-4: Distribution of styles for recording PCR thermal cycles

Indicating the number of notebooks using different patterns of recording PCR thermal cycles by laboratory of origin of the notebook author and by function of notebook author in that laboratory.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician.

The 'specified and identified' style of recording together with its associated cross-referencing schemes was the most common approach, and was used in eighteen (60.0%) of the thirty notebooks in the corpus. The charts indicate a tendency towards the use of the 'specified and identified' style of recording PCR thermal cycles for both postgraduate researchers and technicians. The 'specified and labelled' style of recording was used in

only one notebook by a postgraduate researcher who subsequently switched to the 'specified and identified' style of recording.

4.3.1.3 Recording gel electrophoresis runs

Figure 4-5 overleaf illustrates the two conventional approaches identified within the corpus of laboratory records that characterize how scientists record gel electrophoresis runs.

In each of the examples shown in Figure 4-5, the laboratory records document a gel electrophoresis run to visualize the results of an experiment determining the optimal conditions for a PCR. In both examples, this PCR was designed for use in DNA sequencing work aimed at determining allelic variation for a named gene. In both examples, the focus was on optimizing the choice of primers and the annealing temperature for the reaction. In this sense, the two examples shown in Figure 4-5 represent a continuation of the type of experiment discussed in section 4.3.1.1 with regard to recording experimental purpose. In particular, the same authors produced both sets of records.

Gel electrophoresis employs electrical charge to sort fragments of DNA based on their size and charge. In this way, the technique can be used to visualize the results of a PCR that has been designed to amplify specific DNA sequences of interest. The presence of the target DNA will appear as a band in the resulting gel, and different DNA fragments will appear as bands at different positions on the gel. A ready-to-use DNA ladder containing DNA fragments of known size is loaded onto the gel in order to establish a fixed basis against which to estimate the size of any resulting bands. If no band is visible, then the DNA sequence of interest is considered not to have been present. In the case of a PCR optimization experiment, however, the lack of a band may indicate that the primers being tested were not able to isolate the DNA sequence of interest. The sequence of steps involved in gel electrophoresis consists of preparing an appropriate gel medium, placing the gel medium in a tank that contains a buffer solution designed to maintain a consistent charge on the DNA fragments, loading the DNA samples into separate lanes at one end of the gel, and applying electrical charge for a specified period of time. Electrophoretic forces cause the DNA to migrate through the gel at rates dependent on the fragment size. Once the electrophoresis run has completed, the resulting gel can be visualized. Computing systems termed gel documentation systems were in place in each of the participating laboratories to enable a digital image of the resulting gel to be captured for subsequent analysis.

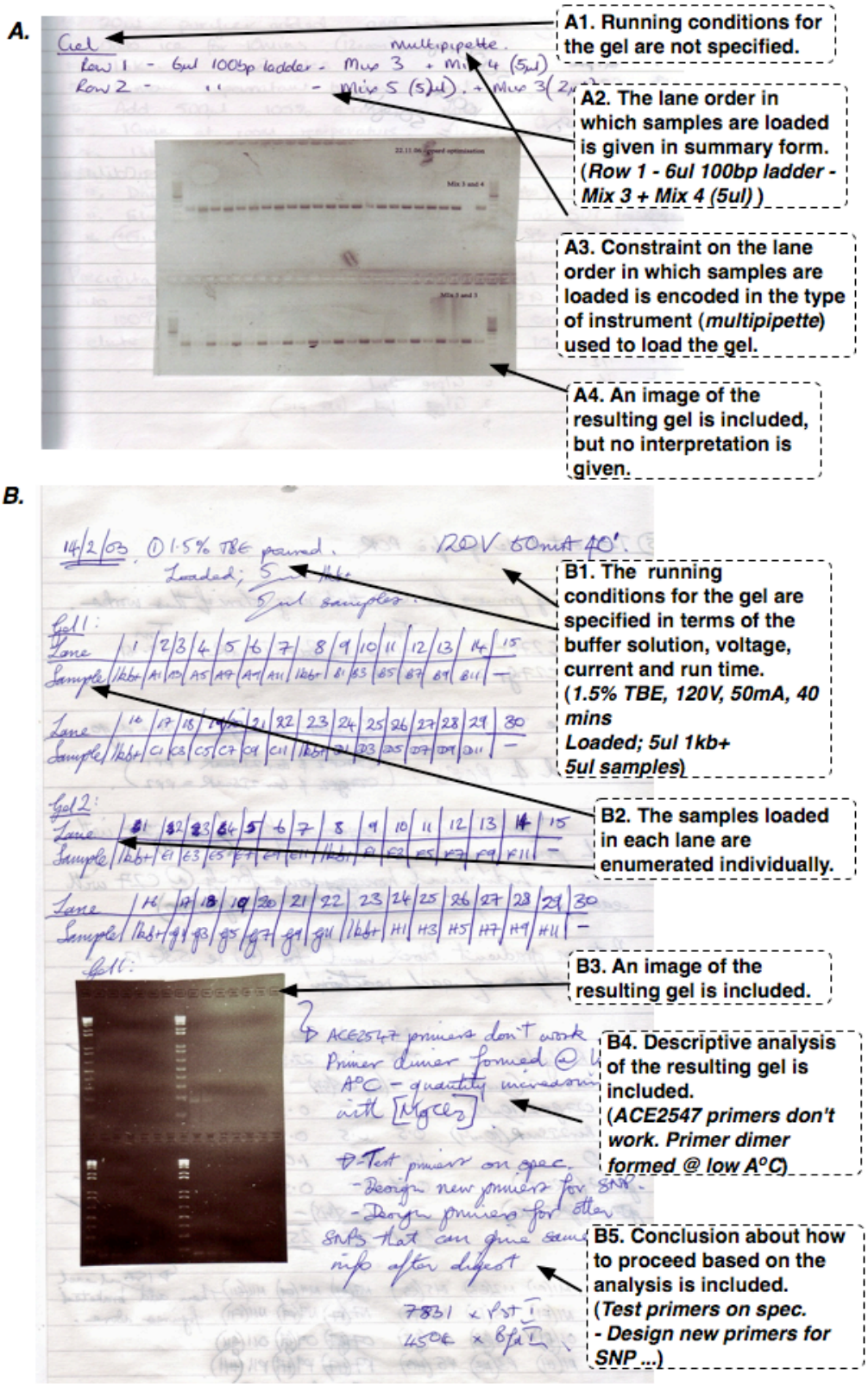


Figure 4-5: Example record entries for gel electrophoresis runs

Showing variation in the representation and content used to record gel electrophoresis runs within experimental records. Gel electrophoresis uses electrical charge to sort fragments of DNA by size. The process involves preparing a gel, loading the samples into separate lanes on the gel, applying electrical charge for a period of time, and visualizing the separation of the fragments across the gel. (*A) and (B).* Scanned excerpts from laboratory records written by two postdoctoral researchers in research laboratory *GS-L1*.

A postdoctoral researcher in research laboratory *GS-L1* recorded the example shown in Figure 4-5A. In this example record, a significant degree of ellipsis is used to achieve an economical style of recording. However, this economy may limit the ability of readers to interpret the results. In particular, the running conditions for the gel in terms of the charge applied, the running time, and the buffer solution are all omitted from the record. These conditions are required in order to be able to repeat the experimental work, but remain unrecorded since they are known to the author of the notebook as the routine settings used in all his/her experiments. More significantly, the loading scheme used to identify the lanes in which individual samples have been loaded onto the gel is summarized rather than itemized. The significance of the loading scheme is that it establishes a basis for interpreting the resulting gel image. Knowledge of the detailed loading order of samples is necessary in order to allow a reader to establish a cohesive link between individual samples and the corresponding bands (or the lack of bands) on the gel. In this example, the omission of a detailed loading scheme is compounded by the lack of a documented interpretation of the results by the author. For the purpose of this study, this style of recording is termed '*loading summary*' recording of a gel electrophoresis run.

A different postdoctoral researcher in the same research laboratory *GS-L1* recorded the example shown in Figure 4-5B. In this example record, the conditions for the gel electrophoresis run are fully specified, and the loading scheme used to load samples onto the gel is fully itemized. The author's interpretation of the resulting gel image is recorded as a narrative block of text together with the proposal for how to proceed on the basis of this interpretation. Although the author's interpretation of the resulting gel is unlikely to be accepted uncritically by any reader given the refined level of scepticism cultivated by molecular biologists, it is important in that it establishes a basis for verification. The granularity of this record supports the gold-standard purpose of laboratory recordkeeping in that it would support replication of the experimental work. For the purpose of this study, this style of recording is termed '*loading and conditions*' recording of a gel electrophoresis run. As mentioned previously in the context of recording styles for experimental purpose, it is interesting to note that these two postdoctoral researchers worked side-by-side at neighbouring benches.

A variation on the 'loading and conditions' style of recording was identified within the laboratory records in the notebook corpus that included an itemized specification of the sample loading scheme but omitted the running conditions for the gel in terms of the

charge, duration, and buffer solution. In this sense, the ability to interpret the resulting gel image in a detailed manner had been prioritized over the ability to repeat the experimental work. For the purpose of this study, this style of recording is termed '*loading*' recording of a gel electrophoresis run.

Figure 4-6 provides an indication of the patterns of use of these approaches to recording gel electrophoresis runs in the notebooks produced by authors in different laboratories and by authors performing different laboratory functions. Since the services currently offered by the service laboratory *GS-L3* did not require the use of manual gel electrophoresis, a 'Not Applicable' category has been added to account for this situation.

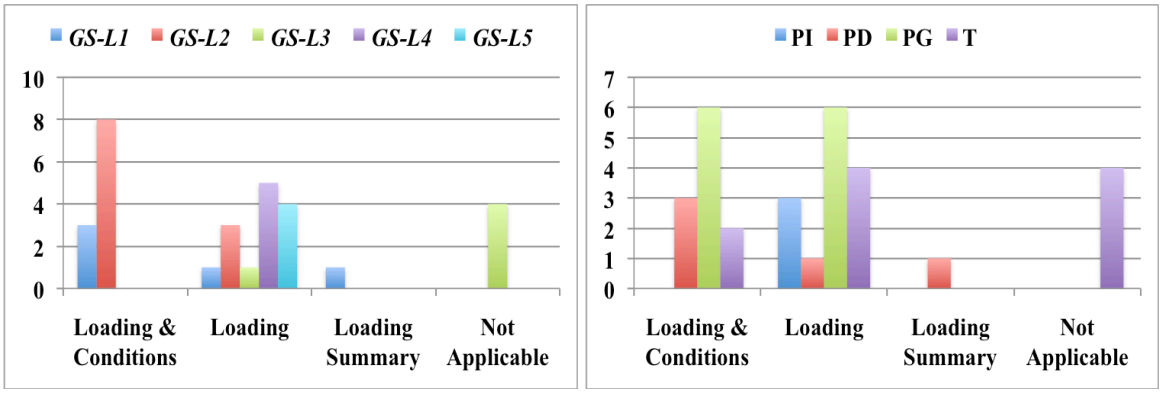


Figure 4-6: Distribution of styles for recording gel electrophoresis runs

Indicating the number of notebooks using different patterns of recording gel electrophoresis runs by laboratory of origin of the notebook author and by function of notebook author in that laboratory.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician.

The 'loading and conditions' style of recording was used in eleven (36.7%) of the thirty notebooks in the corpus, whilst the 'loading' style of recording was used in fourteen (46.7%) of the thirty notebooks. The charts indicate that the use of the 'loading' style of recording gel electrophoresis runs was distributed across all laboratories, whilst the use of the 'loading and conditions' style of recording was focused in specific laboratories.

4.3.2 Temporal variation in records

The following examples survey patterns of temporal variation in the language features used to communicate laboratory records. Diachronic analysis of the laboratory records present in the notebook corpus identified a striking level of consistency over time in the style of

recordkeeping used by researchers and technicians to document their experimental work. In this sense, temporal variation in the style of recordkeeping could be said to be purposive in that it was typically an intentional response to a specific situational context. The following subsections examine temporal variation in the use of language within laboratory notebooks at two levels. The first level is concerned with variation between authors in terms of the manner in which they manage the flow of laboratory work. The second level is concerned with variation within the notebooks of individual authors in terms of recordkeeping responses to specific patterns of laboratory work. In particular, the following subsections examine temporal variation in the use of language to record multi-tasking work, to record routine and non-routine work, and to support current and retrospective use of the records.

4.3.2.1 Chronological and experiment-focused recording

Figure 4-7 overleaf illustrates the two conventional approaches identified within the corpus of laboratory records that characterize how scientists adapt record production to the multi-tasking nature of laboratory work.

As discussed in section 3.3.5.1, the ethnographic study of laboratory recordkeeping highlighted the 'staccato' or intermittent nature of conducting experiments. Many experimental procedures involved an initial period of activity at the laboratory bench during which scientists performed tasks such as preparing samples, preparing reaction mixes, setting up laboratory devices, and loading samples into laboratory devices. These periods of activity were often followed by intervals during which the reactions proceeded, and these intervals varied in duration from minutes to hours to days. In order to manage their workload, the majority of researchers and technicians in the participating laboratories routinely used these intervals as an opportunity to run multiple experiments simultaneously. This multi-tasking approach to laboratory work necessitated the simultaneous construction of laboratory records pertaining to multiple experiments in a single 'linear' laboratory notebook.

The example shown in Figure 4-7A was recorded by a principal investigator in research laboratory *GS-L5*. In this example record, each task performed at the laboratory bench is recorded sequentially in the order in which it was performed. Accordingly, the notebook forms a sequential diary of the work carried out at the laboratory bench, in which successive entries may or may not relate to the same experiment. Any tasks pertaining to

different experiments that were performed on the same day are recorded as separate entries showing the same date but with different experimental purposes. In order to establish a cohesive link between entries pertaining to the same experiment that are now distributed throughout the notebook pages, this approach requires the use of linguistic features to enable cross-referencing. Given the consistent use of date and purpose to document notebook entries, a commonly used primary indexing scheme within laboratory recordkeeping relies on a combination of date and purpose.

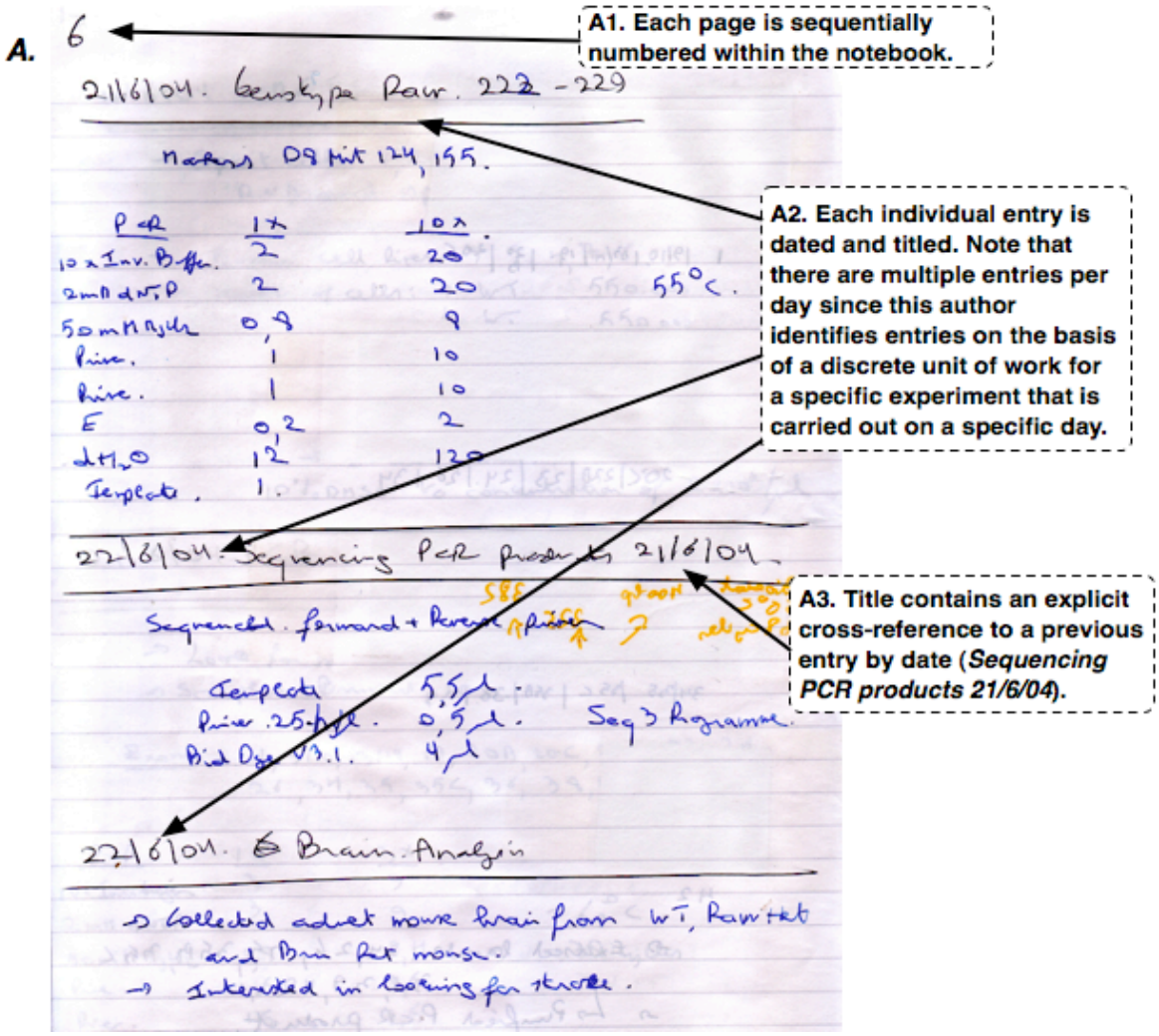


Figure 4-7: Example entries for chronological and experiment-focused records

Showing variation in the representation and content used in experimental records resulting from different approaches to managing experimental workflow. Laboratory work is intermittent in the sense that it involves natural intervals when reactions have started but the scientist must wait for them to complete before moving on to the next step in an experiment. Consequently, scientists may run multiple experiments simultaneously. (*A*). Scanned excerpt from a laboratory record written by a principal investigator in research laboratory *GS-L5* showing interleaved entries for the dates 21/6 - 22/6/2004.

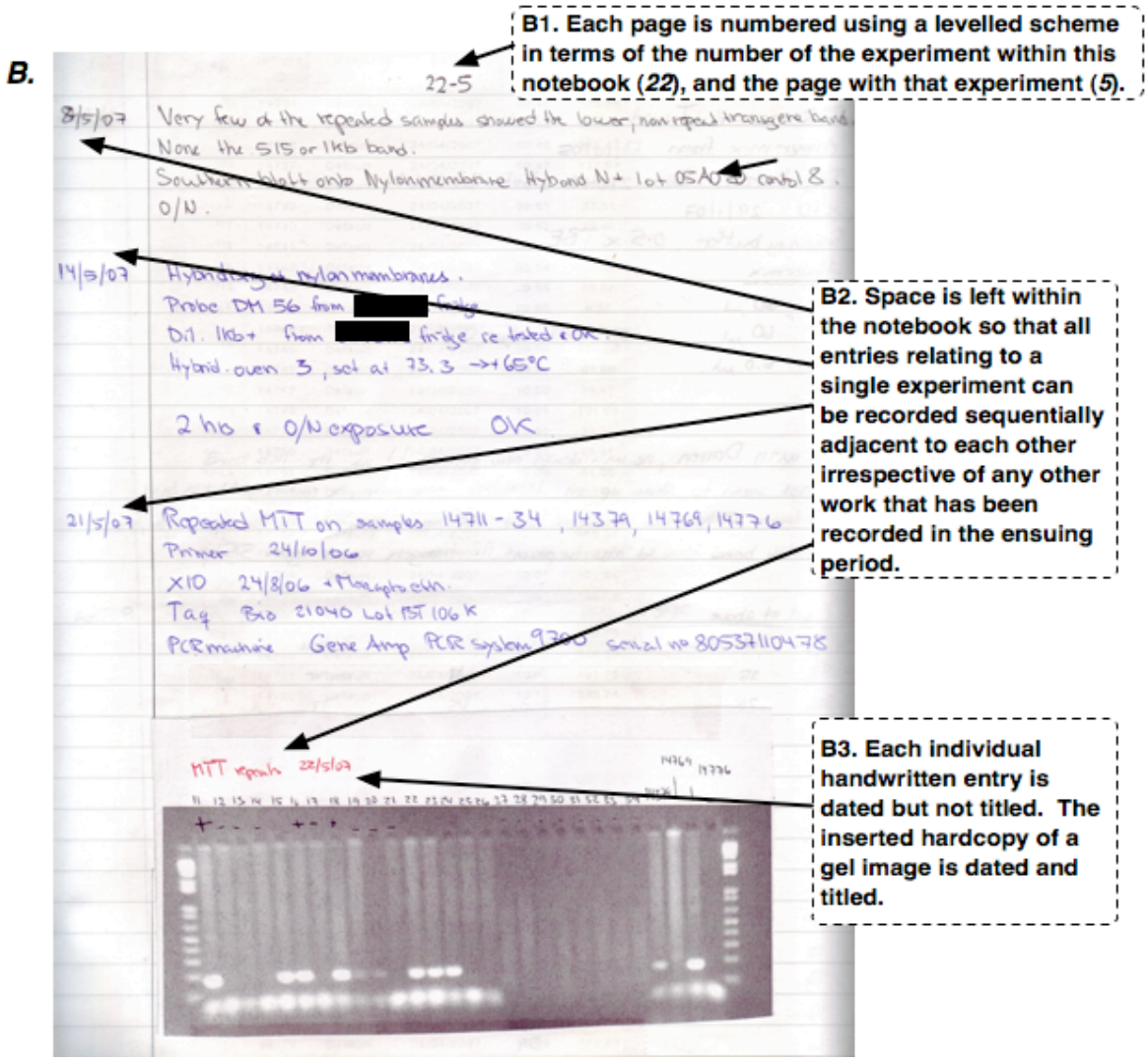


Figure 4-7(Cont'd): Example entries for chronological and experiment-focused records (*B*). Scanned excerpt from a laboratory record written by a technician in research laboratory *GS-L2* showing a continuous entry accumulated over multiple dates.

Although used in both examples in Figure 4-7, another type of indexing scheme that was found on only a limited basis within the notebook corpus was the use of page numbering. For the purpose of this study, this style of recording is termed '*chronological and referenced*' recording of experimental work. The attention paid to cross-referencing in this approach is intended to support both the original author of the record and potentially other categories of reader.

It is important to note, however, that it is not only necessary in the 'chronological and referenced' style of recording to establish references between whole entries in laboratory

records; it is also necessary to establish links at a much finer level of granularity[111]. In particular, it is necessary to establish references between individual samples, reagents, experimental conditions, or reaction products within those entries. This is illustrated in Figure 4-7A in the first entry dated "22/6/04" through the reference to "PCR products 21/6/04" which acts as an anaphoric reference to a specific reaction product identified within the entry for the given date.

Given the need to reconstruct whole experimental records from the individual entries distributed throughout a notebook under this style of recordkeeping, it was perhaps surprising to note another variation on the 'chronological and referenced' style of recording within the laboratory records in the notebook corpus. This variation did not employ explicit cross-referencing schemes based on date, purpose, or page number. Instead, identifying correlations between entries relied upon implicit cues typically using the repetition of the same noun phrase in entry titles. In the records in the notebook corpus, this was often realized through the repetition of a gene name or the names of a batch of samples. In this approach, a reader would be required to infer correlation, for example, between entries entitled "PCR for samples 219-230" and "Sequencing samples 219, 220 and 230" based on the thematic progression signalled by the overlap in sample numbers. This style of recording, which omits explicit referencing schemes, is termed 'chronological' recording of experimental work.

The example shown in Figure 4-7B was recorded by a technician in research laboratory *GS-L2*. In this example record, each individual experiment is recorded as a single continuous block. Sufficient blank space is left between the records of individual experiments so that partial entries may be added to the record of each experiment as and when a task is completed at the laboratory bench. Accordingly, the notebook does not form a chronological sequence of the work carried out at the laboratory bench, but is instead structured into 'sections' for each experiment that contain all entries for work pertaining to that individual experiment. Depending on the complexity of the experiment, these experiment sections could comprise a single page or multiple pages. The example in Figure 4-7B shows a single page within a multi-page experiment section as indicated by the assigned page number. This spatial layout, in which all entries pertaining to any given

[111] Bunton (1999) introduces the term 'scope' to capture the size of a referent item of text, which can range from an entire document to individual sentences. For this study of laboratory records the concept of scope is modified to reflect not only the size of the referent item but also to reflect the domain-specific hierarchy of referent items. For example, the scope of a PCR thermal cycle subsumes multiple items of more limited scope such as individual temperature settings.

experiment are adjacent to each other, reduces the need for detailed cross-referencing between entries. However, the lack of a chronological index in this style of recordkeeping could prove problematic. It was sometimes necessary within the laboratory settings to identify experimental work that was performed during a particular period of time, for example in order to respond to time-dependent faults such as faulty batches of stock reagents that were in use at a given time. More fundamentally, scientists often use date as an important trigger to help recall their work.[112] In some cases such as the notebook from which the example shown in Figure 4-7B is taken, each experiment is numbered and an experimental index is included at the start of notebook to enable ease of navigation to the individual records. For the purpose of this study, this style of recording is termed '*experiment-focused and indexed*' recording of experimental work. A variation on this style of recordkeeping that does not provide an experimental index is termed '*experiment-focused*' recording of experimental work.

Figure 4-8 provides an indication of the patterns of use of these approaches to recording multi-tasking experimental work in the notebooks produced by authors in different laboratories and by authors performing different laboratory functions.

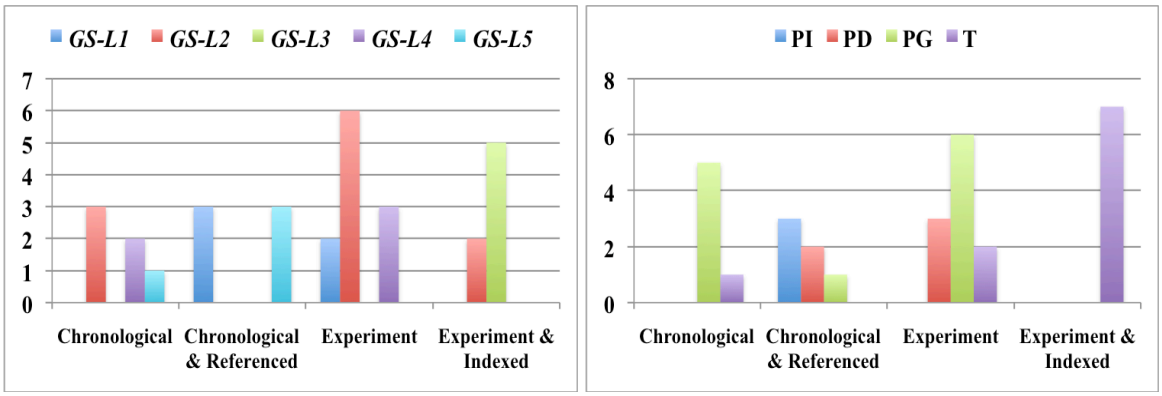


Figure 4-8: Distribution of styles used to record multi-tasking experimental work

Indicating the number of notebooks using different patterns of recording multi-tasking experimental work by laboratory of origin of the notebook author and by function of notebook author in that laboratory.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician.

[112] One of the scientists in a study of work at the Pasteur Institute (Tabard *et al.* 2008) sums up this point by saying "date is something primordial" and advocating the use of chronological indexing on the basis that "that way we're sure to find it".

The ‘experiment-focused’ styles of recording were used in eighteen (60.0%) of the thirty notebooks in the corpus, whilst the ‘chronological’ styles of recording were used in twelve (40.0%) of the thirty notebooks. Given the importance of referencing schemes to enable reconstruction of the distributed records of experiments typical of ‘chronological’ styles of recording, it is interesting to note that only six (50.0%) of the twelve notebooks used referencing schemes. Similarly, only seven (38.9%) of the eighteen notebooks using ‘experiment-focused’ styles of recording also made use of indexing schemes.

4.3.2.2 Recording routine and non-routine work

Figure 4-9 overleaf illustrates the conventional approach identified within the corpus of laboratory records that characterizes how records are adapted over time to document laboratory tasks that have become routine to the author.

As discussed in section 3.3.4.3, the ethnographic study of laboratory recordkeeping identified the potential for breakdowns in interpretation due to temporal variation in the approaches used by some staff when constructing laboratory records. One particular issue in this respect is the recording of work that has become routine to the author in the sense that the technique becomes both well understood and regularly performed.

The academic molecular biology laboratories that participated in this study retain a sharp focus on specialized areas of investigation, and tend to conduct research on the basis of small-scale, highly focused experiments. Most, if not all, of the laboratory work is conducted manually. Many of these manual techniques are well known within the laboratory as the research effort focuses not on methods development but on the generation of data sets and biological model building. These and other aspects of work in the participating laboratories can lead to situations where individuals are required to perform only a small set of core procedures on a repetitive basis. In short, much of the work performed by researchers and technicians can become routine.

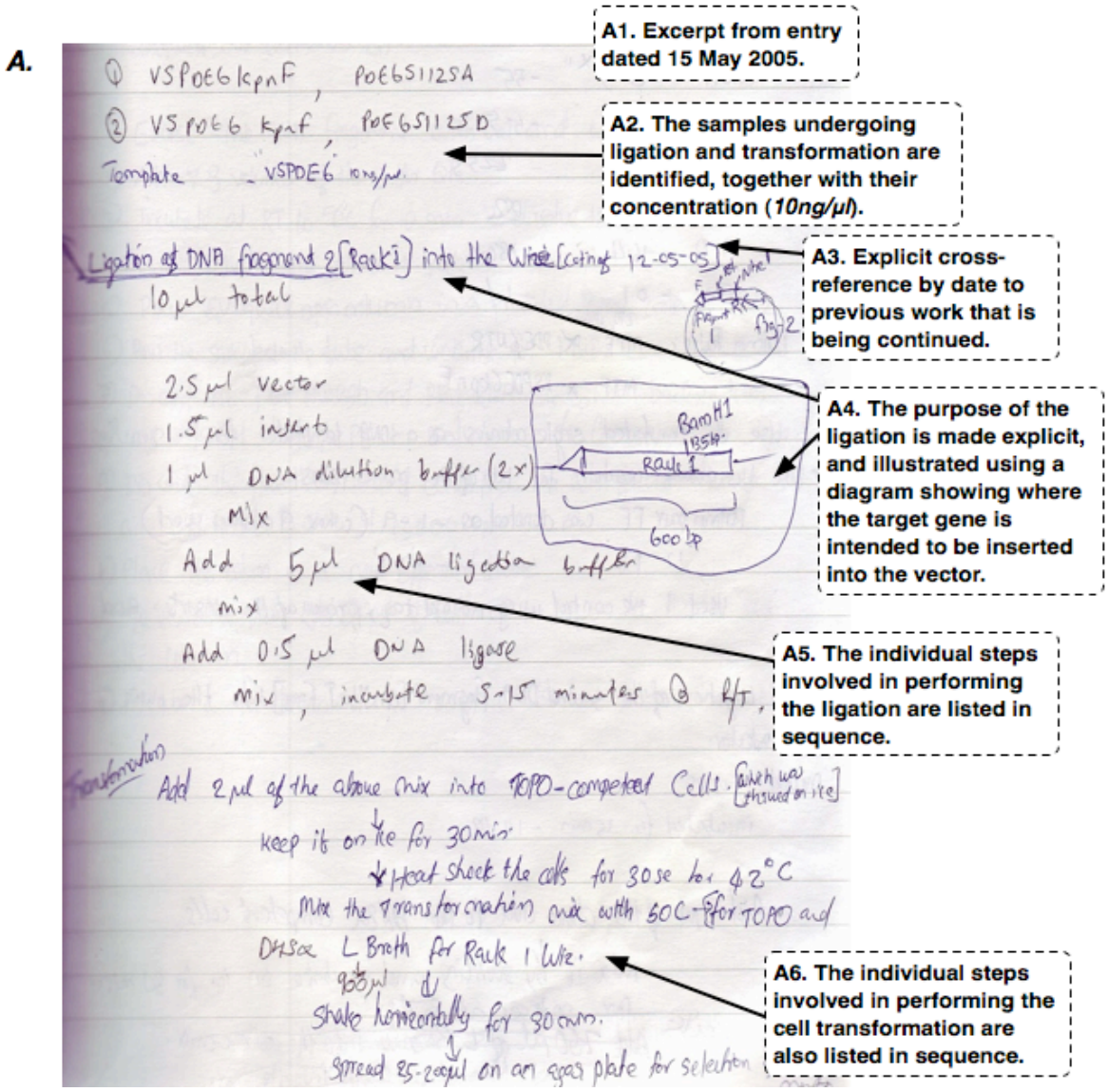


Figure 4-9: Example record entries for routine and non-routine work

Showing variation in the representation and content used in experimental records for routine and non-routine work. The experimental work in both these examples concerns transforming the genetic makeup of bacterial cells in order to insert a DNA fragment of interest. The first step in this work is ligation, which produces a recombined plasmid that contains the DNA fragment of interest in a vector. The second step is transformation, which involves treating bacterial cells to create conditions to allow the plasmid DNA to enter the cells. (*A*). Scanned excerpt from a laboratory record written by a technician in research laboratory *GS-L4* whilst still learning how to perform ligation and transformation using a specific DNA fragment.

Both example records in Figure 4-9 were recorded by the same technician in research laboratory *GS-L4* as records of the same type of experimental work. In particular, the experimental work in both examples is concerned with transforming the genetic makeup of bacterial cells in order to insert a DNA fragment of interest. This work is performed in two

stages, and makes uses of a plasmid[113] cloning vector that can introduce foreign DNA segments into organisms. The first stage, which is termed ligation, produces a recombined plasmid that contains the DNA fragment of interest in a plasmid cloning vector. The second stage, which is termed transformation, involves treating the bacterial cells to create conditions to allow the plasmid DNA to enter the cells.

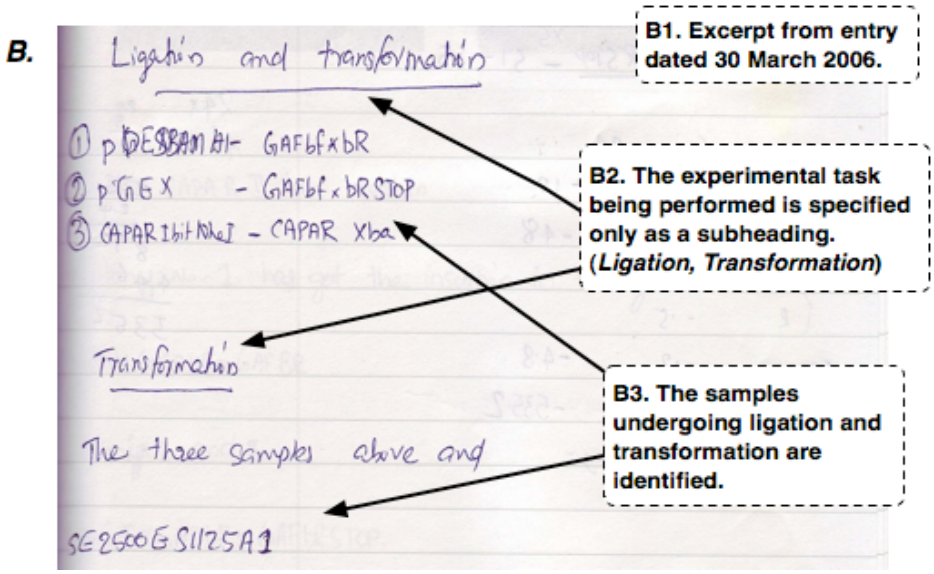


Figure 4-9(Cont'd): Example record entries for routine and non-routine work

***(B).* Scanned excerpt from a laboratory record written by the same technician in research laboratory *GS-L4* after performing this work on multiple occasions.**

As can be seen from the dates attached to the records, a period of approximately ten months elapsed between the executions of the tasks documented in these records. In the example shown in Figure 4-9A, the author is still in the process of learning the laboratory technique for ligation and transformation. At this time, separate headings are used to identify the two stages in the technique. The example record incorporates a significant level of narrative text including directives to specify the sequence of individual steps that are used for each stage. Each of these steps is presented as a separate sentence and includes directives to indicate what is to be done. Volumes of samples and reagents are specified, and duration and temperature for incubation are specified. In addition, the design of the experiment has been supplemented with a graphic illustrating the position at which the DNA fragment of interest will be inserted into the cloning vector. All of these

[113] Plasmids are extrachromosomal DNA elements found in the cells of some prokaryotes such as bacteria. The significance of plasmids in laboratory techniques derives from their role as replicons that are capable of transferring genetic information through autonomous replication within a host independently of the chromosomes.

language features are characteristic of a protocol statement. For the purpose of this study, this style of recording is termed '*protocol-focused*' recording of experimental work.

In the example shown in Figure 4-9B, the technique has now become routine for the author with a consequent change in recording style. All that remains in this example is the headings that were used to separate the two stages in the technique. This extreme degree of ellipsis and abstraction results in a record that appears severely compromised in terms of enabling replication of experimental work. Instead, it functions as a checklist only at a high level of abstraction. Varying levels of abstraction could be applied to transform a laboratory record originally documented in the 'protocol-focused' style of recording, reflecting the author's current level of familiarity with the technique being recorded. It is important to recognize that each of the genre elements of a laboratory record could be independently specified at different levels of abstraction. In particular, it was common for high levels of abstraction to be applied solely to the method elements of a record, whilst the results element continued to be specified in detail.

Two particular variations on this approach to recording were identified within the laboratory records in the notebook corpus. In each case, the style of recording is concerned with a transition from the detailed level of specification embodied in the 'protocol-focused' style of recording to a more abstract level of specification. In the variation illustrated by Figure 4-9B, no reference is used to link the abstracted record of experimental work back to the 'protocol-focused' style of recording. For the purpose of this study, this style of recording is termed '*abstracted'* recording of experimental work. In another variation, which is not illustrated, an explicit reference is included in the abstracted record of experimental work to establish a cohesive link between the abstracted record and the record written in the 'protocol-focused' style of recording used to perform an experiment. Alternative forms for these references included references to previous experiments, and references to information resources external to the laboratory notebook such as Internet websites, local folders, or named individuals within the laboratory. For the purpose of this study, this style of recording is termed '*abstracted and referenced'* recording of experimental work.

Figure 4-10 overleaf provides an indication of the patterns of use of these approaches to recording routine and non-routine experimental work in the notebooks produced by authors in different laboratories and by authors performing different laboratory functions.

The consistent use of the 'protocol-focused' style of recording throughout all entries in notebooks was found in seven (23.3%) of the thirty notebooks in the corpus, whilst the 'abstracted and referenced' style of recording was used in seventeen (56.7%) of the thirty notebooks. The use of 'abstracted' styles of recording was not found in the notebooks produced within the service laboratory *GS-L3*, with the exception of one notebook reserved by a technician to use when testing out new devices and techniques. Notebooks used by technicians in the service laboratory to record work for client orders were recorded in the 'protocol-focused' style of recording.

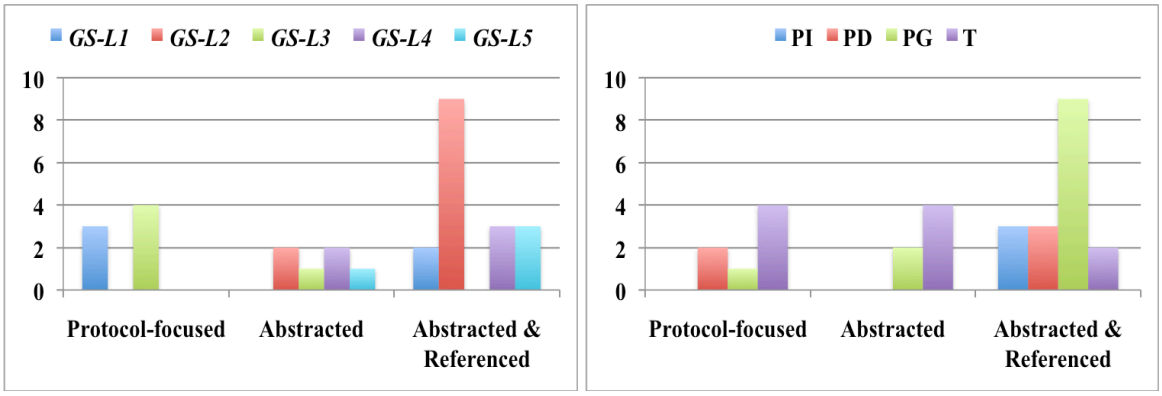


Figure 4-10: Distribution of styles used to record routine and non-routine work

Indicating the number of notebooks using different patterns of recording routine and non-routine experimental work by laboratory of origin of the notebook author and by function of notebook author in that laboratory. The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 4-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician.

4.3.2.3 Current and retrospective recording

Figure 4-11 overleaf illustrates two conventional approaches identified within the corpus of laboratory records that characterize how records could be adapted to prioritize the current use of records over the retrospective use of records by simplifying the production of experimental records at the expense of later interpretation.

As discussed in section 3.3.3, the ethnographic study of laboratory recordkeeping identified a tension between different categories of laboratory staff in terms of the expected readership for laboratory records. The main issue in this respect concerned the construction of laboratory record as either community archive or personal resource. Whilst the legal position was clear in terms of formal ownership of laboratory records in each of the participating laboratories, a more fundamental issue concerns whether the content held

in laboratory records enables them to support future organizational activities in multiple contexts of use. This includes future use of records both by the original author and by other users.

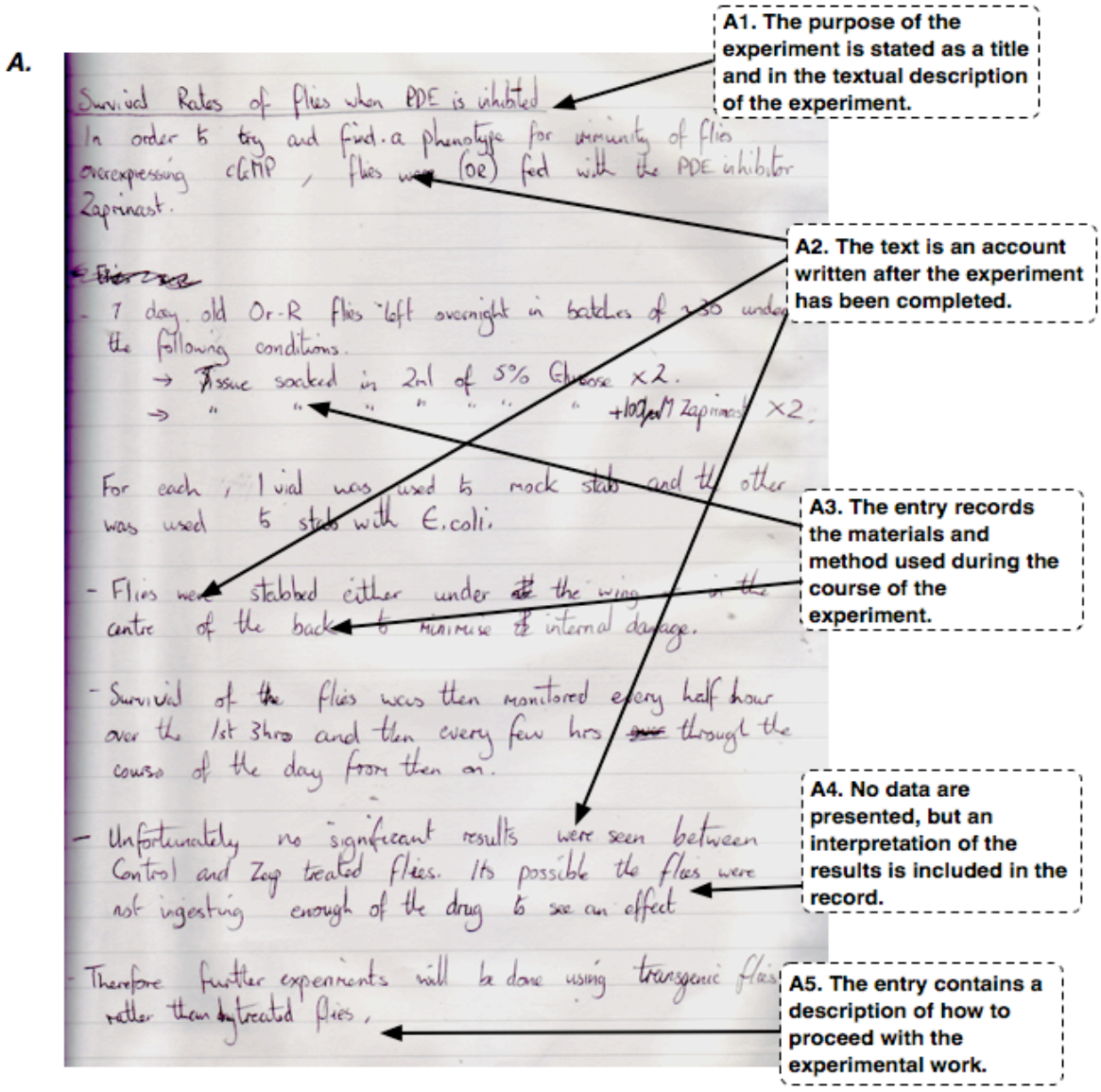


Figure 4-11: Example record entries for current and retrospective use

Showing variation in the representation and content used in experimental records to support current and retrospective use of the record. The experimental work in both these examples concerns a survival immunity assay that is used to investigate the immune response of different genetic strains of *Drosophila melanogaster* to bacterial infection. (*A*). Scanned excerpt from a laboratory record written by a postgraduate researcher in research laboratory *GS-L4*.

In each of the examples shown in Figure 4-11, the records document the same type of experimental work in the form of a survival immunity assay. This is a core laboratory technique in research laboratories such as *GS-L4* that use *Drosophila melanogaster* as a model organism for their research. The assay is used to investigate the immune response

of different genetic strains of *Drosophila melanogaster* to bacterial infection, and involves exposing batches of flies to the infection by stabbing them with an instrument containing a bacterium such as *Escherichia coli*. Data collection involves counting the number of flies that have succumbed to the infection at multiple time points over the duration of the assay.

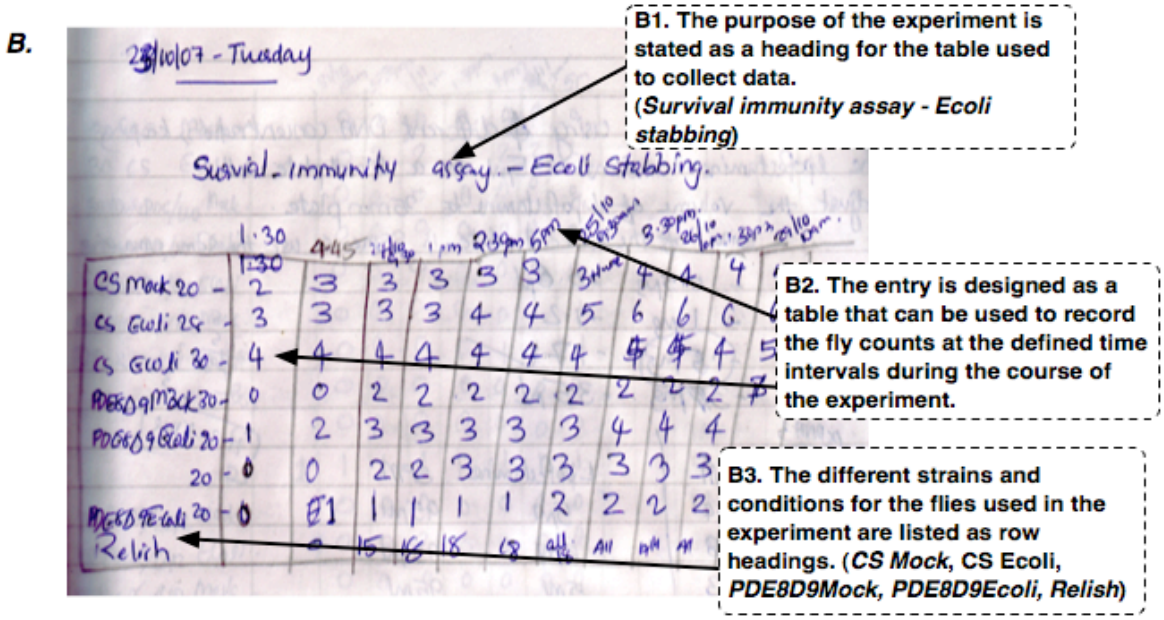


Figure 4-11 (Cont'd): Example entries for current and retrospective recordkeeping *(B).* Scanned excerpt from a laboratory record written by a technician in research laboratory *GS-L4.*

The example shown in Figure 4-11A was recorded by a postgraduate researcher in research laboratory *GS-L4*. This example record begins with a narrative statement of the purpose of the experiment, and continues with a narrative statement specifying the method and materials used for the experiment in which individual steps in the protocol are bulleted as in the 'protocol-focused' style of recording. An interpretation of the results obtained from the assay is provided, together with a statement on how to proceed based on these results. In this sense, the record is able to support retrospective use by specifying both the protocol used and the results obtained in a detailed manner. For the purpose of this study, this style of recording is termed '*report-focused*' recording of experimental work.

The example shown in Figure 4-11B was recorded by a technician in research laboratory *GS-L4*. In contrast to the example shown in Figure 4-11A, the emphasis on language use in this example record is firmly on support for the data collection stage of the experiment. The record is structured as a table containing rows for each of the genetic strains used in the assay, and columns for each of the time points at which counts of the flies are to be

taken. In this sense, the recording style is designed as a tool to facilitate data recording whilst performing the experiment. The purpose of the experiment is recorded in the 'title-focused' recording style, which provides the only contextual cue to the experimental method that would have been used. It is important to recognize here that different forms of assay are possible, each of which would involve data collection in a similar manner to that used in this example. No explicit record of the experimental protocol is given, and no interpretation of the experimental results is stated. For the purpose of this study, this style of recording is termed '*collection-focused*' recording of experimental work. One consequence of this style of recording is that since all data collected during the course of the assay are included in the record, it is possible in some cases for readers of the record to re-compute the results of the experiment. In this sense, the 'collection-focused' style of recording does offer some support for retrospective use.

No distribution charts are included for these styles of recording as use of the 'collection-focused' style of recording was identified only within notebooks in research laboratory *GS-L4* where manual assays of this type were common. It is interesting to note that only the technician notebooks made use of the 'collection-focused approach', whilst the postgraduate researchers preferred the 'report-focused' approach. In the case of the postgraduate researcher, the data collection aspect of the assay relied on the use of a transient form of recording using a notepad with the results stored in digital format in a spreadsheet. Bar charts showing the results of the assay, as generated by the spreadsheet software, were inserted into the notebook entries. Interestingly, the technician also employed the same spreadsheet to generate bar charts showing the results of the assay. These charts were also inserted in the notebook for some entries leading to a variation in recording style termed '*collection-focused with results*' recording of assay records.

4.3.3 Laboratory-level variation in records

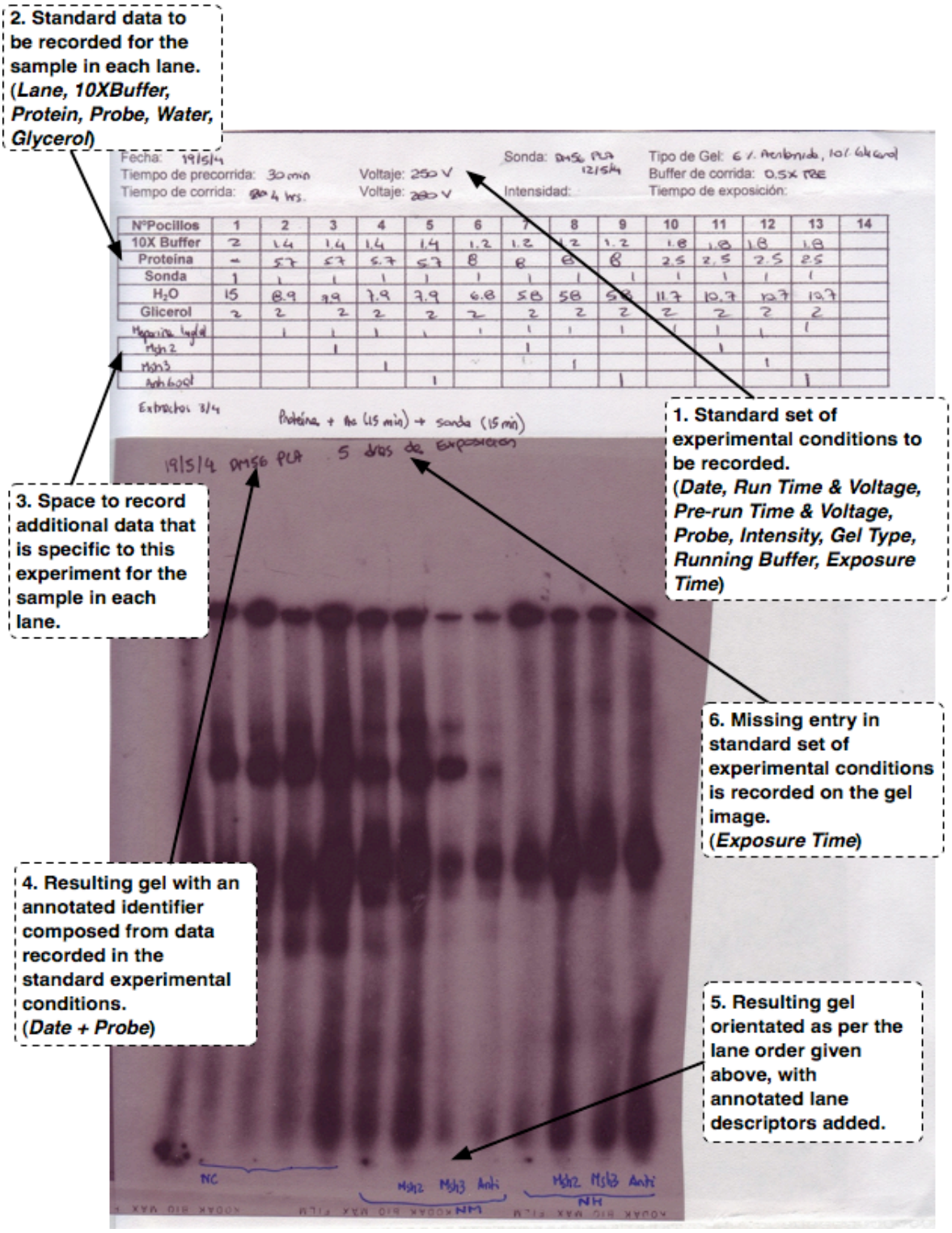
The following examples survey patterns of variation in the language features used to communicate laboratory records that characterize the impact of policy decisions taken by laboratories or individuals within laboratories. In particular, these examples characterize different forms of procedural and information standardization that were identified within the notebook corpus including a template-based approach to recordkeeping, a kit protocol conformant approach to recordkeeping, and a centralized approach to laboratory recordkeeping.

4.3.3.1 Template-based recordkeeping

Figure 4-12 illustrates an approach identified within the corpus of laboratory records that demonstrates the use of a fixed-form template for constructing laboratory records for a specific laboratory technique.

The template shown in Figure 4-12 overleaf has been designed, in particular, to support the performance of electrophoretic mobility shift assays (EMSA). This laboratory technique, which is also known as an EMSA gel or a gel shift assay, is essentially a variation on the gel electrophoresis technique described earlier in section 4.3.1.3. This particular form of assay is used to assess and visualize whether a specific protein is capable of binding to a DNA or RNA sequence. In essence, the assay involves comparing the relative positions of a band produced on the gel by the DNA or RNA sequence on its own, and a band produced by the DNA or RNA in reaction with the protein. If the DNA or RNA sequence is capable of binding to the protein, a larger molecule will be formed and the band corresponding to the combined molecule will appear shifted on the gel due to this change in size.

The template is normative in that it defines the data to be recorded by the scientist whilst performing the steps involved in this laboratory technique. In particular, the top section of the form is used to define the data to be recorded in respect of the running conditions for the gel, which consists of the gel type, the buffer solution used, the charge applied, and the time over which it was applied. The main tabular section of the form performs a dual function. Firstly, it details the reaction mix to be used for each sample through the use of the labels attached to each table that identify the reagents to be added to the mix. It is important to note that a subset of these reagents ("Buffer", "Proteina" = protein, "Sonda" = probe, "H_2O", "Glicerol" = glycerol) are fixed, whilst a subset may be entered by hand to allow the table to be customized to the specific needs of the experiment. Secondly, the columns in the table identify the loading order that is used to load the samples on to the EMSA gel. It is this detailed specification of the loading order that establishes a cohesive link between the data captured in the template form and the gel image that is used to visualize the results of the experiment. For the purpose of this study, this style of recording is termed '*template*' recording of experimental work.



Nº Pocillos	1	2	3	4	5	6	7	8	9	10	11	12	13	14
10X Buffer	2	1.4	1.4	1.4	1.4	1.2	1.2	1.2	1.2	1.8	1.8	1.8	1.8	
Proteína	–	5.7	5.7	5.7	5.7	8	8	8	8	2.5	2.5	2.5	2.5	
Sonda	1	1	1	1	1	1	1	1	1	1	1	1	1	
H_2O	15	8.9	9.9	7.9	7.9	6.8	5.8	5.8	5.8	11.7	10.7	10.7	10.7	
Glicerol	2	2	2	2	2	2	2	2	2	2	2	2	2	
Msh2			1				1				1			
Msh3				1				1				1		
Anti 6x-His					1				1				1	

Figure 4-12: Template-based laboratory record

Showing an example of laboratory recordkeeping in which the content and representation of the record has been embodied in a fixed-form template. This template is design to support a specific laboratory technique termed an electrophoretic mobility shift assay (EMSA), which is used to investigate whether a protein is capable of binding to a DNA or RNA sequence. Note that the form is written in Spanish (*Buffer de corrida* = running buffer, *Fecha* = Date, *Intensidad* = intensity, *Pocillos* = lanes, *Sonda* = probe, *Tiempo de corrida* = running time, Tiempo de exposción = exposure time, *Tiempo de precorrida* = pre-running time, *Tipo de Gel* = gel type, *Voltaje* = voltage).

The example shown in Figure 4-12 was recorded by a postgraduate researcher based in research laboratory *GS-L2* at the time of this study. This template form of recordkeeping was not, however, employed within *GS-L2*. Instead, this example record was identified in a notebook recording work carried out by the postgraduate researcher during a period of research in an academic molecular biology laboratory at a South American university.

No distribution charts are included for this style of recording as the use of a fixed-form template for recordkeeping was only found in a single notebook within the corpus investigated for this study, where it was used consistently for all records in that notebook. It is not the case that the EMSA laboratory technique differed from other laboratory techniques in being uniquely amenable to template-based recordkeeping. It was also not the case that the principal investigator in the laboratory mandated use of a normative form of recordkeeping. Instead, the template grew out of a bricolage effort by an individual researcher who routinely performed the technique and adopted the use of the form as a means of supporting his/her laboratory work. Other researchers who entered this laboratory recognized the benefits of this template-based approach for their own work and copied its use, with the result that use of the template was propagated throughout the research group and beyond in a spontaneous manner.

4.3.3.2 Kit protocol conformance in recordkeeping

Figure 4-13 overleaf illustrates an approach identified within the corpus of laboratory records that characterizes the construction of laboratory records driven by conformance to external standards. In particular, this example demonstrates an approach to constructing laboratory records that assures and communicates conformance with protocol specifications such as those commonly supplied with the commercially available laboratory kits that were routinely used within the research and service laboratories participating in this study.

Many common laboratory procedures have now been reified into a kit form available from commercial vendors. One advantage of these off-the-shelf kits is that they provide a level of procedural standardization within laboratory practice. The kits are typically delivered in a box containing all necessary reagents together with step-by-step instructions in a kit manual.

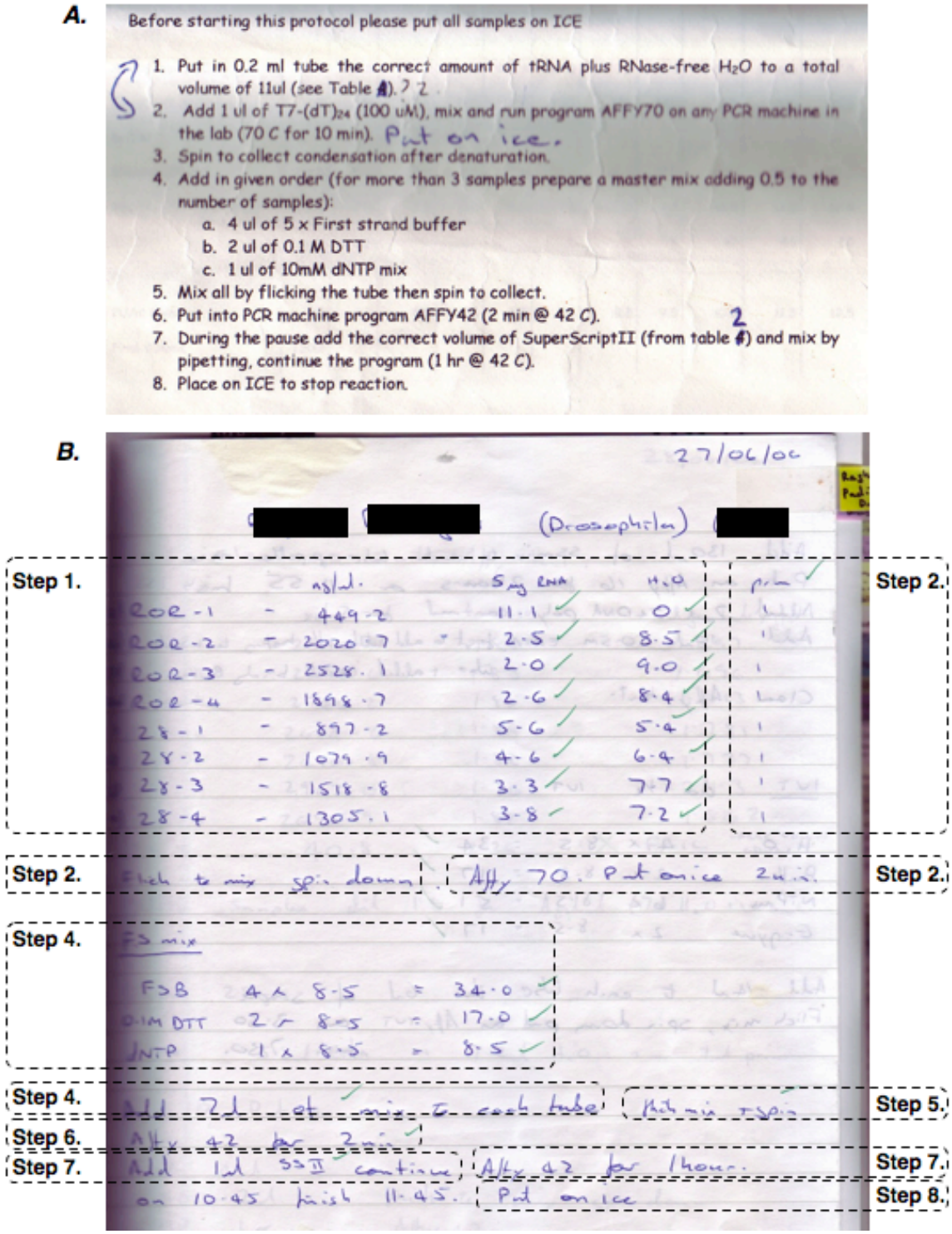


Figure 4-13: Kit protocol conformance in laboratory records

Showing the intertextual mapping between the individual steps listed in a protocol document, and the text written in a laboratory record documenting an execution of that protocol. (*A*). Scanned excerpt from an Affymetrix GeneChip®[114] protocol for use in preparing samples for microarray analysis. (*B*). Scanned excerpt from a laboratory record written by a technician in laboratory *GS-L3* whilst following the protocol. Dashed boxes labelled with the number of a protocol step identify the individual segments of the handwritten record that correspond to each protocol step. A client name in the laboratory record has been blocked out to preserve anonymity.

[114] Product information on GeneChip® is available from Affymetrix Inc at http://www.affymetrix.com [accessed 01 March 2011]. All trademarks are acknowledged.

Although providing an efficient tool for a number of the more common laboratory procedures, kit usage has also raised the potential for 'lost' knowledge within laboratories in the sense that a new generation of laboratory staff rely on the kits and may be unaware how to prepare the reagents or perform the protocol manually[115]. In this sense, a number of laboratory procedures have now achieved the status of "black boxes" (Wiener 1961).

Figure 4-13A shows an excerpt from the step-by-step instructions provided with a kit used during the preparation of samples for microarray analysis, and Figure 4-13B shows an example record written by a technician in service laboratory *GS-L3* whilst using these instructions. It is important to note the correlation between the level of detail specified in the example record and the level of detail specified in the kit instructions. Each of the numbered steps in the sequence of instructions gives rise to an equivalent entry in the written record. In some cases such as step 2, instructions containing multiple clauses give rise to multiple entries. In each case, the text used in the written record is derived from the instruction text so that, for example, step 5 given in the instructions as "Mix all by flicking the tube then spin to collect" becomes "flick mix + spin" in the written record. It is also important to recognize that each of these entries in the written record were accumulated over time during the course of following the instructions, so that the last entry in the written record at any time point identified the current point reached in the execution of the instructions. In this sense, the dynamic process of writing the record served as a means of monitoring progress, whilst the completed textual record continued to serve as an assurance that the instructions had been applied correctly. For the purpose of this study, this style of recording is termed '*standard-conformant*' recording of experimental protocols. This style of recording was consistently used in all records in the notebooks produced by technicians in the service laboratory *GS-L3*.

Figure 4-14 overleaf shows an excerpt from a laboratory record kept by a postdoctoral researcher in research laboratory *GS-L1*. In contrast to the detailed level of specification that characterizes the 'standard-conformant' style of recording in the service laboratory, the use of kit protocols within the research laboratories typically resulted in a minimal record providing only an exophoric reference to the type of kit used, with the specific kit identified using a combination of the manufacturer and kit name. In particular, no statement was recorded of the reagents used in the kit nor of the steps involved in

[115] This phenomenon has led to the introduction of 'kit-scientist' as a term of abuse within some laboratory settings.

performing the kit protocol. For the purpose of this study, this style of recording is termed '*standard-referent*' recording of experimental protocols.

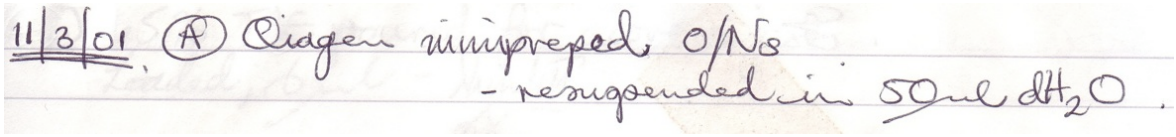


Figure 4-14: Kit protocol reference in laboratory records

Showing a referent form of intertextual mapping between use of laboratory kit protocols, and the text written in a laboratory record documenting an execution of that protocol. Scanned excerpt from a laboratory record written by a postdoctoral researcher in laboratory *GS-L1* whilst using the protocol from a QIAGEN®[116] kit for plasmid preparation.

No distribution charts are included for these styles of recording as the consistent use of the 'standard-conformant' style of recording was identified in the notebook corpus only within notebooks in service laboratory *GS-L3*. The 'standard-conformant' style of recording was used in all records in four of the five notebooks produced by technicians at work in that the service laboratory. It was used less consistently in the other notebook, which was reserved by a technician in the service laboratory for experimental work to test out new devices and techniques, and to design customized forms of microarray experiment. The 'standard-referent' style of recording was dominant across all types of author in the research laboratories, although the 'standard-conformant' style of recording was present in isolated records corresponding to the first use of laboratory kits.

4.3.3.3 Influence of centralized laboratory practices

Figure 4-15 overleaf illustrates an approach identified within the corpus of laboratory records that characterizes the impact of centralized laboratory practices on the laboratory records kept by scientists. In particular, this example demonstrates a situated response to laboratory recordkeeping in a setting that has centralized practices for the naming and storage of samples, for the reagents and conditions to be used in experiments, and for the curation of experimental results.

The example shown in Figure 4-15 was recorded by a postdoctoral researcher based in research laboratory *GS-L2* at the time of this study. This form of recordkeeping was not,

[116] Product information on QIAGEN® kits is available from QIAGEN Inc at http://www.qiagen.com [accessed 01 March 2011]. All trademarks are acknowledged.

however, employed within *GS-L2*. Instead, this example record was identified in a notebook produced by the researcher during the course of postgraduate study in an academic molecular biology laboratory at a North American university.

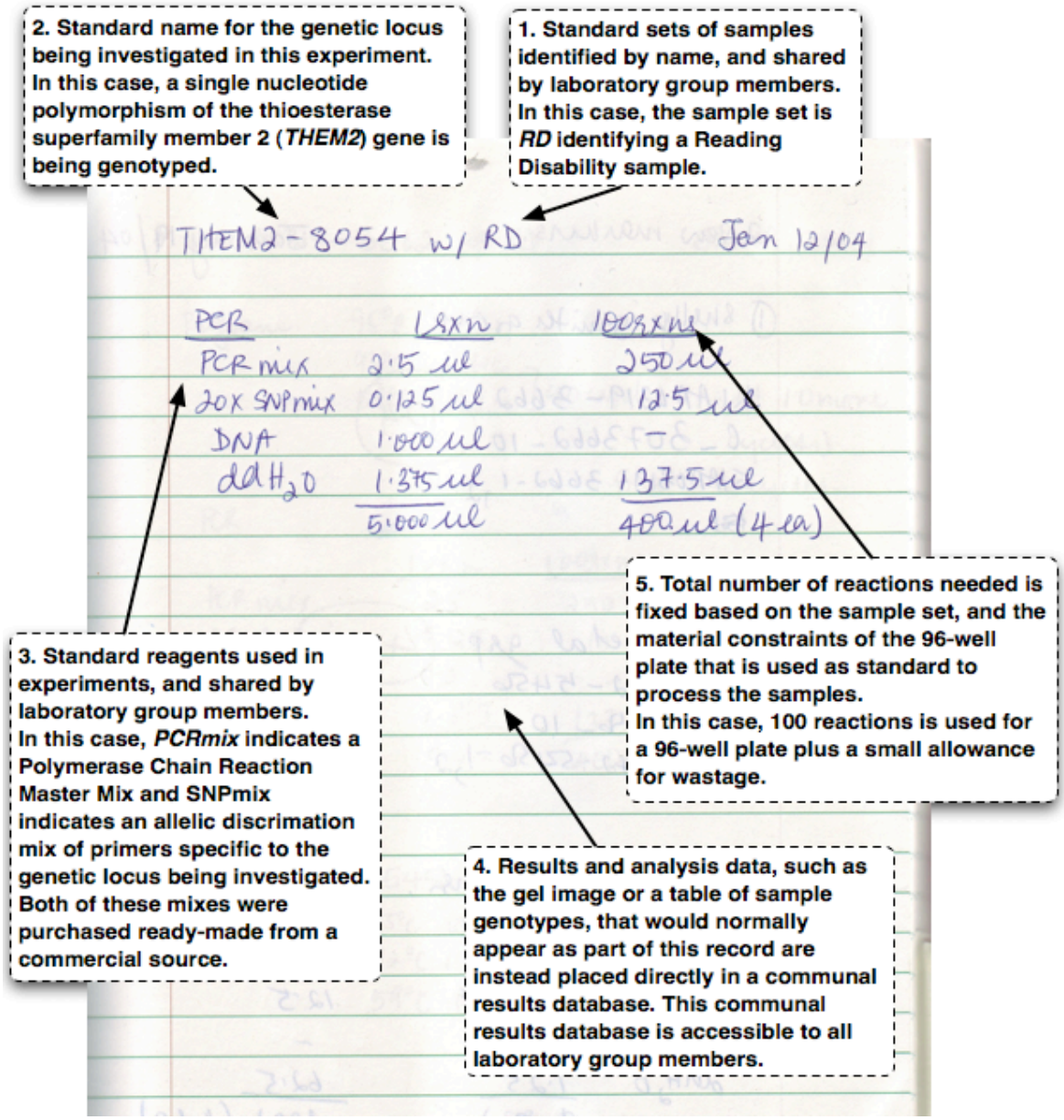


Figure 4-15: Influence of centralized group practices on laboratory records

Showing the influence of different forms of standardized group practice on the records kept by an individual scientist. This record relates to work carried out by a postdoctoral researcher in research laboratory *GS-L2* whilst working in a previous position in a laboratory outside this study. This previous laboratory group employed information, procedural, and standardization in its laboratory practices, and held all experimental results in a central laboratory database. The minimal form of the laboratory record kept by the individual relies on knowledge of the standardized coding schemes in place within the host laboratory. No copies of the experimental results are included in this record since they are held at a location known to the group members and open to the group members.

Although both authors are now based in research laboratory *GS-L2*, it should be noted that the author of this example record is not the same individual who used the 'template' style of recording discussed in section 4.3.3.1.

The type of experimental work carried out in the North American laboratory was broadly similar to the type of experimental work carried out in research laboratories *GS-L1*, *GS-L2*, and *GS-L5* although focused on the genetic basis of a different disease. In particular, the example record shown in Figure 4-15 documents a PCR designed for use in DNA sequencing work aimed at determining allelic variation within a set of samples for a named gene, which is similar to the type of experimental work previously described in the examples from other laboratories such as those in section 4.3.1. In the example record in Figure 4-15, the gene of interest is *THEM2*[117] and the record forms part of a larger project investigating allelic variation at multiple loci in multiple genes across batches of samples. Given the range of genes and the number of samples involved in the project, multiple scientists undertook the experimental work required to generate DNA sequencing data for the sample set. In order to coordinate the work of this group, the principal investigator in this laboratory setting had adopted specific forms of procedural and information standardization. In particular, sequencing results data were stored in digital form in a centralized, shared database that was accessible to the laboratory group. The batches of DNA samples obtained for the project were named according to group-wide naming conventions, and stored in freezer locations according to a group-wide storage policy. Standard sets of primers and other reagents were defined within the group for each target gene, together with standard experimental conditions such as PCR thermal cycles.

The use of these different forms of procedural and information standardization has significantly influenced the style of recording used for the example record shown in Figure 4-15. In particular, the purpose of the experiment is recorded in a highly encoded manner in the 'title-focused' style of recording. Decoding the title requires an awareness of the naming conventions adopted with the laboratory group to specify both the genes of interest and the sample batches obtained for the study. With this awareness, the title can be successfully decoded as DNA sequencing of a named batch of samples ("RD") at a specific locus ("8054") of a specific gene ("THEM2"). Much of the information that would be required for a PCR such as the thermal cycle, and the individual forward and reverse

[117] The *THEM2* or thioesterase superfamily member 2 encoding gene is one of a number of genes that is being investigated in connection with developmental dyslexia or reading disability.

primers is omitted from the record on the basis that this can be assumed from the shared context of the group. Of particular importance is the fact that no resulting gel image is included in the record, which stands in stark contrast to the style of recordkeeping used for this type of experimental work in all other notebooks in the corpus. In this case, the results are directed away from the notebook to the communal repository used by all members of the group to store and to analyse the experimental results in digital form. In this sense, the temporal extent of the activity encountered in the example record has changed since it is now focused solely on the 'materials and methods' aspect of the laboratory work. In terms of the generic "Introduction-Method-Results-Discussion" structure that was realized to varying levels of detail in records within the notebook corpus, this record is now focused solely on the "Introduction" and "Methods" aspects of laboratory work. The "Results" and "Discussion" has been removed to the communal repository. For the purpose of this study, this style of recording is termed '*group-centralized*' recording of experimental work.

No distribution charts are included for this style of recording as it was only found in a single notebook within the corpus investigated for this study.

4.4 Discussion

The following discussion evaluates the results of this genre analysis study of laboratory records with respect to the three questions posed for the investigation. These results are discussed with reference to the findings of the ethnographic study of laboratory recordkeeping as presented in section 3.4.

1. *What roles do laboratory records play in the discourse of academic molecular biology laboratories?*

The ethnographic study highlighted the fact that writing laboratory records involves considerable contextual competency by requiring individual authors to judge when experimental data becomes sufficiently "difficult" or "unusual" to require it to be recorded in the notebook. Individual authors invoked the need to support specific organizational functions as the rationale for making this judgement. In this way, the pattern of language used in constructing the laboratory records was driven by the need to support selected organizational functions. In some cases as in the service laboratory *GS-L3*, the rationale was dependent on the needs of the laboratory as a whole. In the majority of cases within

the research laboratories, the rationale was dependent on the approach preferred by the individual researcher.

Both technicians in the service laboratory relied on the use of a detailed intertextual mapping to demonstrate conformance to a laboratory protocol definition. In some cases, the laboratory protocol formed part of an off-the-shelf laboratory kit; in other cases, the laboratory protocol was developed in-house as a variation on an existing standard. The pattern of language used in these laboratory records served a dual purpose (*cf.* Berkenkotter 2008; Schryer 1993). Firstly, it enabled the act of writing the record to verify that the correct procedure was being applied. Secondly, it imbued the resultant record with evidential value that could be used to serve as proof of conformance for quality assurance procedures. It is interesting to note that this approach was followed even though no supervisory procedures were in place to validate the records.

The role of laboratory records within the research laboratories was considerably more diverse as indicated by the pattern of language use within these settings. This pattern of language use encompassed variation in the level of detail specified within the records, in the types of referencing scheme used to link records and their elements, in the intertextual mappings that shaped the records, and in the spatial arrangement of the records. It is important to note in this respect that the set of entries contained within individual notebooks was strikingly consistent in its use of language. Instead, variation in language use centred on variation between notebooks and between notebook authors.

Once again, the records management viewpoint is that in order for laboratory records to function as an effective written mode of communication, they must be both reliable and usable. The majority of notebooks in the corpus used for this study exhibited summarized forms of recording that routinely omitted selected experimental data. Further, in many cases, few contextual cues were available to enable a reader to reconstruct the missing experimental data. In this sense, the majority of notebooks could not be considered as reliable in that they did not provide full and accurate descriptions of the transaction that they embody. Accordingly, the records would not be useable in serving the 'gold-standard' purpose of enabling other scientists to reproduce the experimental work described in the written record (*e.g.* Ebel *et al.* 2004; Macrina 1995; Wickman 2010). It is important to note that this problem stems from the omission of knowledge such as temperatures, volumes, and concentrations that could be represented in written form rather

than from any concept of tacit knowledge (*e.g.* Collins 1985; Lynch 1997; Polanyi 1967; Schmidt 1997).

The use of summarized forms of recording was, however, broadly consistent with the concept of laboratory notebooks as a personal archive. Given the prior knowledge and contextual awareness of the original author, these summarized forms of recording could be sufficient to enable experiments to be reproduced by the author. At the same time, the use of summarized forms served a dual purpose by enabling a time-efficient approach to recordkeeping.

It is important to note that efficiency is this sense is focused on the author and the act of writing the record. Whilst the use of descriptive economies supports authors in writing the laboratory records in their personal archives, this efficiency typically comes at the cost of internal and external readers attempting to make sense of the records in subsequent contexts of use (*e.g.* Bloor 1999). As Swales (2004:219, emphasis in original), *inter alia*, comments in respect to the description of experimental methods within research articles, "the reader of *fast*, highly condensed Methods sections may in fact be reading such texts much more slowly than he or she would read their more discursive counterparts."[118]

Some postdoctoral and postgraduate researchers did not rely on summarized forms of recording but, instead, consistently recorded detailed statements of individual samples, experimental conditions, and interpretation of results. These styles of recording were used primarily to serve two organizational functions in terms of supporting publication-readiness, and supporting process refinement/error recovery. This latter organizational function warrants particular emphasis given the effort incurred by many researchers in optimizing the temperature, volumes, concentrations, reagents, and other conditions used in general laboratory protocols in order to the meet specific needs of their research work.

[118] The use of the adjective "fast" to describe statements of experimental method derives from Swales and Feak (1994:164-167), who introduced the notion of a "speed" scale against which to categorize descriptions of experimental method. Slow method descriptions characteristically provide explicit detail on procedures, lack presumptions about background knowledge, include explanations and examples, and provide expanded terminology. Fast method descriptions exhibit the opposite characteristics.

2. *What are the structures, content, and representations that characterize the genre of laboratory records in academic molecular biology laboratories, and to what extent do these vary across different contexts of use?*

The specific patterns of variation in the recording styles identified within the notebook corpus are summarized in Table 4-5 overleaf. These encompass the patterns found in both the service-orientated and research-orientated laboratories. As highlighted by this table, the language used in laboratory records varied across multiple dimensions. This encompassed variation in the language used by different authors to document the same structural units such as reaction mixes or statements of purpose, variation in the language used by the same author to record experiments over time, and variation driven by the formal or informal policy in effect within the laboratory in which an author was based.

As previously mentioned in relation to the first research question, the variation in language use across these different recording styles was manifested in the level of detail specified within the records, in the explicitness of the referencing scheme used to link records and their elements, in the intertextual mappings that shaped the records, and in the spatial arrangement of the records. The trade-off exhibited by each of these different styles concerned the level of detail recorded in the record against the time spent recording the entry. Specific forms of abstraction that were used to achieve summarized forms of recording included identifying only a subset of the experimental conditions, identifying sample groups rather than enumerating each individual sample within the group, and identifying only some of the individual steps in a protocol.

It is interesting to note that studies such as Bloor (1999) and Swales (2004:219-224) have reported similar variation in the level of detail used to describe experimental methods within research articles, ranging across a spectrum from highly abstracted (sometimes referred to as ‘clipped’ or ‘fast’) texts to highly detailed (sometimes referred to as ‘elaborated’ or ‘slow’ texts). In contrast to clipped texts, elaborated descriptions of experimental method are characterized by Swales (2004:220), *inter alia*, as avoiding assumptions about any background knowledge on the part of the reader, minimizing the use of shorthand notations, including additional information such as justifications and examples, and employing structural aids such as headed subsections to organize the text.

Table 4-5: Recording styles identified in the corpus of laboratory records

Recording Style
Unit-level variation
‘title-focused’ recording of experimental purpose
‘narrative’ recording of experimental purpose
‘specified’ recording of PCR thermal cycles
‘specified and identified’ recording of PCR thermal cycles
‘annealing only’ recording of PCR thermal cycles
‘loading summary’ recording of gel electrophoresis
‘loading and conditions’ recording of gel electrophoresis
‘loading’ recording of gel electrophoresis
Temporal variation
‘chronological’ recording of experiments
‘chronological and referenced’ recording of experiments
‘experiment-focused’ recording of experiments
‘protocol-focused’ recording of experimental work
‘abstracted’ recording of experimental work
‘abstracted and referenced’ recording of experimental work
‘report-focused’ recording of experimental work
‘collection-focused’ recording of experimental work
Laboratory-level variation
‘template’ recording of experimental work
‘standard-conformant’ recording of experimental work
‘standard-referent’ recording of experimental work
‘group-centralized’ recording of experimental work

Summarizing the range of recording styles identified in the sample set of laboratory records used for the gene analysis study.

The analysis of multiple notebooks written by the same authors indicated that the recording style used by each author remained consistent both within and across notebooks unless specific laboratory-wide policies were in force. In this sense, each individual author had evolved their preferred way of writing laboratory records. In the absence of laboratory-wide policies for recordkeeping, even authors working in close proximity within the same laboratory might employ different approaches to recordkeeping. No specific correlation was found between the different type of functions performed by authors in the laboratory (principal investigator, postdoctoral researcher, postgraduate researcher, technician) and the preferred style of recording used by an author.

Explicit forms of standardization were identified in only two of the notebooks in the corpus, neither of which related directly to work undertaken within the laboratory settings that participated in this study. In one notebook, standardization took the form of a fixed-form template that embodied the recommended data to be recorded for a specific laboratory technique. In the other notebook, a group-centralized approach to laboratory work was in place that propagated multiple forms of information and procedural standardization to constrain reagent naming, sample naming, sample storage procedures, and laboratory protocols across the multiple members of a single research group.

3. *How do readers of laboratory records in academic molecular biology laboratories make sense of laboratory records in different contexts of use?*

The ethnographic study highlighted the fact that the majority of scientists relied on recontextualization of the data held in laboratory records into other genres as the means of retrieving the information held in the laboratory records, and that this recontextualization activity was undertaken in the main by the original author. This genre analysis study indicates the necessity for the recontextualization activity to include retrieval/reconstruction of the experimental data that has been hidden through the use of abstracted forms of recording experimental data in laboratory notebooks. Experimental data in this sense encompasses data relating to experimental methods, experimental results, and other aspects of experimental work. In order to be able to interpret the laboratory records in their direct written form, readers other than the original author will also be required to attempt this process of retrieval and reconstruction. With some of the recording styles identified within the notebook corpus, explicit cues are present in the form of cross-referencing schemes, thematic progression across records, pre-programmed laboratory equipment, and *de facto* laboratory standard practices. However, these explicit cues do not cover all the detail necessary to expand all forms of abstraction used with the laboratory records. In this sense, there were clear limitations on the ability of any individual reader to expand some laboratory records, including those readers who were able to access additional information that could be available from other workers in the original laboratory.

5 A Study of Reading Laboratory Records

This chapter of the thesis reports on a study using reading protocols to investigate how scientists interpret a sample set of authentic laboratory records produced in academic molecular biology laboratories. The chapter is again presented broadly in line with the IMRD structure (*cf.* Thompson 2001) to address the aims of the study, the method used to conduct the study, the results obtained, and the conclusions drawn from these results.

5.1 Aim of the study

The primary aim of this study was to investigate how scientists at work in academic molecular biology laboratories make sense of laboratory records written by other scientists. To this end, the study set out to investigate how scientists read and interpret written laboratory records in different contexts of use, and how structural and linguistic variation in these laboratory records may influence the ability of different readers to make sense of the records.

The motivation for this reading protocol study derived principally from the findings of the ethnographic study reported in chapter 3. These findings indicated that the dominant viewpoint held by laboratory staff other than principal investigators conceptualized laboratory notebooks as a personal resource. This viewpoint contributed to the minimal use by scientists of the records written in laboratory notebooks other than their own notebooks. Instead, other forms of communication such as visual demonstrations, one-to-one meetings, and graduate presentations were the preferred genres for communicating the information held in the notebook records. Only those notebooks written by authors who had since left the laboratory were consulted as written resources, typically by principal investigators, and then only if the information sought was not available in other derived genres such as research articles.

Further motivation for the reading protocol study derived from the findings of the genre analysis study reported in chapter 4. These findings focused on the manner in which the 'personal resource' view of the laboratory notebook was manifested in variation in the language used in the laboratory records written by different scientists. This included variation in the representations and layouts used to document these records, in the

constitutive elements of records, in the use of referencing within the records, and in the framing mechanisms used to contextualize the records.

The joint findings from the ethnographic and genre analysis studies indicate that scientists do not routinely read laboratory records other than their own, and will encounter significant variation in recording style should they eventually be required to read other records. Consequently, this reading protocol study set out to investigate how scientists at work in academic molecular biology laboratories would make sense of authentic laboratory records that exhibit different recording styles when reading the records in different contexts of use.

5.2 Design of the study

The study was designed as a qualitative analysis of reading protocols captured from a range of scientists whilst using a sample set of laboratory records to accomplish prescribed laboratory tasks. These prescribed tasks were designed to simulate typical contexts of record use such as understanding how to perform a laboratory procedure or interpreting the results of an experiment. In order to reflect authentic variation in the language used in laboratory records, the sample records used for the reading protocols were all authentic laboratory records written by researchers and technicians during the course of their work in academic molecular biology laboratories. This design was chosen in order to enable observation and comparison of the processes used by laboratory members to make sense of the laboratory records produced by others, albeit in a less naturalistic setting. As discussed in section 1.4.4, the perceived strengths of the reading protocol analysis approach include the ability to examine the cognitive role of genre in linguistic behaviour, and the ability to support a qualitative comparison of the approaches used by laboratory staff to interpret different styles of laboratory recordkeeping.

5.2.1 Ethical approval procedure

Since the study involved human participants as both authors of the laboratory records to be used as sample texts and as readers providing reading protocols for these sample texts, it was reviewed and approved by the Ethics Committee of the Faculty of Information and Mathematical Sciences at the University of Glasgow under application number FIMS00618. This approval ensured that the study conformed to the code of conduct set

out by the British Psychological Society for studies involving human participants (BPS Ethics Committee 1978). All participants were approached and recruited to the study only after ethical approval had been confirmed. Introductory and debriefing sessions were conducted with each participant in accordance with an interview script. The information sheet sent to prospective participants to describe the study is presented in Appendix 7 of this thesis.

5.2.2 Sample cases

5.2.2.1 Selection policy for readers

The selection of sample readers for this study of scientists' reading practices was driven by similar strategic and practical considerations to those that influenced the design of the previous ethnographic and genre analysis studies of laboratory recordkeeping (see sections 3.2.2.1 and 4.2.2.1).

The key strategic consideration was again to adopt a form of stratified purposeful sampling (Patton 2001:240) to investigate the reading practices of multiple categories of scientist, thus improving the representativeness of the sample set of readers and the scope for observing potential sources of variation. In particular, readers were selected from multiple laboratories rather than a single laboratory, and from both research-orientated and service-orientated laboratories within the university environment. This approach was chosen in order to consider potential variation in the reading practices both between scientists in the same laboratory, and between scientists in different laboratories.

The reading practices of scientists at different stages of their academic career were examined. Academic career stages were again ascribed on the basis of the scientist's function as technician, postgraduate researcher, postdoctoral researcher, or principal investigator in a similar manner to that used for the previous ethnographic and genre analysis studies of laboratory recordkeeping. This approach was chosen in order to consider potential variation in the reading practices between different stakeholder responsibilities, and between novice and experienced users in line with Swales' (2004) gradation of junior and senior researchers (see section 2.6.2). In particular, this enabled a comparison of reading practices associated with laboratory records documenting laboratory procedures that are both familiar and unfamiliar to the reader.

An additional strategic consideration for this study was to include readers from those laboratories that had been observed as part of the previous ethnographic and genre analysis studies of laboratory recordkeeping. Insights gained from these studies enabled the reading protocols produced by the scientists at work in those laboratories to be analysed in a contextualized manner. In this way, the findings of the three studies within the project framework could be integrated on a direct basis. It is important to note that readers from other laboratories were also included in the study.

Including the same laboratories that participated in the previous ethnographic and genre analysis studies again delivered practical benefits by facilitating the recruitment of readers due to the existing professional and personal contacts that had been established with a range of scientists within these laboratories. Interestingly, recruitment of participants to this reading protocol study was greatly supported by the fact that many researchers, both junior and senior, had no experience reading other scientists' records. These individuals were interested in the opportunity to examine other styles of recordkeeping. Given the available access to multiple independent laboratories, no attempt was made to access readers in laboratories outside the same UK university used in the previous studies.

5.2.2.2 Selection policy for authors and records

The selection of sample authors and records for this study of scientists' reading practices was also driven by strategic and practical considerations.

The key strategic consideration was again to adopt a form of purposeful sampling. In this case, however, the authors were selected in order to gain access to a sample set of records that encompassed the various recording styles identified during the genre analysis study as reported in section 4.3. Sample records were drawn from multiple notebooks written by multiple categories of author. In particular, sample records were selected from authors in multiple laboratories rather than a single laboratory, and notebooks were selected from both research-orientated and service-orientated laboratories within the university environment. This approach was chosen in order to enable the capture of reading protocols for sample records produced both in the readers' own laboratories and in other laboratories. Selecting records from multiple laboratories also enabled the inclusion of records documenting a wider range of molecular biology laboratory procedures including common techniques and more specialized techniques. This approach enabled comparison of how scientists make sense of records for both familiar and unfamiliar procedures.

A further strategic consideration for this study was to focus primarily, but not exclusively, on sample records produced by scientists at work in those laboratories that had been observed for the previous ethnographic and genre analysis studies of laboratory recordkeeping. Insights gained from these studies enabled the language used in the sample records to be interpreted in a contextualized manner so that the findings of the three studies could be integrated on a direct basis within the overarching project framework.

Practical considerations for the study derived primarily from the difficulty in gaining access to laboratory notebooks for sample records. Recruiting authors from the same laboratories that participated in the two previous studies again facilitated access to sample records through the contacts that had been established with scientists in these laboratories. This level of trust was essential given the importance attached by individual scientists to their laboratory notebooks as discussed in the findings of the ethnographic study.

5.2.2.3 Sample laboratories

Table 5-1 lists the laboratories from which reader and/or authors were recruited for this reading protocol analysis of laboratory records.

Table 5-1: Laboratories for the reading protocol study

Laboratory	*Type*	*Description*
RS-L1 *(No previous participation)*	Research	A small university research laboratory with approximately 5 members. Members are involved in human genetics research with a specific focus on the use of transgenic models to investigate one type of disease[119].
RS-L2 *(ES-L4, GS-L2)*	Research	A small university research laboratory with approximately 6 members. Members are involved in human genetics research in projects investigating the genetics of human disease with a specific focus on one disorder.
RS-L3 *(GS-L5)*	Research	A small university research laboratory with 2 members. Members are involved in projects investigating the genetics of human disease with a focus on identifying genes and pathways for one type of disease. This is a new laboratory in its first year of operation under a recently promoted principal investigator.

[119] Such models are used in particular within molecular biology laboratories engaged in molecular genetics research in order to investigate the mechanisms associated with the onset of specific diseases, and with the progression of specific diseases. These models can be used to investigate a wide range of disorders including neurological disorders, developmental disorders, vascular disease, myotonic dystrophy, and types of cancer.

Table 5-1(Cont'd): Laboratories for the reading protocol study

Laboratory	*Type*	*Description*
RS-L4 *(ES-L3, GS-L3)*	Service	A common services department housed within a university facility but offering laboratory services and consultancy in sequencing and data analysis to multiple client laboratories within the home university, in other universities, and in other research institutions.
RS-L5 *(ES-L2, GS-L4)*	Research	A large university research laboratory with approximately 20 members formed as a close collaboration of two principal investigators. Members are involved in integrative physiology research using *Drosophila melanogaster* as a model organism for a range of projects including some projects with commercial partners.
RS-L6 *(No previous participation)*	Research	A small university research laboratory with approximately 8 members. Members are involved in human genetics research with a specific focus on investigating the role of signal transduction[120] in types of disease.
RS-L7 *(No previous participation)*	Research	A research institute formed as part of a collaborative endeavour between two universities and a health service. Members are involved in a number of research projects investigating the genetics of human disease with a specific focus on understanding the genetic basis of one type of disease. This institute is involved in collaborative projects with commercial partners working on drug development and target identification.
RS-L8 *(ES-L1, GS-L1)*	Research	A small university research laboratory with approximately 7 members. Members are involved in human genetics research for projects in the field of sports and exercise science with a specific focus on the interaction between environmental factors and hereditary factors on human health and performance.

Summarizing the laboratories that provided authors and/or readers for the reading protocol study in terms of the identifier code assigned to the laboratory, the laboratory type, and a brief description of the laboratory setting.
The identifier code uniquely identifies each laboratory that participated in the study whilst maintaining the anonymity required under the terms of the ethical approval for the study. The laboratory type is used to categorize laboratories into either *Research* laboratories or *Service* laboratories. *Research* laboratories undertake research projects on their own initiative in order to investigate scientific questions of their own choosing, whilst *Service* laboratories provide support services to other laboratories and are commissioned to undertake specialist experimental work on behalf of these client laboratories. The description of the laboratory setting outlines the size of the laboratory, and the broad research interests of the laboratory.

[120] Signal transduction is the mechanism through which signals to cells are converted into specific responses. Deficiencies in signal transduction cascades may impair specific biological processes. For example, deficiences in cAMP signalling are characteristic of several human diseases including cardiovascular disease, renal disease, cancer, diabetes mellitus, and neurological disorders.

All eight of the participating laboratories were housed within a single UK university. It is again important to note, however, that the service laboratory *RS-L4* delivered its services to a number of external clients located throughout the UK. Only two of the research laboratories, *viz*. *RS-L1* and *RS-L5*, made use of the services offered by the service laboratory *RS-L4*. None of the seven research laboratories were involved in any collaborative projects with each other. However, laboratory members in research laboratories *RS-L1, RS-L2* and *RS-L3* were allocated space in the same 'wet' laboratory area and communal office accommodation. In addition, members including the study participants from these three laboratories presented their work at the same weekly laboratory group meeting.

As noted in Table 5-1, three of these laboratories in the form of *RS-L1*, *RS-L6*, and *RS-L7* had not previously participated in either the ethnographic study or the genre analysis study.

5.2.2.4 Sample readers

Table 5-2 overleaf lists the range of laboratory members that participated as readers in the reading protocol study. A more detailed summary of each individual reader including a brief description of the reader's experience in laboratory work is presented in Appendix 8 of this thesis.

All participants who spoke English as a second language regularly read research journals and textbooks in English, regularly used kit manuals and protocols written in English, participated regularly in laboratory meetings held in English, and had completed university-level courses taught in English (see section 2.6.2). Gender was not used to discriminate between those participating in the study as readers, and both male and female scientists were recruited from each laboratory wherever possible (see section 2.6.2).

Individual readers were recruited from research laboratories *RS-L6*, *RS-L7*, and *RS-L8* specifically in order to ensure the inclusion of highly experienced scientists in each of the different laboratory functions within the sample set of readers. Each of these three readers had twenty years or more of experience in a range of bioscience laboratories.

Five readers also participated as authors by providing sample records to be used during the reading protocol study.

Table 5-2: Readers for the reading protocol study

Laboratory	*Readers (N = 15)*							
	Job Function				**Native Language**		**Gender**	
	PI	**PD**	**PG**	**T**	**L1**	**L2**	**M**	**F**
RS-L1	0	0	2	1	2	1	2	1
RS-L2	0	2	0	1	1	2	0	3
RS-L3	1	0	0	1	1	1	1	1
RS-L4	0	0	0	2	1	1	0	2
RS-L5	0	0	1	1	1	1	1	1
RS-L6	0	0	0	1	1	0	1	0
RS-L7	0	1	0	0	1	0	0	1
RS-L8	1	0	0	0	1	0	1	0
All Labs	***2 13.33%***	***3 20.0%***	***3 20.0%***	***7 46.67%***	***9 60.0%***	***6 40.0%***	***6 40.0%***	***9 60.0%***

Showing a numerical breakdown of all participating readers in the reading protocol study by laboratory of origin, by function in that laboratory, by native language, and by gender.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 5-1). The reader's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The reader's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks/writes English as a second language. The reader's gender is indicated using F for a female participant, and M for a male participant.

5.2.2.5 Sample authors

Table 5-3 overleaf lists the range of laboratory members that participated as authors in the reading protocol study. A more detailed summary of each individual author including a brief description of the author's experience in laboratory work is presented in Appendix 8 of this thesis.

All participants who spoke English as a second language had produced written publications in English, had participated regularly in laboratory meetings held in English, and had completed university-level courses taught in English (see section 2.6.2). Gender was not

used to discriminate between those participating in the study as readers, and both male and female scientists were recruited from each laboratory wherever possible (see section 2.6.2).

Table 5-3: Authors for the reading protocol study

Laboratory	*Authors (N = 8)*							
	Job Function				**Native Language**		**Gender**	
	PI	**PD**	**PG**	**T**	**L1**	**L2**	**M**	**F**
RS-L2	0	0	2	0	1	1	0	2
RS-L3	1	0	0	0	0	1	1	0
RS-L4	0	0	0	2	1	1	0	2
RS-L5	0	0	0	1	1	0	1	0
RS-L8	0	2	0	0	2	0	2	0
All Labs	***1*** ***12.5%***	***2*** ***25.0%***	***2*** ***25.0%***	***3*** ***37.5%***	***5*** ***62.5%***	***3*** ***37.5%***	***4*** ***50.0%***	***4*** ***50.0%***

Showing a numerical breakdown of all participating authors in the reading protocol study by laboratory of origin, by job function in that laboratory, by native language, and by gender.
The laboratory of origin is identified using the code number assigned to the laboratory for the genre analysis study (see Table 5-1). The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The author's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks/writes English as a second language. The author's gender is indicated using F for a female participant, and M for a male participant.

5.2.2.6 Sample records

Two specific criteria were used to select the sample set of laboratory records used in the reading protocol study from the authors' notebooks. The first criterion was to select records that exhibited the range of recording styles identified by the genre analysis study reported in chapter 4. The second criterion was to select records that documented a range of laboratory procedures used in academic molecular biology laboratories. In particular, the range of procedures should include core techniques familiar to most laboratory staff, and specialized techniques such as those reserved for work with specific model organisms.

Table 5-4 overleaf summarizes the range of recording styles represented in the sample set of laboratory records that were used for the reading protocol study, whilst Table 5-5

summarizes the range of laboratory procedures described in these laboratory records. A more detailed description of each individual record including a brief description of the record content is presented in Appendix 8 of this thesis.

Table 5-4: Recording styles covered in the records for the reading protocol study

Record Characteristics	Records (N=23)
Unit-level variation	
'title-focused' recording of experimental purpose	13
'narrative' recording of experimental purpose	8
'specified' recording of PCR thermal cycles	2
'specified and identified' recording of PCR thermal cycles	5
'annealing only' recording of PCR thermal cycles	3
'loading summary' recording of gel electrophoresis	4
'loading and conditions' recording of gel electrophoresis	8
'loading' recording of gel electrophoresis	3
Temporal variation	
'chronological' recording of experiments	3
'chronological and referenced' recording of experiments	7
'experiment-focused' recording of experiments	13
'protocol-focused' recording of experimental work	10
'abstracted' recording of experimental work	14
'abstracted and referenced' recording of experimental work	1
'report-focused' recording of experimental work	21
'collection-focused' recording of experimental work	2
Laboratory-level variation	
'template' recording of experimental work	1
'standard-conformant' recording of experimental work	6
'standard-referent' recording of experimental work	3
'group-centralized' recording of experimental work	1

Summarizing the range of recording styles exhibited in the sample set of laboratory records used for the reading protocol study.
Note that specific recording styles do not apply to all records. For example, some records did not include gel images, PCRs, or statements of purpose, and so the associated recording styles do not apply to these records. Note also that multiple recording styles may be present in an individual record. For example, some multiple page records contain multiple entries written in different styles. The names used to identify the recording styles are the terms assigned during the course of the genre analysis study (see section 4.3).

Table 5-5: Procedures covered in the records for the reading protocol study

Laboratory Procedure	Records (N=23)
Core techniques covered	
DNA sequencing of samples involving PCR and gel electrophoresis	5
Optimization of PCR conditions	5
Restriction digest[121]	1
Other techniques covered	
Colony PCR.[122]	2
DNA extraction using a commercial kit	1
Electrophoretic mobility shift assay	1
HEK cell transfection[123]	1
Preparation of samples for a microarray experiment using a commercial kit.	3
Protein kinase assay[124]	1
Stages in IHC[125]	1
Test of commercial kits for IVT[126]	1
Survival immunity assay against *E. coli*	1

Summarizing the range of laboratory procedures documented in the sample set of laboratory records used for the reading protocol study. The laboratory procedures are characterized as either core techniques with which all or most laboratory staff will be familiar, and other techniques that are more specialized in nature or specific to work with particular model organisms.

[121] A restriction digest is a laboratory technique that can be used to prepare DNA samples for further analysis by digesting DNA samples with restriction enzymes that cut DNA into fragments. Specific restriction enzymes are chosen to enable cutting of DNA into fragments at sites that contain specific DNA sequences of interest.

[122] Growing bacterial colonies on suitable culture plates is a basic laboratory technique for producing a population of cells. In molecular biology laboratories, this approach is commonly used to grow populations of cells that have been transformed with a vector containing a DNA sequence of interest. A colony PCR is a laboratory technique that uses a variant of PCR to screen bacterial colonies to determine whether the vector containing the DNA sequence of interest has been inserted into the colony.

[123] Human embryonic kidney (HEK) cells are a cell line now grown in cell culture, and which can be transformed to include specific genes of interest for subsequent analysis.

[124] Protein kinases play a significant role in regulating cellular pathways as enzymes that can cause functional changes in proteins by chemically altering the proteins through a reaction termed phosphorylation. Deregulated kinase activity is associated with a number of diseases including cancer and diabetes mellitus. A kinase assay is a laboratory technique that can be used to quantify kinase function by detecting levels of phosphorylated proteins.

[125] IHC stands for immunohistochemistry, which is a laboratory technique that can be used to assess and visualize the presence of specific proteins within a tissue slice by introducing specific tagged antibodies that will bind to the protein of interest. The location of the protein in the tissue can be visualized using procedures such as diamonobenzidine (DAB) staining that produce colour stains on the tissue.

[126] IVT stands for *in vitro* transcription, which is a laboratory procedure that forms part of the protocol for preparing samples for microarray experiments. The role of the IVT step is to produce RNA by transcribing RNA from DNA templates, in a mirror of a process that takes place *in vivo*.

5.2.3 Experimental procedure

5.2.3.1 Participant contact

All contact with study participants before and during the study was made directly with the participant and independently of the principal investigator in order to minimize the potential for bias on the basis that access to laboratories had been gained initially through contact with the principal investigators in that laboratory. As with the ethnographic and genre analysis studies of laboratory recordkeeping (see sections 3.2.3.1 and 4.2.3.1 respectively), this procedure was followed in order to ensure that participants did not perceive the experimenter as acting in any way on behalf of the principal investigator. In particular, it was important that authors felt able to offer sample records for the reading study that were representative of their actual recordkeeping practices rather than limited examples of what could be termed 'best practice'. Similarly, it was important that readers did not feel pressurized to perform, but were able to respond naturally to the sample records during the course of the reading experiments.

Prior to observing work in each laboratory, one-to-one meetings were arranged with each individual participant to explain the purpose of the study and to gain written consent. In the case of authors, this meeting also involved agreeing a selection of laboratory records for use in the reading protocol study, and arranging convenient dates on which to gain access to the laboratory notebooks in order to make copies of the selected records. In the case of readers, the meeting also involved arranging convenient dates on which to conduct the reading experiments.

Readers were entitled to payment in the form of a £20 gift voucher for participating in this study. No payment was made to authors participating in the study.

5.2.3.2 Preparation of sample record booklets

Each of the laboratory records selected for the study was prepared for use in the reading experiments according to the same conditions. Digital scans of the original notebook pages were made in order to enable editing of the records to render them anonymous. This was achieved by superimposing filled rectangles to blot out individual names, laboratory names, and any other similar references. No other modification was made to the original

text of the laboratory records. In order to maintain consistency with the original means of presentation, the redacted pages were printed, collated, and bound[127] to replicate the format in which the record appeared in the original notebook. In this way, records originally written in A5 format were presented to readers in A5 format; records written in A4 format were presented in A4 format. Multi-page entries were bound to form booklets that maintained the recto/verso page ordering used in the original notebook.

Prior to making use of the records in reading experiments, the redacted version of the record was shown to the author in order to obtain his/her agreement that the record had been rendered anonymous. This meeting was also used to confirm the purpose of the original laboratory record, and to discuss the original context in which the record was constructed.

5.2.3.3 Selection of sample records for a reader

The subset of sample records used in the reading experiments for each individual reader was chosen to satisfy specific criteria. These criteria were designed to ensure that the overall set of reading protocols encompassed the range of recording styles listed in Table 5-5, to ensure that reading protocols were obtained both for laboratory procedures familiar and unfamiliar to the individual reader wherever possible, and to ensure that reading protocols were obtained for work carried out by colleagues in the individual reader's own laboratory wherever possible. Individuals participating jointly as reader and author were not requested to provide reading protocols for their own laboratory records. Attention was paid to randomizing the order in which records exhibiting different recording styles were presented to each reader in order to minimize any potential for bias due to contrast effects. Appendix 8 of this thesis lists the records selected for each individual reader, together with the order in which these records were presented to the readers during the reading experiments.

[127] The term 'bound' is used here with due deference to my colleagues in the Society of Archivists who specialize in record conservation. The use of the term 'binding' is perhaps too strong in this context given that the sample record booklets were bound in a 'Heath Robinson' manner using staples and tape.

5.2.3.4 Definition of reading tasks

Each reading experiment required the reader to make use of the laboratory record to solve a specific reading task. The criteria used to define appropriate reading tasks for this study was driven by the need to simulate reading tasks that were realistic examples of the use of laboratory records in a laboratory setting, and that provided a basis for comparison between multiple readers working with multiple sample records. Accordingly, all readers were required to perform the same task for each record. This task required the reader to identify the purpose of experiment being described in the record, to describe the materials and methods used in the experiment, and to identify any additional information that he/she would require in order to repeat the experiment. The rationale behind these questions was to simulate use of the laboratory record for two purposes. The first purpose was to interpret the record to understand the results obtained by another scientist. The second purpose was to understand the experimental process used by another scientist in order to be able to repeat the experiment.

5.2.3.5 Location for reading experiments

All reading protocols for each reader were collected in the same physical location, which was a small meeting room within the Faculty of Biomedical and Life Sciences at the University of Glasgow. This meeting room was configured in the same manner for each reader, and only the individual reader and the experimenter were present in the meeting room during the collection of reading protocols. During the reading experiments, the reader was positioned at one side of the meeting room table, with the experimenter positioned a short distance away from the reader. A digital video recorder was placed in a fixed position on a mini-tripod, and directed at the desk in front of the reader. A voice recorder was also placed next to the tripod for use as a backup device in case the video recorder failed.

5.2.3.6 Data collection

All reading experiments were conducted according to a predefined script in order to ensure that the same instructions were given to each reader, and that these instructions were neutral in order to avoid introducing expectancy effects. A copy of the script used by the experimenter for this purpose is included in Appendix 9 of this thesis.

The reading experiment for each reader was scheduled to last approximately 90 minutes inclusive of an initial period of training in verbal reporting. During this experiment, reading protocols for multiple sample records were collected from each reader. It is important to note, however, that no time limit was imposed on the reading protocol for each individual sample record, and each reader was free to spend as much or as little time on the verbal reports for each sample record as he/she deemed necessary. Accordingly, the number of reading protocols captured for each individual reader varied from three to six protocols (mean = 4.07, standard deviation = 0.96).

Before collecting reading protocols for the sample laboratory records, a series of 'warm up' tasks was conducted in order to familiarize the readers with both the setting and with the process of verbal reporting. This period of initial training was not recorded. The tasks used during this initial training did not involve laboratory records or any other bioscience-related tasks, but were instead general tasks involving simple mental arithmetic and travel direction problems.[128]

After this period of initial training, both concurrent and retrospective reading protocols were collected from each reader for a subset of sample laboratory records. For each sample record, the reader was asked to read and interpret the record to perform a specific task. Two items were handed to the reader for each sample record, *viz.* a copy of the sample record booklet and a single page task description. This task description contained a short sentence describing the type of laboratory in which the sample record was produced, and a written statement of the task to be performed using the sample record. In each case, the readers were required to perform the same task for each record. Defining the task in this written statement ensured consistency in the questions posed to each reader. Appendix 9 of this thesis includes an example task description and sample record used in the reading protocol study.

A video recorder and audio recorder were used to capture the verbal report and physical interaction of the reader with the sample record booklet as he/she performed the reading task. Successful protocol analysis requires the reader to continue to think aloud whilst performing the reading task. Consequently, if the reader fell silent for a length of time, the experimenter interrupted to remind the reader to think aloud. Aside from this, the

[128] This approach to training readers is in line with Ericsson and Simon (1993), who recommend the use of initial training using general tasks to familiarize readers to the process of 'thinking aloud' in their influential work on verbal reporting.

experimenter did not interrupt during the collection of the concurrent reading protocols. Writing materials were also made available to reader, who could write on a separate notepad or make notes on the sample record booklet whilst performing the task.

After the reader had completed the reading task to his/her satisfaction, the sample record booklet and the task description were removed. Following a brief interlude during which the sample record booklet, task description sheet, and any notes made by the reader were collected, a semi-structured interview was conducted with the reader to capture a retrospective protocol. During this process, the reader was prompted to recall how he/she made sense of the laboratory records to solve the simulation scenario. The retrospective protocol corresponding to this interview was also captured using the video recorder and audio recorder.

All data collected during the study have been rendered anonymous in accordance with the terms of the ethical approval for the study. To facilitate subsequent analysis, the data was collated using version 2.8 of the HyperRESEARCH™ CAQDAS tool. This tool was selected on the basis that it supports effective integration and coding of multimedia data including audio recordings and video recordings as demonstrated during the previous ethnographic study of laboratory recordkeeping.

5.2.3.7 Data analysis

Concurrent and retrospective think-aloud reading protocols may be used to support both quantitative analyses and qualitative, thematic analyses. Consistent with the ethnographic and genre analysis studies, the reading protocols collected for this study have been analysed at an aggregate level of abstraction in order to support a qualitative characterization of how scientists made sense of the laboratory records for the specific reading tasks. The qualitative data collected for the study have therefore been analysed, coded, and categorized in line with the data analysis procedure for developing grounded theory (Charmaz 1983; Glaser and Strauss 1967; Strauss and Corbin 1998). This procedure involved an iterative process of open coding, axial coding, and selective coding in order to classify and interrelate the data obtained during the course of the study.

Open coding involved the systematic examination of the reading protocols collected during the study in order to identify and describe categories of record interpretation and variation between readers in relation to the laboratory records. Axial coding was concerned with

refining the understanding of laboratory record reading behaviour by identifying generalization/specialization relationships to capture semantics links between the categories and their associated subcategories. Finally, selective coding was concerned with integrating and refining the set of categories in order to build an understanding of laboratory record reading behaviour that was derived in an inductive manner from the data collected during the course of the study. As with the ethnographic and genre analysis studies, it is important to note that coding was not a static procedure, but instead proceeded in an iterative manner throughout the course of the study so that categories could be compared, modified, and refined in the light of new observations. This constant comparison approach proved particularly appropriate to the study in terms of facilitating a robust comparison of laboratory recordkeeping across different types of reader.

In order to preserve the rich information content inherent in audio and video data (Goodwin 2000; Heath and Luff 2000; Kress and van Leeuwen 2001), these types of data were not transcribed prior to analysis but were analysed and coded using the multimedia coding facilities in the HyperRESEARCH CAQDAS tool selected for the study.

5.2.3.8 Data validation

Laboratory manuals and a domain specialist were consulted to clarify the materials and methods used for each of the laboratory procedures documented in the sample laboratory records used in the study. This approach was followed in order to establish an independent expert view on the materials and methods necessary to undertake each of these procedures. This expert definition provided a basis against which to compare the responses of individual readers.

Participant validation again played a central role in validating the findings of this study. This not only involved presenting and discussing the findings from the study with individual participants but also, albeit to a much lesser extent, via group presentations in order to evince feedback on insiders' perspectives on the validity of the findings. As for the ethnographic and genre analysis studies, it was again important not to interpret the feedback gained from these sessions, whether supportive or hostile, uncritically as direct validation (or invalidation) of the findings from the study.

5.3 Results

The results of this reading protocol study of laboratory records are presented in the following subsections, arranged around the categories identified during the data analysis and "illustrated by characteristic examples of data" (Glaser and Strauss 1967:5). In this case, the characteristic examples of data consist of transcribed extracts taken from the reading protocols. References to individual participants and notebooks in the study results make use of the identifiers listed in Appendix 8. Excerpts from audio and video recordings used in the study results have been transcribed using the conventions identified in Appendix 10.

5.3.1 Reading records as a novel experience

The recruitment of readers to this study of laboratory records in other scientists' notebooks required noticeably less persuasive effort than the recruitment of participants for the previous ethnographic and genre analysis studies[129]. The results from these two previous studies highlighted the minimal use of the laboratory records written in other scientists' notebooks by most researchers and technicians. This included minimal use of the records produced by other scientists working on the same or related projects within the same research group. Given this background, it was important to note that a number of the participants in this reading study volunteered on the basis that it would be a novel experience and a potentially useful learning experience. Interestingly, those who volunteered on this basis included both junior and senior laboratory staff. These individuals viewed participation in the study as an opportunity to examine other styles of recordkeeping.

The most frequent remark made by readers during the course of the reading experiments was how difficult they found reading the sample records, even those records that documented types of experimental work that were familiar to them. In particular, it was interesting for readers to be confronted with the varying levels of detail used to describe

[129] This is not to suggest that recruiting participants to the ethnographic and genre analysis studies was particularly onerous, and the experimenter did not reach the last resort in the Theodore Roosevelt approach to persuasion of "speaking softly but carrying a large stick". It is simply an acknowledgement that laboratory staff perceived more direct benefit from participating in the reading study compared to the other two studies.

even basic laboratory techniques as illustrated by the comment:

> "it's interesting (.) to see (.) that for PCR people write so little information (.) and I think I'm guilty of that but I tend to write it out full first time and then refer back."
>
> [Postdoctoral Researcher reading record *2L8A1*]

5.3.2 Understanding experimental purpose

Understanding the purpose of an experiment is an essential step in understanding the context in which the experimental work has been undertaken. By defining a set of aims and objectives, the experimental purpose shapes the choice of materials and methods used to carry out the work, shapes the scope of the data to be collected during the course of the work, and establishes a basis for evaluating the results obtained from the work. Given the relevance of experimental purpose to understanding experimental work, it formed the first of the reading tasks to be completed during the reading experiments. The following subsections examine the approaches used by readers to interpret experimental purpose during the course of the reading experiments. These reading experiments involved interpretation of sample records exhibiting diverse styles of recording experimental purpose as indicated in Table 5-4.

5.3.2.1 Interpreting purpose from a direct statement of purpose

All readers were able to isolate and interpret blocks of text containing narrative statements of experimental purpose when these were included in the sample records as in records exhibiting the 'narrative' style of recording experimental purpose. This applied whether or not the text block containing the statement of purpose was directly identified as such through the use of a heading. However, interpreting the experimental purpose and other structural elements within a record in the absence of headings involved a degree of effort in order to isolate logically cohesive "clusters" within the text as illustrated by the comments:

> "this person has mentioned (.) I mean as I said a very good thing (.) the very start he has mentioned the purpose of the experiment which is the objective (.) <u>but</u> it would have been (.) I mean (.) the same piece of document if you just put a heading (.) objective or purpose (.) it would be so clear."

> "it becomes a bit difficult (.) eh from (.) if there were headings (.) because then I have to cut ok (.) cut the purpose here (.) from here I think it's the starting materials and methods (.) then cut it here so I have to do that (.) then if they're intermingled as they were in the first presentation ((the previous sample record)) (.) then (it's a bit careless) it takes time."
>
> [Postgraduate Researcher reading record *1L2A2*]

When a narrative statement of purpose was included in the record, readers examined it prior to moving on to other components of the record. In some cases, however, the scope of the statement of purpose was assessed as too broad to aid directly in the interpretation of the specific experimental work covered by the record. In this sense, the statement of purpose was interpreted more as statement of the long-term goal and not of the experiment *per se* as illustrated by the comment:

> "it's difficult to say (.) because em (.) they've got a broad (.) em mission statement (.) a mission statement even it's so broad (.) to characterize repeat interruptions (.) but em (.) they've not said how they're going to characterize them (.) y'know em (.) the immediate (.) goal of this would appear to be to extract the DNA for different (.) lengths of bands (.) with a PCR (.) a PCR experiment (.) but not sure (.) em what the different PCR length bands would be (.) and I'm not sure where the em (.) not exactly sure what they're going to do with them afterwards."
>
> [Principal Investigator reading record *1L2A2*]

5.3.2.2 Interpreting purpose from a title

The title given to a laboratory record contributed to each reader's interpretation of experimental purpose whether or not a separate statement of purpose was included as a block of narrative text. In particular, the title played a central role in the interpretation of experimental purpose when there was no narrative statement of purpose as in records exhibiting the 'title-focused' style of recording experimental purpose. In most cases, however, the title offered only a limited interpretation and functioned more as a guide for interpreting the remainder of the record as typified by the response:

> "I mean the first thing (.) which (.) I saw was the transfection thing ((the title of the record)) and (.) that was the first impression that I took from this it's something about transfection but it could be anything (.) when I continued reading (.) I mean that was my first impression but not my decision."
>
> [Postgraduate Researcher reading record *4L5A1*]

5.3.2.3 Reconstructing purpose from methods and results

In the absence of a clear statement of experimental purpose, all readers set out to interpret the experimental purpose in an inductive manner based on a combination of the materials and methods that could be identified within the record, and the results that could be identified within the record. Interestingly, the same approach was also followed even when a clear statement of experimental purpose was visible in the laboratory record. In this second case, however, the understanding of purpose that could be reconstructed from method and results was evaluated against the stated purpose as a means of verifying the work that had been undertaken.

Attempting to reconstruct the experimental purpose in this manner was, however, highly dependent on the background knowledge and laboratory experience of the reader as shown in the following two comments made by the same postgraduate researcher attempting to reconstruct the purpose for two laboratory records. The two sample records documented different types of experiment. In the record associated with the first comment, the reader was able to identify a subset of familiar experimental procedures centred on PCR, whilst this was not possible for the second record documenting a survival immunity assay typical of work with *Drosophila melanogaster*:

> "I could identify some reagents (.) dNTPs[130] (.) buffers (.) primers (.) so that's how I assumed they were doing an RT-PCR 'cause I'm familiar with them (.) and then eh I assumed there's something similar to a chip or Southern blotting cause they were using herring sperm which is something you use to blot things."
>
> "the purpose of the experiment (.) is <u>completely</u> unclear (.) it seems like (.) they're (.) stabbing the flies with E. coli and see (.) how their immune system responds that's what I can make out from the title (.) but I don't know (.) because of the timings it seems like they're stabbing the same flies (.) but I don't know if these are (.) different flies (.) if there are different numbers of stabbings (.) or different concentrations or (.) I can't make out what the numbers in the tables are."
>
> [Postgraduate Researcher reading records *1L4A2* and *3L5A1*]

It is also important to note the form of the statements of purpose reconstructed by readers through this process of induction from methods and results. In particular, it is important to

[130] Deoxynucleoside triphosphates (dNTPs) are nucleotides that form the unit building blocks from which DNA is synthesized. Ready-to-use products containing balanced mixes of the dNTPs used in common laboratory techniques such as PCR are available from commercial vendors.

note that these inductively generated statements are essentially statements of what was done rather than why it was done. This applied even in the case of experienced readers who had some knowledge of the techniques involved as typified by the comments:

> "so what I think that they have done in the end is (.) using (.) a kit (.) is they've made coding RNA (.) em (.) which (.) they've done first by doing a first strand synthesis then a second strand synthesis to make a cDNA hybrid into a coding DNA."
>
> [Principal Investigator reading record *2L4A2*]

> "I would say the purpose of that is to (.) label (.) an RNA sample from a (.) bacterium (.) with Cy5 and an Alex- Alexa dye for a microarray (.) but I suppose that is using (.) quite a lot of prior knowledge (.) that they've mentioned array there (.) and they've mentioned RNA (.) and a bacterium and they say which bacterium so (.) you could put it all together but I suppose it doesn't have a specific title."
>
> [Principal Investigator reading record *1L4A1*]

In this sense, the purpose statements were primarily procedure-orientated rather than goal-orientated. Accordingly, the readers were able to produce only incomplete statements of purpose that omitted the wider context and motivation for conducting the experiment. A number of readers clearly expressed an awareness of this limitation in what they could extract solely from the information shown on the notebook pages:

> "I know the purpose of (.) <u>that</u> experiment (.) but I didn't know (.) I kinda wanted to know why (.) that experiment was being done 'cause it was obviously being done for another experiment"
>
> [Postgraduate Researcher reading record *1L8A1*]

5.3.3 Understanding experimental materials and methods

Understanding the materials and methods used to perform an experiment is an essential step in evaluating the results obtained from the experiment, in assessing the quality of the work that has been undertaken, and in reproducing the experiment. Given the importance of experimental materials and methods in understanding experimental work, it formed the second of the reading tasks to be completed during the reading experiments. The following subsections examine the approaches used by the readers to interpret experimental materials and methods during the course of the reading experiments. These reading experiments included interpretation of sample records exhibiting diverse styles of recording experimental materials and methods as indicated in Table 5-4.

5.3.3.1 Interpreting method from a direct statement

All readers were able to isolate and navigate to those entries within the sample records that documented different aspects of experimental materials and methods. This applied across the range of recording styles used in the sample records. In particular, this applied whether or not the descriptions of materials and methods were specified in summarized forms or in greater levels of detail. However, as was the case with understanding experimental purpose, isolating those entries in the sample record that pertained to materials and methods in the absence of headings involved a degree of effort. In particular, the lack of metatextual cues such as headings required readers to try to isolate logically cohesive "clusters" within the text. An alternative approach used by most readers involved searching for "procedural results" in the record, where these procedural results represented intermediate measurements or statements that could be associated with the completion of an experimental task as illustrated by the comment:

> "you get a feel for how they've done it (.) but also how they've sectioned their lab book (.) so you sort of get (.) sort- (.) that block of writing's all (.) sortof anala- (.) looked like a memo to them ((a memo to self)) (.) ok I'll read that (.) and that seems to be the end of a section (.) and then they seem to section (.) even though it all seems to be one stream (.) for me it would look like it was sectioned off into various sections (.) and then at the end they do (.) another test (.) what looked like another test (.) and they give (.) a short (.) four or five words (.) ok I've done all these tests it's fine (.) and then we can move on to the next bit."
>
> [Technician in Research Laboratory reading record *1L4A2*]

Some postgraduate researchers felt challenged by the need to reconstruct a definition of materials and methods from information that had been distributed throughout the record rather than placed in the foreground using a single, cohesive "cluster". In part, this difficulty derived from a mismatch between the laboratory records and the postgraduate researcher's familiarity with the 'gold-standard' genre structure for laboratory records in which materials and methods are presented in an independent section from results. In part, this difficulty arose from the postgraduate researcher's wider exposure to other laboratory genres purposed with the publication of results as illustrated by the comment:

> "when we read eh (.) many journals (.) articles it's already designed (.) so that was really my eh (.) I was expecting (.) that there will be a heading 'materials and methods (.) and I myself write it in this way (.) because my supervisor is very strict about materials and method things so (.) I was expecting that this (.) that I'll start reading it ((the record)) and it will be boom there (.) why is this

> man ((the experimenter)) asking me about materials and methods (.) it should be routine in a very precise manner (.) but it wasn't (.) which confused me (.) which panicked me"
>
> [Postgraduate Researcher reading record *4L5A1*]

Entries corresponding to the core techniques of molecular biology including reaction mixes, PCR thermal cycles, and electrophoresis gel loading provided the scaffolding used by the majority of readers to navigate through the sample records. Isolating each of these types of element within the written record relied on a combination of features. In particular, it involved recognizing the characteristic spatial layout of these entries, and recognizing the characteristic data content of the entries. For example, PCR reaction mixes were typically laid out in three columns, where the first column listed the reagents, the second column listed the single reaction volumes, and the third column listed the scaled-up multiple reaction volumes. In addition, PCR reaction mixes typically included forward and reverse primers in the list of reagents.

Perhaps unsurprisingly, prior experience with different types of experiment was a significant factor influencing the ability of a reader to interpret the materials and method documented in each laboratory record. Within laboratory settings as in many other settings, an individual's prior knowledge can be categorized at a gross level into experience of observing a specific activity or type of work, and experience of performing that activity or type of work. In fact, these categories reflected the typical 'see one, do one' stages used by experts to train novices in experimental procedures observed during course of the ethnographic study as discussed in section 3.3.8.2. It was interesting to observe that both these categories of prior knowledge were successfully employed to aid interpretation of experimental materials and methods. In particular, knowledge of the sequence of activities within different types of experiment facilitated both navigation through the records, and interpretation of the content in the records. In this sense, a procedural model of the specific type of experiment being documented in the record provided a frame for reading and navigating the laboratory record as illustrated by the statement:

> "realized it looked like (.) the beginnings (.) of a microarray experiment (.) so I was trying to remember what I know about microarray experiments (.) having done one so (.) well ok you need a first strand synthesis stage so (.) so this was (.) how I would sort of think (.) ok so I'm looking for a first strand synthesis ok I've got a first strand synthesis (.) that's fine so alright so (.) check (.) all the checks (.) ok they're checking (.) the RNA (.) and then I was looking at what they've actually done (.) they were highlighting problems that they've had (.)

> then ok we've got the second strand synthesis (.) and that's there (.) labelled as second strand synthesis"
>
> [Technician in Research Laboratory reading record *1L4A2*]

The sample set of records used for the reading experiments included various entries recording use of commercial off-the-shelf laboratory kits as part of the direct statement of materials and methods. These kits are an increasingly common feature of laboratory work in most settings. It was interesting to note that the differentiation identified during the genre analysis study between the writing styles employed to document kit use in service and research laboratories (*cf.* 'standard-conformant' recording and 'standard-referent' recording in section 4.3.3.2) was reflected to some degree in different approaches employed to read kit use. In this sense, most of the research staff preferred to adopt an abstraction that encapsulated the detailed sequence of steps involved in using the kit. This is illustrated by the following comment made during a reading experiment by a postdoctoral researcher in a research laboratory reading a record documenting kit use in the service laboratory *RS-L4*:

> "I actually skipped through that ((a 'standard-conformant' record of kit use)) because (.) I was surprised they'd written it all out (.) so (.) em (.) each of the QIAGEN kits is slightly different so if I was going to follow the same kit the GenElute™[131] kit that they used (.) so they have actually said what kit they used (.) what I would (.) if I was trying to follow that I would get the kit booklet."
>
> [Postdoctoral Researcher reading record *1L4A1*]

5.3.3.2 Breakdowns in interpreting method from direct statements

In the majority of reading protocols, the readers were unable to provide what they themselves considered to be a sufficient interpretation of the experimental materials and method used in the sample records. Note that sufficiency in this sense was not necessarily concerned with being able to repeat the work, but involved obtaining a combined understanding of both what was done and why it was done. Interestingly, breakdowns in interpreting the experimental methods occurred with records across the range of recording styles identified in Table 5-4, including styles using both summarized levels of specification and styles using detailed levels of specification. It is also essential to note that breakdowns in interpreting the experimental methods occurred to different degrees.

[131] GenElute™ kits are typically used to purify DNA or RNA samples. Product information on these kits is available from Sigma-Aldrich Inc at http://www.sigmaaldrich.com [accessed 01 March 2011]. All trademarks are acknowledged.

For example, the following comment illustrates a breakdown in interpretation that led to a postdoctoral researcher abandoning use of a sample record. This sample record was written over five pages, and documented multiple tasks that were executed on a single day in the 'chronological and referenced' style of recording. Some of these tasks were steps in the same experiment. Each entry for an individual task exhibited a high level of detail and cross-referencing. It is interesting to note that the reader in this case routinely performed the same type of experimental work that was documented in the sample record. Nevertheless, the reader concluded that:

> "ah I would probably just give up at this point (.) and just do it over (.) myself (.) the way I want to do it ((laughing)) (.) do I have to go through the back ((reader is asking if it is necessary to complete reading the record)) (.) hooray (.) well I mean it looks good this gel looks good (.) I kinda want to know what happens but I just (.) I feel like I want to skip this because it's too difficult to read."
>
> [Postdoctoral Researcher reading record *4L8A2*]

A similar type of breakdown in interpretation occurred in a reading experiment in which a postgraduate researcher was unable to interpret the purpose or method of a record recording work with microarrays in a service laboratory. In this case, the reader was unfamiliar with the type of experimental work documented in the sample record, which was written in the 'standard-conformant' recording style and so again provided a detailed statement of experimental method. It is interesting to note that more experienced readers of the same record, who were also unfamiliar with the technique being described, were readily able to reconstruct the experimental method without understanding the purpose. The postgraduate researcher was unable to recognize sufficient entries within the "fragments" of the record to reconstruct the method as illustrated by:

> "well (.) it doesn't have enough details (.) some of it is eh (.) like eh (.) in fragments described sequentially (.) and there's a big paragraph of something (.) and then there's a table that doesn't have any explanation on it (.) that's what I call messy (.) things didn't seem to be linked together.
>
> [Postgraduate Researcher reading record *1L4A2*]

These two previous examples represent more extreme forms of breakdown where the reader felt unable to achieve any meaningful level of interpretation. In the majority of reading experiments, readers with varying levels of experience were able to reconstruct the experimental materials and methods to some degree of detail. Importantly, in a number of reading experiments that involved techniques familiar to the reader, most readers were also

able to identify the specific gaps within their interpretation. This level of awareness would be important in directing subsequent steps that could be used by the reader to complete his/her understanding of the experimental materials and methods used in the documented work.

The most common problem identified by readers at all levels of experience in relation to interpreting the experimental materials and method documented in laboratory records stemmed from the use of summarized forms of recording. This included summarized recording of individual elements such as 'loading summary' recording of electrophoresis gels, and summarized recording of the experiment as a whole such as 'abstracted' recording of experimental work. In essence, summarized forms of recording gave rise to information gaps. In most cases, readers attempted to fill these information gaps by examining the content of other records identified through explicit cross-referencing schemes. In the absence of any explicit cross-references, readers simply examined adjacent records or sought implicit forms of referencing.

Common forms of implicit cross-referencing that triggered readers during the course of the reading experiments include repetition of the same title in multiple records, use of the same reagents in multiple records, use of the same samples in multiple records, and use of the same equipment in multiple records. For example, the following comment illustrates a situation where a reader attempted to identify the missing temperatures and hold times for a PCR thermal cycle in one record by examining statements describing the use of laboratory equipment in an adjacent record. Note that the PCR thermal cycle in the first record was specified in the 'annealing only' style of recording, and that the reference to the PCR programme in the following comment is taken from the content of the second record.

> "with the PCR programs (.) it just says 55 degrees (.) you need more (.) details probably (.) though I guess (.) you could maybe work something out (.) and he does say sequencing 3 programme ((in the following record)) (.) I guess it could be a PCR machine you go just go to sequencing 3 (.) programme and (use it)."
>
> [Technician in Service Laboratory reading record *2L3A1*]

5.3.3.3 Interpreting method from experimental results

A mutual dependency exists between experimental method and experimental results in that the choice of method determines the type of result data that can be collected, whilst the need for specific types of result data may dictate the choice of experimental method. Given this association, it is perhaps unsurprising that readers also consistently made use of any statement of experimental results that was present in a laboratory record in order to support interpretation of the materials and methods documented in that record. This included the use of both textual and graphical statements of results.

In particular, readers examined the results at two levels. Firstly, the type of the result was used as evidence to support accepting or rejecting any provisional interpretations about the type of experimental work documented in the record. For example, researchers and technicians are aware that manual sequencing using PCR typically produces results in the form of an electrophoresis gel, and so the absence of such a gel in a laboratory record might cause a reader to reassess his/her initial assumption. Secondly, the specific results associated with individual samples established a cohesive link across each stage in the method statement documented in a record. In this sense, readers could trace the progression of any samples listed in the statement of purpose through the intermediate results associated with internal steps in the experimental method and on to the final statement of results generated for the overall procedure. Awareness of the "continuity" of interpretation offered by the samples in the laboratory records is captured by the comment:

> "normally I would run my samples on a gel picture (.) there's no gel picture so (.) y'know but it would be normal to run them on a gel (.) em (.) there's no (.) water control (.) y'know you don't really know what the samples are (.) if those are the samples (.) so there's not that sort of (.) continuity of being able to follow what samples where (.) on sequencing (.) so normally to do sequencing you would run (.) your samples on a gel."
>
> [Postdoctoral Researcher reading record *2L8A1*]

When using data to explicate experimental materials and methods, some readers including most of the experienced readers distinguished between the usefulness of raw data and processed data. In contrast to the preference for abstraction associated with some aspects of experimental recordkeeping, these readers demonstrated a preference for raw data. In contrast to processed data, the raw data offered a level of inscription that was more directly associated with the tasks used to generate the data. This preference is illustrated by the

following comment:

> "I certainly couldn't repeat it (.) em (.) if I were (.) eh (.) I would want the data here (.) because this is deeply processed data (.) which would have gone from (.) counts per minute off a counter (.) y'know there back-calculated into whatever this magic number on the left (.) and then (.) then thrust through a program or a micro (.) that calculates the mean and the standard error of the mean or 95% confidence limits
>
> [Principal Investigator reading record *1L5A1*]

5.3.3.4 Breakdowns in interpreting method from results

Given that readers made use of the result statements in a record as an aid to interpreting materials and methods, any difficulties encountered when interpreting these results impacted on the reader's ability to make sense of the method. Two specific sources of difficulty were evident in the reading protocols. These were difficulties arising from limitations in the reader's awareness of the experimental technique being used to generate the results, and difficulties arising from the style of recordkeeping used by the author to document the results.

Prior experience with different types of experimental technique for visualizing results was a significant factor influencing the ability of readers to evaluate experimental results. This is illustrated by the following example in which a highly experienced reader was unaware of the exact nature of the staining technique used to label images in the sample record.

> "DAB (.) wish I knew what DAB <u>was</u> (.) to really crack this (.) right so (.) he's got things (.) not sure if these things are working (.) and the retina (.) interesting bits of staining there (.) but I don't know if that's the eh melanin that you get in the retina to make it black (.) or the true staining so I think one of these is staining"
>
> [Principal Investigator reading record *1L3A1*]

Limitations in the recording of experimental results by the original author presented a more fundamental problem in that this type of limitation impacted upon all readers of a record, not just those with no prior experience of the visualization technique being used. A particular issue in this respect that became evident during the reading protocols was the use of summarized and/or partial labelling schemes. The "continuity" of interpretation offered by samples was highlighted earlier as an important cohesive link from experimental purpose to method to results. In order to benefit from this link, it is necessary to establish a clear mapping between samples in the handwritten sections of a laboratory record, and the

visualization of those samples in any graphical results such as electrophoresis gel images that have been pasted into the record. In some cases, the use of summarized labelling schemes and/or the lack of an explicit mapping between the text and the graphic gave rise to difficulties for readers as illustrated by the comments:

> "they wrote down four different rows (.) it was a bit confusing (.) I could then make out it was four gels but they were on a (.) different orientation (.) so it took a while for me to (.) to see (.) how the gels were orientated (.) but then I could kinda see the wells (.) so it must have been run that way (.) and then since they labelled rows 1 2 3 and 4 and there were four gels then I could make that out (.) however it would be much easier if they would put the gel (.) on the right position and then perhaps each labelling on the different gel (.) 'cause it could even be that he labelled them (.) on a different order of which I was looking (.) so that's not clear."
>
> [Postgraduate Researcher reading record *4L8A1*]

> "but they did something like A1 through B12 so I'm not really sure how many A samples there were (.) or how many B samples (.) it went up to B12 but maybe they don't have a B10 (.) or something (.) so there's no other labelling of each of the samples and that doesn't make it easy to understand."
>
> [Postgraduate Researcher reading record *4L8A1*]

It is useful to note these two comments reflect issues of both ambiguity and vagueness (*e.g.* Myers 1996) caused by the use of summarized labelling schemes. Ambiguous expressions carry more than one clear and distinct meaning, whereas vague expressions carry only multiple blurred meanings. Faced with an ambiguous entry in the laboratory record, a reader could choose to proceed with one of the possible interpretations; vague entries in the laboratory record presented a different category of problem.

Laboratory records exhibiting the 'collection-focused' style of recording provided a specific example of vagueness in experimental method. By focusing almost exclusively on tables for use in collecting data over time, these records presented no explicit statement of the materials and method used in the experiment. The sample record exhibiting this style of recording that was used in the reading experiments was an example of a record for a survival immunity assay using *Drosophila melangaster*. It was interesting to note that those readers familiar with work in laboratories that used *Drosphila melanogaster* as a model organism were able to interpret the materials and method fully from the 'collection-focused' sample record, whilst other readers were largely unable to interpret any of method.

5.3.3.5 Method as an indicator of quality

It was interesting to note that many readers arrived at subjective assessments during the reading experiments not only of the quality of recordkeeping exhibited by the sample records, but also of the quality of the experimental work performed by the author. This is of particular significance given the potential for changes in laboratory practice such as the adoption of ELN technology to 'publish' the laboratory records to a wider audience.

Some of the milder assessments put forward in the reading protocols are illustrated by the comments overleaf:

> "little bit naïve with the radiation (.) calling it radiation rather than γ ^{32}P ATP[132] and so on ."
>
> [Principal Investigator reading record *1L5A1*]

> "em I think the method is quite good 'cause the've used (.) mastermix so that means that the're cutting down on pipetting error by putting it all together (.) em and it also optimizes each of them"
>
> [Postdoctoral Researcher reading record *1L8A1*]

5.3.4 Supporting reproducibility

The 'gold-standard' definition for a laboratory record is that it should enable another scientist to be able to reproduce the experiment documented in the record. For this reason, the third of the reading tasks to be completed during the reading experiments focused on identifying what would be needed by readers to enable them to repeat the experimental work documented in the sample records. The two previous reading tasks were concerned with understanding the purpose of an experiment, and understanding the materials and method used to perform the experiment. Both of these issues significantly influence the ability of a scientist to repeat an experiment, and in this sense, the findings of this reading study with regard to experimental purpose and method also pertain to this third reading task. The following subsections examine some additional insights gained from the reading experiments.

[132] ^{32}P is a radioactive isotope of phosporus that is used as a label in biochemistry experiments. In this case, the radioactive label is attached to molecules of adenosine triphosphate (ATP) for use in a protein kinase assay.

5.3.4.1 Separation of experimental purpose and method

Readers were able to provide an interpretation of both the experimental purpose and the experimental method for the majority of the sample laboratory records, albeit to varying degrees of detail and completeness. It is interesting to note how readers perceived the interaction between these two aspects of an experiment, particularly in the context of being able to reproduce a body of experimental work.

All readers preferred to understand both the 'why' and the 'how' of experimental work as characterized by the purpose and the method. This applied not only at the scale of the experiment as a whole, but was also recognized as a preferred manner in which to specify the internal steps in the description of experimental methods. This is illustrated in the following example:

> "yeah thes- (.) this is much more detailed I think this (.) this would be much (.) eh you'd be able to follow this (.) it actually all makes sense (.) is sort of step-by-step (.) how you would (.) what you would think and what you would do (.) so the fact they said they'd diluted the RNA by 1 in 2 then actually said (.) how they did it (.) so you can practically (.) follow that."
>
> [Postdoctoral Researcher reading record *1L4A1*]

Understanding purpose did not, however, form a prerequisite for understanding materials and method. In this sense, purpose and method could be considered separable units of interpretation. This was particularly evident in the reading experiments in which readers of varying levels of experience from the research laboratories interpreted records written by technicians in the service laboratory *RS-L4*. The type of experiment documented in these records, *viz*. microarray work, was not the type of work carried out by the research staff although a subset of readers had delivered samples to the service laboratory for processing. However, the records kept by technicians in the service laboratory were written in a detailed manner in the 'protocol-focused' and 'standard-conformant' styles of recording experimental work. In this situation, a number of the readers were able to understand what was being done, but not why it was being done as illustrated by the comment overleaf:

> "there's no (.) there's no information as to what the bigger picture is (.) with this experiment (.) 'cause there's not enough information to know what going except that I've been given four blank RNA samples (.) which is also why I said it I can use the protocol as (.) a general protocol if I want to do a similar thing (.) I can use that (.) probably I can work it out but I have no idea as to the interpretation of what went on"
>
> [Principal Investigator reading record *2L4A2*]

5.3.4.2 Reliability of records

The reading experiments highlighted that all of the sample records were considered by the readers to provide incomplete descriptions of the experimental work that they embodied. These deficiencies translated in the majority of cases to an inability to reproduce the experimental work being documented. In particular, deficiencies were identified by readers in terms of the specific conditions required for a range of experimental tasks including many core laboratory techniques such as PCR as illustrated by the following comment:

> "in terms of the PCR (.) em (.) the Reddymix[133] is something that's bought from a company (.) and (.) it doesn't actually say which company (.) and there's quite a few companies that sell Reddymix (.) so if you wanted to repeat this exactly I think it would be difficult (.) em but they've written out how you would (.) set up (.) em it doesn't (.) as far as I could see (.) actually I would find it very difficult to repeat this PCR because I don't know where that Reddymix comes from (.) it doesn't say what the concentration of (.) a it doesn't say these are primers (.) b it doesn't say what concentration they're at (.) so (.) that would involve some guesswork depending on how experienced you are you may get it right you might not (.) it does tell you the volume they've put in but without knowing the concentration that's meaningless (.) it also doesn't tell you how to (.) what thermal cycling conditions (.) so I wouldn't be able to do this PCR again."
>
> [Postdoctoral Researcher reading record *2L8A1*]

The information required in order to reproduce an experiment varied between readers based on their familiarity with the type of experiment being documented. In this context, it was interesting to note that many sample records documenting techniques that were familiar and even routine to the reader were still considered deficient due to the level of detail specified in the records. This is illustrated by the comment overleaf:

[133] A Reddymix is an off-the-shelf buffer solution specifically designed for use in PCR experiments. This buffer solution contains all the reagents necessary for PCR except for the DNA sample and primers required for a specific experiment.

> "it also comes in the materials and methods thing that (.) which eh (.) technique was used or what kit was used for the gel extraction thing because I totally don't know which kit was that (.) and also what I have seen (.) I mean this is the experiment what I have been doing many times (.) and what I have noticed is each kit really gives you a different amount of DNA (.) which is totally different (.) so (.) it's important."
>
> [Postgraduate Researcher reading record *1L2A2*]

In only a limited number of cases, the type of experiment documented in the sample records was so unfamiliar to the reader that the information required by the reader to make sense of the record exceeded the scope of any laboratory record. This is illustrated by the following comment made by a researcher in the genetics of human disease when confronted by a record produced in a laboratory working in integrative physiology:

> "what would I need to repeat this experiment (.) I would need to study integrative physiology ((laughing))"
>
> [Postgraduate Researcher reading record *3L5A1*]

The majority of readers were not, however, surprised that they would not be able to use the sample laboratory records as a basis for reproducing experimental work. This applied even for those sample records that documented types of experimental work that was familiar to readers. Perhaps more interesting in this context is the set of issues that contributed to this expectation. In addition to issues surrounding economy of effort, some readers again raised the issue of willingness to share on the basis that "some people are hiding something yeah they don't want other people to understand."

5.4 Discussion

The following discussion evaluates the results of this reading protocol study of laboratory records with respect to the three questions posed for the investigation. These results are discussed with reference to the findings of the previous ethnographic and genre analysis studies of laboratory recordkeeping as presented in section 3.4 and section 4.4 respectively.

It is, of course, essential to bear in mind that the reading experiments conducted for this study were artificial in the sense that they provided the readers with only extracted records and limited background on the setting in which the work documented in the records had been conducted. In authentic use, additional contextual information could be available from other workers in the original laboratory or perhaps even from the original author.

1. *What roles do laboratory records play in the discourse of academic molecular biology laboratories?*

The ethnographic study indicated that scientists did not routinely read laboratory records other than those in their own notebooks. In this sense, the findings from this reading protocol study offer insights into the potential role of laboratory records within laboratory settings rather than the actual role.

The central finding from the study was the fact that readers at all levels of experience consistently found the sample laboratory records to be incomplete. In the majority of cases, the records provided an insufficient basis for reproducing experiments. In a number of cases, readers were unable to identify the purpose and/or method used in the experiment. This applied across the range of recording styles identified in the corpus of notebooks used for the genre analysis study, and applied to records documenting both familiar and unfamiliar laboratory procedures. In records management terms, the readers deemed the majority of records to be unreliable. This again questions the utility of laboratory notebooks in their putative role as community archive, and emphasizes the need for laboratory records to be transformed into other public genres as a means of communication both internally and externally to the laboratory of origin. Of course, both these statements are conditioned on the continued use of language and recording styles similar to those identified by the genre analysis study. In particular, less reliance on implicit associations and abstracted forms of specification has the potential to realize more reliable records.

It was also interesting to note the potential for laboratory records to acquire new informational value as an indicator of quality that could be used by readers to assess the work of colleagues through both their approach to recordkeeping and the day-to-day execution of their experimental work. Given that researchers and technicians did not currently expect their records to be read by others, the concept of laboratory records as an indicator of quality could prove influential in motivating a change in both the recording style used by individual scientists or even in the recordkeeping practices recommended by principal investigators. Depending on the extent to which laboratory records are shared with other scientists, value judgements regarding scientists' skill and performance in laboratory recordkeeping could contribute to the professional identity and academic capital of both individual researchers and research groups.

The issue of the willingness of research scientists to share data, both method and results, arose again during the reading experiments, specifically in the semi-structured interviews within the retrospective protocols. In this sense, the reading protocol study corroborated the findings of the other two studies. In particular, the majority of readers indicated that they were not surprised that they would not be able to use the sample laboratory records as a basis for reproducing experimental work, even for types of experiment that were familiar to them. Individual readers also expressed more fundamental concerns that some authors, not specifically those involved in this study, deliberately omitted details from laboratory records as a form of protectionism whereby others would not be able to repeat their work. Different individuals expressed this same contention during the course of the ethnographic study in relation to the statements of experimental methods found in published research articles.

2. *What are the structures, content, and representations that characterize the genre of laboratory records in academic molecular biology laboratories, and to what extent do these vary across different contexts of use?*

The approach used by readers to interpret laboratory records drew on two categories of structuring knowledge in the shape of an IMRD-based genre definition, and domain-specific knowledge of the schemata appropriate for specific types of experimental work, *cf.* Beaufort's (1998) concept of genre knowledge and subject matter knowledge as contributing factors to the shaping of expert texts. This structuring knowledge is instantiated, for example, in the conventional structure of records of PCR-based genotyping experiments, which are typically ordered as an introduction including a statement of samples, a method description including both a PCR reaction mix and a gel loading scheme, a results section including an electrophoresis gel, and a discussion including an interpretation of the gel. Other types of experiment also exhibited consistency in structure expressed as a sequence of core laboratory procedures, each of which assumed specific content and representations. Awareness of specific laboratory techniques provided an important reading frame in which to interpret the records. Readers relied on this domain-specific knowledge as scaffolding to aid navigation through laboratory records. In addition, readers relied on this domain-specific knowledge to identify potential deficiencies in the content of the records for each of the core laboratory procedures.

Few of the laboratory records present in either the notebook corpus used for the genre analysis study or in the sample set used for this reading protocol study followed the IMRD

structure directly. In particular, the readers reported that the experimental materials and methods were typically distributed throughout the record interspersed with intermediate results. In this sense, the genre structure for laboratory records conformed more closely to a structure in the form I-M_1R_1-M_2R_2-M_nR_n-D, where M_1, M_2, … M_n represent the steps in the method and R_1, R_2, … R_n represent the associated intermediate results. This variation on the IMRD structure is motivated in part by the dual prospective/retrospective purposes served by laboratory records, and represents a difference between the genre structure of laboratory records and that more typical of 'reporting' genres such as research articles (*e.g.* Kanoksilapatham 2005; Swales 1990) and laboratory reports (*e.g.* Dudley-Evans 1985; Lobban and Schefter 1992) in which materials and methods are typically separated from results. Whilst experimental research articles and laboratory reports are purposed with reporting work that has already been completed, laboratory records serve an important prospective role in managing the performance of laboratory procedures at the bench in addition to their retrospective role as records of work done.

3. *How do readers of laboratory records in academic molecular biology laboratories make sense of laboratory records in different contexts of use?*

Readers broadly conformed to the sequence of cognitive moves inherent in the IMRD genre to make sense of a record in terms of why the experiment was being performed, how it was being performed, what data was collected during the experiment, and how are these data were being interpreted. Within each of these steps, the readers moved back and forth through the record in order to answer each of the four questions addressed by the IMRD structure[134]. In the absence of a clear statement of experimental purpose, for example, readers set out to interpret the experimental purpose in an inductive manner by examining both the materials and methods that could be identified within the record, and the results that could be identified within the record. In the absence of a clear statement of materials and methods, for example, readers set out to interpret the form and content of the results to determine the type of method that could be used to generate these results. In this sense, the focus of attention was not restricted to specific sections within the record, and the record was consistently read as a whole.

[134] Question-driven information seeking as a model of reading has been proposed, *inter alia*, by Ram (1999:257) on the basis that "understanding is a process of relating what one reads to the questions one already has" and that "the purpose of reading is to find answers to these questions and thus to arrive at a more complete understanding of the issues one is interested in." This question-driven process is iterative in the sense that reading to understand is constructed as a process of "questions + story-in, answered questions + new questions-out".

Broadly consistent spatial layouts were used to represent records of core laboratory procedures such as reaction mixes, PCR thermal cycles, and gel loading schemes. Given this consistency of representation, readers experienced little or no difficulty isolating and interpreting entries corresponding to these core procedures. A typical example of this consistency of representation is given by the three-column table that was routinely used in laboratory records to document reaction mixes by specifying respectively the list of reagents, the unit reaction volumes, and the scaled up multiple reaction volumes.

The most common problem identified by readers at all levels of experience when reconstructing the meaning of laboratory records concerned the use of summarized or abstracted forms of recording. This applied equally across different sections of the record including experimental purpose, materials and methods, and results. In essence, the use of summarized forms of recording gave rise to information gaps, which readers attempted to fill by examining the content of other records. Locating other relevant records to supply additional context proceeded through explicit cross-referencing schemes, adjacency in the notebooks, and implicit forms of referencing. Examples of these implicit forms of referencing included identifying cohesive ties based on the repetition of sample names, the repetition of reagent names, and the identification of experiments with similar purpose statements. Of course, prior knowledge of experimental procedures also played a significant role in filling these information gaps.

Individual labelling of the samples to be processed during experimental work proved important in supporting the interpretation of laboratory records by establishing a cohesive link through all genre sections from introduction to method to results and discussion. In addition, this link could be used to trace samples through each of the separate laboratory procedures and intermediate results in the I-M_1R_1-M_2R_2-M_nR_n-D variant of this structure found in most records. In this sense, individually identified samples, rather than summarized groups of samples, proved a useful tool for readers that enabled ‘continuity of interpretation’ throughout the records. Other important cues that simplified the task of interpreting the laboratory records included the use of metatext such as headings to isolate sections within the record that corresponded to specific laboratory procedures or to specific genre sections.

Readers were concerned, at least for the purposes of this artificial experiment, with the detail that was used to document the experimental work. For example, each reagent, volume, and concentration in a reaction mix was verbalized as part of the process of

reading that reaction mix. It was interesting to note, however, that abstraction also played a role in reading laboratory records in addition to the role it played in writing laboratory records. Perhaps the most striking example of this concerned the reading of protocol statements associated with commercial off-the-shelf laboratory kits. Although the individual steps in using these kits were specified in detail in some recording styles such as the 'standard-conformant' recording style, it was interesting to observe that some readers in the research laboratories abstracted the reading process associated with the kit protocol. Rather than reading the individual steps, the readers interpreted the section of the record containing these steps as a single block to be read as the 'execution of a kit', demonstrating a mismatch between the relevance attached by individual readers to the information to be communicated in laboratory records (*e.g.* Sperber and Wilson 1995).

6 Implications for Electronic Recordkeeping

This chapter of the thesis addresses the secondary goal of the investigation by reflecting on how the findings from the ethnographic, genre analysis, and reading protocol studies of laboratory recordkeeping may inform the development of computing tools such as ELNs for use in academic molecular biology laboratories. In particular, the chapter addresses policy issues concerning data sharing, opportunities for promoting technology use based on the greater willingness of scientists to share methods data in contrast to results data, and the interplay between standardization and personalization in recordkeeping technology.

6.1 Data sharing policy

One of the principal benefits associated with the deployment of computing tools such as ELNs in laboratory settings is the increased ability to share records (*e.g.* Lysakowski 1997) with the resultant potential for increased and novel forms of collaboration (*e.g.* Chin *et al.* 2002; Chin and Lansing 2004; Farooq *et al.* 2005; Tabard *et al.* 2008). In this context, a significant finding from this investigation into laboratory recordkeeping in academic molecular biology laboratories is the minimal use of laboratory records in their direct written form by laboratory staff other than the original author of the record, and the almost exclusive reliance on other communicative genres such as laboratory talks, expert demonstrations, and research articles as the means of communicating selected information held in the records in laboratory notebooks. It is important to note that this finding was consistent across each of the three studies that were undertaken as part of this investigation, and was consistent across each of the laboratories that participated in the studies.

The use of other genres such as research articles in specific contexts to communicate data held at source in laboratory records has also been reported in recent studies of scientific recordkeeping in academic laboratories such as those by Shankar (2007) and Wickman (2010). For example, Wickman (*ibid.*:264) positions laboratory notebooks as occupying "a negotiated space between the scientist's contingent response to exigency in the laboratory and the genre-specific strategies that he deploys to communicate his work outside the laboratory". Shankar (*ibid.*:1462) emphasizes the role of the research article by observing that "many scientists consider the formal publication as the 'true' record of science" on the basis that it "is as fully embedded in webs of meaning and community (in style, content,

reference to other work, purpose) as any document can be." Notwithstanding the role of other genres, it is important to note that both these recent studies in common with other studies of scientific recordkeeping consistently acknowledge the construction of laboratory records as a shared, community resource in their own right that "must be accessible to the larger community of practice" (Shankar *ibid*.:1463).

In the context of ELNs, the important point to address is why academic scientists did not make use of other scientists' notebooks. Based on the evidence of this investigation, it was not the case that research staff simply did not require to read other scientists' laboratory records since they did make considerable use of other spoken and written genres to exchange the information held at source in laboratory notebooks. Further, research staff encountered situations where they were required to repeat or extend experimental work for which the only documentary description was held in laboratory records. It was also not the case that scientists did not have ready access to other scientists' notebooks since the laboratory benches, communal office areas, and principal investigators' offices in each laboratory setting contained shelves full of notebooks within easy reach of all laboratory members. More fundamentally, it was the case that research staff did not write their own laboratory records to be shared by others and did not expect to be able to make sense of other scientists' records. In essence, the scientists constructed their laboratory notebooks as a 'personal resource' rather than a 'community archive'. Further evidence for this view is given both by the lack of indexing, metadata capture, or other curatorial effort within the laboratories aimed at establishing laboratory notebooks as an effective archive, and by the attitude expressed by most research staff including principal investigators that it would breach laboratory etiquette to read the notebook of another scientist who is still based in the laboratory.

The findings of this investigation are in line with the findings of recent surveys investigating attitudes towards data sharing across multiple academic settings (*e.g*. Borgman *et al.* 2007; David 2006; Research Information Network 2008, 2010). These studies have highlighted limited sharing of a range of scientific data, not specifically including laboratory records, in spite of the availability of technology to enable sharing. Instead, the concerns cited by academic scientists relate to policy-related issues including questions of ownership, ethics and publication rights, and scientific issues concerning the definition of meaningful data sets to exchange. Whilst a number of research funding agencies have policies in place to recommend data sharing (BBSRC 2007; MRC 2008),

these are advisory and do not yet require data to be shared across the wider scientific community. In this sense, the findings from this study offer a counterpoint to the findings from other studies, notably a number of studies undertaken as part of technology developments for electronic recordkeeping that report sharing of laboratory records or assume laboratory records to be a shared resource (*e.g.* Arnstein *et al.* 2002; schraefel *et al.* 2004; Tabard *et al.* 2008).

Effective technology design such as the design of information and recordkeeping systems must take account of the work practices that exist within the settings in which the technology is to be deployed (Anderson *et al.* 2008; Bannon and Bødker 1997; Robek *et al.* 1995; Suchman 1987). As Hartswood *et al.* (2008:60), *inter alia*, point out "the 'design problem' is not so much concerned with the creation of new technical artifacts as it is with their effective configuration and integration with existing work practices and the subsequent need for them to co-evolve." In this respect, designers of ELNs should take account of the work practices that are in place within academic bioscience settings with regard to laboratory recordkeeping. In particular, the findings of this investigation suggest that the *de facto* data sharing policy in place within academic molecular biology laboratories runs counter to the deployment of ELNs as a 'community archive', and that this mismatch between scientists' attitude to data sharing and the design goals of ELNs has contributed to the limited uptake of this technology (*cf.* Orlikowski 1993).

6.2 Sharing records of experimental methods

In contrast to the controlled attitude shown by scientists towards sharing experimental data such as the experimental designs and results kept in laboratory records, most laboratory members readily participated in the exchange of experimental methods. As highlighted by the ethnographic study, the preferred approach to exchanging experimental methods between laboratory members was to use visual demonstration. This finding was consistent across each of the laboratories that participated in the study. In practice, this included both participating as an expert to train others in the use of specific laboratory methods, and participating as a trainee to learn new methods. Interestingly, this exchange of experimental methods regularly crossed laboratory boundaries, so that experts in a given laboratory method might be called upon to train staff from other research groups, other divisions, and even other universities.

The findings of this investigation highlighted specific aspects of recordkeeping for experimental methods that would benefit from technology support. The first area of interest concerns the maintenance and visualization of protocol variation histories. The second area of interest concerns the maintenance of visual demonstrations of laboratory procedures.

The joint findings from the ethnographic and genre analysis studies identified that researchers and technicians devoted a significant amount of their time to optimizing and refining the steps defined for existing laboratory methods in order to adapt them for their own specific problems and specific model organisms. This could involve simple changes to individual temperatures or volumes within an existing protocol, or it could involve more significant changes such as the removal or addition of entire steps in the protocol. The need to optimize and refine laboratory methods applied even within the service-orientated laboratory where externally defined protocols were adapted to situated needs. This refinement activity gave rise to a body of knowledge that was captured in the 'personal resource' of the laboratory notebook but could prove useful to other scientists should it be made accessible to them.

Detailed inspection of multiple laboratory records in the corpus used for the genre analysis study identified different categories of information that could be usefully communicated to other scientists. The first category consisted of records of local exigencies such as protocol steps that could be skipped, or temperatures that could be allowed to vary within certain limits without affecting the quality of the experimental results. Interestingly, observation of a training activity within one of the laboratory settings during which a postdoctoral researcher instructed a postgraduate researcher indicated that this category of information was communicated verbally during visual demonstrations. The second category of information that could prove particularly useful to communicate concerns the optimization history of a specific protocol. Given the time devoted to exploring different conditions whilst refining protocols for specific problems, it could prove beneficial to maintain and display a solution/decision tree (*cf*. a genealogy chart) detailing the experimental paths that have been tried along with the results obtained. The potential benefits of such a resource derive from the fact that multiple researchers may be required to undertake the optimization of the same basic protocol, a process that currently involves 'reinventing the

wheel'[135]. The information required for such a resource is currently recorded within laboratory notebooks, but is not necessarily shared with other researchers. In this sense, the focus on a tool to share experimental methods rather than other data might prove a useful technology probe (Hutchison *et al.* 2003) in assessing scientists' attitude towards sharing.

Another form of social media tool that could usefully be integrated into laboratory recordkeeping practices concerns the use of visual records for sharing experimental methods. The findings of this investigation confirmed the genre of visual demonstration as the preferred approach within laboratory settings to communicate and exchange knowledge on laboratory skills and procedures. Indeed, this genre has recently been adopted as the basis for a new form of research journal in terms of a visual journal, JoVE[136], that uses digital video recordings to exchange knowledge about experimental methods within the wider bioscience community. It is interesting to note that visual demonstration by local experts as a basis for communicating knowledge and expertise within localized laboratory communities suffers from the problem that research staff move away from laboratory settings, taking their expert knowledge with them. In this sense, development of a ELN that separates the recording of experimental methods from experimental data, and employs the use of visual demonstration allied to written description of laboratory protocols would establish a more stable basis for preserving and propagating expert knowledge within the laboratory over time.

6.3 Standardizing experimental records

Formal standardization efforts are ongoing within both the molecular biology community and the wider bioscience community in order to facilitate collaboration and data exchange between scientists. These developments have been accepted within the academic research community to a limited degree. This includes terminological standardization to establish a shared vocabulary across bioscience disciplines (*e.g.,* Gene Ontology Consortium 2000), information standards to prescribe formats for data exchange (*e.g.,* Brazma *et al.* 2001),

[135] Solution/decision trees describing the experimental variations explored by individual bioscientists would deliver a useful source of knowledge. Dr Joseph Gray (2011, personal communication), a senior lecturer in biomolecular science at the University of Glasgow, highlighted to me the potentially greater benefits that could accrue from being able to construct and share integrated solution/decision trees that reasoned over and combined the work of multiple bioscientists from different laboratory communities.

[136] JoVE articles are available at http://www.jove.com [accessed 01 March 2011].

and procedural standardization to formalize workflow (*e.g.,* Conery *et al*. 2005). Laboratory recordkeeping has not yet become the focus of formal standardization. In this context, it is perhaps unsurprising that many ELNs rely on the capture of free-form textual descriptions of experimental work, supplemented in some cases with a limited set of metadata to facilitate the subsequent indexing and retrieval of records. Other ELNs, particularly those targeted at industrial laboratories or other settings that must demonstrate conformity to quality assurance procedures, impose a greater degree of rigour through the use of fixed workflow procedures.

The joint findings from the genre analysis and reading protocol studies highlighted that scientists will encounter significant variation in recording style should they eventually be required to read other records, and that the routine use of summarized forms of recording in these records will limit the ability of scientists at all levels of experience to reconstruct the meaning of these laboratory records. These findings identify a requirement for some level of information and procedural standardization within laboratory recordkeeping if the policy intention is to promote sharing of reliable records throughout the laboratory community (*e.g.* Brun *et al.* 2003). Continued use of free-form textual descriptions of experimental work such as those captured by many current ELNs will mean that these tools perform only as "manuscript repositories" (Robek *et al.* 1995:514-515).

Balancing the needs of the individual and the wider community is a concern for all recordkeeping systems, and the issue of localized forms of recordkeeping adversely affecting the intelligibility of records applies both to paper-based records and electronic records. Bannon and Bødker (1997) emphasize that coordinating interpretations is an essential part of the articulation work necessary to construct a Common Information Space (CIS) that encompasses the range of information artefacts accessible to a group. In the context of ELNs, the important point to address is what mechanisms are available to enable the interpretation of laboratory records to be coordinated between author and potential readers. Based on the evidence of this investigation, both reading and writing of laboratory records drew on two categories of textual coordination mechanism in the shape of the IMRD-based genre definition, and domain-specific knowledge of the schemata appropriate for specific types of experimental work. The list of domain-specific schemata (*cf*. the 'template' style of recording experimental work) would require to be configured according to the type of work undertaken by each specific laboratory. For example, the range of schemata appropriate to a laboratory engaged in work on integrative physiology with

Drosophila melanogaster would overlap to a degree and also differ to a degree from that needed for a laboratory working with the genetics of human disease.

Having the opportunity to adapt the behaviour of interactive systems such as ELNs to fit local needs would be critical to deploying them in a range of environments, and would also contribute to the sense of control enjoyed by users. Trigg and Bødker (1994) discuss how off-the-shelf software systems could be tailored by the users in a health and safety inspectorate in order to achieve bounded variety within the organization by defining and sharing customization files in a collaborative manner. Mackay (1990) highlights the social nature of customizing software packages by identifying specific patterns of sharing customizations within an organization, and by identifying the central role of translators, that is, individuals who acted as a central point in defining and distributing customization files throughout the organization. Henderson and Kyng (1991) distinguish between different degrees of customization that can be used to appropriate technology to local needs. The range includes choosing between predefined alternatives, constructing new artefacts from existing pieces, and reprogramming the artefact. Each of these forms of customization could be applied within ELNs in order to manage the textual coordination mechanisms offered by laboratory record schemata.

Enabling a degree of personalization in the choice and content of schemata presented to each individual scientist could enhance the sense of control enjoyed by the laboratory staff using the tools (*e.g.* Greenfield 2006:159-174; Liaskos *et al.* 2005). For example, Shankar (2007:1465) observed during her recent study of recordkeeping in an academic neuroscience laboratory that "standardization of data entry is not the only aim of scientists when creating records – again, there is a poetics to the act of records creation that encompasses and perhaps even exemplifies rich and powerful interactions of the social, personal, and professional". Bailey (2009), *inter alia*, advocates the use of recommender systems to guide users through records management activities, and a recommender system could be usefully applied to ELN systems in order to guide laboratory scientists in personalizing the schemata appropriate to specific laboratory procedures whilst warning of the impact of proposed customizations on prospective readers of the personalized records.

7 Conclusion

This final chapter of the thesis presents a summary of the investigation that was conducted into laboratory recordkeeping in academic molecular biology laboratories. The summary restates the objectives for the investigation, summarizes the key findings obtained from the investigation, identifies limitations on these findings, and outlines potential avenues for future work.

7.1 Academic bioscience laboratory recordkeeping

The work presented in this thesis focused specifically on molecular biology laboratories in the academic sector, and set out with the goal of investigating recordkeeping as a discursive practice in these laboratories by examining the authentic records and practices of a range of scientists in multiple laboratories within a UK university[137].

The motivation for this investigation stemmed from three areas of concern. Firstly molecular biology is a "literate culture" (Smith 1993), in which molecular biologists employ a range of communicative genres. Previous work has examined a number of the genres, both written and spoken, that co-constitute the discourse of academic bioscience laboratories. This includes studies of research articles (Bazerman 1988; Kanoksilapatham 2005; Myers 1990, 1991; Swales 2004; Tarone *et al*. 1998), conferences talks (Dubois 1987; Rowley-Jolivet 2002; Ventola *et al*. 2002), laboratory reports (Braine 1995; Dudley-Evans 1985), peer reviews (Gosden 2003), PhD theses (Bunton 2002; Dong 1998; Dudley-Evans 1991; Hyland 2004), PhD vivas (Grimshaw 1989; Maingueneau 2002), tenure track reports (Hyon 2008), research proposals (Cadman, 2002; Myers 1990), researcher websites (Cronin 2001), and textbooks (Hyland 2000; Love 2002). The laboratory record, however, has received comparatively little attention in spite of its central role in the daily work of laboratory scientists such as molecular biologists. Consequently, part of the motivation for this investigation was to add to the growing body of work on academic research genres by focusing on the laboratory record in molecular biology laboratories.

[137] The university in question is a long-established UK university with a strong teaching and research tradition. In particular, the university is a member of the Russell Group of universities that "represents 20 leading UK universities which are committed to maintaining the very best research, an outstanding teaching and learning experience and unrivalled links with business and the public sector". This quote is taken from the Russell Group website at http://www.russellgroup.ac.uk [accessed 01 March 2011].

Secondly, laboratory recordkeeping across the industrial, health service, and academic sectors has become a focus for technologically driven change aimed at enabling a transition from paper to electronic forms of recordkeeping. This technology, termed electronic laboratory notebooks (ELNs), offers the potential for improved data sharing within laboratory communities (Lysakowski 1997). However, the uptake of ELNs has been markedly low in academic settings (Nature 2005, 2007; Taylor 2006), where paper-based approaches to recordkeeping remain the preferred solution. Previous work on technology-focused projects has examined multiple issues in the design of ELNs to support laboratory work. This includes work on augmented reality systems (Borriello 2006; Mackay *et al.* 2002), on distributed and mobile systems (Arnstein *et al.* 2002; Yeh *et al.* 2006), on integrated workflow systems (Arnstein *et al.* 2002; Frey *et al.* 2004), and on collaborative working systems (Chin *et al.* 2002; Chin and Lansing 2004; Farooq *et al.* 2005; Tabard *et al.* 2008). The focus for this investigation is driven instead by Mirel's (1993) advocacy of the need to understand how users interact with documents in actual work situations as a *sine qua non* for the effective design of document-mediated interaction. Consequently, part of the motivation for this investigation was to add to the growing body of work on the potential for electronic recordkeeping in academic bioscience laboratories by focusing on the language used in constructing laboratory records.

Thirdly, reliable recordkeeping requires balancing the needs of the individual author against those of the wider community. This applies irrespective of whether paper-based or electronic approaches to recordkeeping are in use. In particular, diversity in recordkeeping practices is known to be problematic in the sense that adapting a recordkeeping system to the needs of one user or one group has been shown to render those records less intelligible to other users and groups (Bannon and Bødker 1997; Greenberg 1991). Consequently, part of the motivation for this investigation was to assess the variability present in the structures, content, and representations used by members of molecular biology laboratories when keeping laboratory records, and to assess the types of articulation work used to achieve mutual intelligibility across laboratory members.

A novel aspect of the investigation presented in this thesis lies in its use of a multi-perspective framework to enable an analysis of 'language in use' for laboratory recordkeeping. This framework combined ethnography, genre analysis, and reading protocols to examine both the textual features and the social context governing laboratory recordkeeping from the viewpoint of both the producers and consumers of laboratory records. In this manner, the investigation was positioned within the growing movement in

discourse analysis that advocates the use of multiple research methods, both textual and contextual, to investigate the complexity of language use in social settings (*e.g.* Askehave and Swales 2001; Bhatia *et al.* 2008; Flowerdew 2002b; Lillis 2008; Smart 2008).

In addition, the investigation consistently employed a stratified selection policy for study participants that enabled analysis of recordkeeping by multiple scientists at different career stages performing different job functions in a range of laboratories. This selection policy also positioned the investigation within the movement towards studies of variation in linguistic behaviour (*e.g.* Braine 1995; Bunton 2002; Samraj 2005; Hyland and Bondi 2006; Hyland and Tse 2007).

By identifying patterns of variability in the language used in records by staff within multiple molecular biology laboratories and by examining the behaviour of readers whilst interpreting records produced by multiple authors, this investigation was able to supplement and extend the work reported in other studies of laboratory recordkeeping in academic science. This includes both studies (*e.g.* Arnstein *et al.* 2002; schraefel *et al.* 2004; Tabard *et al.* 2008; Yeh *et al.* 2006) that have focused on the development of ELN technology, and recent ethnographic studies by Shankar (2007) and Wickman (2010) that have focused principally on the context of record production in single laboratories. In line with the findings of the investigation presented in this thesis, these ethnographic studies and technology development studies have reported tensions in laboratory recordkeeping between the needs of individual authors and the needs of the wider community, characterized by what Shankar (2007:1463) terms the "flexibility and autonomy of the academic scientist" in producing laboratory records. Notwithstanding this flexibility and autonomy, studies of recordkeeping to date have broadly assumed, or paid limited attention to, the intentions of authors to share and the ability of authentic laboratory records to function at a community level as both a shared and shareable written knowledge resource. The investigation presented in this thesis suggests that the intentions and practices of scientists in academic molecular biology laboratories in relation to sharing their laboratory records should not be assumed but should be recognized as complex, individualistic, protectionist to a degree, and motivated by personal concerns.

7.2 Summary of findings

In order to understand the role played by laboratory records in the disciplinary discourse of academic molecular biology laboratories, the investigation addressed "three analytically separable elements in processes of meaning-making: the production of the text, the text

itself, and the reception of the text" (Fairclough 2003:10). In this way, the goal of the study was resolved into the following three research questions. The central findings of this investigation are presented with regard to each of these questions in turn.

1. *What roles do laboratory records play in the discourse of academic molecular biology laboratories?*

The key finding from this investigation in terms of the role played by laboratory records concerns the minimal use of laboratory records in their direct written form by laboratory staff other than the original author of the record. This finding was consistent across each of the three studies that were undertaken as part of this investigation, which evaluated laboratory recordkeeping from the three separable but interrelated viewpoints of record use in laboratory settings, textual features of the records, and the ability of readers to interpret the records. Instead, the investigation highlighted an almost exclusive reliance on the recontextualization of selected information from laboratory notebooks into other 'public genres' such as laboratory talks, research articles, and progress reports as the preferred means of communicating the information held in records in laboratory notebooks.

This finding offers a counterpoint to the findings from other studies, notably a number of studies undertaken as part of technology developments for electronic recordkeeping that report sharing of laboratory records or assume a "cyberbolic" view of laboratory records as a shared resource (*e.g.* Arnstein *et al.* 2002; schraefel *et al.* 2004; Tabard *et al.* 2008). In this sense, the findings of this investigation corroborate recent surveys of data sharing within academic settings (Borgman *et al.* 2007; David 2006; Research Information Network 2008, 2010). These surveys report limited sharing of a range of scientific data, including but not limited to laboratory records, in spite of the availability of technology to enable sharing. Specific issues identified by the investigation described in this thesis and by these other surveys include concerns raised by academic scientists over questions of ownership, ethics, recognition and reward policies, publication rights, and the definition of meaningful data sets for exchange. In light of these findings, it is important to recognize that even traditionally public academic genres such as PhD theses (Barnes 2010) and research articles (Maddox 1996) are now becoming subject to injunctions that limit their availability to a wider public.

Laboratory records in their direct written form served a range of organizational functions for the original author both during and after execution of experimental work. This applied

equally to service-orientated and research-orientated laboratories where the laboratory records functioned as a progress monitor to aid the author in managing the workflow of complex laboratory procedures, as a locus for the author to use whilst managing the complexity of daily laboratory practice in collating information from multiple sources such as kit manuals, laboratory devices, reagent information sheets to form 'compound documents', and as a conventional record to be consulted retrospectively by authors to identify the experimental conditions and results for previous work.

2. *What are the structures, content, and representations that characterize the genre of laboratory records in academic molecular biology laboratories, and to what extent do these vary across different contexts of use?*

The key finding of this investigation in terms of the language used in laboratory records was to identify a range of different recording styles that varied in the level of detail specified within the records, in the explicitness of the referencing schemes used to link records and their elements, in the intertextual mappings that shaped the records, and in the spatial arrangement of the records. The trade-off exhibited by each of these different styles concerned the level of detail recorded in the record against the time spent recording the entry. Specific forms of abstraction were identified that produced summarized forms of recording by routinely omitting selected experimental conditions, by referencing samples using a group identity rather than enumerating each individual sample within the group, and by identifying only selected steps in a protocol.

The recording style used by individual authors was found to remain consistent both within and across notebooks unless specific laboratory-wide policies were in force. In this sense, each individual author had evolved their preferred way of writing laboratory records. No specific correlation was found between the different type of functions performed by authors in the laboratory (principal investigator, postdoctoral researcher, postgraduate researcher, technician) and the preferred style of recording used by an author.

Both readers and authors of laboratory records were found to draw on two specific categories of structuring knowledge in the shape of an IMRD-based genre definition for laboratory records, and domain-specific knowledge of the schemata appropriate for specific types of experimental work. This was broadly in line with models of literacy that position both genre knowledge and subject matter knowledge as important contributing factors to the knowledge required for successful domain-specific communication and the

shaping of expert texts (*e.g.* Beaufort 1998). In particular, both researchers and technicians typically realized their laboratory records using a variation on the IMRD structure in which experimental materials and methods were distributed throughout the record and interspersed with intermediate results leading to a structure more accurately characterized as I-M_1R_1-M_2R_2-M_nR_n-D. The intermediate steps in the method (M_1 … M_n) and the associated intermediate results (R_1 … R_n) typically derived from the specific schema associated with the particular type of laboratory work being documented.

3. *How do readers of laboratory records in academic molecular biology laboratories make sense of laboratory records in different contexts of use?*

The key finding of the investigation in this regard was that only those notebooks written by authors who had since left the laboratory were read as written resources, typically by principal investigators, and then only if the information sought was not available in other derived genres such as research articles. In this sense, the majority of scientists did not routinely read notebooks other than their own.

Direct interpretation of the laboratory records, in the limited circumstances when this occurred, relied on prior understanding of the context in which the experimental work was carried out combined with both domain-specific knowledge of the schemata (*e.g.* Spiro 1980) appropriate for specific types of experimental work and a shared notion of the genre of laboratory records. Given the diversity in the structure, content, and representations used by individual laboratory staff, the findings of the investigation indicated that both these coordination mechanisms could and did break down. Multiple researchers experienced breakdowns when interpreting their own laboratory records during the course of experimental work, which served as a striking marker of the limited reliability associated with some laboratory records.

Readers broadly conformed to the sequence of cognitive moves inherent in the IMRD genre to make sense of a record in terms of why the experiment was being performed, how it was being performed, what data was collected during the experiment, and how are these data were being interpreted. Within each of these steps, the readers moved back and forth through the record in order to answer each of the four questions addressed by the IMRD structure, broadly in line with a question-based approach to information seeking (*e.g.* Ram 1999). In the absence of a clear statement of experimental purpose, for example, readers set out to interpret the experimental purpose in an inductive manner by examining the

materials and methods that could be identified within the record, and the results that could be identified within the record. In the absence of a clear statement of materials and methods, for example, readers set out to interpret the form and content of the results to determine the type of method that could be used to generate these results. In this sense, the focus of attention during reading was not restricted to specific genre sections within the record, and the record was consistently read as a whole.

The most common problem identified by readers at all levels of experience when reconstructing the meaning of laboratory records concerned the use of summarized or abstracted forms of recording. This applied equally across different sections of the record including experimental purpose, materials and methods, and results. In essence, the use of summarized forms of recording gave rise to information gaps, which readers attempted to fill by examining the content of other records. Locating other relevant records to supply additional context proceeded by navigating through explicit cross-referencing schemes, adjacency in the notebooks, and implicit forms of referencing.

7.3 Limitations on the studies

The original intention for this investigation had been to recruit laboratories, participants, and records for the ethnographic, genre analysis, and reading protocol studies from sites at multiple universities. However, the difficulty in gaining access to laboratories within multiple universities led to this decision being abandoned in favour of a sole focus on laboratories within a single UK university. This decision could limit the representativeness of the sample set of laboratories, participants, and records used in the studies. To ameliorate the potential effects of this limitation, it was decided to recruit laboratories that did not collaborate with each other.

The original intention had also been to include laboratories, participants, and records from academic research institutes with a commercial focus in order to enable comparison between research laboratories that direct their activities towards publication only and those that direct their work towards commercialization. However, the difficulty in gaining access to laboratories with a commercial focus also led to this decision being abandoned in favour of a sole focus on research laboratories geared towards publication. This decision could again limit the representativeness of the sample set of laboratories, participants, and records used in the studies.

Gaining access to principal investigators and to laboratory notebooks written by principal investigators proved a particular challenge for the studies. This was due principally to the fact that most principal investigators no longer work in the laboratory on a regular basis but are instead involved in administration and directing the work of others. To a much lesser extent, the difficulty in gaining access to principal investigator notebooks also derived from a greater resistance on the part of principal investigators to make their notebooks available. This difficulty again limited the representativeness of the sample set of participants, and records used in the studies.

The reading experiments conducted for this study were artificial in the sense that they provided the readers with only extracted records and limited background on the setting in which the work was conducted. In authentic use, additional contextual information could be available from other workers in the original laboratory or perhaps even from the original author.

7.4 Future directions

Three investigative paths are envisaged through which to build on the findings of this study. The first of these paths continues the thrust of the discourse analytic study of recordkeeping in laboratories and other settings. The second path is concerned with exploring the potential for developing technology solutions to support specific aspects of laboratory recordkeeping. The third path is concerned with the traditional application domain for genre studies in pedagogy by providing training in laboratory recordkeeping.

With regard to the first of these investigative paths, it is important to note that the main emphasis in the study presented in this thesis has been on a qualitative analysis of laboratory recordkeeping in the discourse of academic molecular biology laboratories. This approach was chosen in order to discover and characterize specific patterns of language use. In order to better understand the distribution of these patterns of language use, it is recommended that a quantitative analysis of specific recording styles be conducted against a larger corpus of laboratory notebooks. If access to additional laboratories outside a single UK university could be gained, it may also prove informative to extend the ethnographic and genre analysis studies to incorporate participants from this extended sample. One particular area of interest concerns correlation between the laboratory function performed by an author and the recording style used by an author. No such correlation was found based on the small sample set recruited for the study presented

in this thesis. Examining patterns of use in an extended sample would provide refined statistical evidence for or against this finding, which could prove informative both in understanding the motivation behind the use of specific recording styles and in directing the (re-)training of laboratory members in recordkeeping.

Laboratory settings form only one of a range of professional domains in which recordkeeping plays a role. As Shankar (2007:1465), *inter alia*, points out "understanding recordkeeping as an act of information creation, not just in academic laboratories by individual scientists and groups, but in other organizational settings" would represent a useful return to "first principles" in studies aimed at "understanding how broader professional standards of accountability and reliability become personalized and reified in documentary practices." Given this agenda, the strengths of the multi-perspective approach to genre analysis used for this study of laboratory recordkeeping indicate that the same multi-perspective approach could be successfully applied to investigating recordkeeping in other domains. Whilst it is recognized that reading protocol analysis may not be the appropriate research method to employ in all cases, the inclusion in future studies of some form of reading experiment to elucidate the viewpoint of record consumers is recommended as an important and necessary extension to previous studies that focus on producers and the context of record production.

Acknowledging that the secondary goal of this investigation relates to technology development, one candidate domain to study is that of the design and programming documentation kept by software engineers during the course of software development projects. For example, it would be interesting to investigate the prospective/retrospective functioning of design and programming documentation from the viewpoints of engineers engaged in both design phases and subsequent maintenance phases, and to investigate any co-dependency between personal recordkeeping and different patterns of group working/group accountability. Examining recordkeeping in other domains, both academic and industrial, would also enable a comparative evaluation of the recordkeeping practices at work within academic molecular biology laboratories against those practices in use in other domains, and thus test out Shankar's (2007:1463) statement that the "leeway and power" afforded to academic scientists is "unusual in organizational recordkeeping".

With regard to the second of these investigative paths, it would be useful to build on two particular issues identified during this study concerning the adoption of electronic recordkeeping technology in academic molecular biology laboratories. The first issue

concerns the impact of the data sharing policy in place within the academic molecular biology laboratory settings under which scientists constructed their laboratory records as a 'personal resource' in contrast to a 'community archive' to be shared. It is argued that this mismatch between scientists' attitude to data sharing, particularly record sharing, and the design goals of ELNs has contributed to the limited uptake of ELN technology. The second issue concerns the contrast in attitude exhibited by laboratory members towards sharing experimental methods via demonstration whilst limiting access to the experimental designs and results held in laboratory records.

It is important to recognize that the study presented in this thesis was conducted in laboratories in which ELN systems were not in use. This focus did not result from a sample selection policy that deliberately excluded those laboratories using electronic recordkeeping systems, but instead represented the actual situation observed within the molecular biology laboratories in the university that formed the site of investigation. As previously stated, this university is a long-established UK university that has a strong tradition in teaching, research, and collaboration with business. As a result, it would be highly informative to undertake a survey across multiple laboratory settings, both academic and industrial, in order to identify and characterize those domains in which electronic recordkeeping using ELNs has gained acceptance within bioscience and other laboratories. The identification of a cohort of laboratories that have adopted ELN technology would enable further studies of the recordkeeping practices at work in such settings. In particular, it would enable an investigation of the co-constitutive nature of ELN technology adoption and laboratory recordkeeping, including whether or not the adoption of ELNs can be correlated with a change in data sharing practices such as the sharing of laboratory records within and between laboratory communities.

Focusing technology development on systems to share experimental methods was identified by this investigation as a potential means of both delivering systems to facilitate the work of researchers and technicians in academic molecular biology laboratories, and advocating the benefits of data sharing within laboratory settings. Novel aspects of recordkeeping for experimental methods that have been identified during this study as potentially benefiting from technology support include the maintenance and visualization of protocol variation histories, and the maintenance of visual demonstrations of laboratory procedures. Accordingly, it is recommended that technology prototypes are developed to support either or both of these aspects of recordkeeping for experimental methods. Whilst the primary motivation for developing these technology prototypes is to support the work

of laboratory members, it is worth noting that these tools could also prove informative as a basis for validating the findings of this investigation.

With regard to the third of these investigative paths, the study identified that laboratory members in both the research-orientated and service-orientated laboratories learned and refined their laboratory recordkeeping practices principally through a process of enculturation. In this context, it was interesting to note that a number of the researchers and technicians who participated in the study stated that they had refined elements of their recordkeeping practice based on knowledge gained and examples seen throughout the course of their participation in the study. Based on the feedback received when delivering presentations of the study findings to diverse groups of laboratory staff, it would be challenging but useful to laboratory staff to develop a training course to support this central aspect of laboratory work. In order to enable researchers and technicians to reflect on their own practices, this training course should incorporate the range of recordkeeping approaches observed to be in authentic use within laboratory settings. In contrast to a number of ESP training courses that address the needs of trainees in a field or of L2 speakers, the focus of this training course in laboratory recordkeeping would be on supporting the work of experienced practitioners in academic molecular biology laboratories.

Appendix 1:
Participant Information Sheet for the Ethnographic Study

This appendix contains a copy of the information sheet used to describe the ethnographic study of academic molecular biology laboratory settings to prospective participants. This information sheet was produced in accordance with the terms of the ethical approval for the study.

UNIVERSITY of GLASGOW

IBLS

Ethnographic Study of the Use of Procedural and Information Standards in the Molecular Genetics/Molecular Biology Laboratory

Information Sheet

Aim Of Study

The aim of this study is to investigate the working practices that are used in the molecular genetics/molecular biology laboratory in order to examine how best to provide computing systems to support these practices.

In particular, this study will focus on how you make use of procedural and information standards to carry out your work. Examples of procedural standards might include laboratory protocols for tasks such as sample preparation, and safety plans such as Control of Substances Hazardous to Health (COSHH) forms. Examples of information standards might include common formats such as MIAME for exchanging micro-array experimental data, and proprietary formats for recording experimental data in laboratory notebooks.

What Will Happen

I will accompany you as you go about your day-to-day work so that I can observe and take notes on the working practices of the laboratory.

In order to observe a range of the work that you carry out, the study will involve observing you over a period of approximately two weeks. In order to minimise any potential disruption to your work, I will establish convenient dates on which to accompany you prior to carrying out the observations.

This will include observing you both whilst working at your office desk, and whilst working in the 'wet' laboratories.

In order to clarify any points that arise whilst observing the work, I may occasionally request an informal interview to ask you questions regarding the tasks that are being performed.

What Data Will Be Collected

I will collect the following types of data during the study to help understand the laboratory working practises:

1. Textual notes describing my observations of how you have planned, executed and analysed your experiments,

2. Recordings of informal interviews conducted with you to discuss how you carried out your work,

3. Where appropriate and with your explicit prior agreement, video recordings of you performing specific aspects of your work,

4. Where appropriate and with your explicit prior agreement, copies of any non-proprietary documents such as laboratory protocols or manuals.

Data Privacy

All results will be held in strict confidence, ensuring the privacy of all participants. No personal participant information will be stored with the data. Online data will be stored in a password-protected computer account; paper data will be stored in a locked office.

Assessment Of The Work Practices Not The Participant

Please note that the purpose of this study is to understand the role that the procedural and information standards play in the work practises of the laboratory. You are not being evaluated; instead, it is the laboratory work practices that are being assessed.

Rights Of The Participant

Please note that you are welcome to withdraw from this study at any time, without prejudice.

Please note that you are also welcome to require that any data collected whilst observing your work be discarded.

A feedback message will be sent to all participants by e-mail or post after the study has been completed.

No payment or other inducement will be offered to any participant in this study.

Any Questions

If you have any questions regarding this study, please feel free to contact either my project supervisor or me at:

Investigator: David Wilson
Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

Telephone: 0141 330 6229
E-mail: d.wilson.3@research.gla.ac.uk

Supervisor: Mark Bailey
Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

E-mail: m.bailey@bio.gla.ac.uk

Ethical Guidelines

This study adheres to the BPS ethical guidelines, and has been approved by the DCS Ethics Committee of the University of Glasgow. Whilst you are free to discuss your participation in this study with the experimenter, if you would like to speak to someone not involved in the study, you may contact the chair of the DCS Ethics Committee at hcp@dcs.gla.ac.uk.

Appendix 2:
Additional Data for the Ethnographic Study

This appendix contains supplementary tables for the ethnographic study of academic molecular biology laboratory settings. The appendix contains a table describing the participants who participated in the ethnographic study.

Table A2-1: Participants in the ethnographic study

Participant (N=13)	*Function*	*Language Status*	*Gender*	*Description*
ES-L1R1	PD	L1	M	Experienced researcher; Previously worked in multiple university research laboratories in the UK; Responsible for directing other researchers; Has previously written published articles; Actively participates in weekly seminar series with other laboratories; Consulted by a number of other researchers for advice on experimental statistics.
ES-L1R2	PD	L1	M	Researcher on first postdoctoral position; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Has previously written published articles; Moved into molecular genetics from another bioscience field.
ES-L1R3	PI	L1	M	Highly experienced, eminent researcher; Previously worked and directed in a number university research laboratories in the UK; Responsible for managing industrial collaborations; Responsible for directing other researchers; Has previously written a number of published articles; Actively participates in weekly seminar series with other laboratories; Active in teaching undergraduates and postgraduate students.
ES-L1R4	T	L1	M	Highly experienced technician; Responsible for directing other technicians, and for overseeing a range of undergraduate and junior researcher projects; Previously worked in a number of university research laboratories, including work in a range of bioscience fields.
ES-L2R1	PD	L1	M	Experienced postdoctoral researcher; Previously worked in multiple university research laboratories in the UK; Responsible for directing other researchers; Has previously written published articles.

Table A2-1 (Cont'd): Participants in the ethnographic study

Participant (N=13)	*Function*	*Language Status*	*Gender*	*Description*
ES-L2R2	T	L1	F	Experienced technician; Previously worked in industrial laboratories and university research laboratories; Responsible for work on collaborative research work with industrial partners.
ES-L2R3	PG	L2	M	Recently started as a postgraduate researcher; Previously worked as a technician in a medical diagnostics laboratory; Project work involves collaboration with a research laboratory in another UK university.
ES-L2R4	PD	L1	F	Researcher on first postdoctoral position; Previously worked in other university research laboratories in the UK; Has previously written published articles.
ES-L2R5	T	L2	M	Experienced technician; Responsible for supporting a range of research projects; Worked in same laboratory for a number of years.
ES-L2R6	PI	L1	F	Highly experienced, eminent researcher; Previously worked in a number of university research laboratories in different countries; Responsible for managing industrial collaborations; Responsible for directing other researchers; Has previously written a number of published articles; Actively collaborates with another principal investigator in a joint research group.
ES-L3T1	T	L2	F	Working towards a PhD on a part-time basis in addition to working as a technician; Jointly responsible for managing interaction with clients of the service laboratory.
ES-L3T2	T	L1	F	Experienced technician; Previously worked in industrial laboratories and university research laboratories; Jointly responsible for managing interaction with clients of the service laboratory.

Table A2-1 (Cont'd): Participants in the ethnographic study

Participant (N=13)	*Function*	*Language Status*	*Gender*	*Description*
ES-L4R1	PI	L1	M	Highly experienced, eminent researcher; Previously worked in a number of university research laboratories in different countries; Active in the development of a specialized laboratory technique; Responsible for directing other researchers; Has previously written a number of published articles; Active in teaching undergraduates and postgraduate students.

Summarizing all laboratory members participating in the ethnographic study in terms of the identifier code assigned to the participant, his/her function at the time of the study, his/her native language status, gender, and a brief description of the participant's experience in laboratory work.
The participant is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table 3-1). This approach maintains the anonymity of both participant and laboratory required under the terms of the ethical approval for the study. The participant's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The participant's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks English as a second language. The participant's gender is indicated using F for a female participant, and M for a male participant.

Appendix 3:
Participant Information Sheet for the Genre Analysis Study

This appendix contains a copy of the information sheet used to describe the genre analysis study of laboratory records in academic molecular biology laboratory settings to prospective participants. This information sheet was produced in accordance with the terms of the ethical approval for the study.

UNIVERSITY of GLASGOW

IBLS

Surveying the Grammar of Laboratory Notebooks in the Molecular Genetics/Molecular Biology Laboratory

Information Sheet

Aim Of Study

The aim of this study is to investigate the range of grammars used by scientists in the molecular genetics/molecular biology laboratory to format the experimental records that are written in laboratory notebooks. The results of this study will aid in the development of computing systems to capture these experimental records in a digital manner.

In particular, this study will focus on identifying the structures, content, and notations that you use to record information in your laboratory notebook. In terms of notebook structure, an example area of interest would be whether you index the entries in your notebook so that you can search through them at a later date. In terms of notebook content, an example area of interest would be the range of information that you choose to record in your notebook such as experimental results, protocols, and to-do lists. In terms of notebook notation, an example area of interest would be whether you use shorthand notations to document your experiments.

The study will evaluate both current and past laboratory notebooks written by multiple scientists from multiple laboratories in order to survey the range of approaches taken to documenting experimental records in laboratory notebooks.

What Will Happen

I will perform a content analysis of a selection of your laboratory notebooks. In particular, I will examine a recent notebook and a subset of previous notebooks from your current laboratory and/or any previous laboratories in which you have worked.

In order to minimise any potential disruption to your work, I can either borrow your notebooks for analysis away from your laboratory site, or I can arrange to examine the laboratory notebooks at your own laboratory site. If you would prefer me to work away from your own laboratory, I will arrange convenient dates on which to borrow your notebooks, and I will ensure that the notebooks are returned at a specified time. I anticipate borrowing your notebooks for a period of approximately one week.

In order to clarify any points that may arise whilst analysing your laboratory notebooks, I may occasionally request an informal interview to ask you questions regarding the information that is being recorded in the notebooks.

What Data Will Be Collected

I will collect the following types of data during the study to help understand the grammar used to record experimental records in laboratory notebooks:

1. Textual notes describing my observations of the structures, content, and notations used in your laboratory notebooks,

2. With your prior permission, digital scans and copies of selected pages from your laboratory notebooks, and

3. Recordings of informal interviews conducted with you to discuss how the information is recorded in your notebooks.

Data Privacy

I will keep your laboratory notebooks in a locked office whilst they are in my possession, and I will not show them to any third party without your explicit prior agreement.

All results will be held in strict confidence to ensure the privacy of all participants. No personal participant information will be stored with the data. Online data will be stored in a password-protected computer account; paper data will be stored in a locked office.

Assessment Of The Work Practices Not The Participant

Please note that the purpose of this study is to understand the range of structures, content, and notations used by scientists in the molecular genetics/molecular biology laboratory within their laboratory notebooks. You are not being evaluated; instead, it is the grammar of laboratory notebooks that is being assessed.

Rights Of The Participant

Please note that you are welcome to withdraw from this study at any time, without prejudice. Please note that you are also welcome to require that any data collected whilst observing your work be discarded.

A feedback message will be sent to all participants by e-mail or post after the study has been completed.

No payment or other inducement will be offered to any participant in this study.

Any Questions

If you have any questions regarding this study, please feel free to contact either my project supervisor or me at:

Investigator: David Wilson
Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

Telephone: 0141 330 6229
E-mail: d.wilson.3@research.gla.ac.uk

Supervisor: Mark Bailey
Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

E-mail: m.bailey@bio.gla.ac.uk

Ethical Guidelines

This study adheres to the BPS ethical guidelines, and has been approved by the DCS Ethics Committee of the University of Glasgow. Whilst you are free to discuss your participation in this study with the experimenter, if you would like to speak to someone not involved in the study, you may contact the chair of the DCS Ethics Committee at hcp@dcs.gla.ac.uk.

Appendix 4:
Additional Data for the Genre Analysis Study

This appendix contains supplementary tables for the genre study analysing the laboratory records kept in notebooks by scientists at work in academic molecular biology laboratory settings.

The appendix contains a table describing the authors who provided notebooks for the genre analysis study (p 256), and a table describing the notebooks that were analysed to investigate the genre of laboratory records (p 259).

Table A4-1: Notebook authors in the genre analysis study

Author (N=14)	*Function*	*Language Status*	*Gender*	*Description*
GS-L1R1	PD	L1	M	Experienced researcher; Previously worked in multiple university research laboratories in the UK; Responsible for directing other researchers; Has previously written published articles; Actively participates in weekly seminar series with other laboratories; Consulted by a number of other researchers for advice on experimental statistics.
GS-L1R2	PD	L1	M	Researcher on first postdoctoral position; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Has previously written published articles; Moved into molecular genetics from another bioscience field.
GS-L2R1	PG	L2	F	Researcher just completing her PhD thesis and about to start a first postdoctoral position; Previously worked in university research laboratories in different countries; Has previously written published articles.
GS-L2R2	PD	L1	F	First postdoctoral position; Previously worked in university research laboratories in different countries, including work as a technician and a postgraduate researcher; Has previously written published articles.
GS-L2R3	PG	L1	M	Researcher just completed his PhD thesis and about to start a first postdoctoral position; Previously worked in university research laboratories in different countries; Previously worked in non-bioscience fields including computing; Has previously written published articles.
GS-L2R4	T	L2	F	Experienced technician; Worked in multiple university research laboratories including bioscience fields other than molecular biology.
GS-L2R5	PG	L1	F	Recent graduate just beginning a PhD studentship.

Table A4-1 (Cont'd): Notebook authors in the genre analysis study

Author (N=14)	*Function*	*Language Status*	*Gender*	*Description*
GS-L3T1	T	L1	F	Experienced technician; Previously worked in industrial laboratories and university research laboratories; Jointly responsible for managing interaction with clients of the service laboratory.
GS-L3T2	T	L2	F	Working towards a PhD on a part-time basis in addition to working as a technician; Jointly responsible for managing interaction with clients of the service laboratory; Has previously co-authored published articles.
GS-L4R1	PG	L2	M	Recently started as a postgraduate researcher; Previously worked as a technician in a medical diagnostics laboratory; Project work involves collaboration with a research laboratory in another UK university.
GS-L4R2	T	L2	M	Experienced technician; Responsible for supporting a range of research projects; Worked in same laboratory for a number of years.
GS-L4R3	PG	L1	F	Researcher just completed her PhD thesis; Limited previous experience in any other research laboratories.
GS-L5R1	PI	L2	M	Experienced researcher; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Responsible for managing collaborations with other research groups; Has previously written a number of published articles.

Table A4-1 (Cont'd): Notebook authors in the genre analysis study

Author (N=14)	*Function*	*Language Status*	*Gender*	*Description*
GS-L5R2	T	L1	F	Recently appointed technician in a new research group; Limited previous experience in any other research laboratories.

Summarizing all authors providing notebooks for the genre study in terms of the identifier code assigned to the author, his/her function at the time of the study, his/her native language status, gender, and a brief description of the author's experience in laboratory work.
The notebook author is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table 4-1). This approach maintains the anonymity of both author and laboratory required under the terms of the ethical approval for the study. The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The author's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks English as a second language. The author's gender is indicated using F for a female participant, and M for a male participant.

Table A4-2: Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L1R1	*GS-L1R1-N1*	*GS-L1*	Bound, general-purpose A4 notebook; Principally contains experiments for large-scale genotyping study of disease-related genes; Contains minimal records other than experiment records in the form of a draft for a publication, brief notes from a meeting, and summaries of project statistics.
	GS-L1R1-N2	*GS-L1*	Bound, general-purpose A4 notebook; Principally contains experiments for large-scale genotyping study of sports performance.
GS-L1R2	*GS-L1R2-N1*	Research laboratory in a UK university whilst a postdoctoral researcher.	Bound, general-purpose A4 notebook; Principally contains genotyping study for human disease including significant number of entries relating to design of primers and conditions; Contains limited examples of to-do lists for planning activities; Written in a highly detailed format supporting error recovery and process refinement.
	GS-L1R2-N2	Research laboratory in a UK university whilst a postgraduate researcher.	Bound, general-purpose A4 notebook; Principally contains a range of cell culture, cloning and vector insert techniques; Contains examples of reflective thinking regarding progress and direction of work; Written in a highly detailed format supporting process refinement; Dramatic lexis due to frustration with progress.
	GS-L1R2-N3	*GS-L1*	Bound, general-purpose A4 notebook; Principally contains experiments for large-scale genotyping study of disease-related genes; Written in a highly detailed format supporting process refinement.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L2R1	*GS-L2R1-N1*	Research laboratory in a South American university.	Loose A4 pages in a ring-bound folder; Written in Spanish; Consistent use of template-based records throughout the notebook for some procedures; Principally contains experiments for EMSA, PCR, and restriction digests.
	GS-L2R1-N2	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains entries for genotyping to identify repeats and interrupts for human disease samples; Written at start of PhD research project; Written in a detailed format with significant use of headings and other cues.
	GS-L2R1-N3	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains entries for genotyping to identify repeats and interrupts for human disease samples; Written in middle of PhD research project; Written in a detailed format with significant use of headings and other cues.
GS-L2R2	*GS-L2R2-N1*	Research laboratory in a North American university whilst a postgraduate researcher.	Bound, general-purpose US letter notebook; Written in a laboratory with centralized practices for data management; Principally contains entries for genotyping experiments for human disease samples.
	GS-L2R2-N2	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains entries for genotyping to identify repeats and interrupts for human disease samples; Written in a laboratory without centralized practices after an extended period with lab-centralized practices.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L2R3	*GS-L2R3-N1*	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains a range of cell culture, cloning and vector insert techniques; Written at start of PhD research project.
	GS-L2R3-N2	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains a range of cell culture, cloning and vector insert techniques; Written in middle of PhD research project.
	GS-L2R3-N3	*GS-L2*	Bound, special-purpose lab book with duplicate pages and signature spaces; Principally contains a range of cell culture, cloning and vector insert techniques; Written at end of PhD research project; Consists largely of cut&paste inserts from entries recorded using a word processor.
GS-L2R4	*GS-L2R4-N1*	*GS-L2*	Bound, general-purpose A4 notebook; Contains entries for a range of work including lab administration such as making up stock solutions, and tasks to support the research staff such as genotyping; Contains a manually maintained index of experiments for all entries in the notebook.
	GS-L2R4-N2	*GS-L2*	Bound, general-purpose A4 notebook; Contains entries for a range of work including lab administration such as making up stock solutions, and tasks to support the research staff such as genotyping; Contains a manually maintained index of experiments for all entries in the notebook.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L2R5	*GS-L2R5-N1*	*GS-L2*	Bound, general-purpose A4 notebook; Principally contains entries for genotyping experiments for human disease samples. Written at start of PhD research project during a short-term project in the laboratory.
GS-L3R1	*GS-L3R1-N1*	*GS-L3*	Bound, general-purpose A4 notebook; Principally contains experiments to service client orders for running samples on microarrays, and involves significant use of commercial kits; Written in a highly detailed format to support quality assurance.
	GS-L3R1-N2	*GS-L3*	Bound, general-purpose A4 notebook; Principally contains experiments to service client orders for running samples on microarrays, and involves significant use of commercial kits; Contains work to customize the use of these kits for specific client needs; Written in a highly detailed format to support quality assurance.
	GS-L3R1-N3	*GS-L3*	Bound, general-purpose A4 notebook; Principally contains experiments to service client orders for running samples on microarrays, and involves significant use of commercial kits; Written in a highly detailed format to support quality assurance.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L3R2	*GS-L3R2-N1*	*GS-L3*	Bound, general-purpose A4 notebook; Principally contains experiments to service client orders for running samples on microarrays, and involves significant use of commercial kits; Written in a highly detailed format to support quality assurance.
	GS-L3R2-N2	*GS-L3*	Bound, general-purpose A4 notebook; Principally contains experiments to service client orders for running samples on microarrays, and involves significant use of commercial kits; Written in a highly detailed format to support quality assurance.
GS-L4R1	*GS-L4R1-N1*	*GS-L4*	Bound, general-purpose A4 notebook; Principally contains a range of cell culture, cloning and vector insert techniques; Written at start of PhD research project.
GS-L4R2	*GS-L4R2-N1*	*GS-L4*	Bound, general-purpose A4 notebook; Contains entries for a range of work including some lab administration such as making up stock solutions, and mainly tasks to support the research staff; Contains entries for a range of techniques associated with use of *Drosophila melanogaster* as a model organism; Written in variable style for routine and non-routine tasks.
	GS-L4R2-N2	*GS-L4*	Bound, general-purpose A4 notebook; Contains entries for a range of work including some lab administration such as making up stock solutions, and mainly tasks to support the research staff; Contains entries for a range of techniques associated with use of *Drosophila melanogaster* as a model organism; Written in variable style for routine and non-routine tasks.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L4R3	*GS-L4R3-N1*	*GS-L4*	Bound, general-purpose A4 notebook; Principally contains entries for various assays (lucerifase, ß-galactosidase) of different genetic strains of fly; Includes significant spreadsheet style inserts for assay results; Contains limited example of draft talk; Written at end of PhD research project.
	GS-L4R3-N2	*GS-L4*	Bound, general-purpose A4 notebook; Principally contains entries for various assays (lucerifase, ß-galactosidase)of different genetic strains of fly; Includes significant spreadsheet style inserts for assay results; Written at start of PhD research project.
GS-L5R1	*GS-L5R1-N1*	*GS-L5*	Bound, general-purpose A4 notebook; Contains entries for a range of work including genotyping, IHC, and other staining techniques; Written in an economical style in chronological order with cross-referencing schemes.
	GS-L5R1-N2	*GS-L5*	Bound, general-purpose A4 notebook; Contains entries for a range of work including genotyping, IHC, and other staining techniques; Includes an extensive review of progress to date including results analysis; Written in an economical style in chronological order with cross-referencing schemes.
	GS-L5R1-N3	*GS-L5*	Bound, general-purpose A4 notebook; Contains entries for a range of work including genotyping, IHC, and other staining techniques; Written in an economical style in chronological order with cross-referencing schemes.

Table A4-2 (Cont'd): Notebooks for the genre analysis study

Author	*Notebooks (N=30)*		
	Notebook Id	**Laboratory**	**Description**
GS-L5R2	*GS-L5R2-N1*	*GS-L5*	Bound, general-purpose A4 notebook; Contains entries for a range of work including some lab administration such as making up stock solutions, and mainly tasks to support the research staff; Contains entries for a range of visualization techniques including different types of gel such as Coomassie gels.

Summarizing all notebooks used in the genre analysis study in terms of the author, laboratory of origin, and a brief description of the notebook. A unique identifier code is assigned to each notebook. The notebook author is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table A4-1). The laboratory in which the notebook was originally produced is identified either using the unique laboratory identifier (see Table 4-1) if it was produced in a laboratory that participated in this study, or by a description if it was produced in any other laboratory. The description outlines the container and content in terms of the physical format of the notebook, and the types of entry in the notebook.

Appendix 5:
Content Analysis Framework for Laboratory Notebooks

This appendix contains an extract from a completed content analysis framework (p 267) produced for the genre study of laboratory records, together with a scanned copy of one of the notebook pages (p 285) referenced in the completed framework. Note that only part of the table in section 9 of the completed content analysis framework has been included. The scanned copy of the notebook page has been included in order to enable a comparison between the data collected within the content analysis framework and the record originally written by the notebook author. The same content analysis framework was used to direct the analysis of each notebook in the study in order to ensure a consistent approach to the analysis of all notebooks over the course of the study.

IBLS

Surveying the Grammar of Laboratory Notebooks in the Molecular Genetics/Molecular Biology Laboratory

Notebook Contents Analysis for L1R1-N1

Investigator: David Wilson, Computing Science and IBLS/Molecular Genetics

Address: Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

Telephone: 0141 330 6220

E-mail: d.wilson.3@research.gla.ac.uk

Supervisors: Mark Bailey, IBLS/Molecular Genetics, and
Phil Gray, Computing Science

Aim Of Study

The aim of this study is to investigate the range of representations and grammars used by scientists in the molecular genetics/molecular biology laboratory to format the experimental records that are written in laboratory notebooks. The results of this study will aid in the development of computing systems to capture these experimental records in a digital manner.

In particular, this study will focus on identifying the structures, content, and notations that are used by laboratory staff to record information in laboratory notebooks. In terms of notebook structure, an example area of interest would be whether and how the entries in laboratory notebooks are indexed so that they can be search through at a later date. In terms of notebook content, an example area of interest would be the range of information that is recorded in laboratory notebooks such as experimental results, protocols, and to-do lists. In terms of notebook notation, an example area of interest would be whether shorthand notations are used whilst documenting experiments.

The study will evaluate both current and past laboratory notebooks written by multiple scientists from multiple laboratories in order to survey the range of approaches taken to documenting experimental records in laboratory notebooks.

Role Of Contents Frame

In order to evaluate the range of grammars used by scientists to record experimental work in laboratory notebooks, it is essential to survey a range of these notebooks produced by multiple scientists from multiple laboratories. The contents frame defines a standardized set of issues against which all laboratory notebooks will be evaluated during the course of the study. The document contents frame encompasses different sections identifying the range of issues to be evaluated.

A separate document frame is to be completed for each laboratory notebook in the study.

Section 1 - Participant Information

Participant ID	L1R1
Confirm consent been obtained?	Yes
What was the participant's function in the laboratory when using this notebook?	Postdoctoral researcher
Additional description of the participant *(e.g. years of experience, worked in other labs, experience of other disciplines)*	Senior postdoctoral researcher with 5+ years of experience. Has worked in multiple laboratories including resident periods of work at collaborators' laboratories in Japan and Europe. Used to independent work, and provides supervisory roles for undergraduate and postgraduate research staff. Background in sports science, and subsequent focused on work in molecular biology/genetics. Interested in interactions between genetics and environment in humans for physical performance such as running, and disease-related issues such as obesity.
In what type of laboratory was the notebook produced?	Research
Additional description of the laboratory *(e.g. size, degree of collaboration with other laboratories, layout)*	Academic research laboratory conducting research into genetics and other factors influencing human performance. Example study is looking at the interaction between specific genotype/phenotype dependencies in the area of sports science such as genotype frequency amongst elite runners. Another example study involves looking at the link between genotype and obesity in children. Shared facilities between with two/three research groups. Level of collaborative projects between groups is very limited so it appears to be more of a resource sharing arrangement. Housed in a self-contained facility within the university consisting of a wet lab room, sports science lab with fitness equipment (running machines), and separate offices

	for technician, group leaders and postdocs. Postgraduate researchers share communal writing space. Other share office space with two/three postdoctoral researchers/technicians/visitors per room. Principal investigators have their own office space. Current staff consists of group leader, 2 postdocs, 2 visiting researchers (medics), 4 postgraduate researchers at varying stages of their studentships, and 2 technicians.

Section 2 - Notebook Information

Notebook ID	L1R1-N1
Format of the notebook *(e.g. binding, carbon copy pages, format)*	Bound, hardback, A4, ruled notebook (manuscript book). No carbon copy or duplicate pages.
How has the notebook been named by the participant?	Owner's name, staff number, and the dates covered by the notebook are written on the inside cover.
Which dates does the notebook cover?	20 November 2006 – 06 March 2008
Does this notebook relate to work for the participant's current laboratory?	Yes
Describe any set of which this notebook forms a part.	Not specifically named as part of a project-related set of notebooks, just person-related notebooks.
Where does the participant keep the notebook? *(e.g. in office, on laboratory bench)*	Current notebook is kept either on the participant's desk in the office, or on the participant's lab bench. Previous notebooks are kept in a filing cabinet next to the participant's desk in the office. The office is shared with one other researcher.

Section 3 - Notebook Contents

How many entries are recorded in the notebook?	Notebook is complete, and contains with 82 entries.
Which types of entry are recorded in the notebook? ***(e.g. experimental record, protocol definition, reflective thinking, planning information)***	71 Experimental records; 4 Project plans for experiment; 1 Plan of draft publication; 1 Personal minutes of meeting; 5 Summaries of project statistics.
How are the different types of entry distributed within the notebook? ***(e.g. different sections of the notebook)***	The experimental records form the majority of the notebook. The different types of entry are intermingled – no sections are created in the notebook. Dates of experimental work show intermittent periods when no entries have been recorded in this notebook. Gaps range from 1 week to 5 months between entries.

Section 4 – Experimental Records

What is the range of formats used to document experimental records in the notebook? *(e.g. fully specified with protocol, cross-referenced to previous records, simple status messages)*	Range of formats used in experiments is: 22 Summarized optimization experiment with optimization conditions, PCR mix, summarized gel loading and inserted gel; 26 Summarized genotyping experiment with sample list, PCR mix, summarized gel loading and inserted gel; 16 Continuation of genotyping for additional samples with sample list and inserted gel; 3 Kit tests with more detailed specification of protocol steps; 4 Sample quality tests with sample list, PCR mix, NanoDrop results list.
How are the different types of experimental record formats distributed throughout the notebook? *(e.g. only first use is fully specified)*	The experimental work for each gene is conducted in a block so that the optimization records for a gene precede the genotyping records for that gene. Minimal or no overlap between the records pertaining to work on different genes. Kit tests and sample quality tests are intermixed with the main work on genes and appear to be conducted in a responsive mode.
Are there distinctive subsections in the experimental record formats used in the notebook? *(e.g. aims, materials, methods, results, discussion)* **If so, which subsections are used in the experimental record formats in the notebook?**	Yes Underlined protocol task headings are used to emphasize the spatial layout of the records in order to separate out the major steps involved in the SNP genotyping experiments in the form of sample list, mix, and gel. No results interpretation is listed in the notebook entries.
To what degree is the notation used to document experimental records common to multiple experimental record formats?	Highly consistent series of steps used in documenting the experiments. Gel loading is always summarized. Sample lists for the genotyping experiments are always specified through predefined plates containing sets of samples arranged for a 96 well plate. The association of samples with plates is kept in an external spreadsheet local to the author. No results interpretation is listed in the notebook, as

	the author has offloaded the data generated by the experiments to a personal, Microsoft Excel spreadsheet kept in the author's home directory. A separate spreadsheet is also used to record some detail about the primers used in the experiments, built from data retrieved from shared community databases such as NCBI.
What is the range of the biology being performed in the experimental work recorded in the notebook? ***(e.g. PCR, sequencing, kits)***	Almost all of the experimental work is concerned with SNP genotyping experiments for use in determining allelic variation across a large number of samples. The work is concerned primarily with two genes, the *PPARD* (peroxisome proliferator-activated receptor delta) gene and the *FTO* (fat mass and obesity associated) gene. *PPARD* codes for a nuclear receptor known to be involved in lipid accumulation and implicated as a factor in obesity and diabetes mellitus. *FTO*, as indicated by the choice of name, is similarly implicated as a factor in obesity. Experiments in the notebook are either geared to optimized the experimental conditions (PCR conditions, volumes) for use in SNP genotyping, or to subsequent executions of SNP genotyping experiments on multiple sets of samples. There are also two one-off experiments geared to testing out commercial kits. One of these investigates a Genotek kit to extract DNA from spit samples for use in the project. The other similarly investigates a GenomiPhi kit for DNA amplification.
Scans taken of each experimental record format found in the notebook?	Yes

Section 5 – Protocol Definitions

What is the range of formats used to document protocol records in the notebook? *(e.g. fully-specified, cross-referenced to external source, cut & paste from external source)*	1. Detailed, numbered sequence of steps as in kit instructions; 1 Reference to a standard protocol
How are the different formats for protocol record distributed throughout the notebook? *(e.g. only first use is fully specified*	No standalone protocol records were included in the notebook. Only a single record contained direct protocol references, which were embedded within an experimental record documenting their actual use in extracting DNA from spit samples.
Are there distinctive subsections in the protocol record formats used in the notebook? *(e.g. aims, materials, steps, problems)* **If so, which subsections are used in multiple protocol record formats in the notebook?**	No
What types of supplementary information is recorded? *(e.g. effective temperature ranges, shorter times for steps, steps that can be missed out)*	No contingency information recorded, just the actual execution of the protocol. Only supplementary information was clarification of an instruction to "mix gently by inversion" by adding "(> 5)" to quantify how many inversions.
What types of information about equipment use is recorded? *(e.g. device settings, locations)*	Minimal reference to devices. Only devices referenced are "oven" and "-80ºC" (freezer). Both references include timings with the device use as in "dry in oven at 50ºC for 15mins" and "into -80ºC at 1310 out at 1332".
Scans taken of each protocol record format found in the notebook?	Yes, included in the scan of the experimental record.

Section 6 – Other Types of Record

Are there other types of record beyond experimental and protocol records in the notebook? **If so, what is the range of other types of record in the notebook?** ***(e.g. to-do lists, planning info, reflective thinking)***	Yes 4 Project plans for experiment; 1 Plan of draft publication; 1 Personal minutes of meeting; 5 Summaries of project statistics.
What is the range of formats used to document each of the other types of record in the notebook? ***(e.g. fully-specified, abbreviated, cross-referenced to external source)***	Project plans are essentially to do lists of tasks to be completed, together with some brief resources considerations for a subset of the tasks. No formal project planning approach is used, e.g. Gantt charts, PERT charts, etc. Plan of a draft publication is also presented as a to-do list of ideas/changes to make to present the form of argument supported by the collected data. Minutes of a meeting are textual notes, written loosely in the form of bullet points with some resource calculations. The summaries of project statistics are tabulated lists of summary statistics are might be presented as tables or other formats in published papers – "publication-readiness".
How are the different formats for other records distributed throughout the notebook?	No sectioning is applied within the notebook. Different types of record are intermixed within the notebook.
Are there distinctive subsections in the other record formats used in the notebook? **If so, which subsections are used in multiple other record formats in the notebook?**	No
Scans taken of each type of other record format found in the notebook?	No The minutes of meetings, project plans, and draft publication were considered too sensitive for scanning so only descriptions of these entries are available.

Section 7 – Indexing Schemes

Are the entries in the notebook indexed by date?	Yes
Are the entries in the notebook indexed other than by date? **If indexed other than by date, which criteria are used as the basis for indexing the entries in the notebook?** *(e.g. project, client, protocol)*	Yes, but only implicitly Titles of experiments are repeated throughout the notebook in order to chain related records as in "Ppard optimisation" used as the title for multiple records.
How are the indexes made visible on the notebook? *(e.g. post-it notes, colour-coded inserts)*	No emphasis is given other the position of the "index" as the record title.
At what level of granularity are records indexed? *(whole record, parts, results only)*	Whole records only.
If parts of an entry are indexed separately, which criteria are used as the basis for indexing the parts of each entry in the notebook? *(e.g. project, client, protocol)*	N/A
Scans taken of each of the notebook indexing schemes?	Yes, as part of the experimental records.

Section 8 – Experimental Workflow

Do the experimental records in the notebook contain workflow mark-up ?	Yes
If so, how is the workflow/progress mark-up documented in the notebook?	Some of the records make use of progress ticks alongside sample lists and/or mix lists in order to monitor the adding of reagents and processing of samples in a set.
Are the experimental records in the notebook interleaved or not?	Non-interleaved.
To what extent is the interleaved/non-interleaved scheme used consistently throughout the notebook?	All experimental records are recorded in a non-interleaved manner, irrespective of the duration of the experiment. Blank sections, and in one case two blank pages, were present that had been left as space to record later steps in the record of an experiment.
Scans taken of each of the notebook workflow mark-up schemes?	Yes, as part of the experimental records.

Section 9 – Detailed Description of Exemplar Entries

Entry	Date	Pages	Type	Description	Sections	Indexes/CRs/Inserts	Specification Level
3	22/11/2006	2	ER	Execution of protocol tasks for optimization of PCR targeting the *PPARD* gene using multiple primer pairs. Includes running of a gel to assess results of different mixes applied across a temperature gradient. Main body of entry is written recto, except for the temperatures in the temperature gradient written verso opposite the mix list, and a quality note written verso opposite the gel loading parameters.	Title ("Ppard optimisation"); Protocol task heading ("Mix"); Mix list with units; PCR primer list ("Mix 3 = Ppard 194F + 422R(20) 12 …") with pair names ("Mix 3", "Mix 4", "Mix 5"); PCR gradient with equation ("T = 40 + 14°C") and temperatures; Note ("12 .. 12"); Protocol task heading ("Gel"); Gel loading list ("Row 1 - 6µl 100bp ladder – Mix 3+ Mix 4 (5µl) Multipipette") with amount loaded, ladder, loading device ("Multipipette"), and pattern only; Error report ("Some wells had some evaporation 1 + 12") with affected samples ("1 + 12"); Gel image with printed text annotations	Gel identified by date and same descriptive name used in the title ("22.11.06 ppard optimisation"); Internal CR between error report and gel loading parameters using arrow to point from note to parameters; Inserted printout of gel image.	Abbreviations (F, R, DNA, dH_2O, 100bp); No explicit purpose statement; No explicit results interpretation; Reagents and volumes in mixes stated explicitly; Units stated explicitly for most measurements ("µl, bp"), and for some temperatures ("°C"); Units ("µl") stated explicitly for single reaction amounts in mix but not for multiple reactions; Task headings ("Mix", "Gel") are stated and underlined; No discrete steps in the protocol are stated explicitly; Progress ticks used for some of the temperatures in the gradient temperature list; Structured approach to specifying the gel loading

Entry	Date	Pages	Type	Description	Sections	Indexes/CRs/Inserts	Specification Level
					("22.11.06 ppard optimisation", "Mix 3 and 4", "Mix 5 and 3").		pattern as in (<row> <amount loaded> <ladder> - <samples> <amount loaded>) but no column headings are used to make parameter meaning explicit; No negative control used on gel; No gel running conditions are specified; Gel image is annotated with date as identifier; Different temperatures in the gradient are fully enumerated ("1 40.1 2 40.5" – "12 54.2"); Device used for loading the gel is explicit ("multipipette") – perhaps as a means of describing the loading pattern; Single mix list is used to specify amounts for multiple combinations of primer pairs/samples for variant digests using DNA, F, R as placeholders for sample, forward, and

Entry	Date	Pages	Type	Description	Sections	Indexes/CRs/Inserts	Specification Level
							reverse primers.
4	22/11/2006	1.5	ER	Execution of protocol tasks for Genotek extraction of DNA from sponge/spit samples. Main body of entry is written recto, except for the sample list and volumes used to test the kit written verso opposite the initial sample list with source names.	Title ("YP Children Genotek extraction"); Sample list with source ("4 samples taken - <X> - sponge + spit <Y> - sponge + spit"); Protocol task report ("4 samples/kits into drying oven at 50°C at 1.30pm out at 5.30pm") with time codes and device use"); Sample list with sample label ("<X> spit = G1 <Y> spit = B1 …"); Protocol task report with numbered steps, reagent calculations, device use, and time codes; Note (" – sponge kits have very little left – maybe 500ul?"); Protocol task report with heading ("Precipitation"); , Protocol task report with named protocol ("Standard protocol"), device use, and time	CR to kit/protocol from external supplier ("Genotek extraction"); CR to defined protocol ("Precipitation – Standard protocol").	Abbreviations (TE); No explicit purpose statement; Reagents and volumes in mixes stated explicitly in protocol steps; Units stated explicitly for measurements ("ml, µl,"), durations, and temperatures ("°C"); Units ("µl") stated explicitly for single reaction amounts in mix but not for multiple reactions; Task heading ("Precipitation") is stated for one of the tasks and underlined; Discrete steps are stated explicitly and numbered for the main extraction task based on kit usage; Narrative text is used to document the discrete steps ("2 20ul purifier added and tubes gently inverted", "8 13K for

UNIVERSITY of GLASGOW

IBLS

Entry	Date	Pages	Type	Description	Sections	Indexes/CRs/Inserts	Specification Level
					codes ("into -80°C at 13:10 out at 13:32 … dry in oven at 50oC for 15min"); Protocol task report ("PCR DNA ->"); Protocol task heading ("Mix"); Mix list with units; Sample volume list with sample label ("Do 8 samples with 1ul DNA and 8 with 0.5ul 1->8 = 1ul DNA 9>16 = 0.5ul DNA"); Concentration calculation (5µg = 5000ng 100µl = 50ng/µl").		1minute", "10 Dry off in oven at 50°C (1245-1252)""11 Elute with 100µl 1xTE + incubate at 50°C for 10min"); Labels used for samples are explicitly listed ("<X> spit = G1", "<X> sponge = G2").
5	24/10/2006 (Entered in error as the date should be given as 24/11/2006)	0.5	ER	Execution of protocol tasks for repeat of PCR following on from previous entry. Also includes test of different dilutions of DNA in the PCR for a single test sample. Entry is written verso, immediately below associated previous entry.	Purpose statement ("Redo PCR as G1pre failed. Also testing dilution of DNA (1/2, ¼, 1/8, 1/16) (B2)"); Protocol task heading ("Mix"); Mix list; Sample list sample label ("1 ½ 2 ¼ 3 1/8 4 1/16 5 G1pre 1µl 6 G1pre 2µl 7 G1 1µl (not pre) 8").	Implicit CR to previous entry via sample labels ("G1pre", "G1").	Abbreviations (DNA, dH_2O); Reagents and volumes in mixes stated explicitly except for volume of DNA in mix ("DNA = "); Amount of DNA in mix is deferred to the sample list and is specified for each sample ("5 G1pre 1µl 6 G1pre 2µl 7 G1 1µl"); Units ("µl") stated explicitly

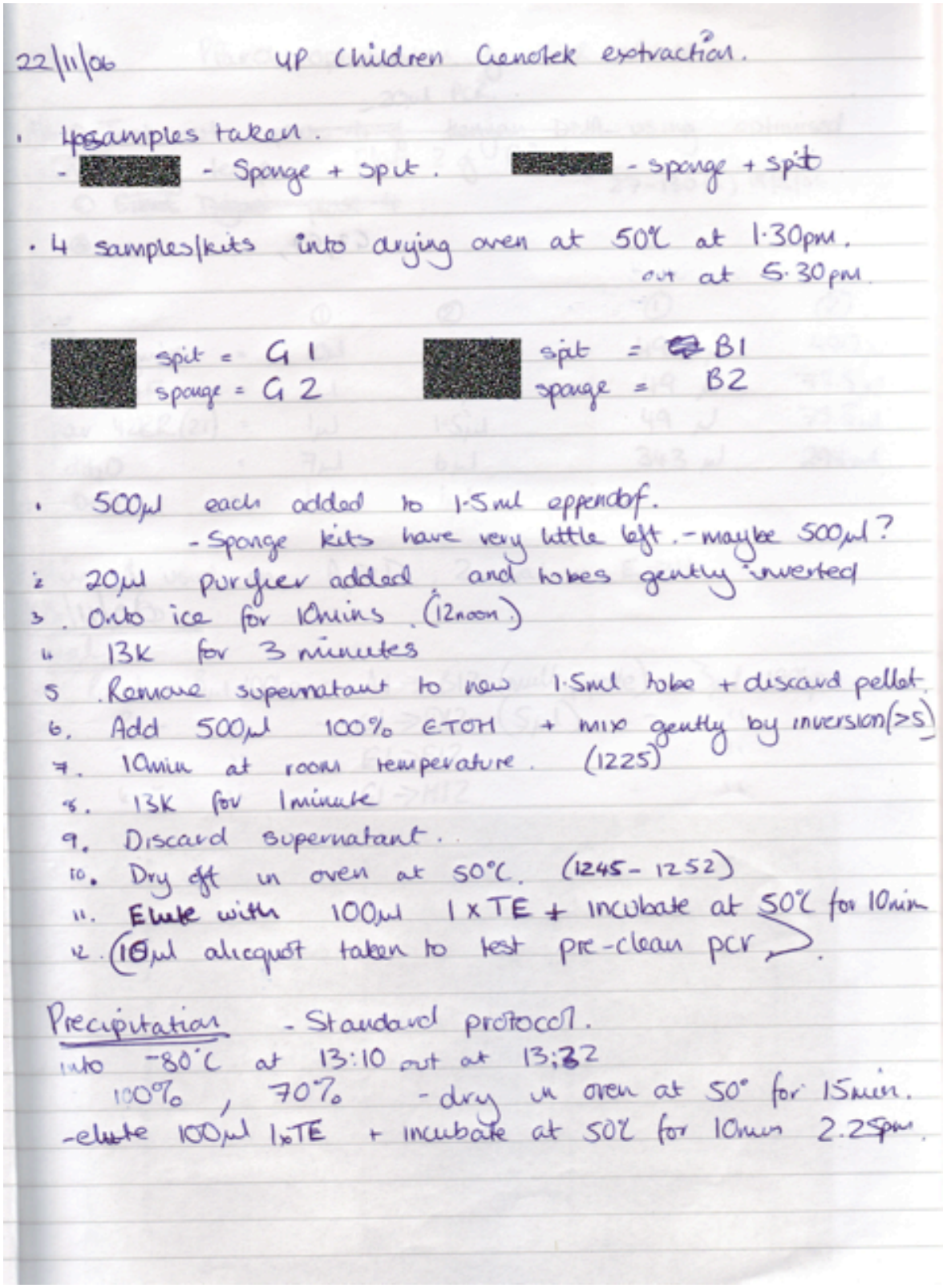
Entry	Date	Pages	Type	Description	Sections	Indexes/CRs/Inserts	Specification Level
				It appears that an area of blank space was left to enable this continuation entry to be filled in.			for single reaction amounts in mix but not for multiple reactions; Task heading ("Mix") is stated and underlined; No discrete steps are stated explicitly; Single mix list is used to specify amounts for the range of samples using DNA, as a placeholder for sample; Labels used for samples are explicitly listed ("B2", "G1pre"); No negative control stated.
6	22/11/2006	0.5	ER	Execution of protocol tasks for optimization of PCR volumes. Entry is dated for the same date as the previous entry. Entry is written recto leaving a blank verso page to separate this entry from the previous entry for the same date.	Title ("PPard optimisation of PCR volume"); Purpose statement ("(1) Test out 20ul PCR Plate 2 of <X> using optimised annealing temp."); Sample list ("27 – 130 (S) 19/5/06") with sample group ("Plate 2 of <X>; Protocol task heading("Mix"); Mix lists with name ("1", "2")	CR between purpose statement and sample list using arrow (purpose statement contains source of samples): CR to colleague supplying samples by name.	Abbreviations (DNA, dH_2O); Reagents and volumes in mixes stated; Units ("µl") stated explicitly for all volumes in mix list; Task heading ("Mix") is stated and underlined; No discrete steps are stated explicitly; Single mix list is used to specify amounts for the range of samples using DNA, as a placeholder

Section 10 – Summary of Abbreviations and Acronyms

Code	Meaning	Origin
100bp	Named ladder mix for gels	Catalogue name
1kb+	Named ladder mix for gels	Catalogue name
Multi; Multipip; MP	Multipipette	Own abbreviation
RM	ReddyMix	Own abbreviation
dH_2O	Distilled water	Common formula
DNA	Deoxyribonucleic acid	Common abbreviation
Eth	Ethiopian	Own abbreviation
mtDNA	Mitochondrial DNA	Common abbreviation
Amp	Amplification (*as in whole genome amplification*)	Common abbreviation
F	Forward (*as in placeholder for forward primer*)	Own abbreviation
R	Reverse (*as in placeholder for reverse primer*)	Own abbreviation
Product	PCR product (*as in placeholder for restriction digest*)	Own abbreviation
C	Control (*as in mtDNA analysis*); Used as an index in published papers.	Own abbreviation
I	International (*as in mtDNA analysis*); Used as an index in published papers.	Own abbreviation
N	National (*as in mtDNA analysis*); Used as an index in published papers.	Own abbreviation
Cush	(*as in mtDNA analysis*); Used as an index in published papers.	Own abbreviation
Sem	(*as in mtDNA analysis*); Used as an index in published papers.	Own abbreviation

Section 11 – Summary of Scanned Images

Image File	Description
221106/001 - 002	Example experimental record used for the design and optimization of the conditions for PCRs to genotype samples. The routine form of record for this purpose in the notebook as in entries 3/82.
221106/003 - 004	Example experimental record showing use of kit to perform experimental work. The detailed level of specifying protocol tasks for an unfamiliar protocol compares with the summary form used for a familiar protocol.
221106/005	Example experimental records used for the design and optimization of the conditions for PCRs to genotype samples. Records show the non-interleaved execution of experimental work completed over multiple days. The routine form of record for this purpose in the notebook as in entries 6/82 and 7/82.
011206/001	Example experimental record for genotyping experiments used after PCR conditions and primer pairs have been optimized. The routine form of record for this purpose in the notebook.
041206/001	Example experimental record for genotyping experiments used after PCR conditions and primer pairs have been optimized. The routine continuation form of record in the notebook used for additional plates following on from a previous setup.
<No Scan>	Shaded entries have not been scanned due to privacy concerns.



22/11/06 4P Children Qiagen extraction.

- 4 samples taken.
  - [redacted] - Sponge + spit. [redacted] - sponge + spit
- 4 samples/kits into drying oven at 50°C at 1.30pm. out at 5.30pm

[redacted] spit = G1, sponge = G2

[redacted] spit = B1, sponge = B2

1. 500µl each added to 1.5ml eppendorf.
   - Sponge kits have very little left. - maybe 500µl?
2. 20µl purifier added and tubes gently inverted
3. Onto ice for 10mins (12noon)
4. 13K for 3 minutes
5. Remove supernatant to new 1.5ml tube + discard pellet.
6. Add 500µl 100% ETOH + mix gently by inversion (>5)
7. 10min at room temperature. (1225)
8. 13K for 1minute
9. Discard supernatant.
10. Dry off in oven at 50°C. (1245 - 1252)
11. Elute with 100µl 1xTE + incubate at 50°C for 10min
12. (10µl aliquot taken to test pre-clean PCR).

Precipitation - Standard protocol.
into -80°C at 13:10 out at 13:32
100%, 70% - dry in oven at 50° for 15min.
- elute 100µl 1xTE + incubate at 50°C for 10min 2.25pm.

Figure A5-1: Scanned page from notebook GS-L1R1-N1

This page corresponds to entry number 4 in the completed content analysis framework for this notebook. Note that two sample donor names in the laboratory record have been blotted out to preserve anonymity.

Appendix 6:
Domain-Specific Analysis Framework for Laboratory Records

This appendix contains a list of the domain-specific semantic units used to describe the composition of laboratory records during the genre analysis study of laboratory records. This list of units was used within the content analysis framework in order to ensure a consistent approach to the analysis of the notebooks used in the study.

IBLS

Surveying the Grammar of Laboratory Notebooks in the Molecular Genetics/Molecular Biology Laboratory

Common Elements of Laboratory Records

Investigator: David Wilson, Computing Science and IBLS/Molecular Genetics

Address: Institute of Biomedical and Life Sciences
Division of Molecular Genetics
Anderson College
University of Glasgow
56 Dumbarton Road
Glasgow G11 6NU
UK

Telephone: 0141 330 6220

E-mail: d.wilson.3@research.gla.ac.uk

Supervisors: Mark Bailey, IBLS/Molecular Genetics, and
Phil Gray, Computing Science

Aim Of Study

The aim of this study is to investigate the range of representations and grammars used by scientists in the molecular genetics/molecular biology laboratory to format the experimental records that are written in laboratory notebooks. The results of this study will aid in the development of computing systems to capture these experimental records in a digital manner.

In particular, this study will focus on identifying the structures, content, and notations that are used by laboratory staff to record information in laboratory notebooks. In terms of notebook structure, an example area of interest would be whether and how the entries in laboratory notebooks are indexed so that they can be search through at a later date. In terms of notebook content, an example area of interest would be the range of information that is recorded in laboratory notebooks such as experimental results, protocols, and to-do lists. In terms of notebook notation, an example area of interest would be whether shorthand notations are used whilst documenting experiments.

The study will evaluate both current and past laboratory notebooks written by multiple scientists from multiple laboratories in order to survey the range of approaches taken to documenting experimental records in laboratory notebooks.

Description Of Grammatical Elements

A contents frame is used to identify a standardized set of issues against which all laboratory notebooks will be evaluated during the course of the study. In order to facilitate comparison between the structures and notations used by different scientists in different notebooks, the following table defines a standard set of grammatical elements used to characterize the entries found in each notebook. This set of grammatical elements has been accrued over time based on the range of grammars found in the laboratory notebooks that have been surveyed during the course of the study.

Each grammatical element is described in terms of the name of the phylum used to represent that element, a textual description of the role played by the element in notebook entries, and an outline of the typical structure of the element in notebooks including optional/mandatory components.

Section 1 - Textual Elements

Element	*Advice statement*
Description	Statement recording advice/proposal received from colleagues or other external sources in relation to any aspect of experimental work such as how to adjust a protocol to improve performance or how to interpret results. The statement may contain an explicit reference to the source of the advice, the date on which the advice was received, and any conditions limiting the applicability of the advice.
Parameterization	Includes any combination of: ***conditions*** (optional), ***date*** (optional), and ***source*** (optional).
Example	***Advice statement*** ("Tam suggested using T_A of 65ºC") ***with source*** ("Tam").

Element	*Concentration calculation*
Description	Textual statement documenting a calculation to compute the dependency between volume, mass, and concentration of one or more samples or other reagents in an experiment. For example, a calculation may be used to work out how to dilute a sample to achieve a required concentration. The statement may be represented as text with standard arithmetic formulae. The calculation may include intermediate steps involved in working out the result, or may simply present the result. The result may be exact or may involve approximations.
Parameterization	Includes any combination of: ***approximation limits*** (optional), ***formulae*** (optional), ***intermediate working*** (optional), and ***samples*** (optional).
Example	***Concentration calculation*** ("20ng/µl in 100µl => v_x = 6.1µl DNA + 93.9 µl TE").

Element	*Error report*
Description	Textual statement reporting an error condition that has been identified in the execution of an experimental task. The error may be localized to specific samples being processed, to specific reagents in the reaction, or to specific steps in the task. The report may also identify the cause of the error.
Parameterization	Includes any combination of: *affected samples* (optional), *affected reagents* (optional), *affected steps* (optional), and *cause* (optional).
Example	*Error report* ("Some wells had some evaporation 1+ 12") *with affected samples* ("1 + 12").

Element	*Gel loading list*
Description	Textual presentation of an association list mapping a set of samples to the lane in which the samples have been loaded in an electrophoresis gel. A gel loading list may be use to present either a fully enumerated list of all samples and the associated lane in which they are loaded, or a description of a scheme or pattern according to which the samples are loaded in the electrophoresis gel. In contrast to fully enumerated sample/lane lists, patterns are not fully enumerated but require implicit contextual knowledge of the loading scheme. The device used to load samples onto the gel as in a multipipette may be specified as they can influence the specific lanes in which samples are loaded. The amount of sample loaded may be specified. The known DNA ladder used to calibrate the resulting gel may also be specified.
Parameterization	Includes any combination of: *amount loaded* (optional), *enumerated list* | *pattern only* (optional), *ladder* (optional), *loading device* (optional).
Example	*Gel loading list* ("Row 3 6ul 100bp – Mix 1+ 2 (multipipette)") *with pattern only, amount loaded* ("6µl"), *ladder* ("100bp"), *and loading device* ("(multipipette)").

Element	*Gel run*
Description	Textual statement identifying the conditions used to run an electrophoresis gel. The conditions may include the type of gel as in "agarose gel", the type of buffer used to run the gel as in "TBE", the time period over which the gel is run, and the electrical charge applied to the gel.
Parameterization	Includes any combination of: *buffer* (optional), *charge* (optional), *duration* (optional), and *gel type* (optional).
Example	*Gel run with buffer* ("2% TBE"), *duration* ("40 mins"), *and charge* ("100V, 60mA").

Element	*Genotyping result table*
Description	Table identifying the genotype identified for a set of samples. The results table may contain one or more samples and the associated genotype obtained for the samples. The structure of the table may conform to the arrangement of the sample set used during the execution of the experimental tasks involving in genotyping as in the 12x8 grid corresponding to 96 well plates. A group name may be used to identify the specific set of samples that have been genotyped. A date may be assigned to timestamp the results. A calculation of the relative frequencies of the different genotypes in the genotype results may also be included with the genotyping results table.
Parameterization	Includes any combination of: *date* (optional), *frequency calculation* (optional), and *sample group* (optional).
Example	*Genotyping result table with frequency calculation* ("DD = 20, II=6, ID=22").

Element	*Mix list*
Description	A tabulated, formatted list of the reagents used in preparing a mix as for a PCR or a restriction digest reaction. The mix list is typically arranged in two columns stating respectively the single reaction amounts for the mix, and the multiple reaction amounts to prepare the required amounts for the number of reaction in the current experiment. The level of detail used to specify the reagents in the mix may vary so that details such as volume, units, concentrations, and commercial source of reagents are not included. A name may be assigned to describe and identify the mix list. Progress ticks may be used to identify/monitor the addition of individual reagents to the mix.
Parameterization	Includes any combination of: *name* (optional), *progress ticks* (optional, graphic), *reagent concentrations* (optional), *reagent source* (optional), and *units* (optional).
Example	*Mix list with name* ("PCR Mix"), *reagent source* ("ABI"), *reagent concentrations* ("0.1M"), *and progress ticks*.

Element	*NanoDrop result table*
Description	Table identifying the measurements obtained from a NanoDrop® device for a list of samples. A NanoDrop is a laboratory device used to quantify the purity of protein and nucleic acid samples. The result table may contain one or more samples together with the associated measurements obtained for these samples. The NanoDrop result table may include only measurement of concentration, or may also include other fluorescence measurements indicating the quality of the sample measurement. The results list may be written by hand, or may be inserted as a printout of a results obtained from a bioinformatics software package.
Parameterization	Includes any combination of: *concentration* (optional), and *quality measures* (optional).
Example	*Nanodrop result table with concentration.*

Element	*Note*
Description	Supplementary statement used to annotate any part of an experimental record. A note may be presented as narrative text, in bullet points, or in any appropriate layout. A note may include a heading.
Parameterization	Includes any combination of: *heading* (optional).
Example	*Note* ("NB> don't use more than 100ng/µl of dig^{n} reaction").

Element	*PCR cycle*
Description	Statement describing the sequenced list of the temperatures used for a PCR, consisting of the standard five steps (initialization, denaturation, annealing, elongation, final elongation) together with the number of iterations of these temperatures. A fully-specified PCR cycle lists all temperatures together with the number of iterations. A summary form of PCR cycle may specify only the annealing temperature, which is most critical in the design of the PCR. A programme name may be used to identify and reference the PCR cycle.
Parameterization	Includes any combination of: *annealing only* | *all temperatures* (optional), and *programme name* (optional).
Example	*PCR cycle with programme name* ("Affy70").

Element	*PCR gradient*
Description	Statement describing the range of temperatures in a temperature gradient used to optimize PCRs. Alternative forms of PCR gradient include specifying the gradient as an equation summarizing the temperatures in the gradient, enumerating the full sequence of temperatures in the gradient, or a combination of these two.
Parameterization	Includes any combination of: *equation* (optional), and *temperatures* (optional).
Example	*PCR gradient with equation* ("T = 40 + 14°C").

Element	*PCR primer list*
Description	Statement describing a list of primer pairs (forward and reverse primers) used to optimize a PCR. The list may contain one or more primer pairs. Each primer pair may be assigned a name to identify the primer combination. The melting temperature for each primer may also be specified as supplementary information. The DNA sequence of the primers may be specified.
Parameterization	Includes any combination of: ***melting temperatures*** (optional), ***pair name*** (optional), and ***DNA sequence*** (optional).
Example	***PCR primer pair*** ("Mix 1 = Ppd 194F + 418R") ***with pair name*** ("Mix 1").

Element	*Plan statement*
Description	Statement identifying a list of experimental tasks that are scheduled for future execution. The list may be ordered, and may contain one or more samples. The planned tasks are typically scheduled in response to a previous experimental outcome, and so the plan statement may include the rationale for the choice of tasks in response to the previous outcome. Similarly, the planned tasks may be defined as a variation on previous experimental tasks reflecting adaptations on the basis of the previous experimental outcome.
Parameterization	Includes any combination of: ***alternatives*** (optional), ***date*** (optional), ***rationale*** (optional), and ***variation*** (optional).
Example	***Plan statement*** ("Try with α–DIG, APconjugated Ab (<X>) tomorrow") ***with date*** ("tomorrow").

Element	*Protocol task heading*
Description	Subsection heading used to identify the experimental task that has been/is being carried out as part of the execution of a laboratory procedure. The granularity used to specify protocol tasks may vary considerably. The heading may be underlined, boxed or otherwise emphasized graphically.
Parameterization	None.
Example	*Protocol task heading* ("Mix").

Element	*Protocol task report*
Description	Description of an experimental task that has been/is being carried out as part of the execution of a laboratory procedure. The granularity used to specify protocol tasks may vary considerably from very detailed specification of individual protocol steps to broad-brush statements such as "did PCR". Individual protocol steps may be numbered. Progress ticks may be used to identify/monitor the completion of individual steps in the execution of the experimental task. The task description may include reagent calculations to specify the amount, concentration, volumes, *etc.* of reagents. Explicit calculation is often used to determine multiple reaction amounts. Use of specific devices and/or types of device may be included such as "used nanodrop". Time codes may be used to denote the start and end times of certain tasks.
Parameterization	Includes any combination of: *device use* (optional), *named protocol* (optional), *numbered steps* (optional), *progress ticks* (optional, graphic), *reagent calculations* (optional), and *time codes* (optional, graphic).
Example	*Protocol task report with numbered steps* ("6. put in oven at 1300"), *device use* ("oven"), *and time codes* ("at 1300").

Element	*Purpose statement*
Description	Textual description of the aim or purpose of an experiment, which may be specified at different levels of detail. This description may vary considerably in length from a few words to multiple lines of text. The purpose statement may contain an indication of the expected outcome of the experiment. If the experiment is being undertaken as part of a sequence of experiments, the purpose statement may identify the overarching project by name.
Parameterization	Includes any combination of: *expected result* (optional), and *project* (optional).
Example	*Purpose statement* ("To characterize repeat interruptions of BC-20").

Element	*Radioactivity calculation*
Description	Textual statement documenting the steps involved in calculating radioactivity levels based on the half-life of a radioactive source. The statement may include standard arithmetic formulae. The calculation may include intermediate steps involved in working out the result, or may simply present the result. The result may be exact or may involve approximations.
Parameterization	Includes any combination of: *approximation limits* (optional), *formulae* (optional), *decay time* (optional), *half-life* (optional), *intermediate working* (optional), and *reagents* (optional).
Example	*Radioactivity calculation with decay time* ("As per today, it has passed 2.5 half-lives so it should be 25µl for 0.37mBq").

Element	*Result interpretation*
Description	Statement describing the interpretation placed on a set of associated experimental results such as an electrophoresis gel or a set of sample genotypes
Parameterization	Includes any combination of: *associated result* (optional).
Example	*Result interpretation* ("PCR has worked nicely")

Element	*Sample list*
Description	Statement identifying a list of samples to be processed during the course of an experiment. The list may be ordered, and may contain one or more samples. The sample list may include a reference to the source supplied the samples as in the name of a collaborator. Additional data such as the concentration, label, or other data may be provided along with each individual sample. A group name may be used to identify the collection of samples in the list as in "Plate A Samples – 30/09/2010". The sample list may include a prediction of the expected result of the experiment presented either as a summary or as a corresponding list indexed to the samples in the sample list.
Parameterization	Includes any combination of: *expected result* (optional), *sample label* (optional), *sample group* (optional), and *source* (optional).
Example	*Sample list with sample label and sample group* ("G1, G2, G3 of Plate C 9/11/04").
Variant Forms	Variant and extended forms of sample list make use of association list mapping samples to associated properties such as temperature, concentration, and others to form: *Sample temperature list;* *Sample concentration list.*

Element	*Storage statement*
Description	Statement identifying the storage conditions and/or location in which samples are to be kept.
Parameterization	Includes any combination of: *conditions* (optional), *location* (optional), and *time period* (optional).
Example	*Storage statement with conditions* ("RT") *and time period* ("O/N").

Element	*Survival assay result table*
Description	Table identifying the measurements obtained for a survival assay. The results table may contain one or more samples and the associated survival assay measurements obtained for the samples at specific stages over a period of time. A group name may be used to identify the specific set of samples that have been assayed. A date may be assigned to timestamp the results. A statistical calculation of the assay results may also be included with the survival assay result table.
Parameterization	Includes any combination of: *date* (optional), *statistical calculation* (optional), and *sample group* (optional).
Example	*Survival assay result table.*

Element	*Title*
Description	Descriptive title associated with an entry in a notebook, often identifying the purpose of the entry.
Parameterization	None.
Example	*Title* ("FTO genotyping").

Element	*To-do list*
Description	Statement identifying a list of tasks to be completed. The list may be ordered, and may contain one or more samples. The to-do list may include a date by which the tasks are to be completed, priorities assigned to the tasks, and resources allocated/reserved for the tasks. Progress ticks may be used to identify/monitor the completion of individual steps in the execution of the to-do list.
Parameterization	Includes any combination of: *date* (optional), *progress ticks* (optional), *resource allocation* (optional), and *task priority* (optional).
Example	*To-do list with date* ("To-do for this week").

Element	*Work allocation*
Description	Statement identifying division of labour for experimental tasks.
Parameterization	Includes any combination of: *allocated resource* (optional), and *time code* (optional).
Example	*Work allocation* ("Mary will wet chips and put in oven tomorrow") *with allocated resource* ("Mary") *and time code* ("tomorrow").

Section 2 - Graphical Elements

Element	*Digest design map*
Description	Diagram illustrating the cutting sites for restriction enzymes as part of the design of a restriction digest. The diagram may be drawn by hand, or may be inserted as a printout of a drawing package.

Element	*Gel image*
Description	Image taken of the results of gel electrophoresis. If the experiment involved radioactive labelling, the image would be an autoradiograph (x-ray) image. Alternatively, the image may be a photograph taken of the gel under ultraviolet lighting using a gel documentation system.

Element	*Kinase assay result graph*
Description	Bar chart graph illustrating the result of a protein kinase assay. The diagram may be drawn by hand, or may be inserted as a printout of a statistical package.

Element	*Plasmid design map*
Description	Diagram illustrating the design of a plasmid used to express a particular gene. The diagram may be drawn by hand, or may be inserted as a printout of a drawing package.

Element	*Primer design map*
Description	Diagram illustrating the forward and reverse primer pairs used to select specific fragments of DNA as part of a PCR. The diagram may be drawn by hand, or may be inserted as a printout of a drawing package.

Element	*Survival assay result graph*
Description	Bar chart graph illustrating the result of a survival assay. The diagram may be drawn by hand, or may be inserted as a printout of a statistical package.

Appendix 7:
Participant Information Sheet for the Reading Protocol Study

This appendix contains a copy of the information sheet used to describe the reading protocol study of laboratory records in academic molecular biology laboratory settings to prospective participants. This information sheet was produced in accordance with the terms of the ethical approval for the study.

Reading and Interpreting Molecular Biology Laboratory Records

Information Sheet

Aim Of Study

This study aims to investigate both how scientists in molecular biology laboratories read and interpret written laboratory records in different contexts of use, and how structural and linguistic variation in these laboratory records may influence the ability of different readers to make sense of the records. In particular, the study will focus on interpreting experimental records such as those documented in laboratory notebooks, and interpreting protocol statements such as those kept in protocol folders.

Previous work has highlighted personal variation in the approaches used by scientists to document their experimental work in laboratory notebooks and protocol statements. This includes variation between scientists working in the same laboratory, and variation between scientists working in different laboratories. For example, individual scientists record different types of entries in their notebooks, use a range of representations to document these records, vary the content of the entries according to different criteria, and employ a range of indexing schemes to organize the records.

Notwithstanding this variability, the records should remain meaningful to other members of the laboratory community over time in order to serve as a reliable written archive. Community in this respect encompasses scientists with both direct and indirect relationships to the original author of the written record. This includes, for example, members of the same group working alongside the original author, collaboration partners in other laboratories, future members of the same laboratory picking up from where the author has left off on a given project, and scientists from other laboratories trying to learn techniques already performed by the author. Different types of community member faced with different contexts of use may have different expectations of the structure and content required of written laboratory records.

The study will investigate the patterns of reading and interpretation that are adopted by scientists in multiple laboratories when using a sample set of laboratory records to accomplish prescribed tasks simulating different contexts of record use. The sample set of records will be selected from actual laboratory notebooks and protocol folders to encompass a range of structural and linguistic forms. Insights gained from this study may inform both the training of laboratory staff in effective recordkeeping, and the design of tools such as electronic laboratory notebooks that aim to facilitate sharing of laboratory records.

Authors and/or Readers - Your Role in the Study

In order to evaluate the impact of the different recordkeeping approaches used by individual scientists to document their experimental work, the study will make use of a sample set of authentic records produced by multiple scientists in multiple laboratories. Some of the participants in this study (*the authors*) will be the original authors of written records selected as sample records for use in testing out reading patterns.

In order to understand the means by which individual scientists make sense of records in different contexts, the study will evaluate how a range of scientists with varying levels of experience read and interpret different forms of laboratory record in the context of multiple simulation scenarios. Making sense of the records produced by others may require scientists to switch between different reading frames, to infer missing information, and to resolve ambiguities in the records. Some participants (*the readers*) will be observed reading and interpreting the sample set of written record in scenarios simulating different contexts of use.

You may choose to participate in this study as author only, reader only, or both author and reader.

Participating as an Author

What Will Happen

If you choose to participate as an author, I will select individual records from your laboratory notebooks and ask for your consent to use these selected records as sample records in reading experiments. For your information, I will describe the types of reading scenarios in which I would like to use your sample record.

Before using any of your records in reading experiments, I will make sure that the records are rendered anonymous by removing any personal information that could be used to identify the source of the record. In order to avoid blank spaces in the text of the example record, I will substitute alternative names or other data as appropriate. I will show you the modified record and ask for your explicit agreement prior to using the modified record in any reading experiments. You will, of course, be able to specify that further modifications be made in order to render the records anonymous.

After you have explicitly agreed to the modified content, I will use the sample record as part of the scenario in multiple reading experiments. Please note that the modified record may be presented to multiple readers in different reading experiments. The reading experiments will be carried out in isolation so you will not be able to attend any of the reading experiments involving other readers.

Schedule and Location for Selecting Sample Records

I will arrange a convenient date and time on which to read through some of your notebooks in order to select appropriate records for use in subsequent reading experiments. In order to minimize any potential disruption to your work, I can either

borrow your notebooks to work through them away from your laboratory site, or I can arrange to look through the laboratory notebooks at your own laboratory site. If you would prefer me to work away from your own laboratory, I will ensure that the notebooks are returned at a specified time. In either case, I anticipate that it will take about 60-90 minutes to select sample records from your notebooks.

After selecting a subset of sample records from your notebooks, I will arrange a short meeting with you to discuss the required modifications to render the records anonymous.

What Data Will Be Collected from Authors

I will collect the following types of data from authors:

1. Digital scans and copies of selected pages from your laboratory notebooks, and

2. Written notes and/or audio recordings of informal interviews conducted with you to discuss the context in which you originally constructed the record in your notebooks.

Participating as a Reader

What Will Happen

If you choose to participate as a reader, I will ask you to perform a range of reading tasks simulating different situations that might occur during your normal work in the laboratory. Each of these tasks will require you to read and interpret example laboratory records.

For example, you might be asked to read through a set of experimental records to identify the conditions used to optimize a PCR and to propose which conditions to vary next. You might be asked to describe how you arrive at a your interpretation of an electrophoresis gel. You might be asked to describe what additional information you would need in order to learn a new technique described in an experimental record.

The reading experiment will take place in a separate room reserved for this purpose. Only two people will be present in the room during the experiment. I will be present to conduct the experiment, and you will be present as the reader.

You will be given a short written description of the task to be performed for each simulation scenario on a separate sheet of paper, together with a copy of the example laboratory records to be used for the task.

During the experiment, I will ask you to 'think aloud' whilst you are reading and interpreting the example record to perform the task. Whilst you are 'thinking aloud', I will make use of a video recorder and/or audio recorder to capture how you make use of the laboratory records.

By 'thinking aloud', I mean that you say aloud everything you are thinking whilst performing the task. I don't want you to try to plan out what you wish to say; instead, I want you to talk aloud constantly from the time you start to perform the task until you are finished with the task. I will be observing you as you work through the task. If you are silent for any long period of time, I will remind you to keep talking.

If you wish, you may also make written notes whilst you perform the reading task. I will provide a notepad and pen for you to use whilst you are interpreting the laboratory records. You may also make notes or comments on your copy of the example laboratory records.

Shortly after you have completed the task to your satisfaction, I will conduct a short interview in which I will ask you to recall how you made sense of the laboratory records to solve the simulation scenario. Once again, I will make use of a video recorder and/or audio recorder to capture your comments.

Before starting on the first reading task using laboratory records, I will present you with some simple warm-up tasks so that you can get used to 'thinking aloud'. These warm-up tasks will not involve using laboratory records but will involve general tasks such as simple mental arithmetic or word puzzles.

Schedule and Location for Reading Experiments

I will arrange a convenient date and time on which to carry out the reading tasks. The reading experiments should take approximately 90 minutes to complete.

I will also arrange a convenient location at which to carry out the reading tasks. Based on your preference, I will reserve a room either at your own laboratory site or away from your own laboratory. Note that only you (*the reader*) and I (*the experimenter*) will be present in the room during the course of the experiment.

What Data Will Be Collected During Reading Experiments

I will collect the following types of data during the study in order to assess the reading patterns associated with laboratory records:

1. Video recording and/or audio recording of you 'thinking aloud' as you perform the reading tasks,

2. Video recording and/or audio recording of our interview after you have completed the reading tasks,

3. Copies of any written notes you make as you perform the reading tasks, and

4. Field notes of my own observations as you perform the reading tasks.

Data Privacy

All results will be held in strict confidence to ensure the privacy of all participants. No personal participant information will be stored with the data so that all data is rendered anonymous. Online data will be stored in a password-protected computer account; paper data will be stored in a locked office.

Rights Of The Participant

Please note that you are entitled to withdraw from this study at any time, without prejudice. Please note that you are also entitled to require that any data relating to your participation in this study be discarded. This applies whether you have participated in the study as an author or as a reader.

A feedback message will be sent to all participants by e-mail or post after the study has been completed.

Payment for Reading Experiments

Readers will be entitled to payment in the form of a £20 voucher for participating in this study. You may choose between a voucher for use at Amazon.co.uk, or Marks and Spencer plc.

No payment will be offered to those participating in the study as authors only.

Assessment Of The Work Practices Not The Participant

Please note that this study aims to understand the patterns of reading and interpretation used by scientists in molecular biology laboratories to make sense of written laboratory records and protocol statements. You are not being evaluated; instead, it is the structural and linguistic frameworks used in written laboratory records that are being assessed.

Ethical Guidelines

This study adheres to the British Psychological Society's code of conduct, and has been approved by the Department of Computing Science (DCS) Ethics Committee of the University of Glasgow. If you would like to discuss your participation in this study with someone other than the experimenter, you may contact the convenor of the DCS Ethics Committee at hcp@dcs.gla.ac.uk.

Any Questions

If you have any questions regarding this study, please feel free to contact either my project supervisor or me at:

Investigator: David Wilson
Institute of Biomedical and Life Sciences
Division of Molecular and Cellular Biology
Davidson Building
Dumbarton Way
University of Glasgow
Glasgow G12 8QQ
UK

Telephone: 0141 330 6220
E-mail: d.wilson.3@research.gla.ac.uk

Supervisor: Dr Mark Bailey
Institute of Biomedical and Life Sciences
Division of Molecular and Cellular Biology
Davidson Building
Dumbarton Way
University of Glasgow
Glasgow G12 8QQ
UK

Telephone: 0141 330 5994
E-mail: m.bailey@bio.gla.ac.uk

Appendix 8:
Additional Data for the Reading Protocol Study

This appendix contains supplementary tables for the reading protocol study investigating how scientists at work in academic molecular biology laboratories make sense of laboratory records written by others.

The appendix contains a table describing the authors who participated in the reading study (p 309), a table describing the readers who participated in the reading study (p 311), a table describing the records used by the readers to produce reading protocols (p 313), and a table of the sequence of protocols produced by each reader (p 325).

Table A8-1: Authors in the reading protocol study

Author (N=8)	*Function*	*Language Status*	*Gender*	*Description*
RS-L2A1	PD	L2	F	First postdoctoral position; Previously worked in university research laboratories in different countries; Has previously written published articles.
RS-L2A2	PD	L1	F	First postdoctoral position; Previously worked in university research laboratories in different countries, including work as a technician and a postgraduate researcher; Has previously written published articles.
RS-L3A1	PI	L2	M	Experienced researcher; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Responsible for managing collaborations with other research groups; Has previously written a number of published articles.
RS-L4A1	T	L1	F	Experienced technician; Previously worked in industrial laboratories and university research laboratories.
RS-L4A2	T	L2	F	Working towards a PhD on a part-time basis in addition to working as a technician.
RS-L5A1	T	L2	M	Experienced technician; Worked in same laboratory for a number of years.
RS-L8A1	PD	L1	M	Researcher on first postdoctoral position; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Has previously written published articles; Moved into molecular genetics from another bioscience field.

Table A8-1 (Cont'd): Authors in the reading protocol study

Author (N=8)	*Function*	*Language Status*	*Gender*	*Description*
RS-L8A2	PD	L1	M	Experienced researcher; Previously worked in multiple university research laboratories in the UK; Responsible for managing collaborations with other research groups; Responsible for directing other researchers; Has previously written a number of published articles.

Summarizing all authors participating in the reading protocol study in terms of the identifier code assigned to the author, his/her function at the time of the study, his/her native language status, gender, and a brief description of the author's experience in laboratory work.
The author is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table 5-1). This approach maintains the anonymity of both author and laboratory required under the terms of the ethical approval for the study. The author's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The author's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks English as a second language. The author's gender is indicated using F for a female participant, and M for a male participant.

Table A8-2: Readers in the reading protocol study

Reader (N=15)	*Function*	*Language Status*	*Gender*	*Description*
RS-L1R1	PG	L1	M	Experienced clinician; Approximately halfway through postgraduate research project; Has previously written published articles.
RS-L1R2	PG	L2	F	Working in a short-term research position whilst applying for PhD studentships.
RS-L1R3	T	L1	M	Experienced technician; Worked in same laboratory for a number of years.
RS-L2R1 *(Author)*	PD	L2	F	First postdoctoral position; Previously worked in university research laboratories in different countries; Has previously written published articles.
RS-L2R2 *(Author)*	PD	L1	F	First postdoctoral position; Previously worked in university research laboratories in different countries, including work as a technician and a postgraduate researcher; Has previously written published articles.
RS-L2R3	T	L2	F	Experienced technician; Worked in multiple university research laboratories including bioscience fields other than molecular biology.
RS-L3R1 *(Author)*	PI	L2	M	Experienced researcher; Previously worked in university research laboratories in different countries; Responsible for directing other researchers; Responsible for managing collaborations with other research groups; Has previously written a number of published articles.
RS-L3R2	T	L1	F	Recently appointed technician in a new research group.
RS-L4R1 *(Author)*	T	L1	F	Experienced technician; Previously worked in industrial laboratories and university research laboratories.
RS-L4R2 *(Author)*	T	L2	F	Working towards a PhD on a part-time basis in addition to working as a technician; Has previously co-authored published articles.

Table A8-2 (Cont'd): Readers in the reading protocol study

Participant	*Function*	*Language Status*	*Gender*	*Description*
RS-L5R1	PG	L1	F	Approximately halfway through a postgraduate research project.
RS-L5R2	T	L1	M	Experienced technician; Working towards a PhD on a part-time basis in addition to working as a technician; Has previously written published articles.
RS-L6R1	T	L1	M	Experienced technician and postdoctoral researcher; Previously worked in multiple university research laboratories in the UK; Has previously written published articles.
RS-L7R1	PI	L1	F	Experienced researcher; Previously worked in multiple university research laboratories in the UK; Responsible for managing industrial collaborations; Responsible for directing other researchers; Has previously written a number of published articles.
RS-L8R1	PI	L1	M	Highly experienced, eminent researcher; Previously worked and directed in a number of university research laboratories in the UK; Responsible for managing industrial collaborations; Responsible for directing other researchers; Has previously written a number of published articles.

Summarizing all readers participating in the reading protocol study in terms of the identifier code assigned to the reader, whether the reader also participated as an author, his/her function at the time of the study, his/her native language status, gender, and a brief description of the reader's experience in laboratory work.
The reader is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table 5-1). This approach maintains the anonymity of both reader and laboratory required under the terms of the ethical approval for the study. The reader's function is indicated using PI for a principal investigator/head of laboratory, PD for a postdoctoral researcher, PG for a postgraduate research student, and T for a laboratory technician. The reader's native language is indicated using L1 for a participant whose first language is English, and L2 for a participant who speaks English as a second language. The reader's gender is indicated using F for a female participant, and M for a male participant.

Table A8-3: Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
1L2A1	*RS-L2A1*	Research laboratory in a South American university.	Written in Spanish.
	EMSA gel.	Single entry on a single A4 page; Non-interleaved; Written over two days.	Annotated gel image.
	Mix list; Gel loading list; Gel run.	None.	Specification level set by a template table; Table entries filled in but no narrative text.
2L2A1	*RS-L2A1*	*RS-L2*	Written in English.
	DNA sequencing of samples involving RP-PCR, gel electrophoresis, and Southern blotting.	Single entry over three A4 pages; Non-interleaved; Written over three days.	Annotated gel image; Pre-printed table for gel loading scheme.
	Title; Purpose statement; Sample list; Concentration calculation; Mix list; PCR cycle; Gel loading list; Gel run; Protocol task report.	Implicit cross-reference to previous work for one task ("hybridized as usual").	Detailed specification as in time and current for gel, and product serial number; Some tasks reported as narrative text; Subheadings/space to emphasize layout.
1L2A2	*RS-L2A2*	*RS-L2*	Written in English.
	DNA sequencing of samples involving RP-PCR, gel electrophoresis, and gel extraction.	Single entry over two A4 pages; Non-interleaved; Written over one day.	Annotated gel image; NanoDrop results.
	Title; Purpose statement; Sample list; Mix lists; PCR cycle; Gel run; Protocol task reports; Results interpretation; Plan statement.	Cross-reference to kit ("QIA gel extraction kit plus protocol"); Cross-reference to PCR program.	Partial specification for some tasks as in no reagent concentrations, and only annealing temperature in PCR; Detailed narrative for purpose, and plan to proceed; Limited signposting in layout.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
2L2A2	*RS-L2A2*	Research laboratory in a North American university, whilst working as PG.	Written in English.
	SNP sequencing of samples involving PCR, and gel electrophoresis; Restriction digests for two sets of samples.	Three entries on three A4 pages, one page per entry; Non-interleaved; Written over one day.	Annotated gel images.
	Title; Mix list; PCR cycle; Protocol task report.	Implicit cross-reference to communal laboratory schemes for sample naming, target gene naming, and experimental results database.	Entries written in line with communal laboratory schemes Minimal narrative; Purpose encoded in title; Reliance on laboratory codes; Partial PCR thermal cycle; No gel image for results only test gel images; No results statement.
1L3A1	*RS-L3A1*	*RS-L3*	Written in English.
	DNA sequencing of samples involving PCR, gel electrophoresis, and gel extraction; Immunohistochemistry of different tissue samples.	Six entries over four A4 pages; Interleaved; Written over two days.	Annotated gel images; Stained tissue images.
	Titles; Purpose statement; Mix lists; PCR cycle; Gel loading list; Gel run; Protocol task report; Results interpretation.	Cross-reference to kit ("Qiagen gel extraction kit"); Cross-reference to PCR program; Cross-reference between interleaved entries by page number and description ("PCR p 169").	Purpose specified in title; Narrative text focused on variation to 'standard' protocol; Partial specification for some tasks as in no reagent concentrations, and only annealing temperature in PCR; Detailed analysis of all result images; Lines and headings to separate interleaved entries.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
2L3A1	*RS-L3A1*	Research laboratory in a UK university.	Written in English.
	DNA sequencing of samples involving PCR, gel electrophoresis, and gel extraction; Preparation of tissue samples.	Four entries over three A4 pages; Interleaved; Written over one day.	Annotated gel images; Tissue images.
	Title; Purpose statement; Mix list; PCR cycle; Gel loading list; Gel run; Protocol task report; Results interpretation.	Cross-reference between interleaved entries by date and description ("Sequencing PCR products 21/6/04").	Purpose specified in title; Narrative text focused on variation to 'standard' protocol; Partial specification for some tasks as in no reagent concentrations, and only annealing temperature in PCR; Detailed analysis of all result images; Lines and headings to separate interleaved entries.
1L4A1	*RS-L4A1*	*RS-L4*	Written in English.
	Preparation of samples for a microarray experiment using a commercial kit.	Single entry over three A4 pages; Non-interleaved; Written over one day.	None.
	Title; Purpose statement; Mix list; PCR cycle; Protocol task report; NanoDrop result list; Results interpretation.	Cross-reference to kit ("Cleaned with MiniElute columns (Qiagen)"); Cross-reference to sample in previous entry by date ("RNA called B 16.05.07").	Title lists client order; NanoDrop results copied rather than inserted as printout; Subheadings used to structure the entry; Detailed description of experimental procedure; Conformance to steps in kit protocol.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
2L4A1	*RS-L4A1*	*RS-L4*	Written in English.
	Test of commercial kits for *in vitro* transcription.	Single entry over three A4 pages; Non-interleaved; Written over one day.	Annotated gel electrophoresis analysis from software tool.
	Title; Purpose statement; Mix list; PCR cycle; Protocol task report; NanoDrop result list; Results interpretation.	Cross-reference to kit; Cross-reference to PCR program ("Affy70").	Detailed explanation of purpose; NanoDrop results copied rather than inserted as printout; Subheadings used to structure the entry; Detailed description of experimental procedure; Kit product expiry and serial numbers noted; Conformance to steps in kit protocol.
1L4A2	*RS-L4A2*	*RS-L4*	Written in English.
	Preparation of samples for a microarray experiment using a commercial kit.	Single entry over six A4 pages; Non-interleaved; Written over three days.	Annotated NanoDrop results.
	Title; Purpose statement; Storage statement; Mix list; PCR cycle; Protocol task report; NanoDrop result list; Results interpretation; Plan statement.	Cross-reference to kit ("Affy cleanup module"); Cross-reference to PCR program ("Affy16").	Title lists client order; Detailed specification level as in concentrations for samples and reagents, and timings for tasks; Subheadings used to structure the entry; Detailed description of experimental procedure conforming to steps in kit protocol; Fully specified names – no acronyms; Written analysis of results and how to proceed with customer.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
2L4A2	*RS-L4A2*	*RS-L4*	Written in English.
	Preparation of samples for a microarray experiment using a commercial kit.	Single entry over seven A5 pages; Non-interleaved; Written over three days.	None.
	Title; Purpose statement; Storage statement; Mix list; PCR cycle; Protocol task report; NanoDrop result list; Results interpretation; Plan statement.	Cross-reference to device by company name ("Agilent"); Cross-reference to PCR program ("Affy70"); Cross-reference to file used to store digital results.	Title lists client order; Detailed specification level as in concentrations for samples and reagents, and timings for tasks; Subheadings used to structure the entry; Detailed description of experimental procedure conforming to steps in kit protocol; Fully specified names – no acronyms; Written analysis of results and how to proceed with customer.
1L5A1	*RS-L5A1*	*RS-L5*	Written in English.
	Protein kinase assay for samples taken from different strains of a model organism.	Three entries written over two, one, and two A4 pages respectively; Non-interleaved; Written on separate days over the course of a month.	Annotated kinase assay results graphs.
	Titles; Sample list; Mix list; Radioactivity calculation; Protocol task report.	Cross-reference to previous entries by date and description ("08/03/08 – Sample Details").	Title offers limited purpose statement; Limited specification of experimental procedure; Detailed reactivity calculation; Explicit linkage between entries; Assay results shown in a graph but no written analysis of results is given.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
2L5A1	*RS-L5A1*	*RS-L5*	Written in English.
	PCR and restriction digest for samples.	Single entry written over two A4 pages; Non-interleaved; Written on one day.	Annotated gel image.
	Title; Purpose statement; Sample list; Mix list; Protocol task report.	Cross-reference to kit by name ("Qiagen PCRpurification kit").	Detailed purpose statement with diagram for vector insert; Detailed specification of experimental procedure confirming to steps in kit protocol; No results analysis.
3L5A1	*RS-L5A1*	*RS-L5*	Written in English.
	Survival immunity assay against *Escherichia coli* stabbing for different strains of a model organism.	Three entries written over one A4 page; Non-interleaved; Written over three days.	None.
	Title; Survival assay results table.	None.	No purpose statement beyond title, No narrative text describing experimental procedure; Filled in hand-drawn tables used to record assay data over time; No results analysis.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
4L5A1	*RS-L5A1*	*RS-L5*	Written in English.
	HEK cell transfromation using different conditions.	Two entries written over two and one A4 page respectively; Non-interleaved; Written on separate days over the course of one week.	None.
	Title; Purpose statement; Mix list; Protocol task report.	None.	Detailed purpose statement; Subheadings used to structure the entry, Detailed description of experimental procedure in narrative text including the equipment to be used; No written analysis of results.
5L5A1	*RS-L5A1*	*RS-L5*	Written in English.
	Colony PCR.	Single entry written on one A4 page; Non-interleaved; Written on one day.	Annotated gel image.
	Mix list.	None.	A minimal record; No purpose statement nor title, No narrative text describing experimental procedure; Use of abbreviations for reagents; No written analysis of results.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
1L8A1	*RS-L8A1*	*RS-L8*	Written in English.
	Optimization of annealing temperature and primers for PCR.	Single entry written on two A4 pages; Non-interleaved; Written on one day.	Annotated gel image.
	Title; Mix list; PCR gradient; Error report; Gel run.	None.	Purpose encoded in title; Limited narrative text describing experimental procedure; Problem with evaporation in wells noted; Temperature gradient specified by equation and itemized; Gel loading is summarized not itemized; Subheadings used to emphasize layout; No written analysis of results.
2L8A1	*RS-L8A1*	*RS-L8*	Written in English.
	Test of DNA extraction from spit samples using a commercial kit.	Single entry written on two A4 pages; Non-interleaved; Written over two days.	None.
	Title; Purpose statement; Sample list; Mix list; Protocol task report; Results interpretation.	Cross-reference to commercial kit ("Genotek").	Detailed statement of purpose; Detailed description of experimental procedure; Specification level conforms to kit protocol; No gel image but written statement of results.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
3L8A1	*RS-L8A1*	*RS-L8*	Written in English.
	DNA sequencing of sample sets involving PCR, and gel electrophoresis.	Two consecutive entries written on one A4 page each; Non-interleaved; Written over four days.	Annotated gel images.
	Title; Sample list; Mix list.	None.	No purpose statement beyond title; No narrative text; Samples identified by cross-reference to a plate; Repetition of a routine task; No analysis of results.
4L8A1	*RS-L8A1*	*RS-L8*	Written in English.
	Optimization of volumes for PCR.	Single entry written on one A4 page; Non-interleaved; Written on one day.	Annotated gel image.
	Title; Purpose statement; Sample list; Mix list; Gel loading list; Gel run.	Cross-reference to sample in previous experiment by name and date.	Narrative text for purpose statement; Set of volumes to be tested are itemized; Gel loading is summarized not itemized; Subheadings used to emphasize layout; No written analysis of results.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
1L8A2	*RS-L8A2*	Research laboratory in a UK university, whilst working as PG.	Written in English.
	Colony PCR.	Single entry written on two A4 pages; Non-interleaved; Written on one day.	Annotated gel image.
	Title; Purpose statement; Sample list; Mix list; Protocol task report; Gel loading list; Gel run; Results interpretation; Plan statement.	Cross-reference to PCR program by name ("Programme 7").	Detailed purpose statement; Detailed experimental procedure; Gel running conditions include buffer, time, and current; Gel loading conditions are itemized; Subheadings used to emphasize layout; Separate stages of the experiment are numbered for subsequent cross-referencing; Written analysis of results.
2L8A2	*RS-L8A2*	Research laboratory in a UK university, whilst working as PD.	Written in English.
	Optimization of annealing temperature and primers for PCR.	Two entries written on three A4 pages; Interleaved; Written on one day.	Annotated gel image.
	Title; Purpose statement; Sample list; Mix list; Protocol task reports; Gel loading list; Gel run; Results interpretation; Plan statement.	Cross-reference to PCR program by name ("Programme 7").	Detailed purpose statement; Detailed experimental procedure; Gel running conditions include buffer, time, and current; Gel loading conditions are itemized; Subheadings used to emphasize layout; Separate stages of the experiment are numbered for subsequent cross-referencing; Written analysis of results.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
3L8A2	*RS-L8A2*	Research laboratory in a UK university, whilst working as PD.	Written in English.
	Optimization of annealing temperature and primers for PCR; Report of meeting with sales representative.	Six entries written on six A4 pages; Interleaved; Written on one day.	Annotated gel images.
	Title; Purpose statement; Sample list; Mix list; PCR cycle; Protocol task report; Gel loading list; Gel run; Results interpretation; Plan statement.	Cross-references to previous experimental work by date, name, and entry number.	Detailed purpose statement; Detailed experimental procedure; Gel running conditions include buffer, time, and current; Gel loading conditions are itemized; Subheadings used to emphasize layout; Separate stages of the experiment are numbered for subsequent cross-referencing; Written analysis of results.

Table A8-3 (Cont'd): Sample records for the reading protocol study

Record (N=23)	*Record Characteristics*		
	Author	**Laboratory**	**Language**
	Experiment Type	**Format**	**Insertions**
	Structure	**Cross-references**	**Text Style**
4L8A2	*RS-L8A2*	Research laboratory in a UK university, whilst working as PD.	Written in English.
	Optimization of annealing temperature and primers for PCR.	Four entries written on five A4 pages; Interleaved; Written on one day.	Annotated gel images.
	Title; Purpose statement; Sample list; Mix list; PCR cycle; Protocol task report; Gel loading list; Gel run; Results interpretation; Plan statement.	Cross-references to previous experimental work by date, name, and entry number as in ("Repeat PCR (4) 13/2/03").	Detailed purpose statement; Detailed experimental procedure; Gel running conditions include buffer, time, and current; Gel loading conditions are itemized; Subheadings used to emphasize layout; Separate stages of the experiment are numbered for subsequent cross-referencing; Written analysis of results.

Summarizing all laboratory records used in the reading protocol study in terms of context, container, and content of the records.
A unique identifier code is assigned to each record. The context of the record is described in terms of the author who produced the record, the laboratory in which the record was produced, and the type of experimental work being recorded. The author of the record is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table A8-1). The laboratory in which the record was originally produced is identified either using the unique laboratory identifier (see Table 5-1) if it was produced in a laboratory that participated in this study, or by a description if it was produced in any other laboratory. The record container is described in terms of the physical format of the record, and the use of any inserted images or device printouts to form compound records. The record content is described in terms of the language used to write the record, the structural elements (see Appendix 6), the use of cross-referencing schemes, and a description of the textual style.

Table A8-4: Sequencing of reading protocols

Reader	*Sequence of Records in Reading Protocols*						*Protocols (N = 61)*
	1st	**2nd**	**3rd**	**4th**	**5th**	**6th**	
RS-L1R1	4L5A1	1L2A2	1L3A1	2L2A1	1L2A1	-	***5 8.2%***
RS-L1R2	2L2A1	1L4A2	4L8A1	2L2A2	1L2A1	3L5A1	***6 9.8%***
RS-L1R3	2L3A1	4L8A1	2L4A2	1L8A2	-	-	***4 6.6%***
RS-L2R1	2L5A1	2L4A1	2L8A2	1L5A1	-	-	***4 6.6%***
RS-L2R2	4L5A1	2L4A1	4L8A2	5L5A1	-	-	***4 6.6%***
RS-L2R3	4L5A1	1L4A2	3L8A1	3L8A2	-	-	***4 6.6%***
RS-L3R1	2L4A2	2L5A1	2L8A2	1L5A1	-	-	***4 6.6%***
RS-L3R2	1L8A2	2L4A1	3L8A1	1L5A1	3L5A1	1L4A1	***6 9.8%***
RS-L4R1	2L3A1	4L8A2	2L4A2	1L8A1	-	-	***4 6.6%***
RS-L4R2	2L8A1	3L8A2	4L5A1	-	-	-	***3 4.9%***
RS-L5R1	1L8A1	1L4A1	3L5A1	1L8A2	-	-	***4 6.6%***
RS-L5R2	1L3A1	2L2A2	2L4A2	-	-	-	***3 4.9%***
RS-L6R1	1L4A2	2L2A2	3L5A1	-	-	-	***3 4.9%***
RS-L7R1	2L8A1	2L3A1	1L4A1	1L2A1	-	-	***4 6.6%***
RS-L8R1	1L3A1	1L2A2	1L5A1	-	-	-	***3 4.9%***

Summarizing the reading protocols collected during the reading study in terms of the reader, and the sequence of records read by each reader. The reader is identified relative to his/her laboratory using the identifier code that uniquely identifies each laboratory participating in the study (see Table A8-2). This approach maintains the anonymity of both reader and laboratory required under the terms of the ethical approval for the study. The records are identified using the assigned identifier code (see Table A8-3).

Appendix 9:
Reading Protocol Script and Example Protocol Data

This appendix contains a copy of the script (p 327) used by the investigator during all reading protocols in order to ensure that consistent instructions and training examples were given to each laboratory member who participated in the study as a reader.

This appendix also contains examples of the two data items provided to a reader for each laboratory record used in the reading study. The first item was a cover sheet containing a brief statement about the laboratory in which the record was produced together with a statement of the questions to be addressed during the reading protocol (p 331). For all records used in the study, the description of the laboratory of origin contained the broad area of interest, and whether the laboratory was research-orientated or service-orientated. The same three questions were posed in the reading protocols for all records. The second item was a colour print of a scanned copy of the laboratory record itself (p 332). In the case of multi-page entries, the printed pages given to each reader were bound and stapled to preserve the same physical layout and *recto-verso* order as in the original notebook. Similarly, pages originally written in A5 notebooks were presented as A5 pages, whilst pages originally written in A4 notebooks were presented as A4 pages.

Reading and Interpreting Molecular Biology Laboratory Records

Reading Protocol Introduction Script

Introduction

In this experiment, I am interested in what you think about when you try to make sense of sample laboratory records taken from other scientist's notebooks in order to perform specific tasks.

In order to do this, I am going to ask you to **'think aloud'** as you read through a set of sample laboratory records. What I mean by 'think aloud' is that I would like you to tell me everything you are thinking from the time you first see the sample record until you have completed the task. I would like you to talk **constantly** from the time I present each task until you have given your final answer. Please don't plan what you say, and don't try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself. It is important for the experiment that you keeping talking aloud. So, if you keep silent for any period of time, I will remind you to keep talking aloud.

Is it clear what I would like you to do?

Recording Options

I would like to use a video recorder and an audio voice recorder to record your 'thinking aloud'.

Are you happy for me to use the video recorder?

Are you happy for me to use the voice recorder?

Script for Practice Problems

To get you used to 'thinking aloud', we can practice with some warm up problems.

Task 1 - Adding numbers

1. First, I would like you to add these two numbers in your head and tell me what you are thinking as you come to an answer.

What is the result of adding 129 and 342? (ANSWER: 471)

2. Now that you have worked it out, I would like to see how much you remember about your thinking from the time you heard the question until you came to an answer. I am interested in what you can actually remember rather than what you think you must have done. If possible, I would like you to tell me what you remember in the same order you worked on the problem. Please say if you are unclear about any of your memories. I don't want you to solve the problem again, please just report everything you can remember about how you solved the task.

Please tell me what you can remember.

Any questions so far?

Task 2 – Route planning (optional)

Now, here is another practice problem.

What is the quickest route to get from this building to the nearest underground station?

Please tell me what you can remember about your thinking.

Task 3 - Listing countries

Here is a final practice problem. Don't worry about keeping count – I will keep count for you.

Name 20 countries in the world.

Now, please tell me what you remember about your thinking.

Reading Laboratory Records

Any questions? Are you clear about what I would like you to do to 'think aloud'.

If so, we can begin with the laboratory records.

Feel free to use the notepad, etc whilst coming to an answer, but please keep 'thinking aloud'.

Here is the first laboratory record.

Here is the next laboratory record ….

Questions for Interpreting Laboratory Records

The following questions appear on the handout for each record.

Identify the purpose of the experiment(s) documented in this record.

Describe the materials and methods used to perform the experiment (s).

What would you need to be able to repeat the experiment.

Voucher for Reading Experiments

Readers are entitled to payment in the form of a £20 voucher for participating in this study. You may choose between a voucher for use at Amazon.co.uk, or Marks and Spencer plc. Which type of voucher would you prefer?

Assessment Of The Work Practices Not The Participant

Please note that this study aims to understand the patterns of reading and interpretation used by scientists in molecular biology laboratories to make sense of written laboratory records and protocol statements. You are not being evaluated; instead, it is the linguistic frameworks used in the written laboratory records that are being assessed.

Ethical Guidelines

This study adheres to the British Psychological Society's code of conduct, and has been approved by the Department of Computing Science (DCS) Ethics Committee of the University of Glasgow. If you would like to discuss your participation in this study with someone other than the experimenter, you may contact the convenor of the DCS Ethics Committee at hcp@dcs.gla.ac.uk.

Do you have my contact details, and those of my supervisor?

Reading and Interpreting Molecular Biology Laboratory Records

Reading Protocol Video

Date:

Participant:

Planned Scenarios:

Approximate Times:

Record	Concurent Protocol	Retrospective Protocol

Reading and Interpreting Molecular Biology Laboratory Records

Sample Record 5L5A1

Background

This laboratory record was written in the notebook of a scientist working in a research laboratory studying integrative physiology using the fruit fly as a model organism.

Task

Please 'think aloud' as you answer the following questions

.

Identify the purpose of the experiment documented in this record.

Describe the materials and methods used to perform the experiment.

What would you need to be able to repeat this experiment?

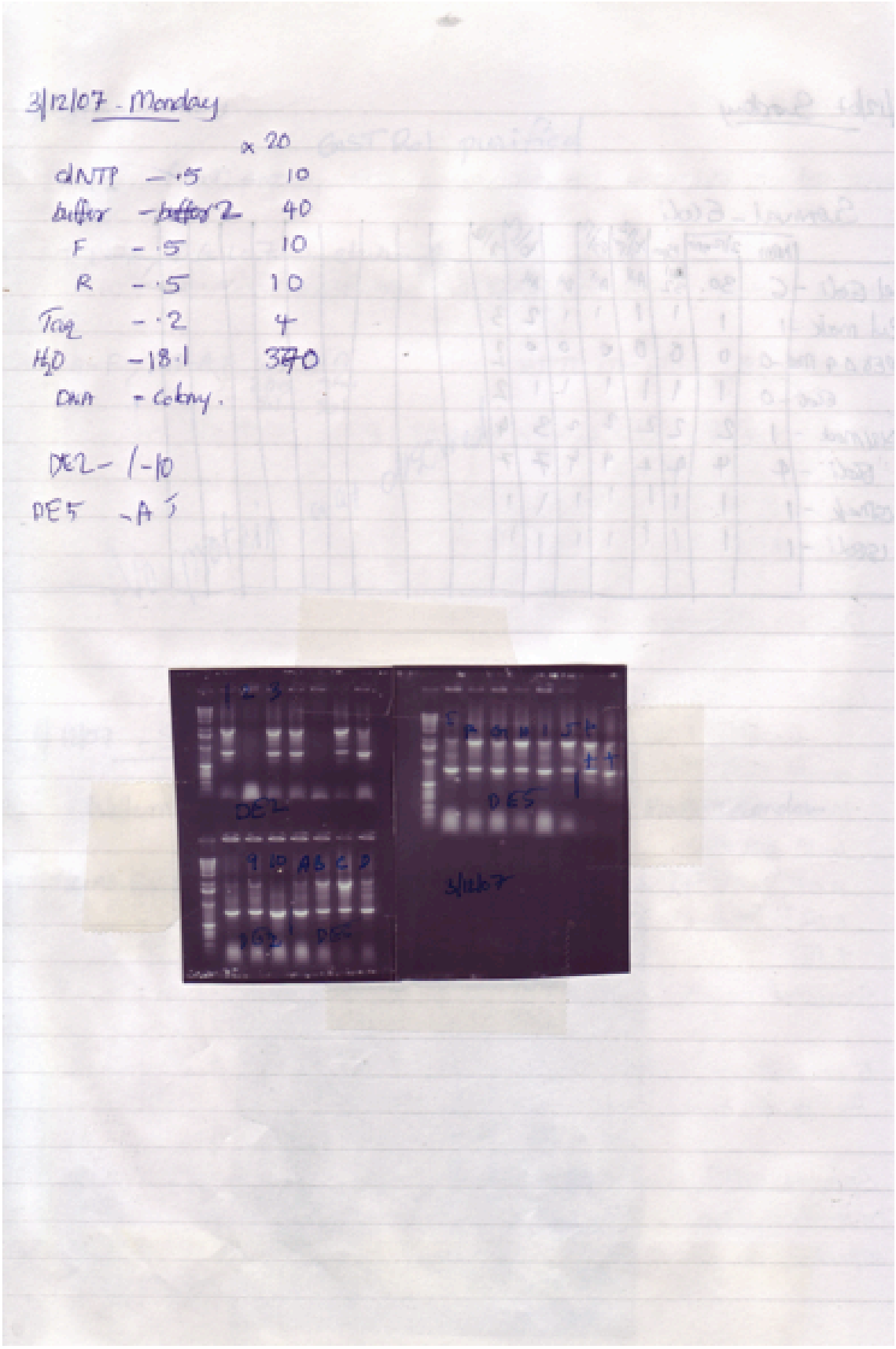


Figure A9-1: Example digital scan of sample record *5L5A1* as used in the reading protocols

Note that whilst this record is a single page record demonstrating a minimal style of recordkeeping, the sample set of records selected for the study encompassed both multi-page records, and records demonstrating styles of recordkeeping that provided varying levels of detail.

Appendix 10:
Transcription Conventions

This appendix lists the basic transcription conventions used in the presentation of speech taken from audio and video recordings for this study of laboratory recordkeeping in academic molecular biology laboratories.

Table A10-1: Transcription conventions

Notation	*Description*	*Meaning*
(( ))	Text in double round brackets	Indicates a note by the transcriber.
(.)	Full stop in round brackets	Indicates a short pause.
tex-	Text followed by a dash	Indicates a word that was cut off.
text?	Text followed by a question mark	Indicates the usual intonation associated with a question.
text	Underlined text	Indicates that the underlined words are stressed by the speaker.
( )	Empty round brackets	Indicates that the transcriber could not determine what was being said.
(text)	Text in round brackets	Indicates text of which the transcriber is unsure.

Summarizing the subset of the Jeffersonian speech transcription conventions as defined in (Jefferson 2004) that are used in all excerpts from audio recordings and video recordings presented in this study of laboratory recordkeeping.

Glossary

allele:

Different individuals may have variant forms of the DNA sequence of a given gene. Each of these possible variant forms of the DNA sequence of a gene is termed an allele. The frequency of these alleles varies across different populations of individuals. The variation between the DNA sequence in different alleles may involve only a single nucleotide or may be more extensive. Humans and other diploid organisms have two chromosomes and therefore two copies of each gene, one per chromosome. Individuals that have the same allele for both copies of the gene are said to be homozygous for that gene. Individuals that have different alleles are said to be heterozygous for that gene. Part of the significance of allelic variation is that different combinations of alleles in individuals may give rise to different observable traits between those individuals.

anaphoric reference:

A linguistics concept that defines a particular type of textual reference based on the relative position of the reference and the referent within a text. Anaphoric references are words or phrases that enable a link to be established across a text where the reference points backwards to a previous part of the text. Anaphoric referencing stands in contrast to cataphoric referencing.

annealing temperature:

abbreviated as **A°C**, or $\mathbf{T_A}$;

Part of the thermal cycle used in a PCR. PCR employs repeated cycles of heating and cooling of DNA fragments in reaction with other chemicals in order to copy and amplify strands of a target DNA sequence of interest. The reaction proceeds through three important steps of denaturation, annealing, and extension in order to melt the DNA and then enzymatically replicate the target DNA sequence. The annealing temperature is the temperature set for the annealing step in the PCR. During this step, recombination of DNA strands that were previously separated through denaturation takes place in the presence of DNA primers in order to amplify the target DNA.

computer-aided qualitative data analysis (CAQDAS):

CAQDAS software systems are tools that can be used by qualitative researchers working across multiple fields of inquiry to facilitate the tasks performed during the process of qualitative data analysis. In particular, these tools offer on a range of services such as tailored user interfaces, databases, and hypermedia linking to support tasks such as collating source data, transcribing audio and video source data, maintaining lists of codes, maintaining mappings between codes and source data, and relating coded source data to theory.

cataphoric reference:

A linguistics concept that defines a particular type of textual reference based on the relative position of the reference and the referent within a text. Cataphoric references are words or phrases that enable a link to be established across a text where the reference points forwards to a later part of the text. Cataphoric referencing stands in contrast to anaphoric referencing.

cell:

The basic functional and structural unit of all living organisms. Humans and higher order organisms are known as eukaryotes, and are composed of cells containing a nucleus, an enclosing cell membrane, and a cytoplasmic region housing organelles such as mitochondria and ribosomes that perform specialized functions within the cell. Bacteria and other simpler organisms, known as prokaryotes, are single-celled organisms. Prokaryotic cells differ in that they do not contain a true, membrane-delimited nucleus but instead contain a nucleoid.

cell culture:

A laboratory technique in which cells are grown and maintained in artificial growth media. The environmental conditions and the composition of the growth medium to be used in cell culture are dependent on the cell type being cultivated.

cell lysis:

A laboratory technique in which cells are disintegrated in a controlled manner in order to gain access to the cell contents such as proteins or DNA without degrading those contents. The product of cell lysis is referred to as a lysate.

centrifuge:

A laboratory device that operates through the sedimentation principle to isolate and separate substances of different density within a tube by spinning the tubes at high speeds. Routinely used within molecular biology laboratories, for example, to isolate DNA pellets from liquid suspensions produced during extraction protocols.

colony PCR:

A laboratory technique that uses a variant of PCR to screen bacterial colonies in order to determine whether the vector containing a DNA sequence of interest has been inserted into the colony. Growing bacterial colonies on suitable culture plates is a basic laboratory technique for producing a population of cells. In molecular biology laboratories, this approach is commonly used to grow populations of cells that have been transfected with a vector containing a DNA sequence of interest.

confocal microscope:

A laboratory device that provides microscopic imaging with the advantage of improved imaging of fluorescently labelled specimens, and the ability to generate a 3-D sequence of optical sections from thick specimens. Often used within molecular biology laboratories to visualize *in vivo* cells or tissues that have been labelled with fluorescent probes.

cyclic adenosine monophosphate (cAMP):

cAMP is a molecule that is known to play an important role in the intracellular signal transduction that regulates a range of functions affecting multiple organ systems. Signal transduction is the mechanism through which signals to cells are converted into specific responses. Deficiencies in cAMP signalling are characteristic of several human diseases including

cardiovascular disease, renal disease, cancer, diabetes mellitus, and neurological disorders.

DNA sequencing:

The process of identifying the structure of a strand of DNA in terms of the ordered sequence of the possible bases of which DNA can be composed.

deoxynucleoside triphosphate (dNTP):

A nucleotide molecule that is composed of the sugar deoxyribose, a base, and a group of three phosphates. Example dNTP molecules are doxyguanosine triphosphate (dGTP), deoxycytidine triphosphate (dCTP), deoxyadenosine triphosphate (dATP),and dTTP (deoxythymidine triphosphate) containing respectively each of the four possible bases (G, C, A, T) found in DNA molecules. dNTPs play an important role in laboratory techniques such as PCR by acting as the building blocks from which DNA strands are synthesized. Ready-to-use products containing balanced mixes of the dNTPs used in common laboratory techniques such as PCR are available from commercial vendors.

deoxyribonucleic acid (DNA):

DNA is a polymer composed of a sequence of nucleotide units. Each of the nucleotide units in DNA molecules contains the sugar dexoyribose and one of four possible bases, namely guanine (G), cytosine (C), adenine (A), and thymine (T). The genetic information carried in DNA is encoded as the ordered sequence of these bases corresponding to the ordered sequence of nucleotides in the DNA polymer. This gives rise to the familiar alphabet soup used to denote DNA sequences as in the example[138] $(GGC)_3G(CCG)_{20}(CCGCTG)_{14}(CTG)_{35}$. DNA molecules usually exist in a double-stranded, anti-parallel form in which the four possible bases occur in two complementary pairs. In this way, nucleotides containing guanine (G) on one strand will be paired with nucleotides containing cytosine (C) on the other stand, whilst nucleotides containing adenine (A) on one strand will be paired with nucleotides containing thymine (T) on the other stand. Note

[138] Thanks are due to Dr Claudia Braida, a friend and colleague, for providing this example of a DNA sequence taken from her recent work characterizing DNA mutations associated with myotonic dystrophy type 1.

that the convention is to specify one only of the two strands of the DNA molecule on the understanding that the complementary, anti-parallel strand reads in reverse.

Drosophila melanogaster:

A species of fly also known as the common fruit fly. *Drosophila melanogaster* is widely used as model organism in biological research, and the genome for this species was published in the journal Science in 2000. The reasons behind the use of *Drosophila melanogaster* as a model organism include the ease and limited costs involved in looking after the flies, the short generation time of the flies avoids delays in experimentation, the reduced requirements for ethical approval inherent in working with flies, and the degree of homology between flies and humans.

electronic laboratory notebook (ELN):

A computer system to support the digital capture and curation of the laboratory records and other data kept by scientists. This includes scientists at work in the biosciences and in other disciplines such as physics.

electrophoretic mobility shift assay (EMSA):

A laboratory technique, also known as a gel shift assay, that uses a variation on gel electrophoresis to assess and visualize whether a specific protein is capable of binding to a DNA or RNA sequence.

endophoric reference:

A linguistics concept that defines a particular type of textual reference based on the relative scope of the reference and the referent within a text. Endophoric references are words or phrases that enable a link to be established in a text where the thing being referred to (the referent) is contained inside the same text. Endophoric referencing stands in contrast to exophoric referencing.

English for Academic Purposes (EAP):

A branch of ESP that is concerned with investigating and teaching the specific form of language use that is appropriate to, and characteristic of,

academic settings. To date, this has focused primarily, but not exclusively, on language use in tertiary education.

English for Occupational Purposes (EOP):

A branch of ESP that is concerned with investigating and teaching the specific form of language use that is appropriate to, and characteristic of, workplace settings. Developments within EOP have focused on specific occupations including air traffic control, accountancy, and healthcare provision.

English for Specific Purposes (ESP):

A movement within English language pedagogy that is concerned with investigating and teaching the specific form of language use that is appropriate to, and characteristic of, specific settings such as workplace settings or academic settings.

Epstein-Barr virus:

abbreviated as **EBV**;

A member of the herpesvirus family, EBV is one of the most common human viruses worldwide. The virus is largely asymptomatic in infants, but can often cause glandular fever (infectious mononucleosis) in adolescents. In immunocompromised individuals, EBV may play a role in causing several types of cancer such as Burkitt's lymphoma and nasopharyngeal carcinoma.

e-Science:

A UK national research programme that aims to facilitate scientific research through the appropriate use of information technology. The goal is to support large-scale collaborative scientific research through the provision of a technological infrastructure that supports large data collections, high performance computing resources, and visualization of complex data.

exophoric reference:

A linguistics concept that defines a particular type of textual reference based on the relative scope of the reference and the referent within a text. Exophoric references are words or phrases that enable a link to be

established in a text where the thing being referred to (the referent) stands outside of the text. Exophoric referencing stands in contrast to endophoric referencing.

flame emission spectroscopy:

Emission spectroscopy is a laboratory technique that is used to determine the elements in a compound by observing the electromagnetic radiation spectra obtained when the compound transitions between energy states. In the case of flame emission spectroscopy, this transition is brought about by subjecting a solution of the compound to high temperatures over a flame burner.

fluid secretion assay:

A laboratory technique, also know as a Ramsay assay, that is commonly used in laboratories working with *Drosophila melanogaster* as a model organism. The assay is used to measure fluid secretion rates from dissected insect tubules that are exposed to altered conditions such as the presence of diuretic hormones. Tubules play an important role in the *Drosophila melanogaste*r osmoregulatory system by acting to regulate the water content within the organism's fluids.

gel documentation system:

abbreviated as **gel doc system**;

A laboratory system that combines a camera, transilluminator, and image capture and analysis software to support the digital capture of gel electrophorerograms.

gel electrophoresis:

Gel electrophoresis is a core molecular biology laboratory technique that uses electrical charge to sort fragments of DNA/RNA based on their size and charge. A gel medium commonly used for this technique is agarose gel, which is the gelling component of agar. The steps involved in gel electrophoresis consist of preparing an appropriate gel medium, loading the DNA/RNA samples into separate lanes at one end of the gel, and applying electrical charge for a period of time. The electrical charge causes the DNA/RNA fragments to migrate across the gel, with smaller molecules

moving faster and therefore migrating further across the gel. Dyes can be then be used to visualize the separation of the DNA/RNA fragments across the gel, typically under UV lighting.

gel extraction:

A laboratory technique used to isolate a DNA fragment by physically excising the segment of a gel that contains a specific band separated during gel electrophoresis so that the specific DNA in the excised gel segment can be extracted.

gene:

The basic compositional unit within the genome defined by its functional attributes and DNA sequence. Humans and higher order organisms, known as eukaryotes, are composed of cells that contain a nucleus. In eukaryotes, genetic information is stored within these cells principally as DNA organized into chromosomes and stored within the nuclei. During the process of cell division, these chromosomes are duplicated so that the DNA can be passed between cells in a process termed DNA replication that effectively transmits genetic information between parents and progeny. In prokaryotes, genetic information is typically stored on a single, circular chromosome arranged in a nucleoid. Specific regions of the DNA held in these chromosomes encode the instructions necessary to perform specific biological functions that will control the behaviour of cells. These specific regions of DNA are termed genes.

generalization/specialization:

A generalization/specialization relationship, also referred to as a supertype/subtype relationship, is an information and type modelling construct that identifies a specific form of relationship between two concepts in which one concept (the specialization or subtype) represents a more specialized form of the other concept (the generalization or supertype). For example, a ‘car’ is a specialized form of the more generic concept ‘vehicle’, and ‘molecular biology’ is a specialized form of the more generic concept ‘biology’. Generalization/specialization relationships are widely used in data analysis techniques such as grounded theory data analysis, and

in systems analysis techniques such as object-orientated modelling and entity-relationship modelling.

genome:

The set of all chromosomes of an organism, which includes all genes and other non-coding regions of DNA, is termed the genome of the organism. More formally, the genome is a single complete set (haploid) of the chromosomes since organisms commonly have two (diploid) or more (polyploid) copies of the chromosomes. Humans, for example, are diploid with two complete sets of chromosomes.

illocutionary act:

A linguistics term deriving from Austin's doctrine of 'illocutionary forces', which is concerned with different types of function of language. Linguistic utterances are characterized in terms of 'speech acts' or the actions that the utterance of a text intentionally or conventionally performs. In particular, utterances are described in terms of three acts, termed the locutionary, illocutionary, and perlocutionary acts. The locutionary act is concerned with the propositional content of an utterance, whilst the illocutionary act is concerned with the action a speaker performs in making the utterance (*e.g.* asserting, commanding, describing, promising, questioning, *etc*.). The perlocutionary act refers to the effect the speaker produces on a hearer (*e.g.* alarming, convincing, *etc*.). Note that each of these actions can apply to spoken, written and other modes of utterance. Theorists such as Austin, Searle, and Habermas have proposed classification schemes to identify and distinguish multiple types of illocutionary act.

immunocytochemistry (ICC):

A laboratory technique that is used to assess and visualize the presence of specific proteins within a cell growing in culture by introducing specific tagged antibodies that will bind to the protein of interest.

immunohistochemistry (IHC):

A laboratory technique that is used to assess and visualize the presence of specific proteins within a tissue slice by introducing specific tagged antibodies that will bind to the protein of interest. The location of the

protein in the tissue can be visualized using procedures such as diaminobenzidine (DAB) staining that produce colour stains on the tissue.

in silico:

Experiments that are not carried out at the laboratory bench, but are instead conducted via computer models or computer simulations.

in vitro:

Experiments that are carried out at the laboratory bench, and are designed to be conducted in a controlled laboratory environment outside living organisms. The controlled test conditions may not correspond to the actual test conditions encountered for *in vivo* experiments.

***in vitro* transcription (IVT)**:

A laboratory technique that produces RNA by transcribing RNA from DNA templates using a process equivalent to that which takes place *in vivo.* In addition to other uses, this technique forms part of the standard protocol used to prepare samples for microarray experiments.

in vivo:

Experiments that are carried out at the laboratory bench, and are designed to be conducted in living organisms.

laboratory information management system (LIMS):

A computer system to support the management of laboratory information and workflow including general administrative tasks such as invoicing, and laboratory-specific tasks such as sample management and laboratory instrument automation.

microarray:

A relatively new technology that is used in molecular biology and related bioscience laboratories to achieve higher throughput for experimental work. Microarray technology enables multiple genetic tests or tests of gene expression level to be performed in parallel so that a set of samples can be tested against multiple probes on a single array. For example, Affymetrix

Inc's GeneChip® Human Mapping 500K Array[139], which is one of a number of commercially available microarrays, is capable of approximately 250,000 genotyping tests per microarray.

myotonic dystrophy type 1:

abbreviated as **DM1**;

An inherited, human disease with symptoms ranging from muscle weakness, myotonia, heart problems, breathing problems, eye problems, and learning difficulties. As the disease is passed on from one generation to the next, the age of onset of the disease decreases and the severity of the symptoms increases resulting in a severe congenital form of DM1.

NanoDrop®:

A type of laboratory instrument that uses spectrophotometry to provide micro-volume sample quantitation. This instrument is commonly used within molecular biology laboratories to measure the concentration and purity of DNA/RNA samples.[140]

nucleotide:

A molecule that is composed of a sugar, a base, and phosphate groups. Nucleotides are the unit molecules, termed monomers, from which polymers including DNA and RNA are formed.

overlap extension polymerase chain reaction:

A variant of the common PCR laboratory technique that is used to introduce mutations into a DNA sequence in a form of site-directed mutagenesis. The reaction is a multi-step process that employs two separate PCRs using complementary conditions to generate two overlapping DNA segments. These overlapping DNA segments can then be used as the templates in a third PCR to form the extended DNA sequence containing the desired mutation.

---

[139] Product information is available from Affymetrix Inc at http://www.affymetrix.com [accessed 01 March 2011]. All trademarks are acknowledged.

[140] Product information is available from Thermo Fisher Scientific Inc at http://www.nandrop.com [accessed 01 March 2011]. All trademarks are acknowledged.

pipette:

A pipette is a laboratory device that is used to transfer accurately a measured volume of a solution. These devices are routinely used in a range of laboratory settings including molecular biology laboratories. Graduated pipettes can be adjusted to handle different calibrated volumes of liquid. Electronic pipettes can be used to automate the process of using a pipette. Multi-channel pipettes can be used to speed up the process of using a pipette to transfer large numbers of samples.

plasmid:

Prokaryotic cells such as those in bacteria typically store their genetic information in a single, circular chromosome supplemented in some organisms by additional smaller, extrachromosomal DNA molecules. These extrachromosomal DNA elements are termed plasmids. The significance of plasmids in laboratory techniques derives from their role as replicons that are capable of transferring genetic information through autonomous replication within a host independently of the chromosomes.

plasmid cloning vector:

abbreviated as **vector**;

A plasmid that accepts foreign DNA and can therefore be used in recombinant laboratory techniques that introduce foreign DNA segments into organisms such as introducing a gene from one organism into another organism's genome.

plasmid preparation:

A laboratory technique that is used to extract and purify the DNA stored in a plasmid. A range of commercial kits is available to perform plasmid preparation, commonly named according to the scale of the expected plasmid yield as either minipreparations, minipreparations, or maxipreparations[141].

[141] Kits such as QIAPrep® Miniprep were routinely used for these techniques in the laboratories that participated in this study. Product information is available from QIAGEN Inc at http://www.qiagen.com/plasmid [accessed 01 March 2011]. All trademarks are acknowledged.

polymerase:

An enzyme that links nucleotide sequences together to form polynucleoide sequences. Polymerases play a central role in the replication of DNA sequences, and in the transcription of RNA sequences from DNA.

polymerase chain reaction (PCR):

A laboratory technique that is used to copy and amplify strands of a DNA sequence of interest. Amplify in this sense refers to bulk replication, and PCR can generate a large number of copies of a target DNA sequence starting from a single or small number of copies. The specific primers used in a given PCR enable a degree of selection over the DNA sequences to be amplified. *Thermus aquaticus* (*Taq*) polymerase is a thermostable polymerase that is widely used in PCR to form new strands of DNA. The reference to chain reaction in the name, PCR, highlights the fact that as the reaction proceeds over time, new strands of DNA formed in any given iteration provide additional templates for further amplification in succeeding iterations. In this way, product accumulates in an exponential manner with each successive cycle.

polymorphism:

A site within the genome subject to sequence variation, where the different variants are termed alleles.

primer:

A short nucleotide sequence that is complementary to a target sequence and enables a polymerase to begin synthesis of a DNA chain.

protein kinase:

An enzyme that plays a significant role in regulating cellular pathways by causing functional changes in proteins through chemically altering the proteins in a reaction termed phosphorylation. Deregulated kinase activity is associated with a number of diseases including cancer and diabetes mellitus.

protein kinase assay:

A laboratory technique that is used to quantify kinase function by detecting levels of phosphorylated proteins.

quantitative polymerase chain reaction (qPCR):

qPCR is a variant of the PCR laboratory technique that is used to both amplify and quantify strands of a DNA sequence of interest. This technique is often used within molecular biology laboratories to provide quantitative data on the expression of a gene over time, perhaps in response to specific stimuli.

reaction mix list:

A reaction mix list identifies the list of the reagents to be combined for a reaction together with the volume and concentration of each reagent to be used in the reaction. A single reaction mix list specifies the volumes for a single reaction only. A multiple reaction mix list is used to scale up the unit reaction volumes for a specified number of reactions.

repeat-primed polymerase chain reaction (RP-PCR):

Unstable triplet repeats in DNA have been associated with the genetics of a number of human diseases including Huntington's disease and myotonic dystrophy type 1. Repeat-primed PCR is a variant of the common PCR laboratory technique that is used to identify the presence in DNA samples of the large expanded triplet repeats that are pathogenic.

restriction digest:

A laboratory technique that is used to prepare DNA samples for further analysis by digesting the samples with restriction enzymes that cut the DNA into fragments. Specific restriction enzymes can be chosen to enable cutting of the DNA into fragments at specific sites that contain a DNA sequence of interest.

ribonucleic acid (RNA):

RNA is a polymer composed of a sequence of nucleotide units. Each of the nucleotide units in RNA molecules contains the sugar ribose and one of four possible bases, namely guanine (G), cytosine (C), adenine (A), and uracil

(U). In a similar manner to DNA sequences, RNA is encoded as an ordered sequence of these bases corresponding to the ordered sequence of nucleotides in the RNA polymer. RNA molecules usually exist in a single-stranded form. In the normal flow of biological information according to the central dogma of molecular biology, RNA is synthesized using a fragment of DNA as a template so that the information held in the DNA is transferred to a messenger RNA (mRNA). This process is termed transcription. Thereafter, protein is synthesized using the mRNA as a template so that the information held in the mRNA is transferred into proteins. This process is termed translation. At this point, the proteins can perform their biological function acting upon the cells of the organism.

single-nucleotide polymorphism (SNP):

A polymorphism where the alleles differ from each other in the nucleotide at a single DNA base position.

Southern blotting:

A laboratory technique that is used to detect the presence of a specific DNA sequence in DNA samples through the use of a hybridization probe that targets the DNA sequence of interest. The fragments in the samples are first separated using gel electrophoresis, and then transferred by blotting onto a sheet of nitrocellulose that is overlaid on the resulting gel. The resulting blot is incubated with a hybridization probe, which can be labelled using fluorescence or radioactivity in order to visualize the presence or absence of the DNA sequence of interest on the blot. The technique is named after its inventor, hence the initial capital in Southern.

survival immunity assay:

A laboratory technique that is commonly used in laboratories working with *Drosophila melanogaster* to investigate the immune response of different genetic strains of flies to bacterial infection. The set of flies in the assay are typically exposed to the infection by being stabbed with an instrument containing a bacterium such as *Escherichia coli*.

transformation:

A process involving the transfer and incorporation of foreign DNA into a

cell resulting in a genetic alteration of the cell. Foreign DNA can be injected into animal cells through microinjection. Some species of bacteria are naturally able to take up foreign DNA. Inducing artificial competence in bacteria to take up foreign DNA is now a standard laboratory technique that is achieved by stressing the bacterial cells through either heat shock or electrical shock in order to create temporary holes in the membranes of the bacterial cells through which plasmid DNA can enter.

vacuum concentrator:

A laboratory device that is used to provide sample concentration and drying by combining a vacuum, centrifuge, and heat to achieve controlled evaporation. Typically used within molecular biology laboratories for the concentration of small samples of DNA, RNA, and protein.

vascular disease:

A disease that is typically caused by hardening of the arteries (artherosclerosis) due to fatty deposits or plaques on the artery lining. The symptoms associated with the disease depend on the part of the body that is affected by the artherosclerosis, which may commonly be the heart, brain, or leg.

References

Abir-Am, P. (2006). Molecular biology and its recent historiography: A transnational quest for the "Big Picture". *History of Science 44(1),* 95-118

Amann, K. and Knorr Cetina, K. D. (1988). The fixation of (visual) evidence. *Human Studies 11(2-3),* 133-169

Ammon, U. (Ed.) (2001). *The dominance of English as a language of science: Effects on other languages and communities*. Berlin: Mouton de Gruyter

Anderson, S., Hardstone, G., Procter, R., and Williams, R. (2008). Down in the data(base)ment: Supporting configuration in organizational information systems. In M. S. Ackerman, C. A. Halverson, T. Erickson, and W. A. Kellogg (Eds.), *Resources, Co-evolution and Artifacts: Theory in CSCW* (pp 221-254). London: Springer-Verlag

Antunes, P., Costa, C. J., and Pino, J. A. (2006). The use of genre analysis in the design of electronic meeting systems. *Information Research 11(3*), paper 251 [available at http://InformationR.net/ir/11-3/paper251.html]

Arnstein, L., Hung, C.-Y., Franza, R., Zhou, Q. H., Borriello, G., Consolvo, S., and Su, J. (2002). Labscape: A smart environment for the cell biology laboratory. *IEEE Pervasive Computing 1(3),* 13-21

Askehave, I. and Swales, J. M. (2001). Genre identification and communicative purpose: A problem and a possible solution. *Applied Linguistics 22(2),* 195-212

Austin, J. L. (1975). *How to do things with words* (2nd Edition). Cambridge, MA: Harvard University Press

Badger, R. (2003). Legal and general: Towards a genre analysis of newspaper law reports. *English for Specific Purposes 22(3),* 249-263

Bakhtin, M. M. (1986). The problem of speech genres. In C. Emerson and M. Holquist (Eds.), *Speech genres and other late essays* (Translated by V. W. McGee) (pp 60-102). Austin, TX: University of Texas Press

Bailey, S. (2009). Forget electronic records management, it's automated records management that we desperately need. *Records Management Journal 19(2)*, 91-97

Ball, A. (2010). *Review of the state of the art of the digital curation of research data*. ERIM Project Document erim1rep091103ab11. Bath: University of Bath [available at http://opus.bath.ac.uk/18774]

Bannon, L. and Bødker, S. (1997). Constructing common information spaces. *Proc. ECSCW 1997,* 81-96

Barabas, C. (1990). *Technical Writing in a Corporate Culture: A Study of the Nature of Information*. Norwood, NJ: Ablex.

Barker, K. (2005). *At the bench: A laboratory navigator.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press

Barnes, T. (2010). *Confidentiality of PhD theses in the UK*. UKCGE Report. Lichfield: UK Council for Graduate Education [available at http://www.ukcge.ac.uk/publications/reports]

Barnes, B., Bloor, D., and Henry, J. (1996). Interpretation. In *Scientific knowledge: A sociological analysis* (pp 18-45). London: Athlone

Bartholomae, D. (1986). Inventing the university. *Journal of Basic Writing 5(1),* 4-23

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science.* Madison, WI: The University of Wisconsin Press

Bazerman, C. (1999). *The languages of Edison's light.* Cambridge, MA: MIT Press

BBSRC (2007). *Data Sharing Policy*. Biotechnology and Biological Sciences Research Council Document 2007-04. Swindon: Biotechnology and Biological Sciences

Research Council
[available at http://www.bbsrc.ac.uk/publications/policy/data_sharing_policy.pdf]

Bearman, D. (1994). *Electronic evidence: Strategies for managing records in contemporary organizations*. Pittsburgh, PA: Archives and Museums Informatics

Beaufort, A. (1998). *Writing in the real world: Making the transition from school to work*. New York, NY: Teachers College Press

Becher, T. and Trowler, P. R. (2001). *Academic tribes and territories: Intellectual enquiry and the culture of disciplines* (2nd Edition). Buckingham: SRHE/Oxford University Press

Bellotti, V. and Bly, S. (1996). Walking away from the desktop computer: Distributed collaboration and mobility in a product design team. *Proc. CSCW 1996*, 209-218

Berg, E. C., Hult, F. M., and King, K. A. (2001). Shaping the climate for language shift? English in Sweden's elite domains. *World Englishes 20(3)*, 305-319

Berger, S. L. and Kimmel, A. R. (1987). *Guide to molecular cloning techniques*. London: Academic Press

Berkenkotter, C. (2008). *Patient tales: Case histories and the uses of narrative in psychiatry.* Columbus, SC: The University of South Carolina Press

Berkenkotter, C. and Huckin, T. (1995). *Genre knowledge in disciplinary communication.* Hillsdale, NJ: Lawrence Erlbaum

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings.* Harlow: Longman

Bhatia, V. K. (2002). A generic view of academic discourse. In J. Flowerdew (Ed.), *Academic discourse* (pp 21-39). Harlow: Pearson Education

Bhatia, V. K. (2008). Genre analysis, ESP, and professional practice. *English for Specific Purposes 27(2),* 161-174

Bhatia, V. K., Flowerdew, J., and Jones, R. H. (2008). Advances in discourse studies. In V. K. Bhatia, J. Flowerdew, and R. H. Jones (Eds.), *Advances in discourse analysis* (pp 1-18). London: Routledge

Biagioli, M. (2003). Rights or rewards? Changing frameworks of scientific authorship. In M. Biagioli and P. Galison (Eds.), *Scientific authorship* (pp 44-49). London: Routledge

Birnholtz, J. P. and Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. *Proc. GROUP 2003*, 339-348

Bloor, M. (1978). On the analysis of observational data: A discussion of the worth and uses of inductive techniques and respondent validation. *Sociology 12(3),* 545-552

Bloor, M. (1999). Variation in the methods section of research articles across disciplines: The case of fast and slow texts. In P. Thompson (Ed.), *Issues in EAP writing research and instruction* (pp 84-106). Reading: University of Reading Centre for Applied Language Studies

Borgman, C. L., Wallis, J. C., and Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries 7(1),* 17-30

Borriello, G. (2006). The invisible assistant. *ACM Queue 4(6),* 44-49

Bourdieu, P. (1991). *The logic of practice.* Standford, CA: Stanford University Press

Bowker, G. C. (2005). *Memory practices in the sciences.* Cambridge, MA: MIT Press

Bowles, M. A. (2010). *The think-aloud controversy in second language research.* Abingdon: Routledge

BPS Ethics Committee. (1978). *Ethical principles for conducting research with human participants.* Leicester: British Psychological Society

Braine, G. (1995). Writing in the natural sciences and engineering. In D. Belcher and G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp 113-133). Norwood, NJ: Ablex

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001). Minimum information about a microarray experiment (MIAME) – Towards standards for microarray data. *Nature Genetics 29(4),* 365-371

Brown, C. M. (2003). The changing face of scientific discourse: Analysis of genomic and proteomic database usage and acceptance. *Journal of the American Society for Information Science and Technology 54(10)*, 926-938

Brown, G. and Yule, G. (1983). *Discourse Analysis,* Cambridge: Cambridge University Press

Brun, C., Dymetman, M., Fanchon, E., Lhomme, S., and Pogodalla, S. (2003). Semantically-based text authoring and the concurrent documentation of experimental protocols. *Proc. DocEng 2003*, 193-202

Bulmer, M. (1984). *The Chicago school of sociology: Institutionalization, diversity, and the rise of sociological research*. Chicago, IL: University of Chicago Press

Bunton, D. (1999). The use of higher level metatext in PhD theses. *English for Specific Purposes 18(Supplement)*, 41-56

Bunton, D. (2002). Generic moves in PhD thesis introductions. In J. Flowerdew (Ed.), *Academic discourse* (pp 57-75). Harlow: Pearson Education

Cadman, K. (2002). English for academic possibilities: The research proposal as a contested site in postgraduate genre pedagogy. *Journal of English for Academic Purposes 1(2)*, 85-104

Cambrioso, A. and Keating, P. (1988). "Going monoclonal": Art, science and magic in the day-to-day use of hybridoma technology. *Social Problems 35(3)*, 244-260

Campbell, J. A. (1990). Scientific discovery and rhetorical invention: The path to Darwin's *Origin*. In H. W. Simons (Ed.), *The rhetorical turn: Invention and persuasion in the conduct of inquiry* (pp. 58-90). Chicago, IL: University of Chicago Press

Canagarajah, A. S. (Ed.) (2005). *Reclaiming the local in language policy and practice*. Mahwah, NJ: Lawrence Erlbaum

Candlin, C. N. (1993). General editor's preface. In V. K. Bhatia, *Analysing genre: Language use in professional settings* (pp ix-xi). Harlow: Longman

Casper, M. J. and Clarke, A. E. (1998). Making the pap smear into the 'right tool' for the job: Cervical cancer screening, 1940-1995. *Social studies of science 28(2)*, 255-290

Charmaz, K. (1983). The grounded theory method: An explication and interpretation. In R. M. Emerson (Ed.), *Contemporary field research: A collection of readings* (pp 109-126). Boston, MA: Little Brown

Chin, G. Jr. and Lansing, C. S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. *Proc. CSCW 2004*, 409-418

Chin, G. Jr., Myers, J., and Hoyt, D. (2002). Social networks in the virtual science laboratory. *Communications of the ACM 45(8)*, 87-92

Chintapalli, V. R., Wang, J., and Dow, J. A. T. (2007). Using FlyAtlas to identify better *Drosophila* models of human disease. *Nature Genetics 39(6)*, 715-720

Cicourel, A. (1968). *The social organization of juvenile justice*. New York, NY: Wiley

Clarke, A. E. and Fujimura, J. M. (Eds.) (1992). *The right tools for the job: At work in twentieth century life sciences*. Princeton, NJ: Princeton University Press

Coffey, A. J. (1993). *Double-entry: The professional and organizational socialization of graduate accountants*. PhD Thesis, University of Wales, Cardiff

Coles, S. J., Frey, J. G., Hursthouse, M. B., Light, M. E., Milsted, A. J., Carr, L. A., De Roure, D., Gutteridge, C. J., Mills, H. R., Meacham, K. E., Surridge, M., Lyon, E., Heery, R., Duke, M., and Day, M. (2006). An e-Science environment for service crystallography - from submission to dissemination. *Journal of Chemical Information and Modeling 46(3),* 1006-1016

Collins, H. M. (1975). The seven sexes: A study in the sociology of a phenomenon, or the replication of experiments in physics. *Sociology 9(2),* 205-224

Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice.* London: SAGE

Conery, J. S., Catchen, J. M., and Lynch, M. (2005). Rule-based workflow management for bioinformatics. *The VLDB Journal 14(3)*, 318-329

Connor, U., Davis, K., and De Rycker, T. (1995). Correctness and clarity in applying for overseas jobs: A cross-cultural analysis of U.S. and Flemish applications. *Text 15(4),* 457–475

Cook, T. (2005). Macroappraisal in theory and practice: Origins, characteristics, and implementation in Canada, 1950-2000. *Archival Science 5(2)*, 101-161

Corbett, J. (1992). Encoding nature: Writing science in English. *FEMS Microbiology Letters 100(1-3)*, 39-44

Corbett, J. (2009). Theme, field and genre: Thematic realisations in academic articles and their popularisations. In G. Forey and G. Thompson (Eds.), *Text type and texture* (pp 70-93). London: Equinox

Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press

Couture, C. (1996). Le concept de document d'archives à l'aube du troisième millénaire. *Archives 27(4)*, 3-19

Couture, C. and Rousseau, J.-Y. (1987). *The life of a document: A global approach to archives and records management* (Translated by D. Homel). Montreal, QC: Véhicule

Crabtree, A., Rodden, T., Tolmie, P., and Button, G. (2009). Ethnography considered harmful. *Proc. CHI 2009,* 879-888

Cragin, M. H. and Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer-Supported Cooperative Work 15(3),* 185-204

Craighead, H. (2006). Future lab-on-a-chip technologies for interrogating individual molecules. *Nature 442*, 387-393

Craik, F. I. M. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour 11(6),* 671-684

Crick, N. (2005). "A capital and novel argument": Charles Darwin's notebooks and the productivity of rhetorical consciousness. *Quarterly Journal of Speech 91(4)*, 337-364

Cronin, B. (2001). Bibliometrics and beyond: Some thoughts on web-based citation analysis. *Journal of Information Science 27(1),* 1-7

Crystal, D. (2003). *English as a global language* (2nd Edition). Cambridge: Cambridge University Press

Cullingford, C. and Swift, H. (2002). The impact of information on individuals and organisations. In G. Trorey and C. Cullingworth (Eds.), *Professional development and institutional needs* (pp 181-192). Aldershot: Ashgate

Cumming, K. (2007). Purposeful data: The roles and purposes of recordkeeping metadata. *Records Management Journal 17(3),* 186-200

de Beaugrande, R.-A. and Dressler, W. (1981). *Introduction to text linguistics*. London: Longman

De Roure, D., Jennings, N., and Shadbolt, N. (2001). *Research agenda for the Semantic Grid: A future e-Science infrastructure*. Technical Report UKeS-2002-2. Edinburgh: National e-Science Centre [available at http://www.nesc.ac.uk/technical_papers/DavidDeRoure.etal.SemanticGrid.pdf]

De Roure, D., Jennings, N., and Shadbolt, N. R. (2005). The semantic grid: Past, present and future. *Proc. IEEE 93(3*), 669-681

David, M. (2005). *Science in society*. Basingstoke: Palgrave Macmillan

David, P. A. (2006). Towards a cyberinfrastructure for enhanced scientific collaboration: Providing its "soft" foundation may be the hardest part. In B. Kahin and D. Foray (Eds.), *Advancing knowledge and the knowledge economy* (pp 431-453). Cambridge, MA: MIT Press

Day, M. (2005). Metadata. In S. Ross and M. Day (Eds.), *DCC digital curation manual*. Edinburgh: Digital Curation Centre [available at http://www.dcc.ac.uk/resource/curation-manual/metadata]

Delanty, G. and Strydom, P. (Eds.) (2003). *Philosophies of social science: The classic and contemporary readings*. Maidenhead: Open University Press

Devitt, A. J. (1997). Genre as a language standard. In W. Bishop and H. Ostrum (Eds.), *Genre and writing* (pp 45-55). Portsmouth, NH: Boynton/Cook

Devitt, A. J. (2000). Integrating rhetorical and literary theories of genre. *College English 62(5)*, 696-718

Dias, P., Freedman, A., Medway, P., and Paré, A. (1999). *Worlds apart: Acting and writing in academic and workplace contexts.* Mahwah, NJ: Lawrence Erlbaum

Dong, Y. R. (1998). Non-native speaker graduate students' thesis/dissertation in science: Self-reports by students and their advisors from two US institutions. *English for Specific Purposes 17(4),* 369-390

Douglas, D. (2000). *Assessing language for specific purposes*. Cambridge: Cambridge University Press

Dubois, B.-L. (1987). "Something on the order of around forty to forty-four": Imprecise numerical expressions in biomedical slide talks. *Language in Society 16(3),* 527-541

Dudley-Evans, T. (1985). *Writing laboratory reports*. Melbourne, VIC: Thomas Nelson

Dudley-Evans, T. (1991). Socialisation into the academic community: Linguistic and stylistic expectation of a PhD thesis as revealed by supervisor comments. In P. Adams, B. Heaton, and P. Howarth (Eds.), *Socio-cultural issues in English for academic purposes* (pp 42-52). London: Macmillan

Eastwood, T. (1992). Towards a social theory of appraisal. In B. L. Craig (Ed.), *The archival imagination: Essays in honour of Hugh A. Taylor* (pp 42-52). Ontario, ON: Association of Canadian Archivists

Ebel, H. F., Bliefert, C., and Russey, W. E. (2004). *The art of scientific writing* (2nd Edition). Weinheim: Wiley-VCH

Edsall, J. T. (1974). History of bioenergetics. *Molecular and Cellular Biochemistry 5(1),* 5-9

Eggins, S. (1994). *An introduction to systemic functional linguistics*. London: Pinter

Eldredge, N. (2005). Darwin's *Other* Books: "Red" and "Transmutation" Notebooks, "Sketch," "Essay," and *Natural Selection*. *PLoS Biology 3(11),* e382. doi:10.1371/journal.pbio.0030382

Ellingsen, G. and Monteiro, E. (2003). A patchwork planet: Integration and cooperation in hospitals. *Computer Supported Cooperative Work 12(1),* 71-95

Ericsson, K. A. and Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (2nd Edition). Cambridge, MA: MIT Press

Fairclough, N. (1992). *Discourse and social change.* Cambridge: Polity Press

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research.* London: Routledge

Farooq, U., Carroll, J. M., and Ganoe, C. H. (2005). Supporting creativity in distributed scientific communities. *Proc. GROUP 2005,* 217-226

Flowerdew, J. (2002a). Genre in the classroom: A linguistic approach. In A. M. Johns (Ed.) *Genres in the classroom: Multiple perspectives* (pp 91-102). Mahwah, NJ: Lawrence Erlbaum

Flowerdew, J. (Ed.) (2002b). *Academic discourse.* Harlow: Pearson Education

Flowerdew, J. and Wan, A. (2010). The linguistic and the contextual in applied genre analysis: The case of the company audit report. *English for Specific Purposes 29(2),* 78-93

Flynn, E. A. (1988). Composing as a woman. *College Composition and Communication 39(4),* 423-435

Foucault, M. (1972). *The archaeology of knowledge.* New York, NY: Pantheon

Freedman, A. and Medway, P. (Eds.) (1994). *Genre and the New Rhetoric.* London: Taylor & Francis

Frey, J. G., Hughes, G. V., Mills, H. R., schraefel, m. c., Smith, G. M., and De Roure, D. (2004). Less is more: Lightweight ontologies and user interfaces for smart labs. *Proc.UK e-Science All Hands Meeting, EPSRC 2004*

Fry, J., Lockyer, S., Oppenheim, C., Houghton, J., and Rasmussen, B. (2008). *Identifying benefits arising from the curation and open sharing of research data produced by UK higher education and research institutes*. JISC Report. London: Joint Information Systems Committee [available at http://www.jisc.ac.uk/publications/reports/2008/databenefitsfinalreport.aspx]

Gagné, R. M. and Smith, E. C. Jr. (1962). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology 63(1)*, 12-18

Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall

Gaudillière, J.-P. and Löwy, I. (1998). *The invisible industrialist: Manufacture and the production of scientific knowledge.* London: Macmillan

Gavaghan, D. J., Simpson, A. C., Lloyd, S., MacRandal, D. F., and Boyd, D. R. S. (2005). Towards a Grid infrastructure to support integrative approaches to biological research. *Phil. Trans. R. Soc. A 363,* 1829-1841.

Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures: Selected essays* (pp 3-30*)*. New York, NY: Basic Books

Geertz, C. (1983). *Local knowledge: Further essays in interpretive anthropology*. New York, NY: Basic Books

Geertz, C. (1988). *Works and lives: The anthropologist as author*. Stanford, CA: Stanford University Press

Gene Ontology Consortium (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics 25(1)*, 25-29

Genette, G. (1997). *Palimpsests: Literature in the second degree* (Translated by C. Newman and C. Doubinsky). Lincoln, NB: University of Nebraska Press

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. New York, NY: University of California Press

Gilbert, G. N. and Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. Cambridge: Cambridge University Press

Glaser, B. G. and Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Hawthorne, NY: Aldine de Gruyter

Goffman, E. (1974). *Frame analysis.* Cambridge, MA: Harvard University Press

Goguen, J. and Linde, C. (1993). Techniques for requirements elicitation. *Proc. RE 1993*, 194-195

Goodwin, C. (2000). Practices of seeing visual analysis: An ethnomethodological approach. In T. van Leeuwen and C. Jewitt (Eds.), *Handbook of visual analysis* (pp 157-182). London: SAGE

Gosden, H. (2003). 'Why not give us the full story?': Functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes 2(2)*, 87-101

Graddol, B. (1999). The decline of the native speaker. *AILA Review 13(1),* 57-68

Greenberg, S. (1991). Personalizable groupware: Accommodating individual roles and group differences. *Proc. ECSCW 1991,* 17-31

Greene, S. and Higgins, L. (1994). "Once upon a time": The use of retrospective accounts in building theory in composition. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp 115-140). Thousand Oaks, CA: SAGE

Greenfield, A. (2006). *Everyware: The dawning age of ubiquitous computing*. Berkeley, CA: New Riders

Greimas, A. J. (1966). *Structural semantics*. Lincoln, NB: University of Nebraska Press

Grimshaw, A. D. (1989). *Collegial discourse: Professional conversation among peers*. Norwood, NJ: Ablex

Grindley, P. (1995). *Standards strategy and policy: Cases and stories*. Oxford: Oxford University Press

Gross, A. G. (2006). *Starring the text: The place of rhetoric in science studies*. Carbondale, IL: Southern Illinois University Press

Gross, A. G., Harmon, J. E., and Reidy, M. (2002). *Communicating science: The scientific article from the 17th century to the present*. Oxford: Oxford University Press

Gruber, H. E. (1989). The evolving systems approach to creative work. In D. B. Wallace and H. E. Gruber (Eds.), *Creative people at work: Twelve cognitive case studies* (pp 3-24). Oxford: Oxford University Press

Habermas, J. (1987). *The theory of communicative action (vol. 2): Lifeworld and system: A critique of functionalist reason* (Translated by T. McCarthy). Cambridge: Polity

Halliday, M. A. K. and Hasan, R. (1989). *Language, context and text: Aspects of language in a social-semiotic perspective* (2nd Edition). Oxford: Oxford University Press

Hammersley, M. and Atkinson, P. (1995). *Ethnography: Principles in practice* (2nd Edition). London: Routledge

Handy, C. (1993). *Understanding organizations*. London: Penguin

Hart, R. P. (1986). Of genre, computers, and the Reagan inaugural. In H. W. Simons and A. A. Aghazarian (Eds.), *Form, genre and the study of political discourse* (pp 278-300). Columbia, SC: University of South Carolina Press

Hartland, R., McKemmish, S., and Upward, F. (2005). Documents. In S. McKemmish, M. Piggot, B. Reed, and F. Upward (Eds.), *Archives: Recordkeeping in society* (pp 75-100). Wagga Wagga, NSW: Centre for Information Studies

Hartswood, M., Proctor, R., Slack, R., Voß, A., Büscher, M., Rouncefield, R., and Rouchy, P. (2008). Co-realization: Towards a principled synthesis of ethnomethodology and participatory design. In M. S. Ackerman, C. A. Halverson, T. Erickson, and W. A. Kellogg (Eds.), *Resources, Co-evolution and Artifacts: Theory in CSCW* (pp 59-94). London: Springer-Verlag

Hayes, J. R. and Bajzek, D. (2008). Understanding and reducing the knowledge effect: Implications for writers. *Written Communication 25(1),* 104-118

Headland, T. N., Pike, K. L., and Harris, M. (1990). *Emics and etics: The insider/outsider debate*. Newbury Park, CA: SAGE

Health and Safety Executive (2005). *The control of substances hazardous to health regulations 2002 (as amended): Approved code of practice and guidance* (5th Edition). London: HSE Books

Heath, C. and Luff, P. (1996). Documents and professional practice: 'Bad' organisational reasons for 'good' clinical records. *Proc. CSCW 1996*, 354-363

Heath, C. and Luff, P. (2000). *Technology in action*. Cambridge: Cambridge University Press

Henderson, A. and Kyng, M. (1991). There's no place like home: Continuing design in use. In J. Greenbaum and M. Kyng (Eds.), *Design at work: Cooperative design of computer systems* (pp 219-240). Hillsdale, NJ: Lawrence Erlbaum

Higgs, E. (1998). Records and recordness: Essences or conventions? In E. Norberg (Ed.), *The concept of record: Report from the second Stockholm conference on archival science and the concept of record* (pp 101-111). Stockholm: Riksarkivet

Hine, C. (2005). Material culture and the shaping of e-Science. *Proc. NCeSS 2005,* [available at http://www.ncess.ac.uk/events/conference/2005/papers/papers/ncess2005_paper_Hine.pdf]

Hine, C. (2006). Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science 36(2),* 269-298

Hoey, M., Sinclair, J., Stubbs, M., Mahlberg, M., and Teubert, W. (2007). *Text, discourse and corpora: Theory and analysis*. London: Continuum

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd Edition). Thousand Oaks, CA: SAGE

Hofstede, G. and Hofstede, G. J. (2007). *Cultures and organizations: Software of the mind* (2nd Edition). New York, NY: McGraw Hill

Holmes, F. L. (2004). *Investigative pathways: Patterns and stages in the careers of experimental scientists*. New Haven, CT: Yale University Press

Holmes, F. L., Renn, J., and Rheinberger, H.-J. (Eds.) (2003). *Reworking the bench: Research notebooks in the history of science*. Dordrecht: Kluwer

Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes 16(4)*, 321-337

Huckin, T. N. and Olsen, L. (1983). *English for Science and Technology: A handbook for non-native speakers (ELT/ESP)*. New York, NY: McGraw-Hill

Hurley, C. (2005). Recordkeeping and accountability. In S. McKemmish, M. Piggot, B. Reed, and F. Upward (Eds.), *Archives: Recordkeeping in society* (pp 197-222). Wagga Wagga, NSW: Centre for Information Studies

Hutchison, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., and Eiderbeck, B. (2003). Technology probes: Inspiring design for and with families. *Proc. CHI 2003,* 17-24

Hutchison, T. and Waters, A. (1987). *English for Specific Purposes: A learning-centred approach*. Cambridge: Cambridge University Press

Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow: Pearson Education

Hyland, K. (2004). Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes 23(3)*, 303-324

Hyland, K. (2006a). English for academic purposes: An advanced resource book. London: Routledge

Hyland, K. (2006b). Disciplinary differences: Language variation in academic discourses. In K. Hyland and M. Bondi (Eds.), *Academic discourse across disciplines* (pp 17-45). Bern: Peter Lang

Hyland, K. and Bondi, M. (Eds.) (2006). *Academic discourse across disciplines*. Bern: Peter Lang

Hyland, K. and Tse, P. (2007). Is there an "academic vocabulary"?. *TESOL Quarterly 41(2)*, 235-253

Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia, PA: University of Philadelphia Press

Hyon, S. (1996). Genre in three traditions: Implications for ESL. *TESOL Quarterly 30(4),* 693-722

Hyon, S. (2008). Convention and inventiveness in an occluded academic genre: A case study of retention-promotion-tenure reports. *English for Specific Purposes 27(2),* 175-192

ISO (2001). *ISO 15489-1:2001 Information and documentation – Records management – Part I: General*. Geneva: International Organization for Standardization

ISO (2003). *ISO 14721:2003 Space data and information transfer systems – Open archival information system – Reference model*. Geneva: International Organization for Standardization

Jacoby, S. W. and Gonzalez, P. (1991). The constitution of expert-novice in scientific discourse. *Issues in Applied Linguistics 2(2)*, 150–181.

Jakobson, R. (1937). *Lectures on sound and meaning*. Cambridge, MA: MIT Press

Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp 13-31). Amsterdam: John Benjamins

Jenkins, J. (2009). *World Englishes: A resource book for students* (2nd Edition). Abingdon: Routledge

Johns, A. M. (1997). *Text, role and context: Developing academic literacies*. Cambridge: Cambridge University Press

Jordan, K. and Lynch, M. (1993). The 'mainstreaming' of a molecular biology tool: A case study of a new technique. In G. Button (Ed.), *Technology In Working Order* (pp 127-146). London: Routledge

Jourdenais, R. (2001). Cognition, instruction, and protocol analysis. In P. Robinson (Ed.), *Cognition and second language instruction* (pp 354-375). Cambridge: Cambridge University Press

Judson, H. F. (1996). *The eighth day of creation: Makers of the revolution in biology*. Plainview, NY: Cold Spring Harbor Laboratory Press

Kachru, B. B. (1992). World Englishes: Approaches, issues, and resources. *Language Teaching 25(1)*, 1-14

Kanare, H. M. (1985). *Writing the laboratory notebook*. Washington, DC: American Chemical Society

Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes 24(3)*, 269–292.

Kay, L. E. (1993). *The molecular vision of life: Caltech, the Rockefeller Foundation, and the rise of the new biology*. Oxford: Oxford University Press

Kaye, J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I., and Pinch, T. (2006). To have and to hold: Exploring the personal archive. *Proc. CHI 2006,* 275-284

Kellenberger, E. (1989). Origins of molecular biology. *Plant Molecular Biology Reporter 7(3),* 231-234

Kevles, D. J. (1998). *The Baltimore case: A trial of politics, science, and character*. New York, NY: W W Norton

Kevles, D. J. and Hood, L. (1992). Reflections. In D. J. Kevles and L. Hood (Eds.), *The code of codes: Scientific and social issues in the Human Genome Project* (pp 300-330). Cambridge, MA: Harvard University Press

Killingsworth, M. J. (1992). Discourse communities local and global. *Rhetoric Review 11(1)*, 110-122

Kim, D. (2008). *English for Occupational Purposes: One language?* London: Continuum

Knorr Cetina, K. D. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon

Knorr Cetina, K. (1999). *Epistemic Cultures*. Cambridge, MA: Harvard University Press

Kress, G. and van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Milton Keynes: Open University Press

Latour, B. and Woolgar, S. (1986). *Laboratory life: The construction of scientific facts* (2nd Edition). Princeton, NJ: Princeton University Press

LeCompte, M. D. and Schensul, J. J. (1999). *Analyzing and interpreting ethnographic data* (*Ethnographer's toolkit vol. 5*). Walnut Creek, CA: Altamira

Lenoir, T. (1998). Shaping biomedicine as an information science. In M. E. Bowden, T. B. Hahn, and R. V. Williams (Eds.), *Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems* (pp 27-45). Medford, NJ: Information Today

Levy, D. M. (2001). *Scrolling forward: Making sense of documents in the digital age*. New York, NY: Arcade

Liaskos, S., Lapouchnian, A., Wang, Y., Yu, Y., and Easterbrook, S. (2005). Configuring common personal software: A requirements-driven approach. *Proc. RE 2005*, 9-18

Licker, M. D. (Ed.) (2003). *Dictionary of bioscience* (2nd Edition). New York, NY: McGraw-Hill

Lillis, T. (2008). Ethnography as method, methodology, and "deep theorizing". *Written Communication 25(3),* 353-388

Linell, P. (1998). Discourse across boundaries: On recontextualizations and the blending of voices in professional discourse. *Text 18(2),* 143-157

Livingstone, D. N. (1995). The spaces of knowledge: Contributions towards a historical geography of science. *Environment and Planning D: Society and Space 13(1)*, 5-34

Lobban, C. S. and Schefter, M. (1992). *Successful lab reports: A manual for science students*. Cambridge: Cambridge University Press

Lofland, J. and Lofland, L. H. (1995). *Analyzing social settings: A guide to qualitative observation and analysis* (3rd Edition). Belmont, CA: Wadsworth

Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organisation. *English for Specific Purposes 23(3),* 280-302

Love A. (2002). Introductory concepts and 'cutting edge' theories: Can the genre of the textbook accommodate both?. In J. Flowerdew (Ed.), *Academic discourse* (pp 76-91). Harlow: Pearson Education

Lynch, M. (1985a). *Art and artifact in laboratory science: A study of shop work and shop talk in a research laboratory.* London: Routledge and Kegan Paul

Lynch, M. (1985b). Discipline and the material form of images: An analysis of scientific visibility. *Social Studies of Science 15(1),* 37-66

Lynch, M. (1991). Laboratory space and the technological complex: An investigation of topical contextures. *Science in Context 4(1)*, 51-78

Lynch, M. (1997). A sociology of knowledge machine. *Ethnographic Studies 2,* 16-38

Lynch, M. (2009). Ethnomethodology and history: Documents and the production of history. *Ethnographic Studies 11,* 87-106

Lynch, M. and Woolgar, S. (1988). Introduction: Sociological orientations to representational practice in science. *Human Studies 11(2-3),* 99-116

Lysakowski, R. (1997). Comparing paper and electronic laboratory notebooks: Part 1. *Scientific Computing and Automation (March 1997)*

McCarthy, L. P. (1991). A psychiatrist using DSM-III: The influence of a charter document in psychiatry. In C. Bazerman and J. Paradis (Eds.), *Textual dynamics of the professions: Historical and contemporary studies of writing in professional communities* (pp 358-378). Madison, WI: University of Wisconsin Press

McKemmish, S., Reed, B., and Piggott, M. (2005). The archives. In S. McKemmish, M. Piggott, B. Reed, and F. Upward (Eds.), *Archives: Recordkeeping in society* (pp 159-195). Wagga Wagga, NSW: Centre for Information Studies

Mackay, W. E. (1990). Patterns of sharing customizable software. *Proc. CSCW 1990*, 209-221

Mackay, W. E., Pothier, G., Letondal, C., Bøegh, K., and Sørensen, H. E. (2002). The missing link: Augmenting biology laboratory notebooks. *Proc. UIST 2002*, 41-50

Macrina, L. N. (1995). *Scientific integrity: Introductory text with cases*. Washington, DC: American Society of Microbiology Press

Maddox, J. (1996). Preface to the expanded edition. In H. F. Judson, *The eighth day of creation: Makers of the revolution in biology* (pp ix-xii). Plainview, NY: Cold Spring Harbor Laboratory Press

Magner, F. L. (2002). *A history of the life sciences* (3rd Edition). New York, NY: Marcel Dekker

Maingueneau, D. (2002). Analysis of an academic genre. *Discourse studies 4(3)*, 319-342

Malinowksi, B. (1922). *Argonauts of the Western Pacific: An account of native enterprise and adventure in the archipelagoes of Melanesian New Guinea*. London: Routledge

Marks, M. R. (1951). Problem solving as a function of the situation. *Journal of Experimental Psychology 41(1)*, 74-80

Martin, J. R. (1992). *English text: System and structure*. Amsterdam: John Benjamins

Martin, J. R., Christie, F., and Rothery, J. (1987). Social processes in education: A reply to Sawyer and Watson (and others). In I. Reid (Ed.), *The place of genre in learning: Current debates* (pp 46-57). Geelong, VIC: Deakin University Press

Matthews, B., Sufi, S., Flannery, D., Lerusse, L., Griffin, T., Gleaves, M., and Kleese, K. (2010). Using a core scientific metadata model in large-scale facilities. *International Journal of Digital Curation 1(5)*, 106-118

Melander, B., Swales, J. M., and Fredrickson, K. M. (1997). Journal abstracts from three academic fields in the United States and Sweden: National or disciplinary proclivities? In A. Duszak (Ed.), *Intellectual styles and cross-cultural communication* (pp. 251–272). Berlin: Mouton De Gruyter.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL: University of Chicago Press

Miller, C. R. (1984). Genre as social action. *Quarterly Journal of Speech 70(2)*, 151-167

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs, CA: Prentice-Hall

Mirel, B. (1993). Beyond the monkey house: Audience analysis in computerized workplaces. In R. Spilka (Ed.), *Writing in the workplace: New research perspectives* (pp 21-40). Carbondale, IL: Southern Illinois University Press

MRC (2008). *MRC Policy on Data Sharing and Preservation*. Medical Research Council Document 2008-12-19. Swindon: Medical Research Council

[available at http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Data sharinginitiative/Policy]

Mulkay, M. (1995). Galileo and the embryos: Religion and science in parliamentary debate over research on human embryos. *Social Studies of Science 25(2)*, 499-532

Muller, H. J. (1936). Physics in the attack on the fundamental problem of genetics. *Scientific Monthly 44(3)*, 210-214

Mühleisen, S. (2003). Towards global diglossia? English in the sciences and humantities. In C. Mair (Ed.), *The politics of English as a world language: New horizons in postcolonial cultural studies* (pp 107-118). Amsterdam: Rodopi

Myers, G. (1990). *Writing biology: Texts in the social construction of scientific knowledge*. Madison, WI: University of Wisconsin Press

Myers, G. (1991). Stories and styles in two molecular biology review articles. In C. Bazerman and J. Paradis (Eds.), *Textual dynamics of the professions* (pp 45-75). Madison, WI: University of Wisconsin Press

Myers, G. (1996). Strategic vagueness in academic writing. In E. Ventola and A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp 3-18). Amsterdam: John Benjamins

Myers, G. (2000). Powerpoints: Technology, lectures, and changing genres. In A. Trosborg (Ed.), *Analysing professional genres* (pp 177-192). Amsterdam: John Benjamins

Nature (2005). Electronic notebooks: A new leaf. *Nature 436,* 20-21

Nature (2007). Share your lab notes. *Nature 447,* 1-2

Nentwich, M. (2004). *Cyberscience: Research in the age of the internet*. Vienna: Austrian Academy of Sciences Press

Neuendorf, K. (2002). *The content analysis guidebook*. Thousand Oaks, CA: SAGE

Newell, A. and Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, CA: Prentice-Hall

Nickerson, C. (1999). The use of English in electronic mail in a multinational corporation. In F. Bargiela-Chiappinni and C. Nickerson (Eds.), *Writing business: Genres, media and discourses* (pp 35-56). Harlow: Pearson Education

Nisbett, R. E. and Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review 84(3)*, 231-259

Noguchi, J. (2006). *The science review article: An opportune genre in the construction of science*. Bern: Peter Lang

Nygren, E. and Henriksson, P. (1992). Reading the medical record I: Analysis of physician's ways of reading the medical record. *Computer Methods and Programs in Biomedicine 39(1),* 1-12

O'Halloran, K. L. (Ed.) (2004). *Multimodal discourse analysis: Systemic functional perspectives.* London: Continuum

Orlikowski, W. J. (1993). Learning from Notes: Organizational issues in groupware implementation. *Information Society 9(3)*, 237-250

Paré, A. (2004). Texts and power: Towards a critical theory of language. In L. Davies and P. Leonard (Eds.), *Social work in a corporate era* (pp 76-90). Aldershot: Ashgate

Patton, M. Q. (2001). *Qualitative research and evaluation methods* (2nd Edition). Thousand Oaks, CA: SAGE

Pennycook, A. (1994). *The cultural politics of English as an international language*. London: Longman

Pettinari, C. J. (1988). *Task, talk, and text in the operating room: A study in medical discourse*. Norwood, NJ: Ablex

Pheysey, D. C. (1993). *Organizational cultures: Types and transformations*. London: Routledge

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago, IL: University of Chicago Press

Phillipson, R. and Skuttnab-Kangas, T. (1999). Englishisation: One dimension of globalisation. *AILA Review 13(1),* 19-36

Polanyi, M. (1967). *The tacit dimension*. New York, NY: Anchor

Price, D. J. de Solla (1963). *Little science, big science … and beyond*. New York, NY: Columbia University Press

Prior, P. (1998). *Writing/disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Erlbaum

Rafoth, B. A. and Rubin, D. L. (1988). *The social construction of written communication*. Norwood, NJ: Ablex

Ram, A. (1999). A theory of questions and question asking. In A. Ram and K. Moorman (Eds.), *Understanding language understanding: Computational models of reading* (pp 253-298). Cambridge, MA: MIT Press

Ravetz, J. R. (1971). *Scientific knowledge and its social problems*. Oxford: Clarendon

Reed, B. (2005). Records. In S. McKemmish, M. Piggot, B. Reed, and F. Upward (Eds.), *Archives: Recordkeeping in society* (pp 101-130). Wagga Wagga, NSW: Centre for Information Studies

Reed, R. H., Holmes, D., Weyers, J. D. B., and Jones, A. M. (2007). *Practical skills in biomolecular sciences* (3rd Edition). Harlow: Pearson Education

Research Information Network. (2008). *To share or not to share: Publication and quality assurance of research data outputs*. RIN Report. London: Research Information Network
[available at http://www.rin.ac.uk/resources/rin-publications]

Research Information Network. (2010). *Patterns of information use and exchange: Case studies of researchers in the life sciences*. RIN Report. London: Research Information Network/The British Library [available at http://www.rin.ac.uk/resources/rin-publications]

Rheinberger, H.-J. (2003). "Discourses of circumstance": A note on the author in science. In M. Biagioli and P. Galison (Eds.), *Scientific authorship* (pp 309-324). London: Routledge

Riley, P. (2007). *Language, culture, and identity*. London: Continuum

Rizova, P. S. (2007). *The secret of success: The double helix of formal and informal structures in an R&D laboratory.* Stanford, CA: Stanford Business Books

Robek, M. F., Brown, G. F., and Stephens, D. O. (1995). *Information and records management: Document-based information systems.* Woodland Hills, CA: Glencoe/McGraw-Hill

Robson, J., Francis, B., and Read, B. (2002). Writes of passage: Stylistic features of male and female undergraduate history essays. *Journal of Further and Higher Education 26(4)*, 351-362

Rooksby, J., Gerry, R. M., and Smith, A. F. (2007). Incident reporting schemes and the need for a good story. *International Journal of Medical Informatics 76S*, S205-S211

Rosmarin, A. (1985). *The power of genre.* Minneapolis, MO: University of Minnesota Press

Rowley-Jolivet, E. (2002). Visual discourse in scientific conference papers: A genre-based study. *English for Specific Purposes 21(1)*, 19-40

Rumelhart, D. E. and Orotny, A. (1977). The representation of knowledge in memory. In R. Anderson, R. Spiro, and W. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, NJ: Erlbaum

Sambrook, J. and Russell, D. W. (2001). Molecular cloning: A laboratory manual (3rd Edition). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press

Samraj, B. (2002). Texts and contextual layers: Academic writing in content courses. In A. M. Johns (Ed.), *Genre in the classroom: Multiple perspectives* (pp 163-176). Mahwah, NJ: Lawrence Erlbaum

Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes 24(2)*, 141-156

Samuels, H. W. (1991). Improving our disposition: Documentation strategy. *Archivaria 33*, 125-140

Sanz, C., Lin, H.-J., Lado, B., Bowden, H. W., and Stafford, C. A. (2009). Concurrent verbalizations, pedagogical implications, and reactivity: Two CALL studies. *Language Learning 59(1)*, 33-71

Saville-Troike, M. (1982). *The ethnography of communication*. Oxford: Blackwell

Schellenberg, T. R. (1956). *Modern archives: Principles and techniques*. Chicago, IL: University of Chicago Press

Schiffrin, D., Tannen, D., and Hamilton, H. E. (Eds.) (2001). *The handbook of discourse analysis*. Oxford: Blackwell

Schmidt, K. (1997). Of maps and scripts: The status of formal constructs in cooperative work. *Proc. GROUP 1997*, 138-147

schraefel, m. c., Hughes, G. V., Mills, H. R., Smith, G., Payne, T. R., and Frey, J. (2004). Breaking the book: Translating the chemistry lab book into a pervasive computing lab environment. *Proc. CHI 2004*, 25-32

Schryer, C. F. (1993). Records as genre. *Written Communication 10(2),* 200-234

Schryer, C. F. (1994). The lab *vs*. the clinic: Sites of competing genres. In A. Freedman and P. Medway (Eds.), *Genre and the New Rhetoric* (pp 105-124). London: Taylor & Francis

Searls, D. B. (1998). Grand challenges in computational biology. In S. L. Salzberg, D. B. Searls, and S. Kasif (Eds.), *Computational methods in molecular biology* (pp 3-9). Amsterdam: Elsevier

Selinker, L. (1979). On the use of informants in discourse analysis and "language for specialized purposes". *International Review of Applied Linguistics in Language Teaching 17(3),* 189-215

Sellen, A. J. and Harper, R. H. R. (2001). *The myth of the paperless office*. Cambridge, MA: MIT Press

Shankar, K. (2007). Order from chaos: The poetics and pragmatics of scientific recordkeeping. *Journal of the American Society for Information Science and Technology 58(10),* 1457-1466

Shapin, S. (1994). *A social history of truth: Civility and science in seventeenth-century England*. Chicago, IL: University of Chicago Press

Shemmings, D. (1991). *Client access to records: Participation in social work.* Aldershot: Avebury

Shepherd, E. and Yeo, G. (2003). *Managing records: A handbook of principles and practice*. London: Facet

Silver, M. (2006). Introducing abstract reasoning: World of reference and writer argument across disciplines. In K. Hyland and M. Bondi (Eds.), *Academic discourse across disciplines* (pp 75-102). Bern: Peter Lang

Simpson, D. and Graham, S. (2002). Appraisal and selection of records: A new approach. *Comma 2002-1/2*, 51-56

Smart, G. (2008). Ethnographic-based discourse analysis: Uses, issues and prospects. In V. K. Bhatia, J. Flowerdew and R. H. Jones (Eds.), *Advances in discourse analysis* (pp 56-66). London: Routledge

Smith, B., Kohler, J., and Kumar, A. (2004). On the application of formal principles to life science data: A case study in the Gene Ontology. *Proc. DILS 2004,* 79-94

Smith, D. E. (1974). The social construction of documentary reality. *Sociological Inquiry 44(4)*, 257-268

Smith, D. E. (1993). "Literary" and business: "Social problems" as social organization. In J. A. Holstein and G. Miller (Eds.), *Reconsidering social constructionism: Debates in social problems theory*. (pp 327-347). New York, NY: Aldine de Gruyter

Sperber, D. and Wilson, D. (1995). *Relevance: Communication and cognition* (2nd Edition). Oxford: Blackwell

Spinuzzi, C. (2003). *Tracing genres through organizations: A sociocultural approach to information design*. Cambridge, MA: MIT Press

Spiro. R. J. (1980). *Schema theory and reading comprehension*. Research Report 191. Urbana, IL: Center for the Study of Reading

Steinle, F. (2003). The practice of studying practice: Analyzing research records of Ampère and Faraday. In F. L. Holmes, J. Renn, and H.-J. Rheinberger (Eds.), *Reworking the bench: Research notebooks in the history of science* (pp 93-118). Dordrecht: Kluwer

Strauss, A. and Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd Edition). Thousand Oaks, CA: SAGE

Sturdy, S. (1998). Reflections: Molecularization, standardization and the history of science. In S. de Chadarevian and H. Kamminga (Eds.), *Molecularizing biology and medicine: New practices and alliances 1910s - 1970s* (pp 273-292). Amsterdam: Harwood Academic

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge: Cambridge University Press

Swain, M. (2006). Verbal protocols: What does it mean for research to use speaking as a data collection tool? In M. Chalhoub-Deville, C. Chapelle, and P. Duff (Eds.),

Inference and generalizability in applied linguistics: Multiple perspectives (pp 97-113). Amsterdam: John Benjamins

Swales, J. M. (1990). *Genre analysis: English in academic research settings*. Cambridge: Cambridge University Press

Swales, J. M. (1996). Occluded genres in the academy: The case of the submission letter. In E. Ventola and A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (pp 45-58). Amsterdam: John Benjamins

Swales, J. M. (1998). *Other floors, other voices: A textography of a small university building*. Mahwah, NJ: Lawrence Erlbaum

Swales, J. M. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press

Swales, J. M. and Feak, C. B. (1994). *Academic writing for graduate students: A course for nonnative speakers of English*. Ann Arbor, MI: University of Michigan Press

Swarts, J. (2006). Coherent fragments: The problem of mobility and genred information. *Written Communication 23(2),* 173-201

Tabard, A., Mackay, W., and Eastmond, E. (2008). From individual to collaborative: The evolution of Prism, a hybrid laboratory notebook. *Proc. CSCW 2008,* 569-578

Tarone, E., Dwyer, S., Gillette, S., and Icke, V. (1998). On the use of the passive in two astrophysics journal papers: With extensions to other languages and fields. *English for Specific Purposes 17(1),* 113-132

Taylor, K. T. (2006). The status of electronic laboratory notebooks for chemistry and biology. *Current Opinion in Drug Discovery & Development 9(3),* 348-353

Taylor, K. T., Essex, J. W., Frey, J. G., Mills, H. R., Hughes, G., and Zaluska, E. J. (2006). The Semantic Grid and chemistry: Experiences with CombeChem. *Journal of Web Semantics 4(2),* 84-101

Timmermans, S. and Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Philadelphia, PA: Temple University Press

Titscher, S., Meyer, M., Wodak, R., and Vetter, E. (2000). *Methods of text and discourse analysis* (Translated by B. Jenner). London: SAGE

Thompson, P. (2001). *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs*. PhD Thesis, University of Reading

Traweek, S. (1988). Beamtimes and lifetimes: The world of high energy physics. Cambridge, MA: Harvard University Press

Trigg, R. H. and Bødker, S. (1994). From implementation to design: Tailoring and the emergence of systematization in CSCW. *Proc. CSCW 1994*, 45-54

Tse, P. and Hyland, K. (2006). Gender and discipline: Exploring metadiscourse variation in academic book reviews. In K. Hyland and M. Bondi (Eds.), *Academic discourse across disciplines* (pp 177-202). Bern: Peter Lang

Tweney, R. D. and Gooding, D. (Eds.) (1991). *Michael Faraday's "chemical notes, hints, suggestions and objects of pursuit" of 1822.* London: Short Run Press

Upward, F. (1996). Structuring the records continuum, Part 1: Postcustodial principles and properties. *Archives and Manuscripts 24(2),* 268-285

Upward, F. (1997). Structuring the records continuum, Part 2: Structuration theory and recordkeeping. *Archives and Manuscripts 25(1),* 11-35

Upward, F. (2000). Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond – A personal reflection. *Records Management Journal 10(3),* 115-139

van Dijk, T. A. (Ed.) (1997). *Discourse studies: A multidisciplinary introduction* (2 vols.). London: SAGE

Valero-Garcés, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish-English economics texts. *English for Specific Purposes 15(4),* 279-294

Valle, E. (1999). A collective intelligence: The life sciences in the Royal Society as a scientific discourse community, 1665 - 1965. *Anglicana Turkuensia 17*, Turku: University of Turku

Vandenbroucke, J. P. (1998). On not being born a native speaker of English. *British Medical Journal 298(6685),* 1461-1462

Ventola, E., Shalom, C., and Thompson, S. (Eds.) (2002). *The language of conferencing*. Frankfurt: Peter Lang

Walne, P. (Ed.) (1988). *ICA Dictionary of archival terminology* (2nd Edition). Munich: K G Saur

Warschauer, M. (2002). Networking into academic discourse. *Journal of English for Academic Purposes 1(1)*, 45-58

Watson, J. D. (1990). The Human Genome project: Past, present, and future. *Science 248(4951),* 44-49

Weissberg, R. (1993). The graduate seminar: Another research process genre. *English for Specific Purposes 12(1),* 23-35

Wellner, P. (1993). Interacting with paper on the DigitalDesk. *Communications of the ACM 36(7),* 87-96

Wickman, C. (2010). Writing material in chemical physics research: The laboratory notebook as locus of technical and textual integration. *Written Communication 27(3),* 259-292

Widdowson, H. G. (2004). *Text, context, pretext: Critical issues in discourse analysis*. Oxford: Blackwell

Wiener, N. (1961). *Cybernetics: Or control and communication in the animal and machine* (2nd Edition). Cambridge, MA: MIT Press

Wilson, K. and Walker, J. (2010). *Principles and techniques of biochemistry and molecular biology* (7th Edition). Cambridge: Cambridge University Press

Witte, S. (1992). Text, context, intertext: Towards a constructivist semiotic of writing. *Written Communication 9(2),* 237-308

Wodak, R. and Meyer, E. (Eds.) (2009). *Methods of critical discourse analysis* (2nd Edition). London: SAGE

Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew and M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp 71-83). Cambridge: Cambridge University Press

Woods, P. (1979). *The divided school*. London: Routledge & Kegan Paul

Wulf, W. (1993). The collaboratory opportunity. *Science 261(5123),* 854-855

Yeh, R. B., Liao, C., Klemmer, S. R., Guimbretière, F., Lee, B., Kakaradiv, B., Stamberger, J., and Paepcke, A. (2006). ButterflyNet: A mobile capture and access system for field biology research. *Proc. CHI 2006*, 571-580

Zuckerman, H. (1988). The sociology of science. In N. J. Smelser (Ed.), *Handbook of sociology* (pp 511-574). Beverly Hills, CA: SAGE