The User-Centered Development and Testing of a Dublin Core Metadata Tool

Catherine Hall, Michael Khoo
College of Information Science & Technology, Drexel University
3141 Chestnut Street, Philadelphia PA 19104, USA
+1 215 895 5912
{ceh48, khoo}@drexel.edu

ABSTRACT
Digital libraries are supported by good quality metadata, and thus by the use of good quality metadata tools. The design of metadata tools can be supported by following user-centered design processes. In this paper we discuss the application and evaluation of several cognitively-based rules, derived from the work of Donald Norman, to the design of a metadata tool for administering Dublin Core metadata. One overall finding was that while the use of the rules supported users in their immediate interactions with the tool interface, they provided less support for the more cognitively intensive tasks associated with developing a wider conceptual understanding of the purpose of metadata. The findings have implications for the wider development of tools to support metadata work in digital libraries and allied contexts.

Categories and Subject Descriptors
H.3.7 [Information Storage and retrieval]: Digital Libraries – collection, standards, user issues.

General Terms
Management, standardization

Keywords
Metadata tools, metadata generation, Dublin Core

1. INTRODUCTION
Digital libraries are supported by good quality metadata. The creation of good quality metadata depends partly in turn on the use of good quality metadata tools.

Several inter-related factors impact the development and use of metadata tools. One is the complexity of metadata concepts and the cataloging tasks to be handled by the tool; another is the expertise of the tool users themselves; and a third is what has been referred to as the metadata ‘gap’ between the rapidly increasing numbers of resources that require metadata, and the limited number of trained metadata creators. Combined, these factors mean that there are increasing groups of users who do not have formal professional training in metadata work, but who also have an interest in creating metadata, who will be using metadata tools. Studies of metadata tools have shown that they can be difficult for non-professionals to use [5]. The findings of these DL studies echo those of a meta-study of bricks-and-mortar library usability studies of Web sites, OPACs, etc. which found that common library terminologies – such as database, library catalog, e-journals, index, inter-library loan, periodical and serial, reference, resource, and subject categories in general – were often little understood by users, and acted as cognitive barriers to use. [10]

Given the complex contexts of metadata creation, metadata tool development benefits from being user-centered, and following iterative design processes, in which designs are improved through iterative cycles of implementation and testing [1]. This paper reports an example of such a development process, an evaluation phase of the ongoing user-centered design and development Erasmus, a metadata tool being developed for use with ipl2 (http://www.ipl.org; formerly the Internet Public Library).

2. USER-CENTERED METADATA TOOLS
Researchers have outlined a variety of requirements for metadata tools. An analysis of metadata records contributed by resource creators found that while the subject domain metadata was generally good, there were often technical errors, especially related to formatting [12]. Recommendations include gaining a better understanding of contributors’ thought processes, better guidelines, and improved interfaces, to “allow contributors to produce high-quality metadata records without having a great deal of organizational expertise.” Despite the provision of a cataloging tool, online work and discussion spaces, face-to-face workshops, conference calls, etc., the cataloging activity of a group of K-12 teachers helping to build a digital library was slow, with the educators having difficulty understanding concepts such as digital library, collection, catalog, and metadata [6, 7]. A study of author-generated records in the National Institute of Environmental Health Sciences found that some users lacked understanding of the tool and metadata, despite having access to a tutorial [4]. “Succinctly explaining the fields of a standardized schema such as Dublin Core to metadata novices appears to be a key challenge for designers,” and that “developing a conceptual understanding of metadata records and their use in retrieval was found to be challenging for users.” [2]

Recommendations that have emerged from this research include that “For decentralized metadata creation to become a reality, better designs are needed to reduce users’ cognitive load and lower barriers to efficient metadata entry” [1, c.f 10]. One approach to better design is to follow an iterative and user-centered approach to metadata tool development, by following the user-centered design approach of Donald Norman [11] who provides a suite of useful cognitive models that can be applied to...
Two inter-related concepts that are used in the current analysis are (a), knowledge in the head and knowledge in the world, and (b), designer images, user images, and system images [11].

Knowledge in the head is supported by processes such as short-term and long-term memory, and declarative and procedural memory. Short-term and declarative memory can be associated with explicitly engaging with an artifact, and also with learning. Long-term and procedural memory is often associated with skill, and with the application of expertise. In contrast to knowledge in the head, knowledge in the world is knowledge that is available to us in the environment. It can take the form of instructions, labels, affordances, and other perceptual cues that indicate how to use an artifact. We rely on knowledge in the world to use artifacts that are unfamiliar, complex, or which otherwise require knowledge apart from that in our heads.

A usable artifact strikes a balance between supporting these two different forms of knowledge, as well as between different forms of memory, and other cognitive processes. In the case of metadata tools, the knowledge in the head required to use the tool might consist of general knowledge of how to use an interface as well as any prior conceptual and technical understanding of metadata. Knowledge in the world, on the other hand, might include more specific information regarding schema requirements, controlled vocabularies, etc., that is required in order successfully to create good quality metadata.

In Norman’s approach, the design of such forms of ‘knowledge in the world’ often takes place in what is called the ‘system image.’ His model of designer image, user image, and system image, states that the interface for a system (such as a metadata tool) should support user interaction with that system by clearly representing the tool’s capabilities in ways that users will understand. The designer’s image is ‘the conceptualization that the designer has in mind,’ and the user’s image is ‘the model … the user develops to explain the operation of the system.’ The two images intersect in the system image, which represents the system components and their relationships to the user, in such a way that takes into account who the users are, and the type of tasks that they are trying to perform. When a system image accurately represents the underlying system model, then a system is easy to use, but when a system image is “incoherent or inappropriate … incomplete or contradictory” then usability problems can arise [c.f. 8].

Drawing on Norman’s approach, user-centered metadata tool design should attend to issues such as knowledge in the head and knowledge in the world, and also the system image. An important general research question for metadata tool design therefore concerns the ways in which specialist metadata knowledge is represented to non-professionals in tool interfaces: How can we represent complex metadata concepts in ways that are not only easy for novices to understand, but which might also anticipate user misunderstandings?

3. APPROACH

In the rest of this paper, we discuss how this research question is being addressed in the ongoing design of a metadata tool for ipl2. ipl2 contains approximately 40,000 resources and receives 8 million visits and 22 million page views a year. In 2009, ipl2 began crosswalking its metadata to Dublin Core and added this metadata to a Fedora Commons database. To support the creation of metadata records by ipl2 staff, volunteers and graduate students, a new web-based tool, *Erasmus* (fig.1), was developed. Initial requirements gathering was undertaken by students in a graduate ‘Introduction to HCI’ class and a prototype was developed which doctoral students evaluated. Many of the usability problems found at this stage were related to how the tool
represented the complex knowledge required for cataloging in an uncluttered and easily accessible form [9].

Following this round of expert usability testing and subsequent modifications, an assignment was designed for graduate students in an online Digital Libraries class to test the tool in a production-like environment. Students were provided with assignment instructions and a copy of the Erasmus “user manual”. Each student was assigned 5 ipl2 links that had been reported as broken. Students were instructed to fix or replace these links where possible and edit and create new metadata as appropriate. Importantly, the student-generated metadata was reviewed by ipl2 staff and judged to be of acceptable quality. However, the focus for this paper is an analysis of student reports in which they summarized their activities, reflected on the process of creating metadata, and evaluated the tool. Over two terms, 46 student reports were collected and anonymized, after which the authors iteratively reviewed the reports and selected sections in which students commented on Erasmus, and also on metadata work in general. These selections were then sorted using an inductive card sort, in order to identify a number of overall themes.

3.1 Results
Four overall themes were identified: overall usability, conceptual difficulties, student suggestions, and educational/career value.

3.1.1 Overall usability
Erasmus was well liked by students. Many positive comments focused on its simplicity and ease of use, interface design, and support for metadata tasks. Importantly, the system was judged to be “practical for a digital library that is maintained by people who are not necessarily all in the same area”. “The Erasmus tool has made this possible so that our inexperience does not jeopardize the integrity of the ipl2 library database”. The process of using the tool to create metadata was described as “stress-free” by one student and “relatively effortless” by another. One student remarked, “considering the actual process of trying to recover records can be complicated, it was reassuring to find Erasmus so easy to use”. Though this was the first experience with metadata creation for many students, others compared Erasmus favorably with data and content management tools they had used previously such as Archon, ContentDM, Solr and library cataloging tools.

The interface was described as “user friendly” and the metadata fields were judged to be “well structured” and “clearly defined”. Two features of Erasmus, in particular, received widespread praise. The first was the dual screen interface that presents the metadata form in the left pane and the resource (website) in the right. As one student noted, “it makes adding correct information much easier than flipping between tabs or windows”. The second feature was the incorporation of drop down ‘tool tips’ for each metadata field. “I appreciated that the Erasmus interface showed brief descriptions of the Dublin Core fields beneath the text entry boxes making it easier to remember the distinct purpose of each metadata element”, said one student. Tool tips also provided formatting instructions for elements as necessary. Although most students felt the provided descriptions and formatting examples were adequate, a few indicated they would have appreciated more guidance; “I felt that the explanation of what was meant to go in each field was obtuse in certain cases”. Several students mentioned that they would have liked to see more controlled vocabularies in use. As one stated, “I like to work from a set vocabulary because I feel more confident that the terms I am using to describe an object are the most semantically accurate”. A small number of technical issues were also reported by students such as browser incompatibility, error messages when saving a new record (even though the record did save), and difficulty loading one metadata record while another was open in a different tab. These issues were not persistent and did not seem to affect completion of the task. Finally, the user manual was very well received by students who found it a helpful go-to resource.

3.1.2 Conceptual difficulties
Most of the difficulties reported with the tool and the assignment were of a conceptual nature. These reported difficulties are consistent with the findings of other studies suggesting that this is a general issue with metadata tools. Observations included the following:

One of the things I struggled with is deciding which metadata is appropriate for a record. Including as much metadata as possible is ideal, however, sometimes the requested metadata is not available and plugging a best guess into a field will not add value to a record.

Throughout the assignment I had trouble deciding which Dublin Core fields were applicable and what to put in those fields.

I had trouble determining whether fields like date and temporal coverage were referring to the website or the history of the paper.

As noted earlier, despite these difficulties, all the metadata created was judged to be of a complete and accurate enough nature to be included in ipl2.

3.1.3 Student Suggestions
Students suggested a number of refinements and additional features they would like to see in future iterations of Erasmus. These included: enhanced search features and particularly the ability to retrieve a record by URL, spell checking tools, more embedded controlled vocabularies, and more examples within the tooltips to “elucidate fields like coverage and temporal coverage that are fairly close in terminology”. A small number of students explicitly asked for examples that helped not just with how to format metadata but where to find it in the source material. Other suggestions included “A section where you can review the work you have completed”, “a log of changes made by one’s account”, and a “live-consultation” or “issue, submission, triage tool”.

3.1.4 Educational/Career value
Students found the ‘hands-on’ process of metadata creation and editing to be very rewarding; “I think this was a good exercise for students in this class to do because it got us to get in the trenches and actually do some metadata updating and editing instead of just writing about metadata schemes as a more theoretical manner”. One student said they enjoyed the assignment as they “got a real feel for metadata entry and the decisions that are involved in creating it correctly” and another commented, “working on the metadata database backend helped me to further understand how metadata can play a role in improving search features in a digital library”.

Looking beyond the classroom, some students also commented on the benefit the exercise had for their future career, “I find the practical application very useful to my professional development and growing skill set as a librarian” said one. Another said, “I
believe this brief encounter with Dublin Core will help me in the future to understand more complex details of other digital collections when I move on from my coursework”

3.2 Discussion

Significant goals with the ongoing development work with Erasmus are both to develop a usable tool, and also to develop and report on a usability method that is appropriate for metadata tools. In this respect, the application of some of Norman’s design principles (knowledge in the head and knowledge in the world, and designer images, user images, and system images) was found to positively support tool development, and also to produce a tool that users generally found to be usable. Many students commented positively on the educational and professional value of the exercise, in particular their increased understanding of metadata concepts and metadata application in a digital library environment. Though students had been introduced to Dublin Core and other metadata schemes earlier in course, for most this was their first time “in the trenches”.

Within this overall result, there was a spectrum of outcomes and successes. More successful design strategies included those that were intended to find a balance between supporting easy use of Erasmus by not making it cluttered, and avoiding putting too much information in the interface so that it became unusable. For instance, the design affordance of the dropdown tooltips was an important one in this regard and between Norman’s ideas of knowledge in the head and knowledge in the world. Tooltips preserved the clean, uncluttered interface design that users found appealing, but also offered support for novice metadata creators. These design approaches also supported a relatively simple yet flexible system image that presented the underlying requirements of Dublin Core to the users in relatively simple ways. At the same time, conceptual difficulties with creating metadata remained a significant issue (although in many cases this was only a perceived difficulty; students created very good metadata but they were sometimes prone to doubt themselves). This latter finding confirms the findings of many of the previous studies of metadata concepts and metadata application in a digital library environment. Though students had been introduced to Dublin Core and other metadata schemes earlier in course, for most this was their first time “in the trenches”.

5. ACKNOWLEDGMENTS

Our thanks to the OCLC Library and Information Science Research Grant Program (LISGRP) for funding the initial development of Erasmus.

6. REFERENCES


