

Quality assurance for digital learning object repositories: issues for the metadata creation process

Sarah Currier^{1*}, Jane Barton¹, Rónán O’Beirne² & Ben Ryan³

¹University of Strathclyde, UK; ²City of Bradford Libraries, UK; ³University of Huddersfield, UK

Metadata enables users to find the resources they require, therefore it is an important component of any digital learning object repository. Much work has already been done within the learning technology community to assure metadata quality, focused on the development of metadata standards, specifications and vocabularies and their implementation within repositories. The metadata creation process has thus far been largely overlooked. There has been an assumption that metadata creation will be straightforward and that where machines cannot generate metadata effectively, authors of learning materials will be the most appropriate metadata creators. However, repositories are reporting difficulties in obtaining good quality metadata from their contributors, and it is becoming apparent that the issue of metadata creation warrants attention. This paper surveys the growing body of evidence, including three UK-based case studies, scopes the issues surrounding human-generated metadata creation and identifies questions for further investigation. Collaborative creation of metadata by resource authors and metadata specialists, and the design of tools and processes, are emerging as key areas for deeper research. Research is also needed into how end users will search learning object repositories.

Introduction

The emergence of the concept of reusable learning objects has been a major recent development in e-learning (Littlejohn, 2003). Much discussion and exploratory work has been undertaken, moving us towards what has been called “the learning object economy” (Downes, 2001; Campbell, 2003), where teachers, course developers and learners can share, reuse and re-purpose digital materials for incorporation into teaching and learning. Some potential benefits of this ‘economy’ include: minimising duplication of effort for individual teachers across subject areas; reducing costs for

*Corresponding author. Centre for Academic Practice, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1QE, UK. Email: sarah.currier@strath.ac.uk

institutions (Duncan, 2003b); and providing access to a wider variety of learning materials. In the past few years, various institutions and projects have been developing repositories for these reusable learning objects (Downes, 2003) supported by international standardization work, notably the suite of specifications produced by the IMS Global Learning Consortium (IMS). Downes (2003) suggests that the next stage of development in this “economy of education” should be the development of a network of distributed learning object repositories.

Because metadata enables users to discover and select digital learning resources suitable to their requirements, it is a vital component of the learning object economy, any future distributed networks and the learning object repositories within them. Extensive groundwork has been carried out in this area, mainly centred upon the development of the IEEE Learning Object Metadata standard, known as ‘the LOM’ (IEEE LTSC, 2002). IEEE worked closely with the interoperability body IMS in creating the LOM; hence it is integral to such learning technology specifications as IMS Content Packaging and IMS Digital Repositories Interoperability. The UK has played a central role worldwide in the ongoing development of good practice, common usage and appropriate vocabularies for the LOM, including Graham and Campbell’s (2003) ‘UK LOM Core’ (originally known as the UK Common Metadata Framework).

So, given the existence of this work, why is there a need for further quality assurance? The key to answering this question involves distinguishing between the concepts of structure and content. The developments above deal primarily with the structure of the metadata; this paper is concerned with the creation of the content of the metadata fields. Once a metadata standard has been implemented within a system, the specified fields must be filled out with real data about real resources; this process brings its own problems. For searchers, these manifest themselves in various ways, including poor recall of available resources and inconsistency of search results. They arise due to errors, omissions and ambiguities in the metadata, many of which are known and understood in other communities of practice with tried and tested solutions.

Within e-learning the problems of metadata creation have yet to be fully addressed. When this paper was first drafted for *ALT-C* in February 2003, almost no formal research had been carried out into the process of filling in metadata fields describing learning objects. However, informal consultation via e-learning metadata forums revealed a significant number of colleagues who shared our concerns. All agreed, usually from personal experience, that the issue of who creates metadata and how has an important impact on the quality of collections of digital materials for resource discovery by end users.

The scope of this paper

This paper surveys the issue of metadata creation for digital learning object repositories with an emphasis on quality assurance, presenting three cases of repositories whose experiences have raised issues for debate and further investigation. We have

limited our scope to the realm of “human metadata generation” (Greenberg & Robertson, 2002), wherein a “person intellectually manages the metadata generation”. The issue of machine-generated metadata continues to be the subject of extensive research within such disciplines as computer and information sciences, data mining and artificial intelligence, and there is much to be learned there for learning object repositories. However, we are not yet in a situation where machines can handle all metadata tasks, particularly in an area where many resources have limited textual content. Those tasks that require human intelligence and creativity can include such areas as subject classification, educational attributes and determining the contributors to a resource. This is the domain with which we are concerned.

We are also only investigating the creation of metadata necessary for resource discovery via searching and browsing, although such metadata can also be used for resource selection. There has been much recent discussion on what has been called secondary or conceptual metadata (McLean & Lynch, 2003), usage data or “third party metadata” (Downes, 2003). This refers primarily (but not exclusively) to metadata created about the use of a resource for teaching and learning, generally encompassing the idea of reviews or comments by users and intended to facilitate selection of appropriate resources. This issue is in the very early stages of investigation and the distinction may not be as clear-cut as we have stated here, but for clarity’s sake we have excluded it from this paper’s scope. However, there will no doubt be implications for quality assurance.

It is worth noting that, although we focus on the creation of metadata necessary for resource discovery, none of the evidence we found in the e-learning domain included research into the ways in which users actually carry out resource discovery. This is a significant gap, perhaps arising from the paucity of working, well-populated repositories; however, other disciplines, such as library and information science, may give preliminary pointers on these issues. This area of research is vital for the formation of priorities and policies for metadata creation. We will return to this in our final section outlining future research questions.

Metadata is powerful

Although many learning resources are available on the web, searching the whole web using a search engine such as Google can prove unsatisfactory. Even with localised or advanced Google-type searching within e-learning, learning objects come in a variety of formats; those which are images (including PDF text files), animations or simulations may have very limited textual content to search. Browsing a directory of web resources which have been selected on the basis of some criteria, typically subject, can also be time-consuming, particularly when the sought-after material exists only as a small chunk embedded within a larger resource. One purpose of digital repositories is to overcome these problems by collecting good quality resources, preferably in small chunks (Duncan, 2003a), together with detailed, consistent information about them, thereby enabling users to conduct precisely targeted searches and to retrieve relevant materials in an efficient and effective manner.

This detailed information about resources, or metadata, is therefore key to unlocking their potential for reuse. At its best, “accurate, consistent, sufficient, and thus reliable” (Greenberg & Robertson, 2002) metadata is a powerful tool that enables the user to discover and retrieve relevant materials quickly and easily and to assess whether they may be suitable for reuse. At worst, poor quality metadata can mean that a resource is essentially invisible within the repository and remains unused.

Metadata within e-learning

The development of interoperability standards and specifications within e-learning has involved, in the main, software and courseware developers, content developers and, to a lesser extent, teachers. Information scientists and librarians, whose expertise lies precisely in the domain we are examining, were simultaneously developing metadata technologies (such as the Z39.50 search protocol), standards (such as MARC and Dublin Core metadata) and practice for web-based and other digital resources. For a long time these two fields remained largely separate (McLean & Lynch, 2003) and opportunities to benefit from the experiences of the library and information science community were often missed.

Consequently, the metadata creation problem space has been elided within e-learning. Downes (2001) stated in his seminal paper on the necessity for a learning object economy:

Whatever the properties, the authoring of metadata itself will be straightforward for most course designers. Because metadata files are machine-writable, authors will simply access a form into which they enter the appropriate metadata information.

This statement encapsulates the (lack of) thinking in this area. IMS and IEEE, in their metadata specifications, have remained agnostic on the matter, offering no guidance on how good quality metadata creation may be ensured (IEEE LTSC, 2002; IMS, 2001; IMS, 2003).

We suggest that there are four erroneous assumptions behind the absence of inquiry into how metadata should best be created within e-learning:

- that, in the context of the culture of the Internet, mediation by controlling authorities is detrimental and undesirable;
- that rigorous metadata creation is too time-consuming and costly, a barrier in an arena where the supposed benefits include savings in time, effort and cost;
- that only authors and/or users of learning materials have the necessary knowledge or expertise to create metadata that will be meaningful to their colleagues; and
- that, given a standard metadata structure, metadata content can be generated or resolved by machine.

We would also put forward a fifth underlying reason, garnered from conversations with e-learning colleagues around the world: that for both technology and pedagogy experts, metadata creation is seen as a tedious chore rather than as a complex intellectual skill which is essential for unlocking access to resources.

However, standards-based learning object repositories are now being more widely implemented and practical problems resulting from poor understanding of the meta-data creation process are beginning to emerge. These experiences challenge the above assumptions and suggest that there is more to the creation of good metadata than simply filling in a form.

Three examples from the UK

We now summarise relevant findings from three UK repositories. They are presented in chronological order, to illustrate the development of understanding in the area of metadata creation.

The Scottish electronic Staff Development Library (SeSDL) taxonomy evaluation

SeSDL was a seminal project that, from 2000 to 2001, investigated the development of a learning object repository based on IMS specifications, including the IMS Learning Resource Meta-data specification (v1.1). Funded by SHEFC's ScotCIT programme and based at the Universities of Strathclyde, Edinburgh and Paisley, its purpose was to encourage the sharing and reuse of staff development materials within HE. The main subject focus was the use of C&IT in teaching and learning.

In planning this early repository, employing an information specialist was not considered. However, when it was discovered that no appropriate, readily available subject classification scheme was available, a librarian was brought in to create the SeSDL Taxonomy. A small-scale peer evaluation of the taxonomy was carried out (Currier, 2001). This evaluation was not designed to test the proficiency of resource authors and users in creating metadata, although it did point to potential problems in the specific area of subject metadata. The data gathered was complex and could no doubt yield more insight with further analysis; here we merely attempt to illustrate in a simple way the difficulties untrained users found in subject classification.

Six consultants drawn from the project's user community were provided with eight learning objects to be classified using the taxonomy. The SeSDL team agreed upon 'ideal' classifications for the eight objects, against which the consultants' classifications would be compared. The team provided as much structure and guidance for the evaluation exercise as possible, while not providing the consultants with training which would bring them too far beyond the skill level of the intended users of SeSDL. However, even with guidance notes, the ability of the consultants to understand and carry out the task varied considerably. One consultant commented in the post-evaluation focus group: "The whole exercise has given me more admiration and respect for librarians" (Currier, 2001).

The SeSDL team assigned a total of 35 classifications to the eight objects, averaging about four classifications per object. In only one instance did all six consultants agree with *one* of the 'ideal' classifications. For five of the eight objects, up to half of

the ‘ideal’ classifications assigned were unused by any of the consultants. In all, only about 50% of the ‘ideal’ classifications had the agreement of more than half the consultants.

A total of 71 ‘non-ideal’ classifications were assigned by the consultants, averaging about nine per object. In only 15% of these cases did more than one consultant agree upon the classification. Only five classifications (7%) had the agreement of three or more consultants. Out of a total of 106 classifications assigned (including ‘ideal’ classifications), only 39 (35%) had the agreement of more than one consultant.

These figures indicate that users of SeSDL will assign a wide variety of classifications to their objects and will do so inconsistently in comparison with each other. For example (Currier, 2001), a learning object consisting of an HTML page defining the terms ‘VLE’ and ‘MLE’ was classified by one consultant as ‘Student-Centred Learning’ and ‘Collaborative Learning’. This appears to reflect their belief that a VLE or MLE should be used in a student-centred way, for collaborative learning. The implication is that a repository user looking under ‘Student-Centred Learning’ in the browse tree would expect to find a learning object defining the term ‘VLE’ there. Tables 1 and 2 show the variety of classifications assigned by all the consultants to this object.

If the learning objects listed under a particular branch of the SeSDL browse tree appear to be randomly or inconsistently classified, this may well influence users’ perception of the quality of the repository as a whole and their willingness to keep searching.

The Evaluation Report (Currier, 2001) concluded with a number of recommendations, the most pertinent of which relate to user support:

- Explain what classification is for using simple, jargon-free language and examples.
- Ensure users understand the availability of multiple classifications, with examples.
- Suggest use of both the upload tool and a paper version of the taxonomy.
- Provide a note-taking facility, or suggest that users take notes offline.
- Suggest users look for other objects of a similar subject to theirs and note the classifications that have been assigned.
- Provide a tutorial in classifying objects, designed to lessen the main barriers to effective classification as highlighted in the evaluation.
- Make alternative terms visible within the upload tool, to assist with understanding the scope of the classifications.
- Provide scope notes for classifications where appropriate.
- Add more ‘See Also’ notes and allow these to be seen within the upload tool. Either include links or automatically bring up the terms referred to.

SeSDL was a project with a finite lifespan; no ongoing funding was available to implement any of these recommendations. However, its experiences have informed subsequent developments worldwide. Perhaps the most pertinent to this enquiry is the final point made in the evaluation (Currier, 2001):

How can online resource provision services which expect their users to classify their own resources best support this so that future users will be able to find what they want? Or is this approach ultimately inadvisable?

Table 1. Number of consultants choosing 'ideal' classifications

'Ideal' classifications	No. of consultants who chose the classification (out of 6)
2.2. Educational technology/virtual learning environments	4
2.2.3 Educational technology/virtual learning environments/managed learning environments	4
1.5.1.1 Educational development/educational environments/electronic classrooms/virtual learning environments	3
1.5.1.1.1 Educational development/educational environments/electronic classrooms/virtual learning environments/managed learning environments	3

Table 2. Overlap between classifications chosen by consultants

Classifications chosen by consultants	No. of other consultants choosing this classification (out of 6)
1.3.4 Educational development/approaches to teaching/student centred learning	0
1.5.1.2 Educational development/educational environments/electronic classrooms/web-based teaching	1
2. Educational technology	0
2.9 Educational technology/Internet	0
2.16.2 Educational technology/software packages/virtual learning environments (intended for resources about the use of specific packages, e.g. Blackboard or WebCT)	0

The Bolton Woods Local History Project

Bolton Woods Community Centre is a UK-Online centre offering C&IT facilities to the local community. Since 1998 it has been a part of community networking project, Shipley Communities Online, which is a partnership offering online learning, occupational guidance services and information and advice on training and work opportunities.

In the Centre's Bolton Woods Local History Project, members of the community create digital resources, mainly family and local history materials, which are shared with their peers for use as informal learning resources. Under the Metadata for Community Content project, a network of community projects, including Shipley Communities Online, were provided with a small repository so learning materials could be shared on a peer-to-peer basis with other communities. Metadata for the resources was required to facilitate this sharing.

Faced with the challenge of creating metadata for a range of materials, the project decided to use the available labour. It became apparent early on that the process of metadata development was one that merited closer attention, so a small study was carried out to investigate whether the creators of resources could also create their own metadata and to compare their ability with information specialists.

Dublin Core metadata was used, with some additional educational elements added for resources used to construct learning pathways (e.g. ‘AudienceLevel’ and ‘Typical-LearningTime’). This proved to be particularly problematic as neither resource creators nor librarians had the pedagogical expertise either to create learning pathways or assign educational metadata.

Four local history enthusiasts with fairly high-level website design skills created community resources for the project and were given the task of creating metadata for their resources. Two qualified librarians working on the project were asked to assign metadata to similar resources. Brief guidance notes were provided, although a metadata tool was not used. Instead, DreamWeaver was employed, as this was familiar to the resource authors taking part.

The study involved another librarian involved in the project observing the efforts of both groups in creating metadata and assigning a subjective score out of five in five key areas of managing metadata: understanding metadata; context of resources; choosing elements; assigning values; and subject classification. The findings were:

- resource creators did not have a good understanding of the purpose of metadata or an appreciation of its value;
- resource creators did understand the context of their resources and focused on these elements within the metadata;
- information specialists had a better understanding of the purpose of metadata and included a wider range of metadata elements;
- information specialists struggled with contextual aspects of the metadata;
- neither the resource creators nor the information specialists handled pedagogic aspects of the resources well.

Table 3 shows the scores gained by the two groups.

The greatest difficulty arose from content creators’ lack of understanding of the rationale for assigning metadata (cf. case three, below). The metaphor of finding a book in a library was useful in explaining its purpose.

Table 3. Comparative assessment of success in creating metadata

Activity	Information specialists Score out of 5	Content creators Score out of 5
Understanding	3	1
Context of resources	1	4
Choosing elements	4	1
Assigning values	4	4
Classification	4	2

Specific issues included the content creators' difficulty with the 'Relation' element, which allows, for instance, a resource to be specified as part of another resource. The 'Rights' element also highlighted problems in understanding IPR issues. Subject classification was particularly difficult for content creators to understand, echoing the findings of the SeSDL case study above.

These findings suggest that a collaborative approach may yield the best results in terms of metadata quality, since it would engage the strengths of both groups. This small study resulted in an improved approach for the project, involving closer collaboration between content creators and metadata specialists. Because neither the content authors nor the librarians were educational professionals, it was noted that further improvement might be facilitated through the involvement of this third group, and that a future study investigating this may be useful.

The Higher Level Skills for Industry repository (HLSI)

The HLSI project is developing a repository for digital learning objects to support the delivery of learning programmes over a wide curriculum area from GCSE to Higher Education. An ongoing development, the project has both fed into and drawn from the development of this paper over the past year. As such, the issues raised and measures taken within HLSI to overcome problems are of great interest and represent a significant potential base for future research.

Based at the University of Huddersfield and funded by local development agency Yorkshire Forward, the project involves 35 partner organisations, with over 300 members of staff actively participating. Learning objects are uploaded to the repository by their authors (generally educational practitioners); they are intended to be shared and reused in e-learning environments across the partnership.

By February 2003, the repository had gathered approximately 6500 learning objects from the initial 12 partner organizations, in a variety of sizes and file formats, together with author-generated IEEE LOM v.1.0 metadata records:

The people who submitted resources also provide the metadata, which gives them some ownership over the records. The drawback is that the quality of metadata varies. (Barker & Ryan, 2003)

Clearly, in the early stages of the project, there was an assumption that authors who submit resources want 'ownership' of the metadata records. This is interesting in light of our initial assumptions and may warrant further investigation, although there was a shift around this issue later in the project.

The problem of metadata quality is explained further (Barker & Ryan, 2003):

The difficulty with this process is making sure the authors understand the purpose of the metadata and the methodology used to enter it. A balance had to be struck between getting high quality metadata and not going above the skill level of those entering the metadata.

At the moment there is quite a large variation in the quality of metadata for the resources. For example some have spelling errors. This affects the performance [of the] repository so several steps are being taken to improve the process.

Table 4. Metadata quality in the HLSI project prior to intervention

Metadata quality	% of metadata records
Good	28
Moderate	26
Poor	32
Unusable	14

As shown in Table 4, an evaluation of the metadata records at that time showed that nearly half (46%) of the metadata records were of poor quality or unusable

Specific problems beyond spelling errors were specifically delineated by Ryan & Walmsley (2003):

- A single metadata record was duplicated, unchanged, for many or all components of a package of educational content.
- The terminology used by the metadata authors was not consistent.
- Some metadata authors described the facets and characteristics of the educational object and not its content, e.g. describing a Flash file about internal combustion as 'Flash file' instead of 'internal combustion'.
- The metadata tool allowed default values for certain fields and these were used inappropriately.

Steps for improving the process have now been under way for some months. Initial measures involved providing more user support through education and documentation and employing a team of information science professionals to improve the existing metadata (Ryan, 2003). Ryan (2003) noted that, by June 2003, 2500 metadata records had been re-edited, taking about 550 hours and costing around £6500 (about £2.60 per record). Subsequently, the partnership expanded and the project now has access to a large number of information science professionals who have adopted the metadata problem and are actively driving improvements forward.

The process of metadata collection has now been split into two stages:

1. The educational practitioner is responsible for entering basic metadata, including title, description, contribution and any technical information they may be aware of.
2. The information scientist is responsible for reviewing the basic metadata and providing additional metadata for subject classification, educational attributes, etc.

This process was created by a group of partners in one of the sub-regions covered by the project; it is now being adopted and actively promoted to the whole partnership, supported by continuous staff development and training.

The information scientists also made a number of comments, suggestions and recommendations that have resulted in new development areas for the project. These include providing:

- spell-checking facilities within the metadata tool;
- functionality for browsing and searching authority lists (centralized coordination of forms of authors' names, for instance) when reviewing and entering advanced metadata;
- a clear separation between basic and advanced metadata, with restricted access to the advanced metadata;
- an on-line thesaurus to support advanced metadata entry;
- updated terminology and option lists to reflect current practice; and
- functionality to record/report/track the review status of metadata records.

These new development areas represent a significant amount of work and reflect the project's change in emphasis from a purely technical implementation of specifications and standards to a more focused approach on addressing a significant problem. The creation of the two-stage process also led to a change in the conception of ownership of metadata records and the responsibility and authority for their quality, without removing the need for the resource author's own expertise.

Specific metadata issues

If the proposed learning object economy bears fruit, we may anticipate repositories and networks holding large collections of tens of thousands of learning objects or more. The above case studies highlight a number of areas where quality of the metadata may impact on the discovery of resources in this economy, which are expanded on below.

Error management

The HLSI case study, with its large numbers of records, found that the issue of errors was significant in their repository. The following quote illustrates (amusingly) a potentially serious problem facing resource discovery. It touches on motivation and support for metadata creation by untrained resource authors, and on the necessity for checking of metadata, whoever it is created by:

Even when there's a positive benefit to creating good metadata, people steadfastly refuse to exercise care and diligence in their metadata creation. Take eBay: every seller there has a damned good reason for double-checking their listings for typos and misspellings. Try searching for "plam" on eBay. Right now, that turns up nine typoed listings for "Plam Pilots". Misspelled listings don't show up in correctly spelled searches and hence garner fewer bids and lower sale-prices. You can almost always get a bargain on a Plam Pilot at eBay. (Doctorow, 2002)

Authors' and other contributors' names: authority control

An obvious example among a multitude of permutations is that of an author's name changing when they get married. In this and similar cases, a search by name cannot retrieve everything by that person unless there is some kind of name management.

Libraries, archives and museums solve this problem by using centralised name authority records, a time consuming and costly exercise. This also applies to names of institutions, conferences, etc.

However, bibliographic indexing services have traditionally not done this, instead relying on the author's name and subject area to do a 'good enough' job. There is no evidence either way for a viable approach for learning object repositories. This is one area where research into how users will search would be useful.

Subject area

This is one of the most complex areas of both metadata creation and resource discovery; there is insufficient space to cover it in depth here. All three case studies showed significant problems when untrained authors attempted to create subject metadata.

There are two main ways in which subject access to resources is provided via metadata (as opposed to free text searching): indexing (e.g. key words) and classification. How may this difficult and complex task best be carried out for maximum resource discoverability by a heterogeneous population of searchers? Should the resource author, who may know their subject area and its terminology well, create the subject metadata? Or should it be a metadata specialist, who may know the specific area less well, but may be better placed to step back and think about all the potential users of a resource, and about consistency of key words and classifications across a repository or network?

Educational metadata

It is commonly thought within e-learning that authors are best placed to create educational metadata. The Bolton Woods case study suggests that those with educational expertise should be involved in this area, where authors themselves do not have such expertise. Nevertheless, there are many successful examples of professional cataloguing of specialised materials, such as music or photographs, so it is possible that a new sub-speciality of metadata experts could emerge in this area.

Accessibility metadata

With the new SENDA (*Special Educational Needs and Disability Act, 2001*) legislation in the UK there has been some interesting recent work around developing metadata to describe the accessibility properties of a resource. However, this may prove problematic for metadata creators who are not experts in accessibility.

Who creates metadata and how?

There is much discussion in e-learning concerning the barriers to uptake of reusable learning objects, often focusing on teachers' unwillingness to engage in the vaunted learning object economy. IPR and the 'not invented here syndrome' are often quoted

as significant hurdles. However, how much of a barrier is the task of having to create metadata when uploading a resource, particularly with poorly designed tools? Moreover, how much of a barrier will it be for teachers searching for resources, if metadata is of such poor quality that they cannot find what they need?

In this section we outline some possible permutations of how metadata may be created and by whom. The task of creating metadata may be divided into two stages: the gathering and recording of the information and the expression of that information as conformant metadata. For example, in recording the names of contributors to a resource, a metadata creator may note illustrators, the authors of any text, any institutions that took part, and perhaps an editor. They may then check the cataloguing guidelines of their repository and enter this data in a way that conforms to the guidelines. For an experienced metadata creator, these stages may happen simultaneously, but it is an important distinction in deciding how metadata tasks may be allocated.

We suggest the following three models for creating metadata: resource author or contributor only; metadata specialist only; and collaborative.

In the first case, the design of metadata tools and user support and training take on a greater weight in terms of quality assurance. Metadata quality in all five of the above-named specific metadata issues may be impacted by inadequate provision here.

In the second case, the trained metadata specialist carrying out the task may be hampered by lack of knowledge about the pedagogical context, history or subject area of the resource, as shown in the Bolton Woods case study.

Within a collaborative model there are a number of possible scenarios. For instance, the author may enter data in certain fields (as in the HLSI case), such as their own name, resource title, institution, educational fields, etc. The metadata specialist may check these for accuracy and conformance and add other selected fields such as subject classification, keywords and accessibility information. This process may be truly collaborative, with the parties communicating directly, or they may work separately, perhaps with the specialist periodically checking records in bulk. In both the Bolton Woods case and the HLSI case, a collaborative method of metadata creation was chosen after practical experience of the difficulties in taking either the first or second approach. Both repositories reported improvements as a result.

There are various communities that have a body of research and experience to draw upon in examining some of these issues. The most obvious is library and information science, with an abundance of relevant peer reviewed journals and conferences. The archive and museum communities may also have something to contribute; for instance, the metadata in museum object records is considered to contain a large portion of a curator's academic knowledge and research (Zorich, 1991). Commercial abstracting and indexing database services have long utilised author-generated metadata, requiring authors to submit abstracts and keywords with papers. Research in this field has found that authors "may lack knowledge of indexing and cataloguing principles and practices, and are more likely to generate insufficient and poor quality metadata that may hamper resource discovery" (Greenberg & Robertson, 2002).

There has been one formal information science study so far on the issue of author generated and collaborative metadata (Greenberg & Robertson, 2002), which looked

at this with regard to Dublin Core metadata in support of the Semantic Web, a development which aims to bring structured knowledge representation to the web's meaningful content. This study concluded that:

... the integration of expert and author generated descriptive metadata can advance and improve the quality of metadata for web content, which in turn could provide useful data for intelligent web agents, ultimately supporting the development of the Semantic Web. [...] If such partnerships are well planned and evaluated, they could make a significant contribution to achieving the Semantic Web.

Conclusions: issues for research

Analysis of the studies above enables several areas to be identified where focused investigation would produce useful information for decision making by developers and managers of repositories:

- What are the important cultural factors that may influence the e-learning community's particular approach to metadata creation? For example, is 'ownership' of metadata by resource authors important? If so, how may this need best be met? In the HLSI case, there was a tradeoff between perceived ownership of metadata by resource authors and the quality of the completed metadata. Further research as this repository progresses may shed more light on optimum management of this balancing act.
- What constitutes good quality metadata, within individual repositories and within the global networked environment? For example, to what extent does metadata that is 'good enough' for local purposes also support effective retrieval by remote users operating in a different contextual setting? Can a set of 'metadata metrics' be agreed within communities and beyond?
- Who is best placed to create the metadata in any given context? For example, to what extent does the type of metadata (subject metadata, educational metadata, etc.) have a bearing? Is a collaborative approach to metadata creation the best way forward? Since the evidence presented suggest that this is the case, how can this approach be managed effectively?
- How can tools be used to facilitate the metadata creation process and how much effect do they have? The HLSI repository improved their tool design as part of a programme of improving metadata quality, while the SeSDL Taxonomy Evaluation stated that tool design will be vital if resource depositors are to create their own metadata. So how can the design of tools encourage the creation of good quality metadata, whoever is creating it?
- To what extent can the provision of guidelines and training improve metadata creation? For example, can information specialists provide adequate guidelines to enable non-specialists to use a taxonomy effectively? Can librarians be trained to create good quality educational metadata?
- What are the costs and benefits associated with the various approaches to metadata creation? For example, are savings at the initial metadata creation stage eroded by subsequent costs such as data cleaning? The HLSI case would suggest this is the

case. In addition, does reducing metadata costs within a repository simply increase the cost, in terms of time and effort, to the end user?

- How will users search for materials within learning object repositories and networks? For example, how important is it to have authority control over the names of authors and contributing institutions? What educational attributes will users search for and how? Answers to these questions will have a profound impact on decisions about metadata creation.

The evidence presented here suggests that a collaborative approach to metadata creation may be necessary, and that good design of processes and tools is important. However, further research needs to be done on specific implementation of these approaches. Other issues have been raised, with no clear answers; particularly important is how end users will search repositories. What is clear is that there is work to be done before the e-learning community has a good understanding of the issues surrounding metadata creation, such that effective policies and practices can be put in place to assure the quality of metadata and hence the ability of teachers and learners to access resources.

Acknowledgements

The authors would like to thank the following colleagues for their contributions to this paper: Phil Barker; Lorna M. Campbell; Jackie Carter; Charles Duncan; Gordon Dunsire; Jane Greenberg; Jessie Hey; Neil McLean; David Nicol; William Nixon; Andy Powell; Pauline Simpson; Stuart Sutton; and Steve Walmsley.

This paper is based on a research paper presented at *ALT-C 2003* and published in the conference's research proceedings.

References

- Barker, E. & Ryan, B. (2003) The Higher Level Skills for Industry Repository, in: P. Barker & E. Barker (Eds) *Case studies in implementing educational metadata standards* (CETIS). Available online: http://metadata.cetis.ac.uk/guides/usage_survey/cs_hlsi.pdf (accessed 4 November 2003).
- Campbell, L. (2003) Engaging with the learning object economy, in: A. Littlejohn (Ed.) (2003) *Reusing online resources: a sustainable approach to e-learning* (London, Kogan Page).
- Currier, S. (2001) *SeSDL taxonomy evaluation report* (Glasgow, University of Strathclyde). Available online: http://www.sesdl.scotcit.ac.uk:8082/taxon_eval/SeSDLTaxFinRep.doc (accessed 4 November 2003).
- Doctorow, C. (2002) Metacrap: putting the torch to seven straw men of the meta-utopia, *E-learning guru newsletter*. Available online: <http://www.e-learningguru.com/articles/metacrap.htm> (accessed 4 November 2003).
- Downes, S. (2001) Learning objects: resources for distance education worldwide, *International review of research in open and distance learning*. Available online: <http://www.irrodl.org/content/v2.1/downes.html> (accessed 4 November 2003).
- Downes, S. (2003) Design and reusability of learning objects in an academic context: a new economy of education? *USDLA Journal* 17(1). Available online: http://www.usdla.org/html/journal/JAN03_Issue/article01.html (accessed 4 November 2003).

- Duncan, C. (2003a) Granularization, in: A. Littlejohn (Ed.) *Reusing online resources: a sustainable approach to e-learning* (London, Kogan Page).
- Duncan, C. (2003b) *The value of managing learning objects: an Intrallect 'white paper'* (Livingston, Intrallect Ltd.). Available online: http://www.intrallect.com/products/intralibrary/loms_value.pdf (accessed 4 November 2003).
- Graham, G. & Campbell, L. (2003) *UK Learning Object Metadata Core Draft 0.1, July 2003*. Available online: <http://www.cetis.ac.uk/profiles/uklomcore> (accessed 4 November 2003).
- Greenberg, J. & Robertson, W. (2003) Semantic web construction: an inquiry of authors' views on collaborative metadata generation, *Proceedings of the International Conference on Dublin Core and Metadata for e-Communities 2002*, 45–52. Available online: <http://www.bncf.net/dc2002/program/ft/paper5.pdf> (accessed 4 November 2003).
- IEEE LTSC (2002) *Draft standard for learning object metadata* (New York, IEEE LTSC). Available online: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf (accessed 4 November 2003).
- IMS Global Learning Consortium, Inc. (2001) *IMS Learning Resource Meta-Data best practice and implementation guide, Version 1.2.1 final specification*. Available online: http://imglobal.org/metadata/imsmdv1p2p1/imsmd_bestv1p2p1.html (accessed 4 November 2003).
- IMS Global Learning Consortium, Inc. (2003) *IMS Digital Repositories Interoperability—core functions best practice guide, Version 1.0 final specification*. Available online: http://imglobal.org/digitalrepositories/driv1p0/imsdri_bestv1p0.html (accessed 4 November 2003).
- Littlejohn, A. (Ed.) (2003) *Reusing online resources: a sustainable approach to e-learning* (London, Kogan Page).
- McLean, N. & Lynch, C. (2003) *Interoperability between information and learning environments: bridging the gaps: a joint white paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information* (Draft version, 28 June 2003). Available online: <http://www.imglobal.org/TAI3/McLean.pdf> (Accessed 25 November 2003).
- Ryan, B. (2003) *Creating, using and re-using learning objects*, presentation, (Huddersfield, HLSI Project). Available online: <http://www.cetis.ac.uk/groups/20010809144711/FR20030807121739> (accessed 4 November 2003).
- Ryan, B. & Walmsley, S. (2003) Implementing metadata collection: a project's problems and solutions, *Learning technology*, 5(1). Available online: http://lttf.ieee.org/learn_tech/issues/january2003/index.html#3 (accessed 4 November 2003).
- Special Educational Needs and Disabilities Act (SENDA)* (2001) (London, HMSO). Available online: <http://www.hmso.gov.uk/acts/acts2001/20010010.htm> (accessed 25 November 2003).
- Zorich, D. (1991) Library and museum information: beauty and the beast, *Spectra*, 1991.