Gaining Expertise in Creating Metadata:

An Exploratory Study

by

Samantha Searle

Submitted to the School of Information Management,

Victoria University of Wellington

in partial fulfilment of the requirements for the degree of

Master of Library and Information Studies

1 July 2002

Abstract

Increasingly, people with little experience of cataloguing, indexing or abstracting are using metadata schemas like the Dublin Core Metadata Elements Set (DC) and the New Zealand Government Locator Service (NZGLS) to describe resources. There is currently little information available about how novices approach the process of metadata creation, and what personal (cognitive) and other factors (particularly organisational) are at work. In this exploratory study, I spoke with novice metadata creators about their skills and knowledge when they began to create metadata and, six weeks later, after they had created records as part of their normal work duties. I asked novices to identify factors that impacted positively or negatively upon their progress, and also sought the opinions of metadata experts who were training and supervising novice creators. The study identified the skills and knowledge that are required to create metadata, and investigated the techniques used to develop expertise. The tools used by metadata creators were evaluated, and the effects of organisational culture were also explored. The insights of the expert and novice participants provide guidance as to how managers can facilitate the production of good quality metadata through developing effective staff training and quality assurance, providing more usable online tools and documentation, and fostering more supportive organisational cultures.

Keywords metadata, resource discovery, expertise, Dublin Core, NZGLS

Acknowledgements

This project would not have been possible without the participation of all the interviewees. Thanks to the interviewees themselves, and also to their managers for their permission to conduct research within their organisations.

Thanks also to my supervisor, Dr Dan Dorner, for his support and insight during the project, and to other staff and students in the MLIS programme in 2001-2002.

VICTORIA UNIVERSITY OF WELLINGTON
School of Information Management

Master of Library & Information Studies

IMPORTANT DISCLAIMER

with respect to a MLIS Research Project (LIBR 550)

' Gaining Expertise in Creating Metadata: An Exploratory Study'

(hereafter referred to as 'The MLIS Research Project')

being undertaken by

Samantha Searle

in partial fulfilment of the requirements of the degree of
Master of Library and Information Studies,
School of Information Management,
Victoria University of Wellington.

Topic Commencement: **November 2001**

1. Victoria University of Wellington and its Council, its members, staff, employees, students and agents undertake no duty of care in contract, tort, or otherwise, to users (whether direct or indirect) of the MLIS Research Project and make no warranties or representations of any kind whatsoever in relation to any of its contents.

2. The MLIS Research Project is only made available on the basis that all users of it, whether direct or indirect, must take appropriate legal or other expert advice in relation to their own circumstances and must rely solely on their own judgement and such legal or other expert advice.

3. Under no circumstances will Victoria University of Wellington and its Council, its members, staff, employees, students or agents be liable in any way whatsoever, whether in contract, tort (including negligence), for breach of any statutory or regulatory duty (to the fullest extent permissible by law), or otherwise, to any user (whether direct or indirect) of the MLIS Research Project for any loss or damage whatsoever arising directly or indirectly as a result of the use in any way of the MLIS Research Project.

4. Each exclusion in the clauses of this disclaimer and each protection given by it is to be construed as a separate exclusion applying and surviving even if for any reason any of the exclusions or protections are held inapplicable in any circumstance.

Table of Contents

Chapter One: Introduction

Libraries and other information services face great challenges in providing access to ever-increasing amounts of information, particularly in electronic formats. The need for less complex (and less costly) standards for identifying and retrieving resources has driven the development of metadata that can be embedded in electronic documents by authors or automatically extracted. More complex uses of metadata have also developed: there are now many examples of 'third party repositories' in which metadata is produced centrally and stored in databases external to the resources being described. These repositories are often specific to particular 'domains' such as heritage insitutions (libraries, archives, museums), education and government.

As different models for creating metadata proliferate – particularly in libraries and other repositories that favour more detailed standards and value the concept of bibliographic control – it is perhaps not surprising that quality issues are arising. In these debates, library professionals posit themselves as the natural protectors of information standards. However, as the number and scale of metadata initiatives continues to increase, more and more novices with little experience of metadata or more traditional types of bibliographic description are becoming involved, a trend that seems likely to continue.

In this context, information professionals may need to accept that arguments about the job titles and professional affiliations of metadata creators are increasingly irrelevant. Attention should be turned instead to investigating how novice metadata creators gain the knowledge and skills required for the tasks they face. For this to happen, there needs to be much greater understanding of what personal (cognitive) and other factors (particularly organisational) are at work during the metadata creation process, and how these might be managed. As Greenberg succinctly states: "The question of who should create metadata

now extends to how can the professionally trained metadata creator help other individuals produce quality metadata records in an efficient fashion?"[1]

In this exploratory study, I spoke with novice metadata creators about their skills and knowledge when they began to create metadata and, six weeks later, after they had created a significant number of records as part of their normal work duties. I asked novices to identify factors that impacted positively or negatively upon their progress. The opinions of metadata experts, currently working as the trainers and supervisors of novice creators, were also sought. By gathering the views of people with different levels of experience, I could build a holistic picture of how novices initially approach metadata creation and what types of training and tools are most helpful for them. This picture has been compared with models of cognitive processes that have been developed for other tasks such as cataloguing, indexing and abstracting; similarities and differences have emerged, which could form the basis for further research in this area. I have also identified a number of organisational factors affecting metadata creation; because these factors can to some extent be controlled and changed, these findings should be of immediate interest to managers of metadata initiatives.

[1] Jane Greenberg, "Metadata Questions in Evolving Internet-Based Educational Terrain," *Journal of Internet Cataloging* 3, no. 1 (2000): 7.

Chapter Two: The Research Problem

Need for the research

Resource description is no longer reserved for highly trained information professionals; increasingly, people with little or no prior experience of cataloguing, indexing or abstracting, are describing resources using schemas like the Dublin Core Metadata Elements Set (DC) and the New Zealand Government Locator Service (NZGLS). Metadata creators can be the authors of electronic documents, specialists in a subject area, or the editors or content managers of websites, intranets and portals. A person in an organisation may be randomly allocated the tasks because metadata supply is mandatory and 'someone has to do it'.

There is currently little information available about how novices approach the process of metadata creation, and what activities and tools might make their transition from novice to expert quicker and easier. Identifying aids and obstacles to the acquisition of skills and knowledge could lead to the development of better staff training and quality assurance techniques, more usable online tools and helpful documentation, and more supportive organisational cultures.

Research questions and objectives

The study focused upon the following three questions:

1. What personal (especially cognitive) factors are involved in learning to create metadata?
2. What other external factors influence the metadata production process?
3. What implications arise for the management of the metadata production process?

The study was conducted with the following six research objectives in mind:

1. To identify possible differences in the skills and knowledge required for metadata creation and required for more traditional types of resource description; and
2. To compare and contrast experts' and novices' perceptions of the cognitive processes involved in metadata creation;
3. To gather information about how novices initially approach metadata creation and to discover if and how the process changes over time as more advanced skills and knowledge are acquired and/or developed;
4. To gather information about how knowledge is acquired by novices, in particular how training and supervision facilitates the transfer of knowledge from experts to novices;
5. To gather information about other factors – social, cultural and environmental – that experts and novices perceive as affecting the metadata production process;
6. To identify possible management issues arising from novice metadata creation and to suggest solutions.

Theoretical framework

The exploratory and multi-faceted nature of the study required drawing out themes and issues, rather than aggregating data into a coherent model. A number of theoretical perspectives were used. Jeng and Weiss' model of cataloguing expertise provided a useful framework for analysing skills and knowledge (especially in Chapters Five, Six, Seven and Eight). The authors move from general theories of expertise to a specific model for cataloguing that evolved from a qualitative study at the U.S. National Agricultural Library: this involved analysis of data gained through unstructured interviews, content analysis of documentation and verbal reports. The model identifies five main task categories of expertise: searching bibliographic databases; determining access points; interpreting bibliographic data, rules and products, identifying and prioritising problems and special cases; and administering performance (recruitment, reviews, performance evaluation and

training experience). These task categories were then analysed according to four skill facets: technology (including computer literacy and search skills), professional tools (such as controlled vocabularies and documentation), communication skills, and subject speciality and foreign language expertise.[2]

Interpretation of the data, especially in Chapter Six, also drew upon a cognitive process model of indexing that focuses on the 'technical reading' that is associated with bibliographic activities. Farrow argues that text comprehension in indexing and abstracting is different from that in normal reading: indexers scan text rapidly for perceptual clues rather than reading; they comprehend text only for the purpose of indexing or abstracting, and so their memory or understanding of the text, the usual psychological measures, are not as relevant as for reading; they comprehend text in order to *produce* an abstract, index entries or classification; and automatic processing beyond fluent reading develops through working repetitively with a narrow range of text types and subjects.[3]

A more practical understanding of indexers and abstractors, and in particular, how to evaluate their work, is found in Lancaster's *Indexing and Abstracting in Theory and Practice*. Lancaster's work, which specifically refers to the guidelines supplied by standards bodies and which deals in some detail with common types of difficulties and errors, was particularly useful for Chapter Seven.

Gaining a holistic picture of how metadata creators learn required a broader approach. A number of organisational factors emerged during the analysis of data that do not fit neatly within a single theoretical framework. Theories of situated cognition advocate taking into account the "objects, tools, people and activities," and recognise that there are many factors that may influence the learning process. These theories, which formed a backdrop for the types of issues discussed in Chapters Nine and Ten are discussed further in the next chapter.

---

[2] Jeng and Weiss. "Modeling Cataloging Expertise", 124-26.

[3] Farrow, John J. "A Cognitive Process Model of Document Indexing." *Journal of Documentation* 47, no. 2 (1991): 151.

Definition of terms

Metadata

Metadata is 'data about data', but a more useful definition is that metadata is a structured set of data about an object or document or resource that assists in its identification, discovery and use.[4] There are different types of metadata, including resource discovery metadata, technical metadata and preservation metadata, and schemas are evolving that attempt to combine all of these types of metadata.[5]

This study focuses on resource discovery metadata, which consists mostly of descriptive information but also some information needed to manage resources (e.g. availability, contact details, copyright and other intellectual property rights). Other types of metadata will require different approaches to management and would need to be the subject of further research.

Cognitive processes

Cognition is the "act or process of knowing which includes awareness and judgment, perceiving, reasoning, and conceiving" while *cognitive processes* are the "mental processes involved in the acquisition, processing, and utilization of knowledge or information."[6] Processes such as classification, comprehension, decision making and problem solving are included within this definition. The terms *social cognition* and *situated cognition* refer to

---

[4] Lee-Smeltzer, 206; Arlene Taylor, *The Organization of Information* (Englewood, Co.: Libraries Unlimited, 1999): 77.

[5] See, for example, the Library of Congress and Digital Library Federation's Metadata Encoding and Transmission Standard (METS), "a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium." Library of Congress, *Metadata Encoding & Transmission Standard (METS): Official Web Site* (2002). http://www.loc.gov/standards/mets/ [Accessed 30 June 2002].

[6] Ovid Technologies, *PsycInfo Database Thesaurus*. http://ovid3.distance.scs.vuw.ac.nz/ [Accessed 16 January 2002].

theories of cognition that assume that learning is context-relevant, developed through social and cultural interactions with other individuals in a particular community of practice.

Experts and novices

The study divided subjects into two groups referred to as experts and novices. For the purposes of the study, a novice is someone approaching the metadata creation process with little or no prior experience, while an expert is considered to be someone with both greater theoretical and practical experience of the processes in question. Chapter Four contains a section on the sample that discusses the characteristics of the participants in more detail.

Limitations and delimitations

The strength of qualitative research is the rich descriptive data that emerges, but obtaining this richness of description required limiting the study to a small sample of three novices and three experts. The amount of time available also limited the scope of the study. Interviews with novices might have been conducted further apart, to allow participants more time to gain new skills and knowledge. It is also possible, however, that conducting interviews in a tight timeframe during participants' initial 'learning curve' has produced more interesting data; had the study progressed further, participants' skills and knowledge may have become second nature and less accessible for retrospective analysis. Finally, the study was limited to Wellington geographically. Although this partly relates to time constraints, Wellington is a fertile site for metadata creation and the base for many of New Zealand's metadata-based initiatives.

Organisation of the report

This chapter is followed by a review of the literature (Chapter Three) and a discussion of the research procedures used for the study (Chapter Four). The results of the research are then broken up into six sections (Chapters Five through Ten), which cover Research Objectives 1-5. Chapters Five, Six and Eight relate to Research Question 1, Chapter Seven relates to Research Questions 1 and 2 and Chapter Nine relates to Research Question 2. Research Question 3 and Research Objective 6, which relate to management issues and potential solutions, are dealt with in the discussion sections at the end of each chapter and in Chapter Eleven.

Copies of the recruitment letter, interview schedule and consent form are attached as appendices to the report.

Chapter Three: Review of the Literature

Metadata: Simplicity vs. complexity

Libraries and other information services face great challenges in providing their users with access to ever-increasing amounts of information, particularly in electronic formats. While 'traditional' bibliographic and technical standards such as the Anglo-American Cataloguing Rules (AACR2) and Machine Readable Cataloguing (MARC) have increased bibliographic control and enabled large-scale resource sharing, critics argue that they are too complex for the quantities and types of information now available.[7] The need for a less complicated resource description standard, one that might be used by laypeople such as the authors of electronic documents, has driven development of the Dublin Core (DC) and similar metadata schema.[8]

In contrast to this aim, DC has also "attracted the attention of formal resource description communities such as museums and libraries": some see the schema as flexible enough to encompass the "additional structure and more elaborate semantics" necessary to provide an "economic alternative to more elaborate description models such as full MARC cataloging."[9] The need for a more extended set of elements has also led to the development of qualifiers (sub-elements) and application profiles (data elements drawn from one or more schemas combined together for a particular local application) that are specific to particular domains.[10]

---

[7] Ingrid Hsieh-Yee, "Organizing Internet Resources: Teaching Cataloging Standards and Beyond," *OCLC Systems and Services* 16, no. 3 (2000): 130-43. For another, more polemic example, see Paul F. Burton, "The Decline and Fall of 'Cat. & Class.'," *Catalogue and Index* 124 (Summer 1997): 9.

[8] Stuart Weibel, "The Dublin Core: A Simple Content Description Model for Electronic Resources," *Bulletin of the American Society for Information Science*, (Oct/Nov 1997): 9.

[9] Ibid., 10.

[10] Dublin Core Metadata Initiative, *Dublin Core Qualifiers* (2000). http://dublincore.org/documents/dcmes-qualifiers/ [Accessed 30 June 2002]; Rachel Heery and Manjula Patel, "Application profiles: Mixing and Matching

This tension between the original conception of DC as a simple description scheme for non-experts and, on the other hand, its potential use as a more complex standard, is the subject of ongoing debate. Weibel and others identify a minimalist approach, which "reflects a strong commitment to the notion that DC's primary motivating characteristic is its simplicity" and a structuralist approach, which forsakes simplicity for "the greater flexibility of a formal means of extending or qualifying elements such that they can be made more useful for the needs of a particular community".[11] This debate has been extended by Medeiros, who perceives a shift from DC's original author-centred approach to a three-level system in which any application of DC can be minimal, qualified or "extreme", [12] and by Gorman, who places minimal and 'enriched' DC records in a spectrum of four types of bibliographic control ranging from unstructured keyword searching to full (i.e. MARC/AACR2) cataloguing.[13]

Who should create metadata?

Running parallel to debates about simplicity and complexity are lively, if somewhat under-theorised, arguments about the skills required to create metadata and who possesses them. Metadata initiatives offer new opportunities for information professionals,[14] however, other

---

Metadata Schemas," *Ariadne* 25 (September 2000); http://www.ariadne.ac.uk/issue25/app-profiles/ [Accessed 30 June 2002].

[11] Stuart Weibel, Renato Iannella, and Warwick Cathro, "The 4th Dublin Core Metadata Workshop Report," *D-Lib Magazine* (June 1997). http://www.dlib.org/dlib/june97/metadata/06weibel.html [Accessed 5 August 2001].

[12] Norm Medeiros, "On the Dublin Core Front," *OCLC Systems and Services* 16, no. 1 (2000): 41-43.

[13] Michael Gorman, "Metadata or Cataloging? A False Choice," *Journal of Internet Cataloging* 2, no. 1 (1999): 20.

[14] See, for example, Burton, "The Decline and Fall"; Gorman, "Metadata or Cataloging?"; Kuang-Hwei (Janet) Lee-Smeltzer, "Finding the needle: controlled vocabularies, resource discovery, and Dublin Core," *Library Collections, Acquisitions, & Technical Services* 24, no. 2 (2000): 205-15; Bruce Mcleod, "Factoring the Librarian back into World Wide Web Cataloguing" (paper presented at the Sixth Australian World Wide Web Conference, 12-17 June 2000, Rihga Colonial Club Resort, Cairns, Australia, 2000); Jessica Milstead and Susan Feldman, "Metadata: Cataloging by Any Other Name," *Online* January (1999). http://www.onlineinc.com/onlinemag/OL1999/milstead1.html [Accessed 29 August 2001]; Margaret E. Phillips, "Managing Chaos in the Cyberworld" (paper presented at the CONSAL 2000, the 11th Congress of Southeast Asian Librarians Conference, Singapore, 26th - 28th April 2000); Lloyd Sokvitne, *Cataloguers may yet inherit the earth* (26 March 2001). http://www.servicetasmania.tas.gov.au/papers/catinherit4.htm [Accessed 8 August 2001]; Lloyd Sokvitne, "An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval" (paper presented at the VALA2000

actors in the information environment are routinely cast as obstacles or dangers, and are assumed to be ill-equipped to handle the required tasks.[15]

Although there is no comparative research on metadata created by information professionals and by other groups, the assumption that librarians are innately better suited to this work has distorted otherwise innovative research. An example of this is Sokvitne's analysis of metadata from twenty Australian organisations, which found that "title metadata added little value, creator/publisher access was flawed by inconsistent name formats, and that the subject terms used were too broad to produced increased precision and decreased recall."[16] While using a useful set of evaluative criteria, bias is evident in the sample selection for the study: libraries are excluded because "it was assumed that the experience and expertise of libraries should enable them to provide high quality metadata". Despite recognising that the research design afforded no insight into who actually created the metadata, Sokvitne argues in a circular fashion that "The poor results do seem to verify that trained metadata cataloguers or professional indexers did not create the metadata." Worryingly, his results have been reported since with no qualification.[17]

The assumption that traditional cataloguing, although costly, is always undertaken by library professionals underpins many of these arguments,[18] but LIS research has consistently revealed increasing use of paraprofessional and non-professional staff for both copy and

---

Conference: Books and Bytes: Technologies for the Hybrid Library, Melbourne, Victorian Association for Library Automation, 2000); Alan R. Thomas, "The Work-Wide Web: A Cataloging Job for Every Librarian," *Cataloging and Classification Quarterly* 24, no. 1/2 (1997): 5-22; Robin Wendler, "Branching Out: Cataloging Skills & Functions in the Digital Age," *Journal of Internet Cataloging* 2, no. 1 (1999): 43-54.

[15] See for example, Paul F. Burton, "Issues and Challenges of Subject Access," *Catalogue and Index* 128 (Summer 1998): 5; Milstead and Feldman, "Metadata: Cataloging by Any Other Name"; and Jessica Milstead and Susan Feldman, "Metadata Projects and Standards," *Online* (January 1999). http://www.onlineinc.com/onlinemag/OL1999/milstead1.html#projects [Accessed 29 August 2001].

[16] Sokvitne, "An Evaluation."

[17] Sokvitne, *Understanding metadata*; Margaret E. Phillips, "Managing Chaos in the Cyberworld" (paper presented at CONSAL 2000, the 11th Congress of Southeast Asian Librarians Conference, Singapore, 26th - 28th April 2000).

[18] Gorman, "Metadata or Cataloging?"

original cataloguing work.[19] Calls for library professionals to create metadata also fail to account for difficulties that professionals encounter with some bibliographic tasks. Subject analysis is persistently problematic in the LIS literature, most obviously in studies devoted to intra- and inter-indexer consistency.[20] It seems unfair to posit subject analysis as difficult only for novices, when decades of research has revealed that "indexing is a subjective rather than an objective process" and that consistency is difficult to achieve even amongst professionals.[21] Similarly, with regard to descriptive cataloguing, a study comparing types of errors made by novice cataloguers and errors found in the OCLC and RLIN databases concluded that some areas within a catalogue record cause difficulty because of their nature, and that both beginners and experienced cataloguers were prone to errors to the same degree.[22]

While the limited scope of some of these studies prevents generalising from their results, they serve to qualify blanket statements about the quality of metadata that novices produce. The research described above blurs the boundaries between occupational groups, suggesting instead a focus on tasks and on the factors that make completing those tasks more or less difficult. This emphasis is also evident in reports of library restructuring and metadata projects that have involved technical services librarians cross-training and team-

---

[19] Although earlier research is available, it has been decided to focus only on recent work dealing with this topic: Sever Michael Bordeianu and Virginia Seiser, "Paraprofessional Cataloguers in ARL Libraries," *College and Research Libraries* 60, no. 6 (1999): 532-40; Lynne C. Howarth, "The Role of the Paraprofessional in Technical Services in Libraries," *Library Trends* 46, no. 3 (Winter 1998): 526-39; D.A. Mohr and A. Schuneman, "Changing Roles: Cataloging by Paraprofessionals in ARL Libraries," *Library Resources and Technical Services* 41, no. 3 (1997): 205-18;Lori Proudfit Robare, "Training Paraprofessionals in Classification and Subject Analysis," *Technical Services Quarterly* 14, no. 2 (1996): 49-63; J.A. Younger, "Support Staff and Librarians in Cataloging," *Cataloging and Classification Quarterly* 23, no. 1 (1996): 27-47. The issue has also been dealt with in a special issue of *Library Trends* 46, no. 3 (1998).

[20] A useful summary of this research, most of which is not recent, is provided in F.W. Lancaster, *Indexing and Abstracting in Theory and Practice* (London: Library Association Publishing, 1998), Chapters 5 and 6.

[21] Ibid., 62-3.

[22] Lisa Romero, "Original Cataloguing Errors: A Comparison of Errors Found in Entry-Level Cataloging with Errors Found in OCLC and RLIN," *Technical Services Quarterly* 12, no. 2 (1994): 27.

working with reference and collection development staff:[23] these emphasise the expanded roles that trained librarians might take in an environment where catalogue or metadata records are produced by non-experts. Professionals are now more likely to be involved in training, quality control and management tasks, rather than in day-to-day technical services work.[24] This has been seen as a generally positive step, and suggests that librarians' skills and knowledge may be more efficiently and effectively applied to design, management, quality assurance and training rather than to routine metadata production.

Beyond absolute professionalism to relative expertise

In spite of protests to the contrary, novices *are* capable of creating metadata using schemes such as DC. Examples are emerging in the literature of the successful use of DC as an introduction to cataloguing standards for library students.[25] Recent research by Gilliland-Swetland *et al* into the abilities of fourth and fifth grade school students to create minimal DC metadata concluded that students could achieve a high level of accuracy (over 90%) in supplying descriptive elements (e.g. Title and Identifier) and subjects from a controlled vocabulary. Problems were only identified with elements describing relationships between resources and elements where students were required to write information themselves, areas which involve more advanced conceptual and communication skills.[26] In another recent study, Greenberg *et al.* found that resource authors in an organisational setting could

---

[23] A. Caldwell and D. Coulombe, "Never the Twain Shall Meet? Collaboration Between Catalogers and Reference Librarians in the OCLC CORC Project at Brown University," *Journal of Internet Cataloging* 4, no. 1/2 (2001): 123-30; Karen Calhoun and John J. Reimer, "CORC: New Tools and Possibilities for Cooperative Electronic Resource Description," *The Journal of Internet Cataloging* 4, no. 1/2 (2001); Howarth, 528.

[24] Mohr and Schuneman, 213; Robare, 52, 59-60.; Younger, 37-38.

[25] C. Glaviano, "Teaching an Information Organization Course with Nordic DC Metadata Creator," *OCLC Systems and Services* 16, no. 1 (2000): 33-40; Hsieh-Yee, "Organizing Internet Resources".

[26] A.J. Gilliland-Swetland, Y.B. Kafai, and W.E. Landis, "Application of Dublin Core Metadata in the Description of Digital Primary Sources in Elementary School Classrooms," *Journal of the American Society for Information Science* 51, no. 2 (2000): 199.

unproblematically create DC metadata of equal, and in some cases better, quality to that produced by metadata professionals.[27]

Lack of explicit debate about pedagogy and learning styles

If some aspects of metadata creation are unproblematic to learn, it is obviously necessary to view expertise in this field as relative; if we consider expert cataloguers, indexers and abstractors to be not born but made, we can begin to examine ways in which their skills and knowledge are attained, and can be shared with novice metadata creators.

Unfortunately, there is little in the LIS literature exploring how these types of tasks should be taught (pedagogical aspects) and can be learned (cognitive processes). Literature on basic competencies/skills and education and of technical services training identifies desirable general skills (such as problem recognition and problem solving, decision making, research, and communication) and specific knowledge (of theories about bibliographic organisation, of how information retrieval systems operate, of the AACR2R rules and their interpretations, MARC, subject headings lists and thesauri; and classification systems).[28] However, educational debates are usually couched in terms of 'practice vs. theory', with most research coming to the vague conclusion that some mixture of experience and theoretical knowledge is preferable.[29] Research into how novices are taught is limited to a handful of

---

[27] Jane Greenberg et al., "Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization," *Journal of Digital information* 2, no. 2 (2001); http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/ [Accessed 30 June 2002].

[28] Debra W. Hill, "Requisite Skills for the Entry-Level Cataloger: A Supervisor's Perspective," *Cataloging and Classification Quarterly* 23, no. 3/4 (1997): 75-83; Judy MacLeod and Daren J. Callahan, "Educators and Practitioners Reply: An Assessment of Cataloging Education," *Library Resources and Technical Services* 39 (1995): 153-65.

[29] Bordeianu and Seiser; Judith Clayden, "Theory Versus Practice in Cataloging Education: Some Australian Experiences," *Journal of Education for Library and Information Science* 36 (1995): 230-8; Lynn Silipigni Connaway, "A Model Curriculum for Cataloging Education: The Library and Information Services Program at the University of Denver," *Technical Services Quarterly* 15, no. 1/2 (1997): 27-41; MacLeod and Callahan, "Educators and Practitioners Reply: An Assessment of Cataloging Education," 153-65.

case studies exploring experiential learning,[30] active learning,[31] and more formal instruction.[32] Although gaining experience in the workplace is widely perceived as the most common and effective mode of learning, the pedagogic implications of teaching by example and giving feedback on work have not been explored. The need for further work on pedagogy is also highlighted by a recent examination of thinking preferences of library students that revealed that the majority were likely to be ill-suited to the analytical and technical aspects of cataloguing.[33]

Models from cognitive psychology and expert systems research

Models from psychology have influenced LIS research into information seeking and searching behaviour, but research into the cognitive processes of those *creating* surrogate information is rare. In 1994, Milstead described cognitive processes in indexing as "the most important need for research [...,] one that appears never to have been directly addressed,"[34] and a recent literature review suggested this need has still not been fully satisfied.[35] Nevertheless a small but influential body of work has developed, which draws upon cognitive psychology to explain how cataloguing,[36] indexing,[37] and to a lesser extent, abstracting[38] and classification[39] are undertaken.

---

[30] LeAnn Garnett, "Dewey, Dale and Bruner: Educational Philosophy, Experiential Learning and Library School Cataloging Instruction," *Journal of Education for Library and Information Science* 38 (1997): 129-36.

31 Lisa Romero, "The Cataloging Laboratory: The Active Learning Theory Applied to the Education of Catalogers," Cataloging and Classification Quarterly 21, no. 1 (1995): 3-17.

[32] Robare, "Training Paraprofessionals."

[33] Ann-Louise De Boer, "Teaching Cataloguing and Classification at the University of Pretoria: Thinking Preferences of Second Year Students," *Libri* 51, no. 2 (2001): 14-23, 121-22.

[34] J.L. Milstead, "Needs for Research in Indexing," *Journal of the American Society for Information Science* 45, no. 8 (1994): 78.

[35] James D. Anderson and Jose Pérez-Carballo, "The Nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part I: Research, and the Nature of Human Indexing," *Information Processing and Management* 37, no. 2 (March 2001): 231-54.

[36] Jeng, Ling Hwey, and Karen B. Weiss. "Modeling Cataloging Expertise: A Feasibility Study." *Information Processing and Management* 30, no. 1 (1994): 119-29; Jeng, Ling Hwey. "Using Verbal Reports to Understand Cataloging Expertise: Two Cases." *Library Resources and Technical Services* 40, no. 4 (1996): 343-58; Sauperl, Alenka, and Jerry D. Saye. "Pebbles for the Mosaic of Cataloging Expertise: What Do Expert Systems for

Of particular note is Farrow's cognitive process model of indexing and Jeng and Weiss's model of cataloguing expertise: these were discussed above in the section on the theoretical framework for this study. Jeng and Weiss's model evolved from Jeng's earlier work which focused upon the visual and linguistic cues of bibliographic data and the formation of 'public' knowledge presented in cataloguing rules and standards.[40] Jeng and other researchers are now revisiting cognitive factors in the context of revitalised attempts to develop expert systems (ES) for bibliographic work.[41] A great deal of LIS research on this topic is available, mostly reporting the results of prototyping projects for cataloguing ES. Successful research has analysed automatic recognition of document characteristics and the use of AACR2R rules as a structure for the knowledge base of systems, but researchers have encountered difficulties with *knowledge acquisition*, the process of "eliciting, analyzing, and interpreting the knowledge which a human expert uses when solving a particular problem, and then transferring this knowledge into a suitable machine representation."[42] In short, cataloguing is a more complex, ambiguous and non-linear task than was expected, and accessing the tacit knowledge of cataloguers more difficult. Gaining further knowledge of

---

Cataloging Reveal about Cataloging Expertise?" *Library Resources and Technical Services* 43, no. 2 (1999): 78-94.

[37] Anderson and Pérez-Carballo; A. Bertrand and J-M. Cellier, "Psychological Approach to Indexing: Effects of the Operator's Expertise upon Indexing Behaviour," *Journal of Information Science* 21, no. 6 (1995): 459-72; C. David et al., "Indexing as Problem Solving: A Cognitive Approach to Consistency" (paper presented at the Annual Conference of the Canadian Association for Information Science: Connectedness: Information, Systems, People, Organizations., Albert University School of Library and Information Studies, Edmonton, Alberta, 1995); Farrow.

[38] B. Endres-Niggemeyer and E. Neugebauer, "Professional Summarizing: No Cognitive Simulation Without Observation," *Journal of the American Society for Information Science* 49, no. 6 (1988): 486-506; M. Pinto and F.W. Lancaster, "Abstracts and Abstracting in Knowledge Discovery," *Library Trends* 48, no. 1 (1999): 234-48.

[39] Clare Beghtol, "Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents," *Journal of Documentation* 42, no. 2 (1986): 84-113.

[40] Jeng and Weiss. "Modeling Cataloging Expertise", 122-23.

[41] Anderson and Pérez-Carballo; Endres-Niggemeyer and Neugebauer; Judy Jeng, "Expert System Applications in Cataloging, Acquisitions and Collection Development: A Status Review," *Technical Services Quarterly* 12, no. 3 (1995): 17-28; William Olmstadt, "Cataloging Expert Systems: Optimism and Frustrated Reality," *Journal of Southern Academic and Special Librarianship*, (2000); http://southernlibrarianship.icaap.org/content/v01n03/olmstadt_w01.html [Accessed 4 January 2002].

[42] Kidd, 1987, qtd in Jeng, "Expert System Applications, 25.

human expertise in cataloguing is now acknowledged as necessary if ES developments are to continue.[43]

Situated cognition: Other factors in metadata creation

Although cognitive processes play a part in metadata creation, these need to be placed in the context of other factors. Theories of *situated cognition* respond to perceived shortcomings in cognition-based analyses, attempting to "shift the focus from the individual as the unit of analysis towards the sociocultural setting in which activities are embedded."[44] Situated cognitionists reject the 'acquisition of knowledge' theory of learning, which proposes that learning is a specific cognitive act separate from doing something. Instead, they argue learning occurs through "opportunities to develop a practice derived from the multiple and varied (but not infinitely varied) circumstances of daily activity."[45]

Theories of situated cognition appear in LIS literature in the context of user education and web-based information provision: the emphasis on the use of 'real world' problems and learning as a group process seems to appeal to librarians with responsibilities for literacy and information literacy strategies. On the technical services side, however, the theories appear to have had little attention. This lack of crossover is disappointing, since situated cognition models seem to correspond with 'common sense' accounts of how indexing, cataloguing and abstracting are learned. Accounts of tailoring apprenticeships,[46] and work drawing on sociocultural analyses of scientists in their workplaces offer interesting points of intersection: one only has to replace "working scientists" with "working cataloguers" in the following paragraph to see parallels:

---

[43] Olmstadt, "Cataloging Expert Systems".

[44] David Kirshner and James A. Whitson, "Editors' Introduction," in *Situated Cognition: Social, Semiotic and Psychological Perspectives*, ed. David Kirshner and James A. Whitson (Mahwah, NJ: Lawrence Erlbaum Associates, 1997): 5.

[45] Jean Lave, "The Culture of Acquisition and the Practice of Understanding," in *Situated Cognition: Social, Semiotic and Psychological Perspectives*, ed. David Kirshner and James A. Whitson (Mahwah, NJ: Lawrence Erlbaum Associates, 1997): 27.

> ... working scientists do not just apply general theories and deduce neat and logical conclusions from them. Rather, they coordinate themselves (in body and mind) with instruments, tools, symbolic and linguistic expressions, other people, objects being studied, and places like laboratories so as to chain all these together into a coherent pattern or configuration that works. Here we see patterns not in the mind, but actual material configurations of objects, tools, people and activities in the world.[47]

A more situated account of metadata creation should therefore need to take into account these "objects, tools, people and activities," recognising that there are many factors that may influence the process.

Conclusion

Novice metadata creators are required to create descriptions using relatively well-defined metadata schemas. Although library professionals are reluctant to accept that novices can produce high quality metadata, there is evidence that learning to create metadata is relatively easy compared to more complex standards. Exactly how that learning can be facilitated is unclear: a small but influential body of literature on cognitive processes in indexing, cataloguing and abstracting provides some insights into how novices' skills and knowledge might develop. Further results may emerge as researchers revitalise interest in expert systems. Theories of situated cognition provide a counterpoint to the individualistic emphasis of cognitivism; having gained attention in LIS as models for *user* education, they could equally serve as a starting point for discussing social and organisational aspects of learning to create metadata. Analysis of factors such as training, supervision and feedback, tools and documentation should provide insight into the experience of novices who, while not necessarily trained cataloguers, are already part of initiatives to enhance access to information by creating resource discovery metadata.

---

[46] Lave, 17-35.

[47] James Paul Gee, "Thinking, Learning and Reading: The Situated Sociocultural Mind," in *Situated Cognition: Social, Semiotic and Psychological Perspectives*, ed. David Kirshner and James A. Whitson (Mahwah, NJ: Lawrence Erlbaum Associates, 1997): 235-59.

Chapter Four: Research Procedures

This study was exploratory and qualitative. I aimed to draw out the tacit knowledge of novices and experts rather than quantify their competencies and academic qualifications. Although in some respects, this process was similar to that of knowledge acquisition, in which engineers tap the knowledge base of practitioners in order to develop expert systems, I took a more phenomenological approach: I was interested not only in what novice metadata creators think and do, but also in the meanings that they attach to their work. As a recent article on qualitative methodologies suggests:

> Many claim that they learn from experience but find it difficult to articulate what has inspired the learning. Delineating the parameters of experience that either inform or serve no useful prompts at all for understanding requires a reflective mode. Becoming reflective in this way suggests delving into the enigmatic and puzzling conditions of real practice. Questioning ordinary work may uncover surprising meanings and inspire unexpected insights.[48]

Sample

The sample was determined by the existing settings and programmes in which metadata was being created in Wellington. The sample was both *purposive*, in that I recruited participants because they had characteristics relevant to the project, and *stratified*, in that I recruited sub-samples of experts and novices.[49]

I identified and recruited three (3) experts through personal contacts within the metadata community. These three participants represented a range of organisational perspectives, having current posts or previous experience in libraries (national and special), archives, government agencies, educational organisations and consultancy. All three were involved to some degree in training and supervising novices. They all had experience of

---

[48] Jinx Stapleton Watson, "Making Sense of the Stories of Experience: Methodology for Research and Teaching," *Journal of Education for Library and Information Science* 42, no. 2 (2001): 139.

[49] G.E. Gorman, Peter Clayton, and with contributions from Mary Lynn Rice-Lively and Lyn Gorman, *Qualitative Research for the Information Professional* (London: Library Association Publishing, 1997), 127-28.

metadata at a strategic level; for example, being involved in the development of NZGLS or participating in international Dublin Core development. At a practical level, all three had also worked in organisations in which a metadata schema had been practically implemented. They were therefore able to provide insight not only into the ways that metadata creation might ideally be achieved, but also into the obstacles and rewards that metadata creators and their managers find at the coalface.

Three (3) novices were recruited in consultation with agencies involved in creating metadata. Subjects had little or no experience of metadata creation and little or no prior knowledge of similar tasks or processes (e.g. cataloguing, indexing, classification, abstracting). Despite the small number of participants, it was possible to gain a broad picture since the roles of individual participants were different and the use of metadata in their agencies varied.

The organisations that participants worked in represented two very different approaches to the creation and use of metadata records. One organisation was responsible for a domain-specific third-party metadata repository built around qualified DC. The scope of this service was very well-defined. Metadata creation was a highly formalised and quality-controlled core business process usually completed by an on-site team of metadata creators. Previous knowledge and work experience in the domain meant that metadata creators had a high awareness of their users' needs, and face-to-face and online feedback was constantly sought. A metadata creation tool/interface, documentation, and training procedures had all been developed in-house.

Two other organisations were public sector agencies, in which a distributed network of individuals and small groups created NZGLS metadata about their own agencies' corporate resources and services. Novice creators received centralised training and helpdesk support, and submitted their metadata using a common web-based interface and centralised quality assurance (QA) process before going live. The functions of the agencies themselves had an impact on the number and type of metadata records. One agency provided information resources but few direct services to the public, so both the size of the agency and its scope

were quite small. The other agency focused on providing information resources and services to a much larger audience, and ensured, through community liaison activities, that groups such as Mäori and Pacific Islanders were able to fully access its services and resources.

The use of two groups, experts and novices, and the deliberate inclusion of different organisational structures for the creation of metadata added value to the project. Although small, the purposive nature of the sample enabled the comparison and contrast of a wide variety of metadata-related practices in a range of workplaces. Data gained from experts supports and validates that gained from the novices and vice versa, while disparities of viewpoints between the two groups highlights areas of contention that might be productively explored in more large-scale research in the future.

Data collection

The most popular and well-tested method for obtaining data about cognitive processes is concurrent verbal reports; subjects talk aloud as they work through a task, with or without the direct participation of the researcher. Although this procedure has been successful in previous studies of cataloguing and indexing,[50] it was not used in this study. Following discussions with the managers of the participating novices, I was concerned that the obtrusive nature of the 'talk aloud' procedure might cause undue stress to the novice metadata creators, who were already under considerable pressure to incorporate new skills into their existing workloads and institutional programmes. A second factor in this decision was David *et al*'s view that concurrent verbal reports cannot provide access to data about "external constraints not immediately visible as the task is being accomplished."[51] Since I was equally interested in these external factors, I needed to find an alternative method.

Data for the study were gathered in three sets of interviews. In the first set, I interviewed metadata experts, seeking their opinions on the competencies required for

---

[50] Jeng, "Modeling Cataloging Expertise"; Jeng, "Using Verbal Reports".

[51] David et al., "Indexing as Problem Solving".

metadata creators, the ways in which novices gain expertise, and effective methods for transferring knowledge to novices through training and supervision. This set of interviews were *dialogic*, in that I "engage[d] in a genuine dialogue and mutually explore[d] the survey topic" with the experts.[52] This interview style, which is regarded as suitable for professional and expert subjects since it avoids artificiality and loss of rapport, worked well.

The second and third sets of interviews were with novice metadata creators. I interviewed novices twice: around the time of their introduction to metadata creation and six weeks later, when they had created a significant number of metadata records. These interviews were semi-structured, and participants were provided beforehand with a schedule. Interviewees were also invited to bring copies of metadata records to the second interview, if these demonstrated particular areas of ease or difficulty. This technique corresponds to what David *et al* describe as *retrospective verbalisation*, "showing subjects the traces... of their own behaviour while questioning them about their interpretation of the process."[53] It was unfortunate that only one of the three novices chose to do this, since the method was more successful in drawing out detailed comment about elements within a metadata record than interviewing alone.

Both sets of interviews included topics for discussion such as training, documentation, use of tools such as thesauri and online metadata creation interfaces, knowledge of the subject area, peer communication networks, understanding of the resources being described and perception of users and their information needs.

Consent to undertake interviews was gained from senior managers of each of the organisations in which the study was conducted, and from the line managers of novices.

---

[52] Margaret Slater, "Qualitative Research," in *Research Methods in Library and Information Studies*, ed. Margaret Slater (London: Library Association, 1990): 115.

[53] David et al., "Indexing as Problem Solving".

Treatment of the data

In accordance with human ethics procedures, interviewees consented to the taping and transcribing of all interviews, and were provided with the opportunity to check their transcripts for any factual errors introduced during transcription. I then analysed the content of the transcripts for common themes and issues that related to the research questions and objectives.

Chapter Five: Requisite Skills and Knowledge

Chapters Five, Six and Seven of this report explore the *personal factors* involved in the creation of metadata. This chapter relates to Research Objective 1, in that skills and knowledge required for metadata creation are identified, as well as compared and contrasted with the skillset required for more traditional types of resource description. The types of job roles that metadata creators fill are also discussed. This chapter also lays the groundwork for Chapter Six, which focuses on the cognitive processes involved in creating metadata.

Skills and knowledge required for metadata creation

Participants identified a number of skills, areas of knowledge/experience, and personal qualities or approaches that were key to successful metadata creation. These are summarised in Table A below; options are ranked according to the number of participants that noted that option, with the more popular options at the top of the each table.

	Experts			Novices		
Key skills/knowledge/experience	**A**	**B**	**C**	**A**	**B**	**C**
Knowledge of subjects/functions	X	X	X	X	X	X
Communication (written)	X	X	X	X	X	X
User perspective	X	X	X		X	X
Information technology	X	X	X	X		X
Communication (verbal)	X		X		X	X
Pays attention to detail	X	X	X		X	
Flexible/pragmatic	X	X	X		X	
Outward-looking		X	X		X	X
Information/content management	X	X	X			
Cataloguing skills	X	X	X			
Search skills	X	X	X			
Overview of information system/environment	X	X	X			
Knowledge of metadata standard/s		X	X	X		
Technical reading skills	X	X			X	
Independent judgement/problem-solving				X	X	
Methodical	X				X	
Consistent	X	X				
Tolerant of routine/repetitive tasks	X			X		
Reflective/analytical		X			X	

Table A: Key skills, areas of knowledge/experience, and qualities

Subject/function knowledge

Specialisation is acknowledged as a desirable characteristic for a cataloguer or indexer: this specialisation can be within a discipline, with a particular type of material or format (e.g. non-books or serials) or can relate to foreign language proficiency.[54] In this study, all participants, both expert and novice, also expected that metadata creators would bring existing specialised knowledge to the process.

For domain-specific subject portals, both experts and novices considered knowledge of the subject area/s or domain vital, and this knowledge was weighted more heavily in recruitment than prior experience of metadata-related tasks (Expert A). The situation was different in settings where metadata was created for an organisation's own resources. Experts asserted the need for knowledge not of a *subject* area but rather of the work of the organisation; in particular knowledge of the *functions* of different business units within the organisation (Experts B and C). This type of knowledge was identified by Experts B and C as different from that traditionally required for cataloguing.

A further layer of functional knowledge was noted by Novices A and C, who thought metadata creators should be aware of the areas of intersection between the functions of their own organisation and those of external organisations. While Novices A and C noted this was particularly relevant to the e-government situation where a distributed larger network of organisations was contributing metadata to a central portal, the need to be outward-looking was generally seen as being a requisite for metadata creation (two experts, two novices).

Communication skills

Jeng's model of cataloguing expertise contains communication as a skill facet.[55] The value of written communication skills is not overly emphasised in the LIS literature in relation to cataloguing. This may be because standards such as AACR2R impose a strict syntax upon

---

[54] Jeng and Weiss, 126.

[55] Ibid., 124.

cataloguing data that is quite different from 'everyday' language. The importance of writing skills in abstracting are more defined: Lancaster argues that a good abstract should be accurate in representing the content of the resource, concise and clear (non-ambiguous).[56]

All participants in this study strongly stressed the need for writing skills. Particular aspects of good writing that were identified were the use of the active rather than passive voice, the development of a consistent vocabulary and the avoidance of redundant words, ambiguous language and jargon. None of these aspects are unique to metadata creation: all are discussed by Lancaster, drawing upon the US national standard for abstracting.[57]

The type of writing required depended upon the organisation. The domain-specific web portal environment required objectivity (Expert A), while the ability to place a particular 'spin' on the Description element within a metadata record was valued within organisations contributing information about their own resources and services to the e-government portal (Expert C). These differences correspond to the typology of abstracts outlined by Lancaster: he notes that 'slanting' is more relevant and feasible where services are pitched to a heterogeneous and less specialised audience.[58]

The value of good writing was usually discussed in terms of user needs. Novice C specified that metadata creators needed to be able to speak the language of their users: "We go out of our way to explain things in plain English, for the common person, because that's how we're trained [....] I observed people sitting there writing in their jargon, thinking 'God, if I was trying to access that information, there's no way I'd get in that way!'"

Unlike written communication, verbal communication skills have been identified as important for cataloguers. Hill argues that "Contrary to popular belief, cataloguers are not eccentric loners who have no need to communicate with their colleagues within and outside their departments."[59] Yet the stereotype that Hill describes was supported by the study, with

---

[56] Lancaster, 107.

[57] Ibid., 107-08.

[58] Ibid., 97.

[59] Hill, 81.

Experts B and C both asserting that many cataloguers would not have experience of the proactive liaison required for some types of metadata:

> The traditional wisdom about cataloguers is that they tend to be backroom people, and this type of work means that you've got to move to the front [....] Communication, *oral* communication and listening skills, are competencies that probably haven't been highly weighted competencies for the librarians here in the past. And that's just because they've not had to go out and do that, they've got the information there in the items they catalogue. (Expert C)

User perspective

According to Lancaster, indexers "need to know much more than the principles of indexing. In particular, they must be thoroughly familiar with the interests of the community served and the information needs of that community."[60] This 'user-centred' approach[61] was supported by the study, with five of the six participants (three experts, two novices) stressing the need for metadata creators to approach their work from the perspective of users: "They're focused primarily on discovery, because you only describe in order to discover. So that's the prime thing, how would someone else find this? And they should have the ability to take more than one viewpoint, so someone might see it this way, somebody else might see it that way. You know, there's more than one customer out there" (Expert A).

I noted above that user needs were described most often in terms of the language that should be used in the metadata record; user perspectives also arose in discussions about the 'digital divide' and access to electronic information. Ways of developing a better understanding of users were also discussed in the context of feedback, which is explored further in Chapter Nine, but it was generally agreed that some knowledge of users and the ability to approach metadata creation from their perspectives were required right from the beginning.

---

[60] Lancaster, 10.

[61] See also R. Fidel, "User-Centered Indexing.," *Journal of the American Society for Information Science* 45, no. 8 (1994): 572-76.

Information technology (IT)

Technology is one of the skill facets noted by Jeng in her model of cataloguing expertise.[62] Five of the six participants in the study (three experts, two novices) stressed that IT skills were key for metadata creation. The level of IT knowledge required was not described as high. 'Familiarity' and 'comfort' with technology were required (Expert A, Novice A), while more advanced skills in programming and web development skills were desirable but not always necessary (Expert A).

The need for IT skills appeared to result partly from problems with the interfaces used. Novice A commented that he "could imagine a lot of people here [in his organisation], for example, having real trouble with it. Finding it really frustrating because they weren't prepared to go and click on something to find out." Interface design issues are discussed further in Chapter Eight.

Information management: general and specific skills

As noted in the literature review, professional literature for librarians and other information professionals consistently highlights the relevance of library-related skills to the developing field of metadata. However, a gap between the importance of many library-related skills to metadata as perceived by librarians and as perceived by other agents in the information environment emerged during interviews. As Table A shows, generic information management skills were thought by all three experts to be key skills. Expert A's comment that "some library training, library education, whatever you like to call it, would be useful. It's about deciding what is good information and how to describe it really" is representative of the view of all three experts. Expert B qualified her view, noting that information professionals also needed "to stay in touch with the people who know the information."

---

[62] Jeng and Weiss, 124.

Experts also identified an overview of the information system or environment as a factor in being able to create good metadata. Awareness of the underlying system design was one part of this overview:

> I think the difference is that if you're used to developing these systems, then when you're filling [in a metadata record], you're putting them in the context of the earlier design work.. This connects up your choices with a larger use, a larger purpose, and I can see the relationships between what is being asked for or I can't. There is that kind of analysis going on. Also, I guess because I am seeing what the designer is trying to do, then I am trying to comply. If you don't have that background, then it can be easy to misunderstand or subvert the intent of the design for capturing useful metadata. (Expert B)

Good search skills were also seen as being part of this overview: according to the experts, metadata creators should not only know how to input metadata, but have an understanding of how metadata is retrieved, so that they can "try to anticipate the behaviour of people" (Expert A). This emphasis on retrieval and user needs is slightly different from Jeng's identification of searching as one of the five key task categories for cataloguing expertise: Jeng's focus is on the cataloguer's own retrieval and evaluation of copy records from bibliographic utilities.[63]

Prior knowledge of the metadata standard itself was noted by two experts and one novice as being important, and a large gap was evident in the importance placed by experts and by novices on cataloguing skills, with all experts and no novices seeing these as valuable. In contrast to the literature reviewed earlier, which assumes that metadata creation is "cataloguing by any another name," [64] experts who had identified cataloguing experience as important still had reservations about cataloguing being seen as the primary skill that was required. Expert B argued that cataloguers may be reluctant to change their practices and might "struggle with a different paradigm, something like NZGLS." Expert C noted that metadata creation required a flexible and pragmatic approach that did not always sit comfortably with the ethos of accuracy and completeness that characterises much

---

[63] Ibid., 125.

[64] Milstead and Feldman, "Metadata: Cataloging by Any Other Name."

cataloguing: “The skills that are required for those people, in my view, is to have a broad perspective and broad understanding of what the purpose of it is and it’s not the same as [cataloguing]. They have to be flexible and the records need to be fit for purpose.”

Although these general and specific information management skills were not explicitly noted by any of the novices, it should not be inferred from this that they do not have these skills; the approach of novices is discussed further in Chapter Seven.

Reading skills

Three participants (two experts, one novice) thought the ability to quickly read and comprehend the ‘gist’ of what was being read was important. This confirms basic models of bibliographic activities, which stress that subject indexing requires a phase of *conceptual* analysis, the examination of a document to determine what it is about (sometimes called ‘technical reading’).[65] In particular, the *speed* of the metadata creator in reading the document was identified as key:

> Good at reading things *quickly*. That’s another skill. Reading things *quickly*, picking what’s important out of something and *quickly* describing it and summing it up basically. I think that’s the other important thing. (Expert A, emphasis added)

> You need to be... quite focused and a good reader, be able to *draw out information quite quickly, pull it together, write it up and then onto the next* metadata [record]. (Novice B, emphasis added)

This corresponds to an aspect of the cognitive process model of indexing outlined by Farrow, who argues that the reading done by indexers differs from other types of reading (e.g. leisure, learning) in that time constraints enable only scanning rather than reading of the text. Other cognitive aspects of this ‘technical reading’ process are taken up in Chapter Six.

---

[65] Lancaster, 8.

The job roles of metadata creators

I noted in the literature review that some professional and scholarly literature suggests that the most appropriately qualified people to create metadata are library and information professionals, cataloguers in particular. I was therefore interested to gain a sense from participants in the study of the types of job roles that metadata creators typically fill within organisations. In the context of skills and knowledge they believed were required, I also wanted to elicit the opinions of experts and novices on what types of professional roles equipped people for creating metadata.

A number of interesting themes emerged. The first of these was that although experts identified information management skills, both generic and specific, as important, they did not believe that only cataloguers or other information professionals could create good metadata. Expert A identified subject specialists, as well as journalists and other writers, as potential metadata creators. Expert B argued that business analysts were well-equipped to deal with the tasks because of their knowledge of the organisation, and that the outward-looking perspective of corporate communications specialists was beneficial.

A second issue that emerged was the sidelining of information professionals within organisations. In his second interview, Novice A stated, “I can also see, in hindsight, we really should have gotten our librarian into this.” When I asked why this had not happened initially, he responded: “I don’t know! I guess there’s a general misunderstanding within lots of organisations about the roles of librarians and they aren’t thinking. They think it’s an IT thing rather than thinking it’s a cataloguing thing or a data thing.” This position was backed up by Expert C, who blamed organisational structures in which libraries are considered add-ons or cost centres. She argued that librarians needed to be more proactive: “The people who’ve been given this work are the communications team or the IT team, and they don’t understand content management. And I think that’s an area that needs to be addressed by the information management people within an organisation, and that they have to put their

hands up and say, we're able to do this and it's core business work" These organisational issues are discussed further in Chapter Ten.

The third theme that emerged draws upon the two already noted. Because information professionals have (or are *perceived* within their organisations as having) some, but not necessarily all, of the skills and knowledge required for creating metadata, it can be useful to bring together teams of people with the required skills, rather than making metadata creation the responsibility of one person or one business unit. Expert C thought that a useful synergy was produced through merging the library and IT areas in some organisations; this synergy could be further enhanced through working with communications specialists. Expert A noted that in the subject portal environment, subjects specialists were vital, but that someone "from a more library-related discipline" was also required: "Given that it's pretty difficult to find people who are both, we balance it out, and you weigh up the pros and cons. People who are subject specialists can by and large be trained to do that sort of thing and then I guess as a specialist in the other aspects of things I then give it the once-over" (Expert A). This quote indicates the validity of a two-tier approach to metadata creation, with information professionals providing internal or external quality assurance (QA) of metadata created by a range of other people with the knowledge of the subject area or functions of the organisation: this is discussed further in Chapter Nine's section on QA.

In general, participants seemed to believe that knowledge and skills were more important than job titles: "I don't think it matters if it's a receptionist or a policy analyst really.... [It's about] the knowledge that you have of what your branch does, and the key is to take that knowledge and turn it into something that makes sense [for the user]." This supports the findings of Greenberg *et al*'s study, in which

> [novice] participants expressed that they are obviously knowledgeable about their work, that they know their immediate and often potential audiences, and that they are aware of the way in which interested people will search for their work. In other words, these participants saw their authorship

> role, together with their command of a discipline's language, as important metadata production factors.[66]

Although other people in the information environment may have relevant skills and knowledge, Expert A also suggested that the common vocabulary and knowledge that develops around job roles determines how metadata creators may be *perceived*:

> I think it's helpful if you've got a library background given that the people you are working with often do, so it doesn't help if you haven't got any, I guess. So that's about how other people perceive you as much as whether you've got any specialist knowledge or not. Other people having confidence in you, so perhaps that's significant. They're pleasantly surprised and much more confident and trusting when they know that person can relate to their issues... they have more confidence when they know it's someone with a like mind that they are talking to.

A team approach to metadata creation can also be of benefit in overcoming this myopic professionalisation: metadata creators become part of a far larger network of different types of professionals if team members can interact with each other as well as move freely within their own communities:

> It's their job to maintain an understanding of the subject areas that they're dealing with and what's the current trends and all that sort of thing. They would be the ones who would go to any meetings or discussions or conferences with other [subject/domain specialists]. They wouldn't send me. And I would be the person who would deal with libraries or other organisations where we're talking about exchanging metadata and all that sort of thing. So we tend to stick to whatever our particular backgrounds are [....] It's not a more or a less thing, it's just the different disciplines and that's the way we organise it. (Expert A)

Discussion

Although LIS literature tends to conflate cataloguing and metadata, the results suggested that creating metadata is *not* 'cataloguing by any other name.' Although creating metadata is similar in some respects to cataloguing, and to indexing and abstracting, it is also a new paradigm for resource description. Concerns were raised that information professionals with a grounding in traditional bibliographic processing, especially cataloguing, may not be the

[66] Greenberg et al.

right people for metadata work unless they could balance attention to detail and their knowledge of standards with the pragmatic requirements of metadata creation. Although the study did not evaluate the actual results of the participating novices, both experts and novices also expressed confidence that novice metadata creators could bring sound skills and knowledge and further develop these on the job.

This suggests that managers need not limit themselves to library professionals in their search for talented metadata creators. They should seek staff with appropriate subject or organisational knowledge, who can demonstrate that they are flexible, pragmatic and outward-looking; metadata creators should also be user-focused and excellent communicators. People with these qualities may not necessarily be cataloguers or any other type of librarian, and could be journalists or other writers, corporate communications specialists, or business analysts.

At the same time, managers should not exclude library professionals, as seems to have happened in a number of organisations where metadata has been perceived as solely an IT or a communications responsibility. Experts believed wholeheartedly that information management skills are required for metadata creation, even if these do need to be combined with other skills and characteristics. Managers must ask themselves, then, how the skills of information professionals are best deployed. The results of the study suggested that a team approach, which included the perspectives of information professionals as well as other types of professionals with subject knowledge and communication skills, could be both efficient and effective.

Chapter Six: Exploration of Cognitive Processes

An objective of this study was to move beyond identifying requisite skills and knowledge to more closely examine experts' and novices' perceptions of the cognitive processes involved in metadata creation. This chapter describes some of the methodological issues that emerged and identifies a number of themes that could be explored in further research.

Methodological issues

In broad terms, all types of bibliographic activity involve describing an 'information package' and selecting access points related to what the document is about. The first stage, descriptive cataloguing, consists of identifying bibliographic elements such as authors, titles, sources, and the extent of the item.[67] The second stage, subject access, is a two-part process: 'conceptual analysis of the content' ("What is it about?") is followed by the 'translation' of the terms in the document into the terms of the information system ("What terms am I allowed to choose that best say what it is about?)"[68]

I noted in my literature review that research into the cognitive processes involved in bibliographic activities is under-developed. Although the model outlined above is generally accepted, there are few specific details of how people actually perform these tasks. The method of verbal reports, also known as the 'talk aloud' method, is one way that researchers have gathered data about the thought processes of working cataloguers. As noted earlier, I rejected this as a means of data collection for this study because the obtrusive nature of the method might have placed unnecessary pressure on participants and because the method's ability to gather information about external factors has been criticised.[69]

Because I also hoped to gather data about external factors, I selected semi-structured interviews as the means of collecting data. I anticipated that details of cognitive processes

---

[67] Lancaster, 1.

[68] Ibid., 8.

[69] David et al.

would emerge in response to open questions about working procedures ("Describe for me how you work through a metadata record") and closed questions such as "Which sections of the documents do you find most useful?" and "Do you work through the record in a linear way, or do you move between different elements in the record?" The interview data, which is discussed below, provided insight into some cognitive points of divergence between metadata creation and other types of bibliographic processing; however, the data were not sufficiently detailed to constitute a coherent model of the cognitive processes involved in metadata creation.

In retrospect, I believe that collecting data about both internal and external factors in the same study was overly ambitious. Although interviews worked well for exploring organisation factors and gaining a useful but impressionistic sense of cognitive differences between metadata creation and other types of process, they were less effective for drawing out detailed descriptions of work practices. Further research, using verbal reporting and focusing specifically on cognitive factors, might balance internal and external factors in a more sophisticated way.

Cognitive processes

Several themes emerged from my analysis of the interview transcripts. The first related to the linear nature of the metadata creation process. Participants stated that they generally worked through a record in the order that the elements appeared. This conforms to a common understanding of the cataloguing process, in which the cataloguer "follows the typical flow of working on a workform [...] in a top-down manner."[70] However, some exceptions to this were noted: three participants (two experts and one novice) described filling in the Description element first at times. Participants also noted occasionally selecting keywords at different times or entering Rights data at the point of discovering it in the resource. These anomalies may indicate a more flexible approach to metadata creation, one identified by Jeng, in which

[70] Jeng, "Using Verbal Reports", 354.

the record is seen as “a workspace that can be used to draft ideas” and store work in progress. Jeng proposes that some cataloguers may “consider the data in the record to be subject to constant change without any fear of losing the integrity of the cataloguing record.”[71] Novice A took this iterative approach: “I always had that sense, and I hope it’s true, that what we were doing wasn’t permanent, that we would be able to go and alter things later.” The interaction between the linear nature of the process and the demands of the interface that creators work with also seemed important: this issue is discussed further in Chapter Eight.

The second theme related to the characteristics of the resources being described. Participants indirectly indicated that characteristics of web resources inflected their cognitive processes in unusual ways. These document characteristics can be described as dynamism, non-standardisation, non-objectivity, adoption of new style standards and lack of boundaries. A further characteristic, which related to the e-services work being undertaken as part of the e-government, was intangibility.

LIS literature has long noted that dynamic web resources pose problems for cataloguers.[72] What is less clear is how descriptive practices might change to accommodate this dynamism: determining the ‘aboutness’ of a changing resource might involve processes quite different from those used for other formats. Expert A provided some insight into this: “You have to be wary of committing to numbers and things, because that will change. A book does only have x number of chapters, but we could say a site has 350 links on it and then another time you’ll go there and there’s not all of those... You have to be cagey about what you put in the Description.”

The non-standard nature of electronic resources also appears in LIS literature.[73] But as with dynamic characteristics, there is little to suggest how non-standardisation might affect

---

[71] Ibid., 355.

[72] Martin Dillon and Erik Jul, "Cataloging Internet Resources: The Convergence of Libraries and Internet Resources," *Cataloging and Classification Quarterly* 22, no. 3/4 (1996): 203.

[73] Ibid., 202.

individual metadata creators' approaches. Cognitive models of bibliographic processing try to explain how documents are read/scanned; recognising structural characteristics of documents is an important aspect of expertise, one which becomes largely unconscious with practice. These "perceptual clues" can be typographic (italics, headings, paragraph beginnings) or verbal (term frequency, connections between similar concepts, and appearance of stock phrases).[74] Following these models, metadata creators would be expected to "scan text selectively, looking for specific perceptual clues."[75] To an extent, this appeared to be true, with Expert A describing how she takes "a quick look at the 'about' page, the introduction. Generally there's an introduction and an 'about' page. That kind of thing... there's certain points that you would look at immediately." However, the lack of standardised web formats appears to complicate this process, particularly if metadata is being created by a third party for external resources. Novice B described her frustration at spending a large amount of time (approximately twice as long as it usually took to create a whole record) trying to find information for a single descriptive element.

The quality of scholarly journals and monographs is determined by notions of objectivity; this is not the case for many web-based resources, and this 'spin' can affect the approach of the metadata creator to analysing the document. Expert A modified her behaviour because "an introduction page on a website is likely to be more of a marketing nature than what you might get in the introduction to a book or a print format." Research by usability guru Jakob Nielsen suggests that this scepticism applies to both technical and non-technical readers of web text. Following a study in which both technical and non-technical readers ranked objective writing higher than 'marketese' in terms of usability, he speculated that "promotional language imposes a cognitive burden on users who have to spend resources on filtering out the hyperbole to get at the facts."[76]

---

[74] Farrow, 154-55.

[75] Ibid., 154.

[76] Jakob Nielsen, *Alertbox for March 15, 1997: Be Succinct! (Writing for the Web)* (1997). http://www.useit.com/alertbox/9703b.html [Accessed 16 June 2002].

Where electronic resources are becoming standardised, these standards may not necessarily be the same as those for existing document types. Metadata creators, some of whom may be used to working with other formats, must now deal with an evolving writing style for the web that advocates greater use of headings, deliberate 'chunking' of text and avoidance of the use of italics (which are notoriously illegible on the screen). Further research is therefore required into whether the meaningfulness of headings and paragraph breaks is diminished in the web environment, and if so, what alternative 'cues' are available to metadata creators.

The final characteristic relating specifically to web-based documents is a lack of boundaries. Although print materials do contain references to external documents in the form of footnotes and bibliographies, hyperlinks provide an immediate context for an electronic resource on screen. Examining this contextual material can be an integral part of creating metadata, particularly in the subject portal environment, with Expert A and Novice C both noting the need to examine linked documents in some detail.

The final characteristic that emerged as part of discussions about the application of NZGLS metadata to government services was intangibility. Two experts (A and C) identified this as a shift away from traditional cataloguing practices: "It's not the traditional describing of a resource that's sitting in front of you which has got a title and an author and all of that information in the item. It's describing a service, which is intangible." Creating descriptive metadata for something that is intangible seems to require a different type of mental activity than identifying the title, author and extent of a given object. Similarly, the analysis required to determine the subject or function of an intangible like a service, rather than a resource, may require different thought processes: what these processes are, and how successfully they might be undertaken by those with experience in describing more traditional resource types, will need to be the subject of further research.

Discussion

One of the three research questions for this exploratory study related to the cognitive factors involved in creating metadata: further research is required to investigate the two themes that were identified. The first theme related to the overall metadata creation process. Although this was largely seen as a linear activity, several participants noted exceptions. This may confirm Jeng's speculation that a more fluid and creative approach to the task than the common top-down understanding of cataloguing is also viable. If some cataloguers or metadata creators 'naturally' approach their work in a less linear way, being forced to conform to a linear process may affect their work; identifying these effects might enable managers to design more suitable workflows and to consider accommodating less linear approaches in the design of tools (interface design is discussed further in Chapter Eight).

The second theme relates to the characteristics of the resources and services that were being described. The LIS literature has already identified some of these characteristics – such as dynamism and non-standardisation – as barriers to providing access to electronic resources, however little is known about their effects upon metadata creators as they approach their work. It is generally accepted that the structural properties of documents act as important visual cues in the reading/scanning process, so the effect that the absence or transformation of these cues may have on cognitive processes is something that deserves further attention from researchers.

A further question for future research is whether generic web reading mirrors the type of technical reading that usually takes place as part of bibliographic processing. Morkes and Nielsen's analysis of web reading found little difference between technical and non-technical readers; they noted that web readers scanned rather than read, that they focused upon headings, bullet points and emphasised text (bold or coloured fonts rather than the traditional italics), and that they preferred an 'inverted pyramid' structure in which the most important

information is placed first.[77] Morkes and Nielsen also speculated upon the cognitive 'burden' imposed upon readers, both general and technical, who are forced to deal with Internet 'spin' rather than more factual information.

The study elicited comments from metadata creators, both experts and novices, that suggest that cognitive processes for metadata creation might differ from those that have been outlined in previous models of bibliographic activity, where the range of resources covered has been small and standardised. Morkes' and Nielsen's findings suggest that these differences may not be limited to metadata creators, but may be generic to web users. Further research into metadata creators' handling of on-screen reading compared to print-based formats is required, to determine whether their experiences are work-specific or represent the evolution of a cognitive process as much related to generic web literacy as to specialised bibliographic tasks.

[77] John Morkes and Jakob Nielsen, *Concise, Scannable, and Objective: How to Write for the Web* (1997). http://www.useit.com/papers/webwriting/writing.html [Accessed 16 June 2002].

Chapter Seven: Novices' Approaches to Metadata, and Problem Areas in the Metadata Record

This chapter continues from Chapters Five and Six. The results presented still relate to the skills and knowledge that *individuals* require to create metadata. While the previous chapters focused on metadata creation in general, this chapter explores in more depth the experiences of the novice participants. The first part of the chapter presents data that relate to how novices initially approached metadata creation. A general overview is presented of novices' existing skills, as described by the novices themselves and as demonstrated throughout the interviews. Particular ways in which the process changed over time as novices grappled with new skills are also discussed. The second part of the chapter identifies specific areas in metadata records that were identified by experts or by novices themselves as problematic.

Overview of novices' approach to metadata

Novices selected for the study had little or no prior experience of creating metadata, or of other bibliographic tasks like cataloguing, indexing or abstracting. Two of the three novices were completely new to creating metadata, while one novice (A) had limited experience in applying HTML metatags to webpages. At the time of their first interviews, the three novices had attended initial training (one-on-one or group) and had created a small number of records. At the time of their second interviews, after six weeks had passed, Novice A had attended further training and created more records, Novice B had created more records but had not received any further formal training, and Novice C had attended more training and edited existing records but had not created any further records.

The degree to which the novices' job roles focused upon metadata creation varied. Novice A already worked in his organisation and was given the tasks largely because of his knowledge of the web. Although he was the only person creating metadata in his organisation, he consulted with others in the organisation about the descriptions of services

and resources. He focused on web-based and not print resources. Like Novice A, Novice C was solely responsible for creating her organisation's metadata; she was given the role partly because of her availability but also because of her communications expertise. Novice B had recently been appointed to a role in her organisation in which metadata occupied around half her working week. Although responsible for her own area of expertise, she worked with other metadata creators and was involved in maintaining email, web and personal contact with portal users. More information about organisational factors and their effects on the novices is given in Chapter Ten.

All three novices demonstrated a basic understanding of the purpose of metadata, in that they identified their aim as increasing the ability of users to access resources or services. Novices B and C expressed this aim in simple terms, while Novice A demonstrated an advanced understanding of how metadata would facilitate access to users:

> My understanding is that the main benefit for a website is to make it easier to find specific information. [Metadata] groups specific information, if different sites are using the same metadata then they may become grouped in terms of how users find them. So that's part of it, to improve the searchability of them, but also I guess, to provide more information about that actual page. Stuff like the date it was created and modified, that isn't actually visible, to record all that. (Novice A)

Novice A also appeared to have the most sophisticated understanding of how the information system worked, noting details such as the search engine functionality that would enable users' natural language search terms to be mapped to keywords in the controlled vocabularies and how the user interface would present "a visual catalogue" of search results to users.

In literature about librarians' suitability for metadata tasks, adhering to bibliographic standards is often pitched as a skill that that non-librarians do not possess. Although the three novices had little or no experience of metadata schemas, in discussions about similarities between metadata-related tasks and previous work experience two out of three identified other formal frameworks in which they had operated. For Novice A, complying with the NZGLS metadata schema was similar to meeting the W3C (World Wide Web Consortium) guidelines for accessibility "in terms of not taking shortcuts and doing it

properly." Novice B drew a parallel between conforming to a metadata schema and working as a teacher within a national curriculum.

None of the novices thought that creating metadata would be a particularly difficult task: Novice A's remark that "Doing the description wasn't rocket science or anything. I didn't have a huge problem with it" is representative of the confidence that novices had. However, all three novices acknowledged in their second interviews that the task was more complex than they had imagined, or that their understanding of what they were doing had deepened:

> Just seeing the bigger picture of metadata and how it's helpful. I mean at the beginning... I was feeling okay with it. But I actually feel it's getting beneath the surface to that deeper understanding of how useful it can be. (Novice B)

> It wasn't until I actually got into describing them, and looking up things, that I thought, 'This is hard work.' It's far deeper than just data entry, describe what you see and put it on the system. Once I got into it and discussed it, I saw the potential. (Novice C).

When asked to identify the skills they were bringing to the task of metadata creation, all three primarily identified existing specialised subject/function knowledge. Novices A and C rated knowledge of their organisations' work as a key factor, and compared themselves favourably to colleagues in this regard:

> I've probably got a broader sense of what people do in the [organisation] than a lot of people. You know, they sit around in their own bits, whereas I have to deal with everybody. (Novice A)

> I probably have one of the better overall views on what the [organisation] does, which is because policy sections and operations sections are very silo-affected. Because of who I target – I target Mäori – I have to have a general view of what everybody does [....] I think that if you asked people in policy what other sections did they wouldn't have the foggiest idea. (Novice C)

Novice B identified recent work experience in the subject area that the portal covered as central to her ability to create metadata. Apart from their subject/function knowledge, the novices also mentioned other skills they brought to the task. Novice A noted good writing skills and an understanding of the web; he also thought his impatience with filling in web-based forms was a negative characteristic. Novice B emphasised writing and listening skills, her willingness to learn, and her focused and methodical nature. Novice C emphasised her

skills in written and oral communication, her approachability and her forthrightness in dealing with other staff in the organisation.

All three novices were very focused on the needs of their users. This was particularly evident in comments about the type of language that needed to be used in the metadata records. Novices A and C also provided very specific examples of a typical user query that might be assisted by the metadata they would be creating, and Novice B demonstrated an in-depth understanding of, and empathy with, the wider professional concerns of her portal's target user group. Novice C raised issues relating to specific users such as Mäori and Pacific Islanders, stressing that their access to the metadata she was creating would be affected by 'digital divide' issues: she noted the importance of 'agents' such as Citizens Advice Bureaux advisors and public librarians, whose needs as intermediaries would need to be addressed (a point that was also raised by Expert A).

In contrast to Novice A and C, who did not receive feedback from users during the period of the study, Novice B had direct face-to-face and email contact with users of the information service. This deepened her understanding of users' needs, and also impacted upon her sense of the value of her work: "Without getting out there and meeting the people who are using it and the people who will be influenced by the metadata, you may pay lip-service to it. You may say, 'Yes, I know how important the job is,' but it really hits you when you go out there."

The previous chapter noted that metadata creators needed to take a pragmatic approach. Novice A, in particular, appeared to place his metadata tasks in a cost-benefit framework. He observed a great variation in the numbers of 'hits' to various web resources in his agency – one resource with wide general appeal had 20,000 hits in one month while others had a potential audience of around 10 per annum – and pondered "whether it was worth putting the same effort into [the second resource]." Novice A also noted that there needed to be a balance between creating information resources and creating the metadata that describes them: "I've sort of found that a few institutions seem to be much more obsessed with metadata than actual data. You know, they're putting a lot of energy into

getting their metadata databases [...] together and then when you look at what they're actually doing, there's nothing there!"

Problematic areas for novice metadata creators

This section outlines specific parts in metadata records that novices themselves or the experts, having trained and supervised novices, perceived as problem areas. Not every element appears as some were not noted as problem areas by any participants. Since no content analysis was undertaken of the records, the fact that some elements have not been perceived as problematic should not be taken to mean that these were completed in full and accurately all of the time. Interviews with a larger number of metadata creators combined with content analysis of the records would be required before conclusions of this kind could be made.

Subject and Function

Both DC and NZGLS contain a Subject element. NZGLS also contains a Function element. Best practice guidelines for DC suggest the use of a controlled vocabulary for the Subject element; the choice of which thesaurus or subject headings list to use is determined by the needs of the organisation and its users. Government agencies conforming to NZGLS must use the Subjects of New Zealand (SONZ) and Functions of New Zealand (FONZ) thesauri, which were specially developed for use within NZGLS and are built into the interface used for creating metadata. Organisations can use additional schemes; the National Library of New Zealand, for example, combines SONZ and FONZ headings with LCSH subject headings.

None of the novices had experience with using a controlled vocabulary before, although Novice A previously had taken a relatively advanced approach to providing ad hoc metatag keywords on webpages – "I like to put keywords in that are misspellings, alternative spellings and alternative versions of things that people will know to search for" – and thought that he would find choosing from the thesauri "quite limiting."

Expert A described assigning subject keywords as “difficult, as in time-consuming” for novices. In her view, novices needed to understand the importance of not having “an endless amount of keywords” since this would supply users with too many possibilities. Although this problem of ‘over-indexing’ (use of an excessive number of keywords) was found in an evaluation of Australian government metadata,[78] none of the novices in this study noted this as an issue.

Expert A believed novices improved over time as they developed an overview of the controlled vocabulary: “You have to know a bit in your mind what’s there. We do have the keywords all built into the system, but sometimes it’s not always easy to know what’s there or what’s on offer.” Expert B also noted difficulties that novices could have, but argued that these were sometimes due to insufficient knowledge of how the information system worked, for example, in translating users’ search words into the controlled vocabulary’s preferred terms:

> If I can be reassured that [...] if anyone types in ‘licensing’ they’ll find what I’ve put in [‘registering’], then I’ll probably be a lot more relaxed about it. Rather than going, “I have to call it such-and such, because that’s what everyone calls it’[....] The two extremes are the people who don’t care about the result and the people who care very intensely about the result. And those second group of people will be sitting and thinking about all the different ways in which the data might never be found. (Expert B)

Lancaster notes a problem arising from indexers being “too expert” in the subject area: “they tend to interpret too much and perhaps go beyond the claims of the author.”[79] Novice B described this in her own initial approach: “There was confusion with what I was reading, because I had to keep remembering, what is this specifically about, what is it saying, not what it can be. Implying really, and I was doing a bit of implying, and seeing in a more roundabout kind of way how this material could be used.”

---

[78] Sokvitne, "An Evaluation", 10.

[79] Lancaster, 82.

Both Novices A and C noted some initial problems selecting indexing terms at the correct level of specificity, another 'error' outlined by Lancaster[80] and noted by Sokvitne in his evaluation of Australian government metadata:[81]

> The other thing that I had a problem with was choosing the right subject and function terms. Sometimes I'd be choosing things that were too detailed and I was told to put it in a more general thing. Or vice versa, I'd put something more general and there was a more specific term I could have used. Terms like 'accrediting' and 'registering', I was getting a bit confused sometimes. (Novice A, second interview)

> We changed just about all of them! And I was sitting there thinking I was such a dummy! [...] I think it was that I took a too narrow view on the work that we do [....] It's just not having that overview of what's actually in the [controlled vocabulary]. In fact, [the broad term] is a good term, but in respect to that particular service, there are much more defined related areas. (Novice C, second interview)

Another problem that appears in Lancaster's typology and that Novice A candidly noted was using an inappropriate term because of lack of subject knowledge or carelessness[82]: "I still feel that they seemed okay... But some of them, I guess part of it's laziness. You see a couple of things that look right and you bung those in. You don't go and look at the whole list." This tendency to take the first appropriate word, because of a lack of knowledge of what was in the controlled vocabulary as a whole, was also noted by Novice C: "Once I had tweaked onto [a relevant, but too broad, term], that was the standard one that we used all the way through."

The mandatory combination of a subject and a function keyword was anticipated by Expert B as a problem specific to the NZGLS metadata schema: "If people have been used to what they've always thought of as a subject thesaurus and they look at a functions thesaurus [...] they'll get distracted into a battle of how the thing is designed versus what they're trying to describe." However, because of the small size of the functions thesaurus, and the likelihood that organisations would only use a limited set of terms, she believed that

---

[80] Ibid., 80.

[81] Sokvitne, "An Evaluation", 10.

[82] Lancaster, 79.

novices would quickly learn. This appeared to be true for Novice C: “It took me ages to get my head around the Subject and the Function. Why do they have to do these things! Especially the difference [between Subject and Function. But] once I’d established where I was going, and why I needed to take a SONZ and a FONZ that was fine.”

Novices’ descriptions of their difficulties with assigning Subjects raised a number of issues that are discussed in more detail in later chapters. Firstly, some issues seemed to be related to the metadata interface and the format of vocabulary that was used (online or print): these issues are discussed further in Chapter Eight. Secondly, novices stressed the positive role of feedback in alerting them to alternative terms that might be more suitable. Novice C described the process that she had been through with assigning Subjects and Functions: “That was really quite good, again having someone else externally check.... [The narrower term] wasn’t even in my mindset. I knew what I thought the subjects and functions were, but when [the QA person] came up with the recommendations, of course they were [better]. So some of that was my ignorance about how deep those fields went to.” Feedback is discussed further in Chapter Nine.

Description

The Description element in both DC and NZGLS consists of a free text description or abstract of the item being described. Experts A and B believed that the Description element was quickly and easily picked up by metadata novices. Expert A noted that as novices gained experience, their use of this element might also become more sophisticated: “In time, they remember things like putting in words that people might search by. Proper names and that sort of thing.”

Novice B, who worked in an environment where conciseness was highly valued, had little problem conforming to the prescribed rules for length; this also helped her to focus her attention on eliminating redundant words that had already appeared as Subjects:

> I think that probably that takes me the longest part, to be as concise as possible, and often when I get to the end of the metadata record, I come back to it again and often change sentence starts, or

> condense two sentences into one. I pull out a few keywords, and think, oh well if I'm using these keywords I can take out five or six words out of the Description and can be more precise with what the article is about.

The problem noted in the Subject element – incorrect level of specificity – appeared to arise for novices in their Descriptions as well: "I needed to go back and redo some of my descriptions because I got a bit carried away or I didn't have enough information or whatever" (Novice C). A further issue identified by Novice C was the danger that extensive knowledge of the content of the resources could mean that a user perspective was difficult to maintain:

> I thought I had the Descriptions pretty much sussed, but when [the QA person] came up and we went through, she broke it down even further. We sort of pride ourselves on lowering the level of our language for people, so it was quite refreshing to have that perspective from someone that doesn't work in our area of work. She was saying, maybe we should put an example up And I sort of thought, yeah, well from a consumer or user's point of view, that is the type of thing they would be looking for. (Novice C)

The electronic format of many of the resources being described meant that the Description element could be populated using cut-and-pasted text from the resource itself: "The writing didn't really come into it that much because a lot of it was just copying and pasting things that had already been created" (Novice A). The degree to which this can be viewed as a problem depends on the value that one places on metadata creators' interpretation of content. For Lancaster, there is little value in changing the author's words just for the sake of stylistic considerations: "An abstract is utilitarian and need not be a work of art."[83] The costs involved in original writing vs. cut-and-paste also need to be weighed, especially in contexts where a great deal of repetition is involved: "I already knew about cutting and pasting, but you could copy a whole service and then just re-jig it slightly, so that was awesome. Because you have a primary function or role – and while it may sway, you may have different ways of fulfilling that role – it doesn't change" (Novice C).

---

[83] Ibid., 108.

Date

Novice A noted problems with the Date field: “It’s all the fiddly bits I find I’m not quite clear on. Information about the date; is it the date that the page was created or the date the information was put together? Is the update date when you’ve added a full stop that was missing?” This problem seems to relate to lack of clarity about the definitions of qualifiers within the element, such as between Date.Created and Date.Modified, rather than any difficulty understanding the purpose of the unqualified element, Date.

Because publication date is such an important piece of bibliographic data, Expert C anticipated that experienced cataloguers might struggle with this element if the date information was not readily identifiable; this related to the non-standard nature of what was being described rather than any conceptual problems. No other problems with the date field were noted by any other experts or novices.

Type

The Type element in DC and in NZGLS refers to the type of resource being described. Two of the three experts (Experts A and C) identified Type as a problematic area. Issues related to the encoding schemes that were used, specifically to the lack of fit between what needed to be described and the terms supplied by the scheme. For Expert C, these problems were particularly pronounced, since not only were the terms in the encoding scheme limited in number and seen as inadequate, there was no immediate way to change the scheme:

> It’s been really interesting watching [a novice] get to grips with the encoding schemes [...] They aren’t really that useful [... and] we’ve had to discuss whether or not we change them, because we don’t like them, or whether we stick with [them]. Of course we’re sticking with [them] but that is going to be quite hard to explain.... The encoding scheme for Type is the main one [...because it] doesn’t have Policy as a term. Most of the information that we put up [...] is policies and guidelines, and the closest one is Advice. And although Advice is different from Policy, that is the term we will have to use. (Expert C)

Expert A’s portal environment supported a more extensive Type scheme; while this made lack of fit less of a problem, the use of more terms meant a similar need for an overview of

the available options: “I think it’s just a matter of knowing that the more you do it, the more you get a picture of it [....] knowing them and knowing what each of them means.”

Novices A and C did not detail any issues with the Type element, however since their NZGLS work had mostly been with e-services rather than resources, the Type element would usually have been given the value ‘service’, removing some of the need to select from the more extensive resource type list. Novice B, who worked with resources, noted that experience with a limited number of document types had improved her abilities in this area: “I’ve become used to the same types of material that I’m reading. I can pretty much determine that it’s going to be a Professional Reading, or a Collection, that’s becoming easier.” This corresponds with the arguments noted in the previous chapter that expertise is developed through repetitively working within a narrow range of text types.

Relation

The Relation element is used in DC and NZGLS to indicate that a relationship exists between the resource being described and another resource. The element can be used in conjunction with qualifiers such as Has.Part and Is.Part.Of to indicate the nature of the relationship. In the hands of a metadata creator with good knowledge of other resources, the Relation element is very powerful, enabling direct relationships to be established that can assist users to find information that is relevant to their query. Expert C considered use of the Relation field “critical.”

Two of the three novices (A and C) had some difficulties with Relation. For Novice A, this difficulty seemed to be caused by the conceptual difference between a service and a webpage that described that service:

> All the related things, things that have Is.Part.Of [...] what I did wrong the first time was that I was putting URLs that went directly to a webpage and identifying them for what was a service record. And what I was supposed to be doing was creating a whole new resource, so that webpage was a resource and in the service, I’d put that it had a relation to that resource.

Unlike other elements, Relation appeared to cause difficulties even after its correct use had been re-explained through the feedback process:

> I still get a bit confused with terms like Is.Part, Has.Part, all those sort of things [....] Probably that is the hardest [....] The Relation. The other stuff is pretty straightforward [....] In the Relation, the one that springs to mind straight away is the different between Has.Part and Is.Part.Of. I phoned up the helpdesk and we had a *long* discussion about it, and I just couldn't visualise it, and then they said, have a look at the manual, and I looked at the manual and I still didn't see the difference. (Novice C)

Despite these difficulties, Relation was a source of interest and excitement for Novice A, whose experience of linking a resource from his own organisation with resources from other agencies increased his awareness of how metadata could facilitate information sharing across different organisations:

> As I did it I got a sense that it was quite exciting in some ways to create a Has.Part or whatever Relation with a resource from another [organisation]. Also realising that maybe more than one of the resources I'd created could be linked to from one of the services I'd created. That's how it all crosses over and interleaves and it made it all a bit more interesting in some ways.

Discussion

The results of the study suggested that novice metadata creators bring a range of prior skills and knowledge to their work. Their knowledge of the subject area or of the work of the organisation and their understanding of users may be well-developed. With little training, novices are able to grasp the purpose of metadata in at least a basic way, and novices with more experience of IT, and of the Internet in particular, may have a sophisticated understanding of how information systems work. Novices may also bring knowledge of the relative importance of different resources and services, which enables them to quickly prioritise their work and to maintain a balance between the need to create good metadata and the high costs of doing so.

An interesting feature that emerged was the parallel between working with metadata schemas and adhering to standards or frameworks in other disciplines or domains. Because creating metadata of a consistently high quality requires understanding the metadata

schema, managers may find it helpful to identify other types of standards that novices have experienced and to incorporate discussions about these into early training sessions. Novices' understanding of the importance of conforming to the metadata standard might be enhanced if good examples of standards-based tasks from their earlier work experience can be drawn upon.

Although novices in the study were confident about their abilities, all three recognised that their understanding of what they were doing deepened over time. The importance of feedback mechanisms is discussed further in Chapter Nine, but it is important to note here that novices' confidence in the early stages of creating metadata should not necessarily be taken at face value by managers. Novices may believe that they are creating adequate metadata (and they may indeed be doing so), but they still need to develop their skills and knowledge through close interaction with a supervisor or quality assurer in the early stages.

Monitoring the results of novice metadata creators' work is also initially very important; although this may be labour-intensive, ultimately time and money will be saved, as ongoing quality assurance can be targeted to recurrent problem areas rather than to every element of every record. The novices in this study seemed to have little trouble with most elements, but identified a small number, in particular Subject and Function, Type and Relation, that were picked up at the QA stage. Rather than being sensitive to this type of criticism, participants welcomed constructive feedback on their work and developed positive relationships with the people supervising or quality assuring their work that assisted in the development of their skills and knowledge over time; this is discussed further in Chapter Nine.

Chapter Eight: Evaluation of the Metadata Creator's Toolkit

Having outlined how the novices in the study approached metadata creation and some of the areas in a metadata record that proved to be obstacles, I want to turn to the tools that are used. The metadata creator's toolkit consists of the software interface used to create and edit metadata, the hardware and network that the software runs on, the controlled vocabularies in print or electronic form (these may or may not be integrated into the interface), and other related documentation.

Software (the interface)

A range of interfaces for creating and editing metadata have been developed,[84] and a small body of literature evaluating these is emerging.[85] These studies suggest that the tools available to creators are a crucial part of metadata creation: just as automated library systems and networked bibliographic utilities have enabled paraprofessionals to take on expanded roles in traditional cataloguing environments,[86] interfaces can assist (or hinder) metadata creators in their efforts. The results of the study suggest that the ideal software interface would: be online and complete; reduce the amount of text that needs to be manually entered; automatically validate entries; be navigable using the mouse and the keyboard; allow creators to view and print the record in its entirety; have a logical, linear element order as the default but allow customisation; give creators control over how and when their work is saved; integrate the feedback process into the record; and be continuously improving through feedback from metadata creators.

---

[84] DESIRE (Development of a European Service for Information on Research and Education), *DESIRE Information Gateways Handbook* (26 April 2000 1999). http://www.desire.org/handbook/welcome.html [Accessed 4 August 2001].

[85] Edmunds and Brisson; Preben Hansen, "Evaluation and Design Issues of Nordic DC Metadata Creation Tool" (paper presented at the International Conference on Dublin Core and Metadata Applications, National Institute of Informatics,Tokyo, Japan, 2001).

[86] Howarth, 532; Mohr and Schuneman, 206.

Experts B and C both thought that the interface should be online and complete: it should incorporate everything necessary for creating metadata, including the controlled vocabularies and encoding schemes, the documentation and context-sensitive help. Given that some users, both expert and novice, continue to rely on paper versions of documentation and encoding schemes (see below), care would also need to be taken to provide a printable version of these resources.

Interfaces can be built to assist metadata novices by reducing the amount of text that needs to be typed manually: as Expert B described, "When you've got to enter text, anything can go wrong, and does." Better accuracy can be achieved through the use of drop-down menus, automatic content generation (e.g. automated insertion of the URL of a webpage) and cut-and-paste editing (surprisingly, problems with cut-and-paste were noted by a number of participants). Experts A and B took the reduction of manual entry one step further, identifying a need for templates; particularly when undertaking work on large amounts of similar resources, this would make metadata creation less "tiresome" (Expert A) and "more enjoyable" (Expert B).

Expert B noted that the software should be programmed to validate the content that is entered: "I want immediate QA so that you're selecting something that does exist, or it can pick up a spelling mistake[...] or if the date or the format's wrong." Examples of validation are spell-checking, URL checking and mapping to a controlled vocabulary. An expert analysis of CORC's metadata system, which includes some of this functionality, evaluated the tool as likely to produce far more accurate and consistent results than manual processes.[87]

Interfaces should also provide more than one way to navigate through the record: Expert A noted that the interface that had been developed in-house in her organisation forced the metadata creator to use the mouse because it did not allow 'tabbing' through the fields. In addition to being an occupational health and safety issue, this also affected the

[87] Edmunds and Brisson, 105.

speed at which creators could work, particularly when doing large numbers of repetitive records.

Interfaces should allow creators to view and to print the work that they have done in its entirety: Novice A described his frustration at the layout of his interface:

> This printout that I've got here has got everything in little boxes, because you couldn't get a printable page view of what you were typing in. You could never see the whole thing, which for me, well, it was a bit of a nuisance. I imagine that for other people who have got a different person doing the agency QA, the easiest way of doing that is to hand them the hard copy. So that sort of thing could be a real nuisance.

Experts A and B both noted that the fields should be in a logical order:

> It's much better to have it one thing after the other, and then it's easy to check it, it's easy to check once you've done it if you've missed anything. (Expert A).

> The fields are in a logical order so that the decisions that are made in an earlier field can be employed again as you fill out later fields. (Expert B)

While this corresponds with common sense understandings of cataloguing as a linear process, as described in Chapter Six, this is not always the case. Further research needs to determine how people work through the metadata records: in the interim, interface developers should not assume that all creators work through records in a linear fashion. If possible, creators should have the option to 'change views' or customise their interface. This flexibility would be particularly useful where prioritisation is built into the metadata schema: this was noted by Novice C, who disliked having mandatory elements dotted throughout the record rather than in one place.

Ideally, creators should control how and when their work is saved to a database because, as Expert B noted, "there's nothing worse than filling a form in and having it disappearing away from you, and you don't know what you did." This may not be possible if there are limited network connections: Novice C noted her frustration at the amount of time she could leave her work before her session 'timed out'. The time-out function combined with the lack of control over the layout of the mandatory elements within the field meant that "if

you got called away to a meeting and you'd entered all this data but you hadn't filled in all the mandatory forms, you had lost all that session's work."

Two of the three novices received feedback on their work that was integrated with the record itself (i.e. through a Notes field not available to public users) and commented upon how helpful this was. This integration is discussed further below in the section on feedback in Chapter Eight.

Finally, novices welcomed the opportunity to be able to comment upon interface problems at the developmental stage: "I was impressed [... that] they were open to that sort of criticism, which was good because quite often when you're doing that sort of stuff you get a bit precious about it. 'Oh well, we've tried our best...' But they were really open to it, and they have tried to fix up a lot of those glitches" (Novice C, second interview).

Hardware and network

Participants identified two interrelated issues with the hardware and network: speed and quality of connection. These comments correspond to the findings of an expert evaluation of OCLC's CORC interface, which noted slow online response times and "intermittent bugginess" as the most serious flaws for users, even preventing the tool being worked with at all.[88]

The time taken to download a page and move within the record is important because of targets that metadata creators often work to: "Speed is important. It is important that you have a system that works quickly because you often have lots to do and you've got to get through it in a limited amount of time" (Expert A). This could be particularly problematic for off-site metadata creators: "From time to time, we do have problems loading the pages, whether it's slow or peculiarities like that [...] That can be a big issue, as far as working on-site and working at home[...] *People have had to modify the way they did things because the pages were so so slow loading*" (Expert A, emphasis added).

---

[88] Edmunds and Brisson, 105.

A number of the experts and novices had experienced system 'crashes'. Expert C perceived crashes as a real issue for novices, who might be put off if their early experiences with were poor: "It would be very hard [for a novice to deal with system crashes], when they are very low in confidence and uncertain and don't have the complex step-by-step manuals that they're used to [....] There is a risk that people could be switched off if that happens a lot." This quote, like the one above from Expert A, indicates that bugs like this might especially affect isolated metadata creators, for example, those working alone in an organisation or working off-site.

Controlled vocabularies/encoding schemes

Creating metadata requires the use of a number of controlled vocabularies or encoding schemes. Many information professionals are well-versed in the use of controlled vocabularies, and this has been one argument used to suggest that novices are less suited to metadata creation. In contrast to this view, Expert C made the point that increasing web literacy means that "People are used to searching from lists on the internet [...] I think as long as it's pretty similar to what people are using normally in their lives, I think generally they would be okay."

The results of the study suggest that controlled vocabularies and encoding schemes should be: simple and no larger than required; online and integrated into the interface but also available in a printable version; available in different views; complete with scope notes or definitions; customisable for local use; and continuously improving through feedback from metadata creators.

Where people with different skill levels and backgrounds will be choosing from encoding schemes, it is important that these should be simple and not overly large. This point was emphasised by Expert C: "I think having that simple one, which could be used by communications or IT people was a good approach." Expert B noted that the FONZ thesaurus for NZGLS, with 200 terms, did not cause too many problems for people once they had grasped the conceptual difference between subject and function.

I outlined above the need for controlled vocabularies to be available online and fully integrated with other functions. Novice A, who worked with a print version of the thesauri and had some difficulties choosing terms at the right level, found the process much easier when the full functionality of the online versions of the thesauri was explained to her: “

> I did have a hard copy of both the FONZ and the SONZ. It was just not a good tool for me [....] But one of the trainers] came up and saw us entering the terms, and said, ‘You can do all that online,’ and I thought that was much easier because you’re not having to move [....] Having it in its current form is even better. You just type it in and it throws up related terms or better terms or whatever.

Although Novice A described using the print versions of the thesauri “a bad mistake”, the results suggested that a printable version should be made available, particularly to novices. Some people appear to have a strong personal preference for browsing using a print version:

> Particularly for new people who aren’t very familiar with it, it’s easier to scan that than it is to scroll through lists of words on the interface [....] And generally, yes, people hang onto these when they first start. Some do and some don’t, and that’s kind of a personal feature, you might find. I know one person who’s been doing it longer than just about anyone does like to have the keywords on paper. He’s certainly very high-tech in other respects, but he finds it useful to have them on a list to refer to. (Expert A)

The danger in developers of the system providing a print version is that metadata creators may be using an out-of-date list: a number of participants noted this, including Novice C, who described her earlier use of the print versions as “a dumb thing because then if they want to update it, that version is way out of date.”

Expert B thought that controlled vocabularies needed to be presented in more than one way where this is relevant to the underlying structure: “You can have the lists or the tools available with different views, so that you can go down it hierarchically if you want so that you can see the structure, so that you can see it in a straight list if you want, you can do it alphabetically. You know, so there’s more than one way to get into it.”

Although no participants explicitly mentioned it, lack of definitions or scope notes for the terms in controlled vocabularies, encoding schemes and qualifier lists seemed to cause some of the difficulties with metadata records that were noted in the previous chapter. In

particular, more explanatory text appeared to be required for the encoding scheme for resource Type and the list of qualifiers in Relation.

Just as the interface would ideally be customisable for local use, the controlled vocabularies could also be edited to create a local version: “In my ideal world, you would do a mapping exercise[....] You’d sit down and look at the thesauri on their own and highlight or mark or note the ones that apply to you. So that way when you go to describe something, you’ve already got your own little truncated thesauri” (Expert B). Again, the need to ensure the truncated local version remained up-to-date was emphasised.

Finally, as with the interface, controlled vocabularies should be continuously improving through feedback from metadata creators and from users: this is discussed further in the section on Feedback in the following chapter.

Documentation

According to Jula Hakala, tools alone “do not provide enough support for metadata providers”: user guides, and other documents are needed to specify what kind of metadata is wanted.[89] The results of the study suggest that documents such as manuals are required as a reference and for interpretation of “subtleties” and should preferably be accessed through keyword searching rather than read in their entirety (Expert B). This more extensive documentation is necessary, however, since it allows metadata creators with more experience to gain further knowledge of the system:

> I think [the needs of novices and experts] are different. I think a novice is looking to, just to do something more concrete, and the expert is interested in how the system works. So they should have that option. The novice may be looking for help to do something, but the expert may be inquiring, wanting to engage more with the developers or the central systems, to provide feedback. Because they’ll have their own opinion on how it’s done or how it ought to be done. (Expert B)

---

[89] Jula Hakala, "Internet Metadata and Library Cataloguing," *International Cataloguing and Bibliographic Control* 28, no. 1 (1999): 25.

Formal documentation should be supplemented with other types of information that are more immediately accessible. Participants remarked that these resources could include frequently asked questions (Expert C), step-by-step instructions (Expert C), reference to common problems that arise (Expert A), statements about overall priorities (Novice A) and as many and varied examples as possible (Novice A). This level of documentation is not only required for novices: Expert C observed that "experts may be wanting to do it *too* thoroughly" and could therefore also benefit from guidelines that suggest a more minimal approach.

As with controlled vocabularies and encoding schemes, participants believed that some users would prefer printed version to online and that their needs must also be met. Expert C thought that the need for printed manuals would diminish over time as metadata creators began to appreciate the currency and superior functionality (e.g. searchability) of online versions.

Discussion

The results of the study suggest that good tools and documentation can make metadata creators' jobs easier, and participants in the study identified the positive and negative aspects of the existing tools that they are working with, providing a useful template for an 'ideal' toolkit.

With regard to interfaces and controlled vocabularies, managers should not assume that a one-size-fits-all solution will work for everybody. Metadata creators may have their own ways of working, and should be given the maximum amount of customisation and control that is possible. Obviously managers can only meet the needs of individuals within the confines of the information system. Although Novice C's frustration at the short 'time-out' period of her session was real, managers of that system are likely to already have weighed this against the cost of maintaining more server connection capacity than is required, or the frustration of other users as they wait for forgetful metadata creators' sessions to end.

The results also suggest that the creation and maintenance of documentation should be informed by the needs of metadata creators. Again, a one-size-fits-all approach may not

work, and managers should encourage metadata creators to find (or create) the types of documentation that supports their work most efficiently. This may involve some risk: allowing metadata creators to work with customised versions or printed copies of controlled vocabularies means that the versions and printouts must be monitored to ensure that they are current.

Managers need to ensure that the development of tools and documentation is not a one-way process; staff will be more willing to work with 'buggy' tools and unhelpful documentation if they are aware that their feedback will lead to improvements. Managers should also be aware, though, that the toolkit, particularly the interface, contributes significantly to the initial impression novices receive. Although novices, like experts, may appreciate the opportunity to give feedback, there is a risk that exposing them to poorly developed tools may turn them off altogether. This is especially true at the early stages in a novice's development, when, as Expert B observed, "they're trying to do two things at the same time. They're trying to familiarise themselves with the descriptive standards, and they're trying to use the tools, and that combination can be a bit steep straight up."

Particular combinations of interface, hardware and network issues may also make it difficult for metadata creators who work from home or alone in an agency in a distributed environment. Managers need to be aware that the productivity of metadata creators might decrease if they are working in situations where the speed and functionality of their tools are affected by poor internet connections or obsolete technology. These productivity issues should also be considered in the context of other human resources and organisational issues: the benefits to employees of choosing their own work environment may be deemed more important.

Chapter Nine: Methods of Gaining Metadata Skills and Knowledge

This chapter follows on from Chapters Five through Eights above, and focuses on Research Objective 4, which was to gather information about how knowledge is acquired by novices. The transfer of knowledge from experts to novices through formal processes such as training, quality assurance and feedback are explored. Less formal ways of gaining knowledge – for example, networking and peer support, are also discussed.

Training

All three experts had trained novices, although the level of formality involved in this training varied from environment to environment. Experts A and B detailed the usual methods they used to train novices. Both experts stressed starting with an overview:

> Whoever is going to do the training has to get the context of what it's all for, and how what they do fits into the bigger picture. [... The metadata scheme] didn't come full-formed out of thin air. How was it derived? What is [the organisation] planning to do with all these descriptions? Who might the real customer be? And that kind of thing, so that they understand their part in the bigger picture. (Expert B)

Both experts also identified the phase after the overview as an element by element 'walk-through' of the metadata record: "After a few days when they've got a picture of things, then usually I would have a session with them, say for an hour, going through the actual record, element by element, in what you do and some of the problems that you can encounter" (Expert A).

After the overview and walk-through, Expert A and Expert B's training techniques differed. Expert A, who undertook most training on a one-to-one basis, believed that the formal component of training should end at this point, with the novice then encouraged to go away and try to put what they have learned into practice: "We try very much not to take more than an hour because that's really as much as you can take in when it's all abstract, when you haven't done it. The best thing is to have an introduction and then go away and do it." Expert B preferred a more structured approach, in which novices worked through a series of

set examples that increased in complexity. Expert B also advocated interspersing creating metadata with searching for the resource once the metadata had been created: “Put it in, now see if you can find it or see if someone else can find it. It’s better to see if someone else can find it. Because you know what you were thinking, but they don’t [...] And it’s very salutary to have someone else try to find what you’ve described for them.”

The novices reinforced the importance of an initial overview, however the level at which this should be pitched appeared to be something that needed to be carefully considered by trainers. Although their training had included an overview of some type, all three novices felt dissatisfied at its level. Novice A felt that more information about how the metadata would be used was required:

> [The training was] a little bit disassociated in some ways, in that I didn’t really realise what we were doing until a few weeks later when I was at another demonstration of the portal. A fuller demo of the portal itself and how it was actually going to be working... They broke it all into little parts, without actually giving us the big picture. And that had an effect in terms of how important, you know, the importance I placed on things.

Novice B indicated that “a general intro” to metadata might have enabled her to gain “the bigger picture”; she thought the overview could detail how metadata was used across different domains and compare what was happening in different organisations. Novice C noted that some of the terminology used in one training session had been difficult; but, as someone who was heavily involved in training others (not for metadata), she also expressed doubt that another method of training would have been more effective: “It was just more the terminology that I was going, ‘What the hell are they talking about? What the hell is a metadata? What’s a FONZ? What’s a SONZ?’ But I don’t know if there’s any other way you could do that.”

Practice and experience

I noted in my introduction that most literature on teaching cataloguing, indexing and abstracting advocates a mix of theory and practice.[90] This view was reiterated in this study. All experts in this study stressed that experiential learning, during training and afterwards, was the most effective way for novices to learn to create metadata. Formal training had a place, but moving beyond the level of abstraction that was involved in overviews of standards and systems to creating actual records was necessary: "The best thing is to have an introduction and then go away and do it" (Expert A). Novices also observed that deeper understanding of what they are doing came with the experience of creating records: "Okay, you wade your way through [the training], but some of that [understanding] came later when I actually got through to doing it; what was expected and how you'd describe and where you'd go to access [the controlled vocabulary] and all the rest of it" (Novice C).

A recent article on qualitative methodologies in LIS research observes that "Many claim that they learn from experience but find it difficult to articulate what has inspired the learning."[91] While expert and novice participants in the study stated that their performance improved over time as they gained more experience creating metadata, it was difficult to determine exactly how practice led to improvement. However, two broad themes were identified that suggest how experience may contribute to skills development and greater knowledge; these themes relate to learning by example, and learning by repetition.

Learning by example

Jeng asserts that cataloguers "learn to be experts from working on examples." All three novice metadata creators considered examples in training and documentation (e.g. manuals)

---

[90] Judith Clayden, "Theory Versus Practice in Cataloging Education: Some Australian Experiences," *Journal of Education for Library and Information Science* 36 (1995): 230-8; Lynn Silipigni Connaway, "A Model Curriculum for Cataloging Education: The Library and Information Services Program at the University of Denver," *Technical Services Quarterly* 15, no. 1/2 (1997): 27-41; Michael. Gorman, "How Cataloging and Classification Should be Taught," *American Libraries* 23, (September 1992): 694-97.

very important. A distinction needs to be drawn between the activity of copying from an example and using examples as the basis for developing one's own practice as a metadata creator: Jeng notes that learning from examples "requires high-level thinking strategies on the learner's part": examples need to be used to generate explanations that work for broad categories of resources rather than for individual resources.[92]

Novices appeared to be working with examples at a higher level of categorisation rather than simply copying them. Novice A described working from examples as a normal part of his work: "I looked at what they'd done, I looked at the code on their site and used that as a guideline[....] which is basically how so much is done in web work, you know, you look at what someone's done." He went so far as to describe the entire intellectual process of metadata creation as "adapting an example to my special set of circumstances", but also emphasised his own knowledge about the work of the organisation and the need to interpret the original example in terms of user requirements: "You've got to create it, you're not just copying [....] We have to make the judgments about what part of that information is the most important part, and how people might want to find out about it. I think it is definitely more than just doing data entry."

Novice C, in contrast, did not adapt examples as part of her usual practice, but used examples from other organisations when she was grappling with a particular problem. Her use of examples appeared to be related to obtaining a solution to a *type* of problem, rather than the individual problem itself: when she came to types of service that were atypical within her organisation – for example, those requiring fees to be paid – she viewed records for similar services in other organisations in an attempt to clarify her own approach to the problem.

---

[91] Watson, 139.

[92] Jeng and Weiss, 127.

Learning by repetition

The second experience-related theme that emerged from interviews with both experts and novices was repetition. Repetition is an important factor in literature on expertise: in his discussion of expert systems research in the field of cataloguing, Olmstadt draws on psychological theories that assert that expertise is "not the result of superior performance in unusual tasks", but develops instead through "deliberate practice on representative tasks in the domain."[93] More specifically, in the case of indexing, Farrow asserts that "indexers work within a narrow range of text types and subject fields, and the consequent repetitive element in their work leads to automatic processing beyond that associated with normal fluent reading."[94]

Creating metadata can be a repetitive task that not all people are suited to: Experts A and B observed that many people consider metadata creation to be a boring activity. Expert A argued that being able to undertake repetitive tasks, and to balance these with other types of work, was important: "If you have to make changes to things or just if you have to do a whole bunch of stuff that's very similar but you have to do individual records for everything, it's not the most exciting aspect of the job. You have to get people who have a broad view of things and think, 'Well, yes, this is okay. I don't mind doing this' and you balance it in amongst your other stuff."

As noted earlier, although Novice A realised how important it was to create metadata to the required standard, he described himself as impatient with "filling out endless forms that I can't quite see the results of." In his second interview, he reinforced this tension between the desire to work to the standard and his boredom with the repetition: "I thought, well, you might as well do it properly and it's not that big a hassle once you get in the right frame of mind. Numb mind!"

---

[93] Olmstadt.

[94] Farrow, 153.

All three novices also indirectly asserted the value of repetition in their discussions of how they organised their metadata creation work. All three intuitively organised their so that they dealt with groups of related resources: “It’s useful to do it in batches. You start off with all this glee, but you get it back four or five times and then you just want to get it out of the way. There’s only so much creativity you can put into it” (Novice C, second interview).

The importance of repetition in learning to create metadata was most obvious in the case of Novice B. As the only novice in the study whose job role was directly related to metadata, she observed that doing different types of work diminished her ability to meet her targets: “I was focusing on a couple of areas at the same time so I wasn’t able to stick to that target. That actually disheartened me for several weeks.” Novice B understood that repetitive practice was necessary in order to improve: “If I was doing metadata *all* the time, and that was *all* I was doing, my progress would probably be a lot faster.” However she also stressed the importance of balancing repetitive tasks with other types of work: “I also enjoy the variety. I don’t think I could do metadata only.” Novice B was the only novice to directly state that working with limited types of documents had enhanced her skills. She had organised her workload to deal with batches of resources in similar subject areas, since this enhanced her ability to learn the sections of the controlled vocabulary that applied to these topics: “The challenges for me were just remembering the keywords. So I’d do a batch of [resources on one topic] just to feed it through and get familiar with the vocabulary[....] I spend a day looking at one lot of metadata and then the next day I choose another [subject area....] That seems to assist my revision and recall of the different terms.”

Feedback

Feedback emerged through interviews as one of the most important ways that novices learned. Participants identified several different types of feedback, that could be categorised as internal and external, direct and indirect. Table B represents these different feedback types in a matrix depending on these characteristics.

	Internal	External
Direct	Internal QA (supervision) Peer support	External QA Helpdesk support User feedback (verbal/written contact with individual users)
Mediated	User feedback (activity logs, statistics)	User feedback (activity logs, statistics)

Table B: Types of feedback

Internal quality assurance (supervision)

The first and most immediate feedback that novices receive is from their manager or supervisor. Expert A indicated that supervision would initially be intensive but would probably decrease over time: ”I will have a look at them and go back and have a chat with them about certain things. They’ll remember things from that because they’ve done it as well as read about it and talked about it. That process will probably happen a couple more times, but usually after that it’s okay.” Experts A and B both stressed that novices should feel able to make mistakes during this early phase, a point that was supported by Novice B: “I mean, we all need feedback, don’t we? And I would feel more comfortable if someone was able to look at my work and say, ‘Yep, that’s okay.’ At least in this beginning stage, that’s what you’re really interested in, isn’t it? [...] I always prefer to do things to the best of my ability and I wouldn’t just like to be making things go live and then have it backfire.”

Novice B was the only novice to have direct supervision within her own organisation (Novices A and C's experiences with external QA processes are discussed below). She sometimes appeared reluctant to take problems to her supervisor informally, because of her supervisor's own workload: "[She] did also have her own work so she couldn't always stop and help me." It was important to Novice B to be able to work independently while knowing that her work would be reviewed: "I think I'd rather write something and get it wrong and that becomes a learning point rather than seeking help the whole time [....] I value the feedback, and trying to nut it out myself rather than getting help the whole time."

Novice B also highlighted how the feedback process could be usefully integrated with the design of the interface through addition of a notes field for internal use: "When I print off the metadata that I've done, there's lots of notes around my entries and [my supervisor is] quite thorough with the feedback so that's good." In addition to tying specific comments to the actual record, which Novice B found helpful, this method seemed efficient in that the supervisor could QA in batches and then provide feedback based on recurring, rather than one-off, problems.

Peer support

Communication amongst peers is an important way in which knowledge is transferred; where knowledge is tacit, peer communication is even more valuable, since it amounts to "the formalization of private knowledge into public knowledge" that can be more easily shared.[95] The lack of immediate peer support in some metadata creation environments is a human resources management concern: "Problems can include a feeling of isolation for the distributed team members. The work relies on self-motivation and good morale and this can be hard to achieve."[96] Metadata creators might need to be supported by central staff, but may also desire contact with each other: communication via mailing lists and personal

---

[95] Jeng and Weiss, 127.

[96] Emma Worsfold, *Distributed and Part-Automated Cataloguing: A DESIRE Issues Paper* (March 1998). http://www.sosig.ac.uk/desire/cat/cataloguing.html [Accessed 18 August 2001].

email/telephone contact is described as essential and some face-to-face contact is also desirable.[97]

Peer support was an important method of gaining expertise for Novice B. She identified two types of peer feedback. The first of these was one-on-one feedback from another metadata creator, whose area of expertise overlapped with hers: “You want to have a buddy, I think, that you can go to. For at least a month or six weeks [I thought, ...] but in actual fact it took me a lot longer than that.” Novice B and her buddy had recently moved to adjacent desks, and this resulted in a more collaborative approach: “There’s lots of fun and good laughs and I have a good working relationship with [her] and feel I can interrupt her and ask her at any time about anything. She often confers with me [...] She’ll often say, ‘Hey, how does this sound?’ And together we work through it and rearrange the words. I mean, that’s quite good.”

The second type of peer relationship that Novice B identified was group meetings for all the metadata creators in the organisation to share ideas. Although none of these had taken place while Novice B had been working, she was looking forward to attending this type of forum. Expert A had also identified this type of meeting as important: group meetings allowed the sharing of ideas and collective local decision making, but could also be used as a forum for providing feedback to strategic decision makers in the organisation.

Novices A and C were solely responsible for creating metadata in their own organisations, and therefore did not have access to peers in their own workplace. This situation changed for Novice C, when she suddenly learned her agency should have two people for metadata, one to perform internal QA. Because Novice C had already attended training and created metadata for the agency, she became the quality assurer, while another person was taken on board as an ‘inputter’. Novice C did not find this two-tier system ideal, and asserted that a ‘cross-QA’ system would have been better:

---

[97] DESIRE; Sladen, 78-79; Worsfold.

> Well, if I knew at the beginning of the process that was the way it would be, we probably would have both gone on all of the training, and just because of the huge workload, we would have split it and then QAed each other's work[....] Especially in an organisation where there are few people dedicated to it, you'd be better off getting a complete overview and finding out the best way to apply it in your small team.

External quality assurance (QA)

In a distributed metadata creation environment in which metadata is contributed by a range of creators to a centralised repository, QA can take place external to the organisation that the metadata creator works in. Novices A and C had experienced an external QA process, which both had found extremely rewarding and which contributed greatly to their understanding of metadata. In his second interview, Novice A, who had worked closely with a single quality assurer, noted that the external QA process had provided him with motivation and reassurance: "There was this sort of dialogue with the QA person. That's been really key. I could have been left to flounder away with it [and...] it would have gone a lot further down on my list of priority things to do because it would have become something I didn't feel confident about doing, so then you don't. Whereas with this, I knew even if I got it wrong, someone could tell me."

Some LIS research suggests that successful feedback should focus on improving processes rather than on uncritical provision of corrections.[98] This was not the experience of Novices A and C, who noted that their external quality assurers had in some cases rewritten their text or provided them with alternatives that they had cut-and-pasted verbatim into their own records. Novice C stated that her creativity could only be sustained for so long, after which she was happy to accept the suggestions made by the quality assurers. Novice A felt that the quality assurers were better qualified to deal with more 'conceptual' requirements and had been able to take a more objective user perspective. These comments seemed to

[98] Robare, 58.

support a report in the LIS literature about the success of a two-tier workflow at Education Queensland where author-created descriptive metadata was validated by librarians.[99]

As with internal QA, integration of the feedback process into the design of the interface was welcomed. Novice B noted the usefulness of being able to print off the record that he had created with the quality assurer's comments included within the record. For Expert B, the feedback should go both ways, providing an opportunity for the novices to present their own thoughts and ideas: "There's the chat, the follow-up of 'How's it going? Are there issues?' Because often people don't comply, not because they think it's wrong, but simply because it's too hard or they don't know enough about it or something like that. There may be reasons."

Helpdesk support

Helpdesk support was vital for Novices A and C, who were working without direct supervision or peer support in their own organisations. Just as Novice A had developed an excellent relationship with his assigned external quality assurer, Novice C had found the support of the helpdesk vital in working through issues: "There have been many and varied questions that I've phoned the helpdesk about, and they've been awesome. Really helpful. I've called up and said, 'This is my dumb question for the day,' and they've said, 'No questions are dumb, and we can help you.' It makes you feel better about it. The support has been excellent."

User feedback

Novice metadata creators need to receive feedback from the users of the information service. User feedback cuts across the categories in the matrix above; it can be both internal and external, direct or mediated. The type of user feedback that is available to the novice metadata creator will depend upon the environment they work in and the tools available for gathering information from users.

---

[99] Thornely, 122.

Working for a domain-based portal, Novice A quickly developed direct relationships with users of the service. She received feedback from users at workshops and other events as well as via email, and described her evolving relationships within her user community as giving her 'ownership' of her role. This type of direct relationship with users is less likely to occur in more distributed environments, such as that developing for e-government in New Zealand. In this context, user feedback is more likely to be gained in a mediated or aggregated way, for example, through the use of web logs to identify users' activity on a site and their choice of search terms. Expert B identified the need for this data to be collected, not only so that metadata creators could improve their own performance but so that system and tools could be refined:

> You want that whole feedback system, so that the [metadata creator] learns but also the standard improves and the lists like the thesauri can adapt. Because the whole point is if people keep asking for something then we should make sure there's a way to find it. You can get lost in the amount of information that will be available via the portal. So if someone is constantly searching for something, does that thing go onto, say, the searching tool that sits in the front, or does it get added into the thesaurus or the lists at the back so that it will come up? There's various strategies to make it work.

Novices A and C had not received any feedback of this kind, having created their records well in advance of the 'go-live' date for the service. For Novice A, this was frustrating, in that the results of his work were not immediately available to him: "[Other things that I do] are much easier to test; there's tools and you run it through and it passes or you can see where it's wrong. Also you can see the results, and that's much more a definite thing, whereas metadata seems quite ethereal to me. Whether it's there or not, the webpage still seems to work fine, you know?" Novice C was also looking forward to receiving feedback about user behaviour with regard to her records: "Well, you have to [see what users are searching for], don't you? There's no point in throwing something out there and then saying, 'Well, go for it. Next project!'"

Further development: lack of current opportunities

Results of the study suggest that there are few professional development opportunities for metadata creators who want to increase their skills and knowledge beyond what can be gained through peer interaction and user feedback. When asked about development opportunities, Expert A said "We don't go to courses about it because there aren't any, you know!" Novice C also identified lack of training opportunities as an issue for librarians who may want to cross-train: "There's a gap there for people who have been trained to catalogue traditionally to learn how to up skill [....] I think that there's going to be a big need for much more training in this area for current librarians, and that needs to be well-publicised."

Although they did not indicate a need for any further training, Novices A and B both regarded the metadata work they had already undertaken as important for their careers. Despite initial reservations, by the time of his second interview, Novice A believed that metadata was "the kind of thing I need to know about", while Novice B had a keen awareness of the opportunities that might be available to someone with her newly-gained metadata experience:

> I can see myself moving onto something else again. Not tomorrow, but within a couple of years, because the professional development has been so valuable on the job here[....] Prior to coming to work here, I was chasing a couple of other opportunities for work. One was advisory work, and I'm actually pleased now that neither of those two eventuated [because....] you're using what you know already. Whereas I've gotten the opportunity to add to my bank of knowledge, while I'm putting it out there online.

Given their increased awareness of the career opportunities that might be available to them, it is reasonable to expect that metadata creators like Novices A and B will, in time, seek further professional development. The lack of current opportunities is an issue that will need to be addressed.

Discussion

The study identified a number of different techniques that could be used to transfer knowledge from experts to novices. These were training, practice and experience, and

feedback. Feedback could come directly or indirectly from a number of sources, including supervisors, peers, external quality assurers, helpdesk staff and users. No one of these techniques seemed to be more important than the others, and it is necessary for managers to strike the correct balance between the different techniques for their own organisations and for individual staff members. This may involve some experimentation, and should be based firmly on the needs of the metadata creators themselves, as identified through performance reviews or other mechanisms.

Two models for quality assurance emerged from the interviews. The two-tier model, which has been discussed at other points throughout this report and which is common for many types of metadata initiative, involves metadata creators' work being quality assured, either internally or externally, by an individual or unit with advanced knowledge of the metadata schema and of the information system as a whole. This system worked well in both types of setting covered by the study: in the centralised subject portal environment, one person quality assured the metadata that had been created by a team of subject specialists, and in the distributed e-government environment, a team of quality assurers and helpdesk staff were available to the distributed network of metadata creators.

The perceived benefits of the two-tier model were an increase in the overall quality and consistency of metadata, achieved through the sound application of information management techniques once the bulk of the content had been created. However, this approach is resource-intensive, and may not be relevant to smaller organisations. An alternative that was noted by Novice C is a cross-QA system, in which one or more people create metadata and also QA the metadata created by their colleagues: the potential of this model for smaller organisations should be considered by managers and explored in further research.

Chapter Ten: Organisational Factors Affecting Metadata Creation

> I think organisational attitudes are the most important thing because you'll get what you expect. Or you tend to get less than what you expect. But if you lower your expectations, then you'll get even lower than that. If the organisation sees this as important, and has ways to demonstrate its importance, then it's far more likely to happen. (Expert B)

In their comparison of human and machine indexing, Anderson and Pérez-Carballo argue that "It is hard to imagine that fundamental (essential) human cognitive processes do not play a large role." They note, however that application of these processes is "just as surely influenced, even determined by social forces and contexts."[100] While previous Chapters have focused on Research Question 1 and explored personal factors involved in the creation of metadata, this chapter moves outwards to the substance of Research Question 2, the *external*, in particular organisational, factors that affect the metadata creation process.

Metadata creators do not work in a vacuum, and as the study progressed, the importance of organisational culture became obvious. All novices were confident that they could create good metadata, however the degree to which they were motivated to put time and effort into the activity varied. Although a number of organisational factors were identified and are outlined below, these can be summarised into a simple statement: *The value that an organisation places on metadata is likely to have an effect on the way that metadata is created.*

Recognising, managing and rewarding metadata creators

Metadata creators are more likely to have a sense of ownership of their work if there are sound reasons for their being in the role, if their contribution is acknowledged, managed and rewarded, and if they feel that they are gaining personal benefit from it. For this to happen, metadata tasks and roles need to be well-defined, but the results of the study suggest that this does not always happen. Although Expert B argued that "the person who has the most

[100] Anderson and Pérez-Carballo, 246.

time is not necessarily the person that you most want," the study suggested that availability, rather than skillset or knowledge, was sometimes the reason a person had been allocated the task. Expert C described her own situation as follows:

> [Metadata] has absolutely nothing to do with my title and function! But I have an interest in – well not even so much an interest – but I have an ability to wade my way through computers and projects and stuff. I happened also to be free at the time, well freer than other parts of our team. So I mean that's how I fell into the role basically [....] How do I feel about being given these tasks? Given that my primary function is heading in a completely different area, sometimes I feel a bit, what's the word? If you're talking about prioritising the work, the priority for this is way down the bottom.

She also mentioned a further staff member in the organisation being recruited for metadata work because it would have "the least detrimental impact on the team if she got pulled away from her work."

Metadata creators should be acknowledged for their work. A lack of formal recognition within a person's job role could have a negative impact upon their motivation to create high quality metadata. Expert C observed this tendency:

> I think another issue [....] is the concern about how do they fit it into their already busy lives, why is it being done, and who's going to maintain it. All those ongoing management issues could well impact on how well they are doing it. So what that means is that this work has to be recognised in the job description for it to be fully accepted [....] It then becomes part of their performance review process. Otherwise, they may be going along to a performance review and saying, well actually I haven't done those things that you wanted me to do, because I've had to do this other stuff that is actually not in my job description. (Expert C)

Neither Novice A nor C initially had metadata within their job descriptions and both had metadata tasks 'lumped' on top of their normal work duties. By the end of the study, both were expecting some credit for the work that they had done. Without this recognition, the situation that Expert C envisaged would have arisen for Novice A, particularly if she had not actually enjoyed the work: "It's all reliant on people's goodwill. And I think if it hadn't been the fact that it had been quite removed from my current role, I may have lost interest really quickly [....] *I would have said, 'Get stuffed. I've got other stuff to do. This is my job description*' (emphasis added).

Novice B's situation contrasted with that of Novices A and C, in that metadata was included in her job description. This affected the way that she viewed metadata tasks: where Novice A and C gave metadata a relatively low priority and had to fit creating metadata in around their 'normal' work duties, Novice B placed a high priority on it: "It's just a necessary, a must. When the job description was shown to me, I mean, I could see that 18 hours out of 35 was for metadata, so I thought, yep, that's quite an important part of my role here, and I need to give it due priority."

Job descriptions are also important in publicly defining the role of metadata creators for other staff in the organisation whose knowledge of metadata may be limited:

> I think it is important [...] for others to understand that what [metadata creators] are doing is *important*, and carries a lot of weight in the scale of things. It's easy for someone who hasn't been involved and doesn't understand, it's easy for them to think that it's a data entry job and that can be a real barrier [...] It *isn't* a data entry job, and it's important that a certain value is put on it. It isn't rocket science, but it's not just data entry [...] There is certain intellectual input that has to go into it. (Expert A)

The above quotes from Experts A and C imply that if metadata work is seen as valuable, an organisation will ensure that the creation process is not just recognised, but properly managed. Mechanisms such as performance reviews, feedback (as discussed in Chapter Nine) and targets give metadata creators, as individuals and groups, a sense of the priority they should give to the work. The importance of these mechanisms was noted by Expert A:

> I think it's quite good to perhaps set some sorts of goals [...] for how many new records a person will be expected to cover in a week, say, or what percentage of their job is a responsibility for this. So that they have a bit of a feel for it, and to get a measure for how much time they should spend on records [....] You define the expectations because you then have to pace yourselves, know that 'No, I can't spend an hour doing this, I have to do x amount.' And it helps you manage your time, that's a time management thing. (Expert A)

Rewarding metadata creators for their achievements is another way for organisations to demonstrate their value. Novice C was pleased to receive an award because metadata creation in her organisation had been a longer and more complex process than initially

thought: “I was like, ‘Well, if you guys want to acknowledge me, that’s fine!’ [....] Nobody really had any idea of the impact that it would have on them, or what was involved in it. It was, ‘Oh yeah, it’s pretty easy, just describe your services and whack them on a portal, the people will come...’”

Expert B observed that good metadata creation emerged when work processes were tied to personal benefit, but that with metadata “the benefit might not be immediate or it might go to somebody else.” One immediate personal benefit is career advancement. Expert C believed that “You’ll eventually have a critical mass... at some point everyone will say, oh my gosh, that’s something we do have to know about, and we can see it’s important for our future career progression as well.” This appeared to be true for the participants in the study. Novices A and C thought that doing metadata work had added to their skillsets: Novice A, in particular, believed metadata work had enhanced her career development opportunities.

Knowledge of metadata within organisations

Metadata is not well understood across all sections of some organisations. Novices A and C expressed doubt that many people at their workplaces really understood what they (the metadata creators) were doing. Both noted that strategic people within the organisation, such as CEOs and managers, had some knowledge, but that the bulk of employees did not. Interestingly, neither novice particularly saw this as a problem: Novice A argued that he had “shielded” other staff from metadata and that involving more people would have made his metadata tasks more, not less, onerous because of the increased consultation that would have been involved. Novice C thought the lack of knowledge about metadata was “fair enough, actually. It’s not that they don’t have an interest, it’s either 1) they don’t fully understand or appreciate what’s trying to be achieved here, or 2) if they do, they don’t have the time to get really into it.” Both noted that they expected staff awareness of metadata to increase once the portal went live: they both thought that it would be easier to explain metadata once it was less abstract and could be seen in operation.

Knowledge within the organisation needs to be both top-down and bottom up. Expert C asserted that organisations needed systems in place to ensure that metadata creators are "feeding what they are doing to their manager who has a slightly bigger picture, who then is feeding that up as well to the senior management team so they can see the trends. Of course, the reverse also applies – the senior management will also be setting changed outputs which the managers and their staff will have to implement." This flow of information appeared to be happening, to some degree, at the organisations of all three novice creators, who described receiving and giving feedback in a range of ways, including individual meetings with their managers, regular or special group meetings, training, and internal communications such as newsletters.

Evaluating metadata

The results suggested that just as individuals perform better when they have specific targets to meet, organisations should be gathering data to assist them in evaluating their performance and improving their services to users. Expert B detailed one strategy, which involved organisations benchmarking their 'hit rates,' the percentage of searchers who found what they were after in relation to particular topics that were in the organisation's area:

> Are people finding what they want? That is really the only thing that counts. Well, plus of course how useful it was to them. You could enquire of them but people typically, users don't want to tell you, they're not inclined to say. But certainly if they describe something and it is found for them, then that would be a success factor we could measure[....] You've got the describers being given that information as well, so that they can see, because otherwise I don't see how the describer can do their job well. If they've got no feel for how it's being used, they can't get the description right. (Expert B)

Expert B also detailed how other types of evaluation noted in the section in Chapter Nine on user feedback (activity logs, statistics) could enhance organisational knowledge of metadata, while Expert C stressed the strategic importance of performance indicators such as departmental forecast report outputs.

Costs vs. benefits

Results of the study suggested that some organisations tend to perceive metadata-related work in terms of cost rather than benefits: “It’s not perceived as real work, perhaps. You know, it’s seen as an overhead on what else has to be done” (Expert B). This organisational attitude seems to foster a culture of compliance, in which individual metadata creators fail to take ownership of their work because their sense of the costs and benefits of metadata is not well-balanced. Expert B posed the question: “When a novice is approaching the task, what is their purpose? Is their purpose to create useful information for later retrieval, or is their purpose to get the hell out of there fast and get on with whatever they consider to be more important?”

At an organisational level, organisations need to understand the costs of metadata in context. Firstly, costs decrease every time an organisation re-uses the metadata it has created: Expert C noted that “You could easily have the same metadata being used in different databases and documents and your intranet and whatever, but only if the organisation perceives that it’s not doing it. You know, perceives that, understands that a unified system is valuable.” Re-purposing data internally is not really relevant to third party metadata repositories that are cataloguing external resources, but is a valuable strategy for organisations that are providing metadata for their own resources, for example, agencies involved in e-government. The results of the study suggest that this integration is not yet happening in many organisations, but that as time goes on, more integration will be likely. Experts B and C both believed that within two years, there would be far more compatibility between the metadata describing documents in internal document management systems and intranets and that used for external resources. Expert B also thought that sharing metadata within the organisation enhanced the motivation of metadata creators and the perceptions of other business units:

> If people can internalise some of the those descriptions, so that they’re using them more than once, then they’ll get better and better at it. They’ll get more and more motivated [....] If that piece of information is leveraged and used here, here, here and here [then] look! The marketing people can

> do this with it, and the HR people can do this with it, and your group can do this with it, and the people in the group next to you can do this with it... Then they can see that it's going to be used many times, and they can also see if someone else is describing something they'll be able to use it. And so they start to see we're all in this together.

In addition to the reduced costs (lack of duplication) and benefits (competitive advantage, innovation) that can be gained by leveraging information internally, some organisations benefit from sharing metadata with other organisations nationally or internationally where this fits with their mission:

> How do we stand up in the wider world of the Internet? Given that we are on the Internet and the ideal is that we want to share our information with other people. And so, how do we stand up in other people's eyes and how can we benefit from other people's doing the same? How can *this*, what we are doing here, have added value? And the added value is that by working in with other systems who are doing the same sort of things, we can benefit from the work that they are doing and they can benefit from what we are doing. And so what we have to do is communicate with other organisations about what we are doing and what they are doing so that we can capitalise on that. (Expert A)

This section has noted a mix of personal and organisational costs and benefits; it is important to note that these are inextricably related. Greenberg *et al* argue that metadata creators in organisations "may be reluctant to participate in a metadata project because they view an executive recommendation for author-generated metadata as a bureaucratic order or extra chore as opposed to an option that has rewarding benefits."[101] Novice A demonstrated how a well-defined sense of organisational costs and benefits flowed down to individual work practices. He candidly admitted that initially, when the benefits of metadata were not obvious to him, the metadata he created was at a 'compliance' level (i.e. the minimum required):

> [Metadata] that was optional I just ignored because I just wanted to get it done, you know. I didn't have enough of a sense of ownership of it to feel that this was something that I was doing for us as much as it was something that I was doing for them[....] There was the sense that we were doing it to go along with the project, you know, rather than feeling it was an essential thing to do. And that if we'd done absolutely nothing at all, that it wouldn't actually have mattered that much! (Novice A)

[101] Greenberg et al.

However, by the end of the study, Novice A was far more enthusiastic about metadata. He had thought much more about his organisation's resources ("I can see that we're going to have to think about metadata quite a lot, and make those [websites] work together and make the most of how they do work together") and was excited by the possibilities of connecting those resources with related resources in other agencies: "The fact that all those institutions are using either Dublin Core or something that can be converted into Dublin Core meant it was really easy for them to link in together. That was quite an eye-opener for me really. It's all okay in theory, you know, wonderful [but] now I can see what the point of it all is!"

Discussion

If an organisation requires high quality metadata, then it should value metadata creators and the work that they do. Managers need to assure metadata creators of the value of their work through formalising metadata tasks as part of job roles and rewarding and publicly recognising their efforts. Managers also need to foster positive attitudes to metadata throughout organisations, and act as conduits for knowledge about metadata to be passed between the different levels within the organisation. Metadata-friendly managers should see metadata in terms of benefits rather than costs, and ensure that metadata is leveraged across the organisation and shared with other organisations. They should measure their metadata-related performance using appropriate indicators and use this to improve the services offered to users. Their focus should be not on compliance but on interoperability, and metadata creation should be integrated with other aspects of the business and tied to critical processes rather than being treated as a one-off project.

The results of the study suggest that organisations in New Zealand recognise some of the pitfalls and poor practices that were identified by the experts and novices and are addressing these issues. Some metadata creators are relatively isolated within their organisations and have undertaken metadata tasks on top of their normal duties, with little acknowledgement of the affect on their workload: however, their organisations appeared to be recognising and rewarding this work. Although some metadata creators were selected

because they could be easily withdrawn from the organisation's 'real' business, they are likely to become more valued as the benefits of their good work become obvious to the organisation as a whole, through improved relationships with customers or users. Similarly, while knowledge about metadata is not currently dispersed equally within all organisations, as the benefits of metadata are more widely seen, the process of creating it is likely to become better planned, more integrated into business processes and better understood across the organisation as a whole. As this happens, metadata creators will become more valued, and consequently will place a higher priority on their metadata tasks.

Chapter Eleven: Conclusion

I undertook this exploratory study because I was interested in finding out about the experiences of the people with little or no prior experience of cataloguing, indexing or abstracting, who are creating metadata. In particular, I wanted to explore the view prevalent in LIS literature – that metadata creators who are not library professionals cannot create metadata of sufficient quality – from the perspective of the novice creators themselves and the supervisors and trainers who are responsible for their development. I looked at personal (especially cognitive) factors involved in learning to create metadata, other external factors that influence the metadata production process, and some of the issues around managing the metadata production process. By gathering the views of people with different levels of experience, I have been able to paint a picture of how novices initially approach metadata creation and what types of training and tools are most helpful for them.

The transcripts of interviews with metadata experts and novices yielded information about the range of skills and the types of knowledge that creating metadata requires. The results suggested some tensions between qualities required for creating metadata and the skillsets and knowledge base of information professionals. In particular, the study showed that library professionals can be perceived as inflexible and unable to adapt to the pragmatic demands of metadata creation. This perception is partially responsible for the sidelining of information professionals in some organisations with metadata initiatives underway. Librarians have not helped themselves by insisting that metadata is just cataloguing: it is perhaps time to focus on the parts of metadata creation that we are good at, but to admit that other types of skills and knowledge are also required. A more team-oriented approach and models for achieving the optimum mix of skills have been discussed in this report.

The study showed that some of the cognitive processes involved in creating metadata are similar to those that have been modelled for cataloguing, indexing and abstracting, but that some others appeared to be quite distinct. The need for further research into a number of areas was identified. Finding out whether metadata creators work in a linear fashion or in a

less structured way might impact upon how workflows are managed and tools are developed in future. Research is also needed into the particularities of web-based resources as opposed to more traditional document formats, and the effect of these characteristics upon technical reading processes.

Regardless of what they bring to their work initially, metadata novices require a toolkit and training, feedback and further development. Participants in the study provided valuable insight into what works and what does not in terms of the software interfaces, hardware and networks, controlled vocabularies and encoding schemes, and documentation that they had used. I was also able to glean from their interview transcripts an understanding of the role of different types of training and feedback. A need for training and further development was identified not just for novices, but for library professionals with experience of more traditional types of bibliographic processing: this reiterated that creating metadata is not simply cataloguing, and that some up-skilling or cross-training is required, even for experienced cataloguers and indexers. The study also identified the currently limited range of opportunities for metadata creators to develop their skills and knowledge once they have attained a level of competency. A number of participants believed that creating metadata would enhance their career opportunities; if so, demand for more formal professional development is likely to increase, something which must be considered collectively by the range of organisations involved in metadata initiatives in New Zealand.

Creating metadata not only involves cognitive processes; a range of business processes and organisational factors also come into play. The importance of these factors cannot be underestimated. Poor organisational understandings of metadata, insufficient integration with other core business processes, a culture of compliance rather than interoperability, and a lack of validation of the work that metadata creators do are just as, if not more, likely to affect the quality of metadata than the skill level of the person creating metadata. Managers have an important role to play in ensuring that metadata creators have access to a good toolkit and the training and feedback that they need to develop their expertise in the field; but it is equally important that managers ensure that metadata creators

know that they are not engaged in a pointless administrative exercise but are creating a valuable asset that will enable the organisation to improve its internal processes, share resources with others, and help users to find the information that they are seeking.

Word count=27,296 (including participant quotes)

Works Consulted

Anderson, James D., and Jose Pérez-Carballo. "The Nature of Indexing: How Humans and Machines Analyze Messages and Texts for Retrieval. Part I: Research, and the Nature of Human Indexing." *Information Processing and Management* 37, no. 2 (March 2001): 231-54.

Beghtol, Clare. "Bibliographic Classification Theory and Text Linguistics: Aboutness Analysis, Intertextuality and the Cognitive Act of Classifying Documents." *Journal of Documentation* 42, no. 2 (1986): 84-113.

Bertrand, A., and J-M. Cellier. "Psychological Approach to Indexing: Effects of the Operator's Expertise upon Indexing Behaviour." *Journal of Information Science* 21, no. 6 (1995): 459-72.

Bordeianu, Sever Michael, and Virginia Seiser. "Paraprofessional Cataloguers in ARL Libraries." *College and Research Libraries* 60, no. 6 (1999): 532-40.

Burton, Paul F. "The Decline and Fall of 'Cat. & Class.'" *Catalogue and Index* 124, no. Summer (1997): 9.

———. "Issues and Challenges of Subject Access." *Catalogue and Index* 128, no. Summer (1998): 1-7.

Caldwell, A., and D. Coulombe. "Never the Twain Shall Meet? Collaboration Between Catalogers and Reference Librarians in the OCLC CORC Project at Brown University." *Journal of Internet Cataloging* 4, no. 1/2 (2001): 123-30.

Calhoun, Karen, and John J. Reimer. "CORC: New Tools and Possibilities for Cooperative Electronic Resource Description." *The Journal of Internet Cataloging* 4, no. 1/2 (2001).

Clayden, Judith. "Theory Versus Practice in Cataloging Education: Some Australian Experiences." *Journal of Education for Library and Information Science* 36 (1995): 230-8.

Connaway, Lynn Silipigni. "A Model Curriculum for Cataloging Education: The Library and Information Services Program at the University of Denver." *Technical Services Quarterly* 15, no. 1/2 (1997): 27-41.

David, C., L. Giroux, S. Bertrand-Gastaldy, and D. Lanteigne. "Indexing as Problem Solving: A Cognitive Approach to Consistency." Paper presented at the Annual Conference of the Canadian Association for Information Science: Connectedness: Information, Systems, People, Organizations., Albert University School of Library and Information Studies, Edmonton, Alberta 1995. http://www.ualberta.ca/dept/slis/cais/ [accessed 15 January 2002].

De Boer, Ann-Louise. "Teaching Cataloguing and Classification at the University of Pretoria: Thinking Preferences of Second Year Students." *Libri* 51, no. 2 (2001): 14-23.

DESIRE (Development of a European Service for Information on Research and Education). *DESIRE Information Gateways Handbook* (26 April 2000). http://www.desire.org/handbook/welcome.html [accessed 4 August 2001].

Dillon, Martin, and Erik Jul. "Cataloging Internet Resources: The Convergence of Libraries and Internet Resources." *Cataloging and Classification Quarterly* 22, no. 3/4 (1996): 197-238.

Dublin Core Metadata Initiative. *Dublin Core Qualifiers* http://dublincore.org/documents/dcmes-qualifiers/ [accessed 30 June 2002].

Edmunds, Jeff, and Roger Brisson. "Cataloging in CORC: A Work in Progress." *Journal of Internet Cataloging* 4, no. 1/2 (2001): 89-109.

Endres-Niggemeyer, B., and E. Neugebauer. "Professional Summarizing: No Cognitive Simulation Without Observation." *Journal of the American Society for Information Science* 49, no. 6 (1988): 486-506.

Farrow, John J. "A Cognitive Process Model of Document Indexing." *Journal of Documentation* 47, no. 2 (1991): 149-66.

Fidel, R. "User-Centered Indexing." *Journal of the American Society for Information Science* 45, no. 8 (1994): 572-76.

Garnett, LeAnn. "Dewey, Dale and Bruner: Educational Philogophy, Experiential Learning and Library School Cataloging Instruction." *Journal of Education for Library and Information Science* 38 (1997): 129-36.

Gee, James Paul. "Thinking, Learning and Reading: The Situated Sociocultural Mind." In *Situated Cognition: Social, Semiotic and Psychological Perspectives*, edited by David Kirshner and James A. Whitson, 235-59. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

Gilliland-Swetland, A.J., Y.B. Kafai, and W.E. Landis. "Application of Dublin Core Metadata in the Description of Digital Primary Sources in Elementary School Classrooms." *Journal of the American Society for Information Science* 51, no. 2 (2000): 193-201.

Glaviano, C. "Teaching an Information Organization Course with Nordic DC Metadata Creator." *OCLC Systems and Services* 16, no. 1 (2000): 33-40.

Gorman, G.E., Peter Clayton, and with contributions from Mary Lynn Rice-Lively and Lyn Gorman. *Qualitative Research for the Information Professional*. London: Library Association Publishing, 1997.

Gorman, Michael. "How Cataloging and Classification Should be Taught." *American Libraries* 23, no. September (1992): 694+.

———. "Metadata or Cataloging? A False Choice." *Journal of Internet Cataloging* 2, no. 1 (1999): 5-22.

Greenberg, Jane. "Metadata Questions in Evolving Internet-Based Educational Terrain." *Journal of Internet Cataloging* 3, no. 1 (2000): 1-11.

Greenberg, Jane, Maria Cristina Pattuelli, Bijan Parsia, and W. Davenport Robertson. "Author-Generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization." *Journal of Digital information* 2, no. 2 (2001). http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/ [accessed 30 June 2002].

Hakala, Jula. "Internet Metadata and Library Cataloguing." *International Cataloguing and Bibliographic Control* 28, no. 1 (1999): 21-25.

Hansen, Preben. "Evaluation and Design Issues of Nordic DC Metadata Creation Tool." Paper presented at the International Conference on Dublin Core and Metadata Applications, National Institute of Informatics,Tokyo, Japan, 2001. http://www.nii.ac.jp/dc2001/proceedings/product/paper-33.pdf [accessed 8 April 2002].

Heery, Rachel, and Manjula Patel. "Application profiles: Mixing and Matching Metadata Schemas." *Ariadne* 25 (September 2000). http://www.ariadne.ac.uk/issue25/app-profiles/ [accessed 30 June 2002].

Hill, Debra W. "Requisite Skills for the Entry-Level Cataloger: A Supervisor's Perspective." *Cataloging and Classification Quarterly* 23, no. 3/4 (1997): 75-83.

Howarth, Lynne C. "The Role of the Paraprofessional in Technical Services in Libraries." *Library Trends* 46, no. 3 Winter (1998): 526-39.

Hsieh-Yee, Ingrid. "Organizing Internet Resources: Teaching Cataloging Standards and Beyond." *OCLC Systems and Services* 16, no. 3 (2000): 130-43.

Jeng, Judy. "Expert System Applications in Cataloging, Acquisitions and Collection Development: A Status Review." *Technical Services Quarterly* 12, no. 3 (1995): 17-28.

Jeng, Ling Hwey. "Using Verbal Reports to Understand Cataloging Expertise: Two Cases." *Library Resources and Technical Services* 40, no. 4 (1996): 343-58.

Jeng, Ling Hwey, and Karen B. Weiss. "Modeling Cataloging Expertise: A Feasibility Study." *Information Processing and Management* 30, no. 1 (1994): 119-29.

Kirshner, David, and James A. Whitson. "Editors' Introduction." In *Situated Cognition: Social, Semiotic and Psychological Perspectives*, edited by David Kirshner and James A. Whitson, 1-16. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

Lancaster, F.W. *Indexing and Abstracting in Theory and Practice*. London: Library Association Publishing, 1998.

Lave, Jean. "The Culture of Acquisition and the Practice of Understanding." In *Situated Cognition: Social, Semiotic and Psychological Perspectives*, edited by David Kirshner and James A. Whitson, 17-35. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

Lee-Smeltzer, Kuang-Hwei (Janet). "Finding the needle: controlled vocabularies, resource discovery, and Dublin Core." *Library Collections, Acquisitions, & Technical Services* 24, no. 2 (2000): 205-15.

Library of Congress. *Metadata Encoding & Transmission Standard (METS): Official Web Site* http://www.loc.gov/standards/mets/ [accessed 30 June 2002].

MacLeod, Judy, and Daren J. Callahan. "Educators and Practitioners Reply: An Assessment of Cataloging Education." *Library Resources and Technical Services* 39 (1995): 153-65.

Medeiros, Norm. "On the Dublin Core Front." *OCLC Systems and Services* 16, no. 1 (2000): 41-43.

Milstead, J.L. "Needs for Research in Indexing." *Journal of the American Society for Information Science* 45, no. 8 (1994): 577-82.

Milstead, Jessica , and Susan Feldman. "Metadata: Cataloging by Any Other Name." *Online* January (1999). http://www.onlinemag.net/OL1999/milstead1.html [accessed 15 June 2002].

Milstead, Jessica, and Susan Feldman. "Metadata Projects and Standards." *Online* January (1999). http://www.onlineinc.com/onlinemag/OL1999/milstead1.html#projects [accessed 29 August 2001].

Mohr, D.A., and A. Schuneman. "Changing Roles: Cataloging by Paraprofessionals in ARL Libraries." *Library Resources and Technical Services* 41, no. 3 (1997): 205-18.

Morkes, John, and Jakob Nielsen. *Concise, Scannable, and Objective: How to Write for the Web* http://www.useit.com/papers/webwriting/writing.html [accessed 16 June 2002].

Nielsen, Jakob. *Alertbox for March 15, 1997: Be Succinct! (Writing for the Web)* http://www.useit.com/alertbox/9703b.html [accessed 16 June 2002].

Olmstadt, William. "Cataloging Expert Systems: Optimism and Frustrated Reality." *Journal of Southern Academic and Special Librarianship*, (2000). http://southernlibrarianship.icaap.org/content/v01n03/olmstadt_w01.html [accessed 4 January 2002].

Ovid Technologies. *PsycInfo Database Thesaurus* http://ovid3.distance.scs.vuw.ac.nz/ [accessed 16 January 2002].

Phillips, Margaret E. "Managing Chaos in the Cyberworld." Paper presented at the CONSAL 2000, the 11th Congress of Southeast Asian Librarians Conference, Singapore, 26th - 28th April 2000. http://www.nla.gov.au/nla/staffpaper/mphillips5.html [accessed 18 August 2001].

Pinto, M., and F.W. Lancaster. "Abstracts and Abstracting in Knowledge Discovery." *Library Trends* 48, no. 1 (1999): 234-48.

Robare, Lori Proudfit. "Training Paraprofessionals in Classification and Subject Analysis." *Technical Services Quarterly* 14, no. 2 (1996): 49-63.

Romero, Lisa. "The Cataloging Laboratory: The Active Learning Theory Applied to the Education of Catalogers." *Cataloging and Classification Quarterly* 21, no. 1 (1995): 3-17.

———. "Original Cataloguing Errors: A Comparison of Errors Found in Entry-Level Cataloging with Errors Found in OCLC and RLIN." *Technical Services Quarterly* 12, no. 2 (1994): 13-27.

Sladen, Catherine. "Biz/ed: Trusting the Providers." In *Electronic Library and Visual Information Research - ELVIRA 4*, edited by Claire Davies and Anne Ramsden, 75-81. London: Aslib, 1997.

Slater, Margaret. "Qualitative Research." In *Research Methods in Library and Information Studies*, edited by Margaret Slater, 107-27. London: Library Association, 1990.

Sokvitne, Lloyd. "An Evaluation of the Effectiveness of Current Dublin Core Metadata for Retrieval." Paper presented at the VALA2000 Conference: Books and Byes: Technologies for the Hybrid Library, Melbourne, Victorian Association for Library Automation 2000. http://www.vala.org.au/vala2000/2000pdf/Sokvitne.PDF [accessed 18 August 2001].

———. *Understanding metadata* (26 March 2001). http://www.servicetasmania.tas.gov.au/papers/understanding_metadata.htm [accessed 8 August 2001].

Taylor, Arlene. *The Organization of Information*. Englewood, Co.: Libraries Unlimited, 1999.

Thornely, J. "Metadata and the Deployment of Dublin Core at State Library of Queensland and Education Queensland, Australia." *OCLC Systems and Services* 16, no. 3 (2000): 118-29.

Watson, Jinx Stapleton. "Making Sense of the Stories of Experience: Methodology for Research and Teaching." *Journal of Education for Library and Information Science* 42, no. 2 (2001): 137-48.

Weibel, Stuart. "The Dublin Core: A Simple Content Description Model for Electronic Resources." *Bulletin of the American Society for Information Science*, (Oct/Nov 1997): 9-11.

Weibel, Stuart, Renato Iannella, and Warwick Cathro. "The 4th Dublin Core Metadata Workshop Report." *D-Lib Magazine* June (1997). http://www.dlib.org/dlib/june97/metadata/06weibel.html [accessed 5 August 2001].

Worsfold, Emma. *Distributed and Part-Automated Cataloguing: A DESIRE Issues Paper* (March 1998). http://www.sosig.ac.uk/desire/cat/cataloguing.html [accessed 18 August 2001].

Younger, J.A. "Support Staff and Librarians in Cataloging." *Cataloging and Classification Quarterly* 23, no. 1 (1996): 27-47.

Appendix A: Sample Recruitment Letter (Novice Participant)

134 Washington Ave
Brooklyn, Wellington
Tel: 04 389 2218
Email: samsearle@paradise.net.nz

[Date]

[Name]
[Address]

Dear [Name]

I am writing to ask you to participate in a research study that I am conducting as part of my Master of Library and Information Studies programme at Victoria University.

I am investigating how people new to creating metadata go about this process, and what support might be put in place to make their transition from novice to expert as quick and easy as possible. As someone who has recently begun creating metadata, your help in identifying aids and obstacles that you have encountered would be greatly appreciated.

Your participation in the project will consist of two interviews. Each interview should take no more than an hour and will be recorded on audio tape. The interviews will then be transcribed and you will be given the opportunity to check the accuracy of the transcripts.

The interviews will cover topics such as:

- your general impressions of the metadata creation process
- what processes you go through when creating metadata
- what areas you have found easy or difficult
- what skills and knowledge you feel you are gaining
- what factors in your environment impact upon how you create metadata

You will be free to bring up any other topics that you think are important and you can bring metadata records that you have created to the interviews if these will help you describe how you create metadata.

Permission to conduct this research has been gained from your employer [insert name]. Participation in the study is voluntary, and you may withdraw from the study at any time prior to the final analysis of the data: you do not have to provide any reasons for withdrawing. Access to the data collected during the study will be restricted to my supervisor and myself, and will be stored securely and then destroyed at the end of the project. Your name and the name of your organisation will be kept confidential, and you will only be identified in the report using a letter code.

I hope that you will be able to participate in the study. Currently, there is very little research in this area and your input could lead to more effective training, documentation and tools.

An information sheet that provides further information about the study is enclosed. If you would like to participate in the project, please sign the consent form and return it in the pre-paid envelope provided. I will then contact you to arrange an interview time.

If you require any further information, please contact me by email at samsearle@paradise.net.nz or by telephone on 04 389 2218. You can also obtain information from my supervisor, Dr Dan Dorner (dan.dorner@vuw.ac.nz, phone 04 463 5781).

Yours sincerely

Samantha Searle
Investigator

Appendix B: Sample Permission Letter (Manager of Organisation)

134 Washington Ave
Brooklyn, Wellington
Tel: 04 389 2218
Email: samsearle@paradise.net.nz

[Date]

[Name]
[Address]

Dear [Name]

I am writing to ask your permission to interview your staff member, [insert name of participant], for a research project that I am conducting as part of the Master of Library and Information Studies programme at Victoria University.

The study will investigate how people new to creating metadata go about this process, and what support might be put in place to make their transition from novice to expert as quick and easy as possible. As someone who [trains and/or supervises metadata novices / has recently begun creating metadata], [insert name of participant]'s contribution would be greatly appreciated.

[insert name of participant]'s participation in the project will consist of an interview. The interview should take no more than an hour and will be recorded on audio tape. The interview will then be transcribed and [insert name of participant] will be given the opportunity to check the accuracy of the transcript.

The interview will cover topics such as:

- the participant's general impressions of metadata
- the processes that the participant goes through when creating metadata
- what areas within a metadata record the participant has found easy or difficult
- what skills and knowledge the participant feels she/he is gaining
- what factors in the participant's environment impact upon how she/he creates metadata

The research will be confidential: neither [insert name of participant]'s name nor the name of your organisation will be used. Access to the data collected during the study will be restricted to my supervisor and myself, and will be stored securely and destroyed at the end of the project.

I hope that you will grant permission for [insert name of participant] to participate in the study. Currently, there is very little research in this area and [insert name of participant]'s input could lead to more effective training, documentation and tools.

An information sheet that provides further information about the study is enclosed. I would be very grateful if you could take the time to read about the project and contact me regarding permission to conduct the research. I will then contact [insert name of participant] to gain [her/his] consent and arrange an interview time.

If you require any further information, please contact me by email at samsearle@paradise.net.nz or by telephone on 04 389 2218. You can also obtain information from my supervisor, Dr Dan Dorner (dan.dorner@vuw.ac.nz, phone 04 463 5781).

Yours sincerely

Samantha Searle
Investigator

Appendix C: Information Sheet

This study is being conducted as part of the requirements for the Master of Library and Information Studies programme at Victoria University of Wellington and it complies with the University's human ethics guidelines.

The aim of the study is to find out how novices approach the process of creating metadata, and what support might be put in place to make their transition from novice to expert as quick and easy as possible. Effective staff training, helpful documentation, usable online tools, and more supportive peer networks might be facilitated by the identification of aids and obstacles to the acquisition of skills and knowledge.

Participants in the study will be asked to take part in one or two interviews. Metadata experts who are involved in training and supervision will be interviewed once: opinions will be sought on the skills required for metadata creators, the ways in which novices gain expertise, and the methods used by experts to transfer their knowledge to novices. Metadata novices will be interviewed twice, at the time of their introduction to metadata creation and at a later date when they have created a significant number of metadata records. They will be asked about their general perceptions of the metadata creation process, what areas they have found easy or difficult, what skills and knowledge they feel they are gaining, and whether other factors impact upon how they create metadata. Interviewees will also be invited to bring copies of metadata records they have created to the interview.

The research will be **confidential**. The researcher will identify individuals using letter codes in the final report (for example, Novice A), and will not use the name of any people or organisations. Data that the researcher believes could lead to the identification of an individual or an organisation will not be reported.

During the course of the study, data will be restricted to the researcher and her supervisor, and will be stored securely. All data collected will be destroyed at the end of the project.

Copies of the research paper will be deposited in the Victoria University of Wellington Library. Results of the research may be presented in academic or professional journals or at conferences. A summary of the research may be posted to the DC-ANZ email list [Dublin Core Australia and New Zealand] and copies of the final report may be supplied on request to interested metadata professionals in New Zealand and overseas.

During the study, participants will be given the opportunity to view their interview transcripts. At the conclusion of the study, participants will be given a summary of the research and invited to provide feedback to the researcher.

Further information can be obtained from the researcher, Samantha Searle, who can be contacted by email at samsearle@paradise.net.nz, or by telephone on 04 389 2218. Information can also be obtained from the supervisor of the project, Dr Dan Dorner (dan.dorner@vuw.ac.nz, phone 04 463 5781).

Appendix D: Consent Form

I have been provided with adequate information relating to the nature and objectives of this research project, I have understood that information and have been given the opportunity to seek further clarification or explanations;

I understand that I may withdraw from this study at any time prior to the final analysis of the data;

I understand that the research is confidential, that neither my name nor the name of my organisation will be used, and that other information that might identify me or my organisation will not be reported;

I understand that the information I have provided will be used only for this research project and that any further use will require my written consent;

I understand that the results of the study may be published in academic or professional journals and presented at academic or professional conferences, that a summary of the research may be posted to academic or professional email lists, and that copies of the final report will be deposited in the Victoria University of Wellington Library and may be supplied on request to academics and professionals in New Zealand and overseas;

I understand that I will have the right to check the transcript of my interview and to provide feedback to the researcher at any time.

Signature _____

Name _____

Date _____

☐ I would like to receive a summary of the research. (please tick)

Appendix E: Sample Interview Schedule (Expert)

Can you talk me step by step through the way that you personally would create a metadata record for a webpage or a document? (RQ 1 & 3, RO 2 & 6)

How would you expect the process that an expert like yourself goes through to be similar/different from how a novice would approach the tasks? (RQ 1 & 3, RO 2 & 6)

Which sections of the metadata record do you think are the most easy and the most difficult to gain expertise in, and why? (RQ 1 & 3, RO 2 & 6)

What skills and knowledge you would look for when recruiting people to do metadata? What personal qualities would you look for also? (RQ 1 & 3, RO 1 & 6)

What training process do you usually provide for someone starting to create metadata? How did you develop this training process? How effective do you think that it is? (RQ 1 & 3, RO 4 & 6)

Once someone has been trained and has started creating metadata, how are their skills and knowledge further developed? [Try to draw out information about further training, supervision, feedback, appraisal, practice] (RQ 1 & 3, RO 4 & 6)

In your experience, what are the job roles within organisations that usually incorporate metadata? How appropriate do you think people in these roles are? What other types of role/types of people are well-suited to metadata tasks? (RQ 1, 2 & 3, RO 1 & 6)

How well do you think metadata tasks are integrated into the overall workflow of the teams/departments you work with? (RQ 2 & 3, RO 5 & 6)

A number of systems/tools are required when creating metadata, e.g. Internet connection, metadata interface, online and/or print thesauri. What difficulties do you think arise for people when they start using these tools? (RQ 2 & 3, RO 5 & 6)

What sort of documentation do you think is most effective in assisting those new to metadata creation? How are the needs of experts different from novices with regards to documentation? (RQ 2 & 3, RO 5 & 6)

How would you describe the relationship between people who create metadata and the users of the portal/service? How does this relationship develop over time? (RQ 1, 2 & 3, RO 1, 2, 5 & 6)

How do other environmental and organisational factors affect the ability of people to create metadata? (e.g. time constraints, workflows, financial support, organisational attitudes?) (RQ 2 & 3, RO 5 & 6)

Appendix F: Sample Interview Schedule (Novice)

What is your role in your organisation and how does creating metadata fit into that role? How do you feel about being given these tasks? (RQ 1, 2 & 3, RO 1, 3, & 6)

Why is your organisation creating metadata? How do you think the metadata you create will be used? (RQ 1 & 3, RO 3 & 6)

What training have you been given? What are your impressions of the training? (RQ 1 & 3, RO 2, 4 & 6)

Apart from your training, what other methods have you used to gain knowledge about metadata? (RQ 1 & 3, RO 2, 4 & 6)

Looking at the various elements in the NZGLS scheme, can you talk me through the way that you think you are going to create metadata records? (RQ 1, 2 & 3, RO 2, 3 & 6)

Which sections of the metadata record do you think are going to be the most easy and the most difficult to complete, and why? (RQ 1 & 3, RO 2 & 6)

What do you think you will do if you have any problems? (RQ 1, 2 & 3, RO 2, 3, 4, 5 & 6)

Do you think that creating metadata will be similar to other job tasks that you have done? How? (RQ 1 & 3, RO 2 & 6)

What skills/knowledge or personal qualities do you have that you think will assist in creating metadata? (RQ 1 & 3, RO 1 & 6)

Can you describe the systems/tools that you will be using to create metadata (e.g. metadata interface, SONZ and FONZ)? What are your initial impressions of these tools? (RQ 1 & 3, RO 3, 5 & 6)

Can you describe the type of documentation that you have been given/will refer to? What are your initial impressions of the documentation? (RQ 1 & 3, RO 2 & 6)

What is the general attitude of people in your organisation to metadata issues? (RQ 2 & 3, RO 5 & 6)

What types of support will you receive in your workplace when you are creating metadata? (RQ 2 & 3, RO 5 & 6)

How will you be working with other people in your organisation to create metadata? How will you be working with people from other organisations when you create metadata? (RQ 2 & 3, RO 5 & 6)

How much control will you have over how to fit metadata tasks into your personal workload? How will metadata tasks be incorporated into the workflow in your team/department? (RQ 2 & 3, RO 5 & 6)