

Author: R. John Robertson
Contact details:
Centre for Digital Library Research
Department of Computer and Information Sciences
University of Strathclyde, Livingstone Tower
26 Richmond Street, Glasgow, G1 1XH, UK
Tel: +44 (0) 141 548 5854
Email: robert.robertson@cis.strath.ac.uk

Title: Metadata quality: implications for library and information science professionals

Paper type: Viewpoint

Structured abstract:

Purpose

In contrast to recent studies noting the necessity of library and information science skills on digital library and repository projects, this study examines the impact of metadata quality requirements on how library and information science professionals apply their skills outside of a library setting.

Methodology/Approach

The paper reviews the concept of metadata quality and examines the implications of this for LIS professionals by reviewing the differences between the context of the library community and other relevant communities of practice.

Findings

The paper argues that, although much needed, library and information science skills require contextualisation before application outside of library settings.

Research limitations/implications

Many of the new opportunities for and settings of library and information science skills are immature – consequently this analysis may date as the context of these settings mature. Current trends however, suggest that it will not.

Practical implications

Training in library and information science skills should take account of how they might apply differently outside of libraries.

Librarians co-operating with colleagues outside of the library should appreciate the potential metadata ‘compromises’ they might have to make and why they are necessary.

Originality/value of paper

The paper provides a food for thought for the increasing number of library and information science professionals working outside library settings.

Keywords:

Library and information science careers; Metadata; Metadata quality; Institutional repositories; Learning object repositories; Digital Libraries

Metadata quality: implications for library and information science professionals

A first glance at the title might lead you to suspect that a mistake had been made in proofreading, that the title should have been the implications of library and information science (LIS) for metadata quality. There is much such a piece could say in stressing the role of LIS professionals in ensuring the quality of metadata in various settings and its importance for the discovery and management of resources within them. The title is the right way round however, and, after some preliminary stage setting through looking at metadata quality, this article suggests that there are important lessons the LIS community needs to learn as they engage with these new settings of metadata. These settings of metadata quality include digital libraries, institutional repositories, subject repositories, learning object repositories and

the wider web environment. This article focuses on the more controlled settings of the first four: digital libraries – which contain a collection of digitized or born-digital resources; institutional repositories – which contain collections of the research output of an institution; subject repositories - which contain collections of research output but are subject discipline based; and learning object repositories – which contain collections of learning objects (collectively hereafter digital repositories). The role of and need for LIS skills within these settings is beginning to be understood and advocated.¹

Metadata

There is a sense in which the LIS community is in the metadata business. Whether defined simply as 'data about data' or more precisely as:

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. (National Information Standards Organization, 2004, p. 1)

it is clear that the library community has been creating metadata records in the form of catalogue entries as long as there have been libraries. Metadata though refers to much more than library catalogue records and the community involved in its creation is much wider than the just the library community. This wider community of researchers and practitioners developing and implementing these digital repositories is often not based in a library per se and involves core support skills from computer science and learning technology as well as LIS, together with enthusiasts and from a great diversity of other disciplines as well.

Metadata quality

Within this community of interest, there are those, often from an LIS background, involved in examining the issue of metadata quality through attempting to describe and evaluate the characteristics of a metadata record in an effort to inform and assess the processes by which it is created and the functionality it can support. In their discussion of metadata for the semantic web, Greenberg and Robertson (2002, p45) succinctly describe quality metadata as "accurate, consistent [and] sufficient". The idea of sufficiency is double-edged though and refers back to the primary and over-riding definition for quality in any setting: fitness for purpose - as true for metadata as it is for designing a car or boiling an egg.

Supporting the development of quality metadata is perhaps one of the most important role for LIS professionals. Although not always understood or appreciated skills that fall under the LIS remit are key to the future success of digital repository endeavours. These key skills include: an awareness of how to address the aforementioned aspects of quality (e.g. why clear guidelines for title construction aid; how authority files aid accuracy); understanding of the implications of making compromises in metadata quality within large systems; and teaching 'e-literacy' (i.e. including both ICT literacy and Information Literacy (Secker and Price, 2004 p98)).

The library setting

The knowledge that LIS professionals have in these areas arises from the communal experience of the library community past and present. Although the profession has undergone rapid change with the advent of computers and electronic resources, the information skills remain constant.

Within libraries there is a clear purpose: to find, retrieve and manage the stored objects. Whether books, videos, journals, minutes or anything else, this is accomplished through a central catalogue and the record is created according to specific rules. The commonly used rules for creating such metadata are currently the 2nd edition of the Anglo-American Cataloguing rules (AACR2) which set out in some detail how what is recorded should be expressed. To give an example of such a rule (from the concise version):

3A4. Ceased Serials. If the serial has ceased publication, give the designation and/or date of the first issue or part followed by the designation and/or date of the last issue (Gorman, 2004).

¹ See for example: Currier et al. 2004; El-Sherbini and Klim, 2004

Likewise there are extensive rules about applying particular classification schemes. Although within any given library the implementation of these rules and the completeness of a record will be interpreted through local priorities and resource constraints, there is an acknowledgement that a, nearly, perfect record is possible. There are also mechanisms which allow libraries to buy or exchange this agreed 'perfect' minimal record from external sources to reduce the volume and cost of in-house cataloguing. Mechanisms such as this can exist because the library community has shared purpose and conception of metadata quality, which allows an agreed 'level' for exchange. As the metadata for a digital object contains basic bibliographic metadata there is a sense in which digital objects can be viewed as library objects (El-Sherbini and Klim, p245-246). This is not however, the whole story.

Implications of defining metadata quality outside the library

To return to the title, the question that metadata quality poses to the assumptions of the LIS community relates to the purpose of the metadata created within the setting of a digital repository. As outlined above within the library community, the purpose is understood and the context is clearly limited. Yet this purpose and context is not necessarily transferable to a given digital repository. The clear rules and implicit context of libraries mean that LIS professionals working within the new digital repository environment are in danger of sustaining assumptions about how their skills should be implemented from the library setting into their new setting unquestioned.

That different settings and purposes require different types of metadata quality should be no surprise as there are already other domains of knowledge management which have very different standards and purposes. The museum and the archive communities take a different approach to what represents quality in metadata. Museums record extensive detail about the provenance of an object as a necessary part of their purpose. Archives record extensive information but often only at the collection, rather than object, level due to the volume of materials they manage. These different purposes have existed side by side within the traditional knowledge management domain with little transference between. The metadata record for the same book will look very different in each setting and no one option is objectively better.

With the creation of digital collections in libraries, archives, and museums there has been a greater degree of collaboration between these traditional domains. The distinct metadata requirements for their peculiar purposes however, remain an issue in their collaboration (Spinazzé, 2004). These differences are relatively acknowledged but the question of purpose becomes even more important as other functions (and the metadata to support them) are now also required of digital repositories. These new functions may include: the generation of faculty web pages; departmental research output reporting; publishing documents to facilitate Freedom of Information act compliance; internal asset management; long term preservation of assets. These new functions may be required by administrators, funding bodies, or may be needed to get individuals to participate in the repository (Harnad et al 2003; Foster and Gibbons, 2005). For a repository to support such a range of purposes it may have to compromise on the support it can give to any one of them.

A development in digital repositories which illustrates the effect of different purposes and contexts on LIS assumptions is that metadata records are often required at a greater degree of granularity. Asset management or research reporting requires metadata records at the level of individual article or minutes – a far cry from a single entry for an entire run of a journal (as assumed in AACR2).

A thread running throughout digital repository development is that the metadata requirements to support preservation, accessibility, or intellectual property rights are as yet not fully known. Irrespective of the chosen future solution to these issues, they will require new skill sets. Whilst not every record may require input, every repository will have to consult or employ these skill sets.

The situation becomes more complex when the metadata requirements of those digital repositories storing new types of materials are considered. In the instance of learning objects,

a metadata record using the IEEE LOM standard (IEEE, 2004) is as complex as a MARC record but has a smaller bibliographic description and supports extensive educational description of the nature and use of the resource. By implication, such a record requires different skills to create its different parts. The use and life expectancy of such learning resources is however, a very unknown quantity and it remains to be seen how justifiable an investment in extensive and precise cataloguing is.

In the instance of large data sets which require complex metadata that is not knowable or obvious from the data itself, some metadata creation by its creators is needed. This development which is a necessity for data sets may also prove to be beneficial in the cataloguing of more simple resources, such as a web page, as the creator may know information that a cataloguer may not. The potential role of resource creators in metadata creation is another area of extensive investigation within digital repository research, which doesn't mesh with normal library practice (Greenberg and Robertson 2002).

Summary observations and implications

From this brief exploration of the purpose and context of the new settings of metadata creation some observations can be made and implications drawn.

Observation: The purpose of a repository may not solely be that of a library

Implication: The metadata required to support such multiple purposes, will require the use of new or multiple standards, and may demand compromising on library metadata guidelines.

Observation: Different purposes may require different approaches to granularity and thus increase the metadata creation workload

Implication: The granularity required for a given purpose and the scale of the digital repository may influence what metadata can be provided and how it is created.

Observation: Some of the resources in a repository may have a very short or an unknown life-cycle but require very complex metadata.

Implication: The nature of the resource being described should influence how much metadata is created.

Observation: The requirements placed on metadata are increasing and require skills other than the core LIS abilities.

Implication: LIS can't provide all the answers. Repositories are growing increasingly complex, and require cross-domain partnerships to function effectively and efficiently.

Observation: Resource creators may have to be involved in the metadata creation process

Implication: The metadata creation process may require a multi-agent workflow which includes agents without the LIS skill set.

Concluding thoughts

These implications suggest that metadata creation, though greatly benefiting from the LIS skill set, are often not undertaken within the existing library model. LIS professionals working with digital repositories need to step back and assess the context in which they are working and the requirements that context imposes on them. If they do not there is a risk that digital repositories will evolve without informed LIS input and will suffer until they reinvent the wheel, or there is a risk that repositories within LIS control are unable to survive in the marketplace as their cost per record becomes unsustainable.

Providing quality metadata within a digital repository cannot normally be about describing everything - such an approach is unsustainable. The challenge of metadata quality for LIS is that the core skills of the profession need to be applied in settings that are outside of familiar territory and require different rules. LIS skills are vital for digital repositories to function well but they need to operate with the appropriate frame of reference. Consequently LIS training, whether initial education or CPD, needs an awareness of the contexts of LIS work outside of the traditional library setting and how different settings change the rules.

References

Currier, S., Barton, J., O'Beirne, R. & Ryan, B. (2004), 'Quality Assurance for Digital Learning Object Repositories: Issues for the Metadata Creation Process', *Alt-J*, vol. 12, no. 1, pp. 5-20.

- El-Sherbini, M. & Klim, G. (2004), 'Metadata and cataloging practices', *The Electronic Library*, vol. 22, no. 3, pp. 238-248. Available from: <http://www.emeraldinsight.com/rpsv/~1091/v22n3/s4/p238> . Last accessed 16th Feb 2005.
- Foster, N.F. & Gibbons, S. (2005), 'Understanding Faculty to Improve Content Recruitment for Institutional Repositories', *d-Lib Magazine*, vol. 11, no. 1, Available from: <http://www.dlib.org/dlib/january05/foster/01foster.html>. Last accessed 15th February 2005.
- Gorman, M. (2004), *The concise AACR2: based on AACR2 2002 revision, 2004 update*. 4th ed. London: Chartered Institute of Library and Information Professionals.
- Greenberg, J. & Robertson, D.W. (2002), 'Semantic Web Construction: An Inquiry of Author's Views on Collaborative Metadata Generation', *Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002*, Firenze University Press, pp. 45-52. Available from: <http://www.bncf.net/dc2002/program/ft/paper5.pdf>. Last accessed 21st December 2004.
- Harnad, S., Carr, L., Brody, T. & Oppenheim, C. (2003), 'Mandated online RAE CVs Linked to University Eprint Archives', *Ariadne*, no. 35. Available from: <http://www.ariadne.ac.uk/issue35/harnad/intro.htm>. Last accessed 15th February 2005.
- IEEE. (2002), *Final 1484.12.1-2002 LOM Draft Standard*, Available from: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf. Last accessed 1st September 2005.
- National Information Standards Organization. (2004). *Understanding Metadata*. NISO Press. Available from: <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>. Last accessed 21st December 2004.
- Secker, J. & Price, G. (2004), 'Developing the e-literacy of academics: case studies from LSE and the Institute of Education, University of London', *Journal of eLiteracy*, vol. 1, pp. 97-108 Available from: http://www.jelit.org/archive/00000038/01/JeLit_18.pdf. Last accessed 16th Feb 2005.
- Spinazzé, A. (2004), 'Museums and Metadata: A Shifting Paradigm', in D.I. Hillmann & E.L. Westbrooks (eds), *Metadata in Practice*, American Library Association, Chicago, pp. 37-50.