

Iterative Design of Metadata Creation Tools for Resource Authors

Jane Greenberg, Abe Crystal
University of North Carolina at Chapel Hill, USA
janeg@ils.unc.edu, acrystal@email.unc.edu

W. Davenport Robertson, Ellen Leadem
National Institute of Environmental Health Sciences, USA
{robert11,leadem}@niehs.nih.gov

Abstract

The National Institute of Environmental Health Sciences (NIEHS), as part of the Metadata Generation Research Project, is exploring resource authors' abilities for creating acceptable metadata. As part of this work, two different versions of the NIEHS Dublin Core-based metadata generation tool have been tested. An iterative design approach, supported by cognitive walkthroughs, guided the design of the NIEHS metadata application. This paper reports on the design process and on two author surveys that gathered feedback on the usability of both versions of the NIEHS metadata application (Version 1.0 and 2.0). The results show a slight improvement between Version 1.0 and Version 2.0. The paper concludes by summarizing key findings and identifying further research needs specific to author generated metadata tools.

Keywords: Metadata generation tools, metadata applications, Dublin Core, Author-generated metadata

1. Introduction

Resource authors (hereafter referred to as authors) and other metadata creators have access to an array of metadata generation tools. These tools include *templates* (simple forms), *editors* (forms enhanced with documentation and often automatic processes), and *generators* (applications that merely require the submission of a URL or other Web address for metadata generation) [1].

Editors, among the most popular of these tools, may provide metadata creators with access to a metadata schema or schemas, metadata element definitions, examples of how the metadata will appear when encoded in HTML or XML, and examples of metadata consistent with a number of different schemas (e.g., Dublin Core, Text Encoding Initiative (TEI)-header, or the Government Information Locator Service (GILS) schema). Editors may also include such features as drop-down menus or radio buttons for selecting values, access to controlled vocabularies or classification schemas, guidelines or specific examples for content syntax (e.g., surname, forename, or yyyy—mm—date [year-month-date]), and even tutorials defining metadata and its applications. The Nordic Metadata Project

(http://www.lub.lu.se/metadata/DC_creator.html) and The Gateway to Educational Materials Cataloging Module (http://geminfo.org/Workbench/Workbench_cataloging.html) are both metadata editors and employ many of the features identified here. These features are intended to facilitate the creation of high-quality metadata and, at the same time, make the metadata generation task easier and more efficient. Although this goal is emphasized in design decisions, little is known about authors' perceptions of these features or how they actually impact author-generated metadata. In fact, research investigating the usability of metadata generation tools appears to be limited, despite the importance of this activity.

If digital initiatives within organizations, or those of broader scope such as the Open Archives Initiative (<http://www.openarchives.org/>) and D-Space (<http://www.dspace.org>), are to engage authors in the metadata generation process, researchers need to study authors' use of metadata generation tools and solicit feedback on their usefulness and usability. The research presented in this paper examines this issue by surveying authors (specifically, scientists) at the National Institute of Environmental Health Sciences (NIEHS). These authors have engaged in metadata generation as part of NIEHS' metadata initiative and the Metadata Generation Research Project at the University of North Carolina (<http://ils.unc.edu/~janeg/mgr>).

In this paper, we report on the design process and on two author surveys that gathered feedback on the usability of both versions of the NIEHS metadata application. These data extend our previous analyses of author-generated metadata quality [9], collaborative metadata generation [5] and interface design for metadata creation [10] to provide a more detailed view of metadata creation from the resource author's perspective.

This paper begins with an overview of NIEHS' interest in author-generated metadata and a brief discussion of author-generated metadata tools. Research goals specific to the evaluation of the NIEHS metadata generation tool are presented, followed by an account of the design and testing of two versions of the NIEHS metadata application. The paper concludes by summarizing key findings and identifying further research needs specific to author-generated metadata tools.

2. Author-Generated Metadata at The National Institute of Environmental Health Sciences (NIEHS)

NIEHS is one of 27 Institutes and Centers of the National Institutes of Health (NIH), which is a component of the U.S. Department of Health and Human Services (DHHS). The mission of NIEHS is to reduce the burden of human illness and dysfunction from environmental causes. The NIEHS Strategic Plan states that one goal is to “enhance understanding of environmental health sciences and its importance to human health among scientists, policy makers, and the American public.” The Web is central to NIEHS’ plans for meeting this goal, and the integration of metadata is crucial to the success of the Institute’s Web strategy.

NIEHS maintains over 25,000 Web pages (<http://www.niehs.nih.gov/>), but very few of them contain metadata. Over the last few years NIEHS staff have periodically examined their Web search logs to ascertain how Web page visitors (both internal users and the general public) search the Web site. An individual visiting the NIEHS home page can search using the NIEHS site’s *full-text* search engine, but the limited amount of resource metadata and the current absence of a search engine that indexes metadata constrain the retrieval results. Often, a searcher retrieves hundreds of hits only mentioning the topic in passing, thus demonstrating that poor precision is a severe problem. In some cases, the primary Web page that NIEHS considers authoritative and desires for a searcher to retrieve first is buried so far down the list of retrieval results that is likely never viewed. Likewise, poor recall is also a problem—in some cases, the most useful page is not retrieved at all, because the user chose a search term that doesn’t appear in the title or text of the relevant Web page. These retrieval problems are in line with Weibel’s argument that, “resource discovery is the most pressing need that metadata can satisfy.” It is for this reason that the NIEHS Library initiated a project to enhance access to NIEHS Web resources [2]. The goal of the project was to create metadata that could improve discovery of the most appropriate environmental health information for both NIEHS staff and the public. NIEHS sought assistance from the School of Information and Library Science at the University of North Carolina at Chapel Hill, and representatives from the two groups formed a metadata research team.

A key question the team faced at the outset was, “Who will create the metadata?” [3]. The Library staff was relatively small and lacked the resources to create metadata internally. The logical alternative was to investigate whether the creators of Web content could also create the metadata themselves. Another alternative to consider was whether the metadata could be automatically generated. The team decided to concentrate on first testing the hypothesis that resource authors (e.g. authors of a resource’s

intellectual content) could successfully create metadata. In order to launch the project, the team mapped out the NIEHS Dublin Core metadata schema, which later became NIEHS application profile. The development and testing of a metadata creation application for resource authors, based on both schemas, has been a major aspect of the project and is the focus of this paper.

3. Metadata Creation Tools for Resource Authors

In an effort to improve resource discovery, many organizations similar to NIEHS are instituting frameworks to support metadata creation by individual authors [4]. For many authors this is an extension of familiar activities. Researchers, for example, regularly produce abstracts, keywords, and other types of metadata for their scientific publications. When writing for the Web, authors can provide metadata via a template or editor, while relying on information architects or other specialists to ensure the metadata is used effectively.

Metadata creation can be a complex process, especially for authors, given their limited skill and knowledge in this area compared to catalogers. Therefore authors could benefit from support offered by professionals [5], automatic processes [6], and well-designed metadata creation tool interfaces. However, there has been little research into appropriate designs for these interfaces. Thus, while new metadata creation applications are proliferating (e.g. The OAI Provider Service Template C, Version 2.0 available at: <http://oai.grainger.uiuc.edu/ProviderTools/TemplateCReadme.htm>), there is little evidence as to whether these tools are effectively supporting authors, specifically resource authors lacking professional metadata creation education.

Instead, much metadata research has continued the tradition of cataloging research, which primarily focused on users’ interactions with the online library catalog, rather than on the catalogers’ interactions with cataloging modules. As an extension of these early practices, current metadata research activities emphasize interoperability and standardization, again with little focus on the usability of metadata creation applications. These current emphases are undoubtedly important, however, they will have widespread implications only inasmuch as metadata is actually being created. The challenge of enabling metadata creation has motivated us to examine the design of metadata creation tools, particularly to support decentralized, author-generated metadata.

4. Research Goals and Objectives

The goal of the research presented here was to examine the usability of both versions of the NIEHS metadata generation tool (Version 1.0 and 2.0) in the context of an iterative design process. As part of this process, the

metadata research team conducted cognitive walkthroughs (Wharton et. al. 1994), analyzing scenarios that an author, generating metadata, could potentially face. The metadata research team also conducted two author surveys, one associated with each version of the NIEHS metadata generation tool. The surveys enabled the metadata research team to examine the following topics:

1. Usability of the NIEHS metadata generation tool (metadata application).
2. Time required to learn to use the metadata application.
3. Usefulness of included help text and examples.
4. Comprehensibility of terminology used throughout metadata application.

A series of open ended questions also allowed the metadata research team to gather data on how effectively the metadata application supported authors' understanding of metadata creation as an intellectual activity, and of the use of metadata in a decentralized information environment.

5. Designing and Testing the NIEHS Metadata Generation Tool

5.1. Designing the NIEHS Metadata Generation Tool, Version 1.0

The NIEHS Metadata Generation Tool, Version 1.0, was based on the NIEHS Dublin Core metadata schema [7]. The Dublin Core metadata element set was initially chosen for the NIEHS metadata project because it was developed to support author-generated metadata and because of its underlying principles of simplicity, commonly-understood semantics, international scope and extensibility. The NIEHS Dublin Core metadata schema incorporated all fifteen Dublin Core elements, as well as the element "Audience," which was not a valid Dublin Core element at that time. The NIEHS schema diverged from Dublin Core, Version 1.1, in the areas of obligation and repeatability for selected elements. All Dublin Core elements are optional and repeatable, whereas the metadata research team determined that certain elements needed to be required (e.g., authors' name) and that some could not be repeatable (e.g., rights management). The metadata research team also created *value lists* for element refinement (qualification) for selected elements (e.g., language and audience). All of these developments were incorporated into the design of Version 1.0 of the NIEHS Metadata Generation Tool.

In designing the initial tool, the team analyzed a sample of the available Dublin Core metadata creation tools so as to gain a better sense of the possible design approaches. In particular, the team closely examined Metabot, Reggie, Medical Metadata Creator, and the Nordic Metadata Project. Information on a number of these applications can be found at: <http://www.dublincore.org/tools/>

Each of these tools specialized in varying aspects of metadata generation—from creating and managing only

HTML META tags within Web pages to creating surrogate records in HTML and XML for varying metadata schemas. The metadata produced by most of these tools could be stored in a database and indexed separately from the Web pages, or placed in the header section of Web pages. The team noted strengths and weaknesses of each tool. In particular, they found the time saved in automating metadata generation to be outweighed by the time required to subsequently review and edit the resulting records. The most common areas of concern with automatic tools were the generation of irrelevant keywords, problems with pre-set stop-word lists, the inability to accommodate different punctuation conventions when storing subject keywords of other lists of terms, and limited access to controlled vocabularies such as the *Library of Congress Subject Headings* or *Medical Subject Headings*.

Given the difficulties with automatic generation, the team chose to concentrate its efforts on supporting manual creation by resource authors. The team believed that metadata generation should be made as simple as possible for the creators, without sacrificing the potential for richly descriptive metadata records. Based on this philosophy, and an analysis of the available tools, the team developed a list of design objectives to guide development:

- Ø *Simplicity*. Data entry and editing, navigation, and form submission should be simple, intuitive and give clear feedback.
- Ø *Size or extent*. The data input form should be as compressed as possible in order to reduce scrolling and enable the creator to gauge the "extent" of the task in a single glance.
- Ø *Layout*. Metadata elements should be ordered in a way that made sense to the user, but also in a way that would likely aid authors by providing values for the mandatory elements first, in case authors did not complete the input form
- Ø *Context sensitive help*. Help should be embedded at key points in the tool's interface and also hyperlinked to each metadata element in order to reduce the need to shift to a different context to find information.
- Ø *Color*. Should be selected carefully to enhance the creator's ability to scan the form and focus on particular elements.

With these objectives in mind, the team re-examined the applications noted above. This analysis guided the choice of interface designs (particularly for the metadata input form), interaction styles and widgets. The team then created prototype designs that were analyzed collectively with cognitive walkthrough exercises [8]. The team considered different scenarios that authors would face in working with the metadata application to create metadata for the different types of Web pages that comprise the NIEHS Web site (e.g., departmental, informational, fact-sheets, and so forth). These exercises aided the refinement

of the application design. The final design for Version 1.0 incorporated the following features:

- Ø *Single-page design.* The Web form, the central component of the metadata application, was designed to fit as much as possible on one screen (see APPENDIX, Figure 1).
- Ø *Metadata element ordering.* Mandatory elements appeared at the top of the input Web form, while optional elements appeared at the bottom. Authors were not informed of this distinction, because the goal was to induce authors to contribute as much metadata as possible.
- Ø *Catch page.* If authors failed to submit all the mandatory elements, they were presented with a secondary Web form (a catch page) and informed they needed to submit values for these elements in order to submit their metadata record. Only mandatory metadata elements appeared on the secondary Web form.
- Ø *Drop-down menus.* Drop-down menus were favored over radio buttons (used in an initial prototype) because of their small footprint, which facilitated a highly compressed page layout.
- Ø *Value lists.* The drop-down menus contained lists of terms from which the author can choose. The feature allowed for metadata element refinement, simplified the input process, and helped authors to avoid potential typographic errors.
- Ø *Metadata element definitions.* Each element was hyperlinked to its definition, enabling rapid access to information about unfamiliar elements.
- Ø *Example input.* Selected examples of correct data entry for content syntax (e.g., YYYY-MM-DD for “date”) and form completion (e.g., place on subject keyword per line) were provided.
- Ø *Default values.* Appropriate default values were provided for metadata elements where the data would always be the same. This metadata was not made visible to authors, but generated automatically to ensure the creation of valid records.

5.2. Evaluating Version 1.0 of the NIEHS Metadata Generation Tool

5.2.1. Baseline Study Research Methods

When version 1.0 of the NIEHS web form was completed, a pilot test and a stress test were run at the University of North Carolina to make sure the system could accommodate use by multiple authors at once and to make final revisions to the tool design. The team then set out to conduct the baseline study to test the feasibility of author-generated metadata and survey authors about the usability of the tool.

This research took place in the NIEHS computing laboratory. Each user session lasted approximately one hour. During the first half-hour participants completed both a pre-test questionnaire and a metadata tutorial. A member of the metadata research team presented the tutorial in person. The tutorial introduced authors (participants) to the concept of metadata and explained the features of the metadata application. During the second half-hour, participants created metadata records using the metadata application. Screen-capture software was used to record the process of creating metadata, including authors’ interacting with the metadata application and navigating the Web page they were cataloging. After submitting their metadata, participants completed a post-test questionnaire. A more detailed account of the research method and results of metadata quality evaluations are reported in [9].

5.2.2. Results

The post-metadata creation questionnaire addressed ease of use, learning time, usefulness of help text and examples, and overall comprehensibility of the NIEHS metadata generation tool, Version 1.0. This questionnaire used the term “Web form” rather than “NIEHS metadata generation tool” for simplicity. Survey results found that five of the six participants (83.3%) found working with the metadata application was easy, with the remaining participant giving it a standard (average) score. In the area of time required to use the application, all of the participants indicated that minimal time was required, with two participants selecting the value of 4 (33.3%) and four participants selecting the value of 5 (66.7%) on a semantic differential scale where 1 indicated “slow” and 5 indicated “fast.”

In examining the usefulness of the help text, one participant found it to be standard, three participants found it to be above standard, and two participants indicated that the text was highly useful. Finally, participants also indicated that terminology used in the design of the metadata applications was understandable, with two participants selecting the value of 4 (33.3%) and four participants selecting the value of 5 (66.7%) on a semantic differential scale where 1 indicated “confusing” and 5 indicated “understandable.”

Comments from the post-metadata creation survey indicate that authors were generally pleased with the usability of the tool. One author described it as “self-explanatory and intuitive,” and another said it was simply “easy to use.” However, detailed analysis of authors’ interactions with the metadata application (using the screen captures) uncovered a number of subtler usability problems [10]. For example, participants had difficulties cutting-and-pasting source code and HTML text into the web form.

Additional responses on the post-metadata creation questionnaire further indicate the problems observed via the screen captures. For example, authors reported a need for a better understanding of the metadata record and its purpose,

with one calling for a “philosophic” overview. Participants also reported confusion or uncertainty regarding specific fields and requested greater assistance in determining appropriate inputs. One author noted that it “wasn’t obvious what some categories included.” This problem was especially acute for intellectual fields such as *subject*, for which participants had difficulty determining appropriate granularity and scope.

5.3. Iterative Design and Evaluation of the NIEHS Metadata Generation Tool, Version 2.0

5.3.1. Revisions and Naturalistic Study Research Methods

The results of the baseline study indicated that authors (scientists at NIEHS) could in fact create acceptable metadata and provided a foundation for further investigation of the metadata creation process in a more *naturalistic setting*. The metadata research team sought to examine authors generating metadata outside the structured setting of a laboratory exercise, in the comfort of their own offices without any overt monitoring by the metadata research team. To prepare for the naturalistic testing, the metadata research team first revised the NIEHS Dublin Core-based metadata schema and the metadata application. These activities were based on an analysis and evaluation of the baseline study and new developments within the Dublin Core Metadata Initiative (DCMI).

The NIEHS Dublin Core metadata schema was revised to meet the standards of an application profile [11]. The results of this revision are reported by Harper, et al. [12].

Version 1.0 of the metadata application was revised to Version 2.0 (*see APPENDIX Figure 2 for revised Web form*), in order to support the NIEHS application profile and the following design considerations:

- Ø *Online Tutorial*. An online tutorial was developed to introduce authors to the concept of metadata and the metadata application.
- Ø *Visual Examples*. Hyperlinked definitions of metadata elements were enhanced with screen captures to clearly illustrate the input for each element.
- Ø *Subject Guidelines*. The tutorial and the examples provided for the subject metadata elements provided special instruction on how authors could achieve subject indexing “exhaustivity” and “specificity.” These technical terms, however, were not used; rather, participants were encouraged to cover all the subjects noted in the resource and to be as specific as possible in their assignment of subject keywords.
- Ø *Logical Ordering of Metadata Elements*. Mandatory and optional metadata elements were integrated in a more logical order. The metadata input form was redesigned to better match the

order in which a user might read a Web page while searching for information to record as metadata.

- Ø *Enhanced Catch Page*. The revised catch page included a secondary form for capturing metadata when the “relation element” was used on the initial input screen. This was necessitated by the NIEHS application profile’s merging of the “relation” and “source” fields. In addition, the change helped to avoid overwhelming authors, as they were no longer required to complete this element on the initial screen (*see APPENDIX Figure 3*).

The naturalistic study was conducted as an experiment. The research framework was based on the baseline study, however the setting was labeled as “naturalistic” as the authors participated from the comfort of their own offices during a time that was convenient to them, rather than in a structured and timed setting. The NIEHS Library Director recruited participants via e-mail. The research design included an online metadata tutorial that introduced authors to the concept of metadata and the metadata application. Online participant profile and post-metadata questionnaires were also part of the design.

However, because participants used their own computers, in their own offices, it was impossible to modify the operating system to capture detailed user-interface activity as was done in the baseline study. Also, the revised metadata schema and different methodology make it difficult to directly compare the results of the two studies.

5.3.2. Results

Following the baseline study’s test design, a post-metadata creation questionnaire was implemented to gather author feedback about the NIEHS metadata generation tool and gain insight into their views on metadata creation. Again, for simplicity, the phrase “web form” was used on the post-metadata to refer to the metadata application. Twenty-eight of the 29 participants completed the post-metadata questionnaire.

A five-point semantic differential scale, on which 1 indicated “difficult” and 5 indicated “easy,” was used to collect participants’ perceptions about working with the metadata generation tool. The majority of participants indicated that the metadata generation tool was easy to use, with 11 participants (39.3%) selecting the value 5—the highest value, and 7 participants (25.0%) selecting the value 4. Results from all participants’ responses to this question are given in **Table 1**.

Table 1. Ease of Working with the NIEHS Metadata Generation Tool

Scale	↓ ↓	Value	Frequency	Percentage
		1	1	3.6%
		2	1	3.6%

	3	8	28.6%
	4	7	25.0%
	5	11	39.3%

Participants also found that the time required to learn to use the metadata generation tool was minimal. A semantic differential scale, on which 1 indicated “slow” and 5 indicated “fast,” was used to gather feedback on the learnability of the tool. Fifty percent (14 participants) indicated that the time requirement was minimal (fast). Results are presented in **Table 2**.

Table 2. Time required to learn the NIEHS Metadata Generation Tool

Scale 5 → (slow) ← 1 (fast)	Value	Frequency	Percentage
	1	1	3.6%
	2	1	3.6%
	3	9	32.1%
	4	3	10.7%
	5	14	50.0%

The post-metadata creation questionnaire also gathered data about the help text and terminology used on the metadata generation tool. We hoped to discover whether the textual information included in the form’s interface and available via hyperlinked definitions was helpful and if the terminology used was understandable. A mean result of 3.3 was found when asking participants about help text usefulness (results based on a semantic differential scale, on which 1 indicated “useless” and 5 indicated “helpful”). The value of 3 was selected by the highest number of participants (10 participants, 37.0%). It should be noted that only 27 participants answered this question. A mean result of 3.0 was also found when asking participants about the overall clarity of the metadata generation tool’s terminology. The mode was 3.0 and selected by 9 (32.1%) of the participants. Overall, the results of these two questions indicate that the help text and terminology were of acceptable utility.

Qualitative results support the generally positive impressions participants had of version 2.0. One author described the form as “straightforward and easy to fill out,” while others said it was “easy,” “easy to do,” and “easy enough.” This ease-of-use apparently emerged from the tool’s design, which was described as “nicely explained and organized.”

However, as in Version 1.0, less obvious usability problems were also evident. These centered around the issues of conceptual understanding and motivation. One author reported that he “didn’t understand the meaning or need for some information.” This was a common complaint, despite the presence of the metadata tutorial, which several authors praised. This indicates the importance of appropriate information design in the

metadata creation interface in order to facilitate conceptual understanding during the creation process. In particular, examples of useful metadata could be helpful; several authors requested “more concrete examples.” Examples were provided for each metadata element via screen captures and could be viewed if the author selected the corresponding hyperlink. The research team surmised that authors wanted examples of completed metadata records.

A critical problem for decentralized metadata creation, particularly in a high-productivity research organization like NIEHS, is motivating resource authors to provide useful metadata. Comments from the authors indicate the magnitude of this problem. One noted skeptically that metadata creation “seems like just one more thing added to all the other ‘one more things’ that eventually take up all your time.” Another put it bluntly: “not my job or responsibility or interest.” A third took a pragmatic perspective, saying that “although it was easy, I’d have to have an incentive to think about it.” Given these motivational concerns, it is imperative that metadata creation be as simple and rewarding for resource authors as possible.

Our data indicate that metadata creation using Version 2.0 was fairly simple for most authors. In addition the issue of simplicity can be addressed by continuing to refine the usability of metadata creation applications, as in the iterative design process reported here. The issue of incentives is more difficult, but can be partially addressed by helping authors understand the usefulness of the metadata records they are creating. Some authors grasped this usefulness immediately and reported being engaged in a useful exercise, important to NIEHS’ institutional objectives. “I hope [it] will help people searching for information to find our site,” one said. Another concluded that the metadata initiative “should be a great help,” while a third author placed the issue in larger context, saying “I am glad you are doing this; we and our web site have much catching up to do to make all the info we have easily accessible to the taxpayers.”

In contrast to these authors, however, others were puzzled as to the importance of metadata for NIEHS. “If the form is adopted by NIEHS, how exactly will it be used?” one asked. Another author thought the creation process would make sense “only if you can explain things a little better,” while another simply reported that “I don’t understand the value of this exercise.” “It may be useful to someone,” one author said skeptically, “I’m not sure who.” While motivation is fundamentally an organizational issue, we believe tool design must contribute to authors’ understanding of metadata and how rich metadata repositories can contribute to resource discovery. Our data suggest that Version 2.0 was partially successful at accomplishing this objective.

6. Conclusions and Future Research

Our general conclusion is that metadata creation is a complex task that can greatly benefit from well-designed metadata applications. Iterative design of tools with close attention to local needs can help create useful applications. Related to this goal, we are developing a model of the metadata creation process that can guide design. The model seeks to integrate different metadata generation processes and creator-skills (e.g., input from different people, such as authors and metadata professionals). Our hope is to facilitate the creation of metadata applications that dramatically simplify metadata creation for resource authors, enabling large-scale decentralized metadata production.

We see several immediate opportunities for further research in this area. Researchers could seek to apply formal user-centered design methods [13, 14] to the design of metadata creation tools. These methods could produce more refined metadata applications and present authors with usable interfaces and a more comprehensive understanding of the process of metadata creation. Cross-sectional approaches (even in naturalistic settings) only provide one view of the problem, however. Longitudinal studies of metadata creation within organizations could lead to new insights into how organizations and resource authors develop the metadata creation process over time. This view is particularly important for understanding author motivation, a key issue. Finally, we believe that research needs to be conducted in different organizational contexts with different classes of authors in order to identify institutional factors that influence metadata creation.

Acknowledgements

We would like to thank Microsoft Research and OCLC, Online Computer Library Center for funding that made this research possible. We would also like to acknowledge Michelle Mascaro, Master's student at the School of Information and Library Science, UNC, and Oknam Park, recent graduate, for their assistance with encoding statistics; NIEHS scientists for participating in the studies reported here, and former research team members Bijan Parsia, Corey Harper, and Cristina Pauttuelli for their assistance with the project.

References

- [1] Greenberg, J., (2003). Metadata Generation: Processes, People and Tools. *Bulletin of the American Society for Information Science and Technology*, 29(2): 18-21.
- [2] S. Weibel. Metadata: the Foundations of Resource Description. *D-Lib Magazine*, July 1995. Retrieved from: <http://www.dlib.org/dlib/July95/07weibel.html> .
- [3] Thomas, C. F., & Griffin, L. S. (1998) Who will create metadata for the internet? *First Monday*, 3(12) Retrieved from: http://www.firstmonday.dk/issues/issue3_12/thomas/ .
- [4] Onyancha, I; Keizer J.; & and Katz, S. (2001). A Dublin Core Application Profile in the Agricultural Domain. DC-2001 Proceedings of the International Conference on Dublin Core and Metadata Applications 2001, National Institute of Informatics, Tokyo, Japan. Published on October 24, 2001 by the National Institute of Informatics (ISBN 4-924600-98-9), pp. 193-199. 185-192.
- [5] Greenberg, J. & W. Davenport Robertson (2002). Semantic Web Construction: An Inquiry of Authors' Views on Collaborative Metadata Generation. DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence. Proceedings for the International Conference on Dublin Core and Metadata for e-Communities, 2002, Florence, Italy. October 13-17. Firenze University Press, pp. 45-52.
- [6] Liddy, E.D., Sutton, S. A., Paik, W. Allen, E., Harwell, S. Monsour, M., Turner, A, & Liddy J. (2001). Breaking the Metadata Generation Bottleneck: Preliminary Findings. *JCDL 2001*: 464.
- [7] Robertson, W. D., Leadem, E. M., Dube, J., & Greenberg, J. (2001). Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema. DC-2001 Proceedings of the International Conference on Dublin Core and Metadata Applications 2001, National Institute of Informatics, Tokyo, Japan. Published on October 24, 2001 by the National Institute of Informatics (ISBN 4-924600-98-9), pp. 193-199.
- [8] Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994) The cognitive walkthrough method: A practitioner's guide. In Nielsen, J., and Mack, R. (Eds.) *Usability Inspection Methods*. (pp.105-140) NY: Wiley.
- [9] Greenberg, J., Pattuelli, M. C., Parsia, B., & W. D. Robertson. (2001). Author-generated Dublin Core Metadata for Web Resources: A Baseline Study in an Organization. *Journal of Digital Information (JoDI)*, 2(2): <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Greenberg/>.
- [10] Crystal, A. (2003). Interface Design for Metadata Creation. In *Extended Abstracts of CHI 2003* (Ft. Lauderdale, April 5 - 10 2003). Available at: http://ils.unc.edu/~acrystal/chi03_crystal.pdf.
- [11] Heery, R. & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne*, 25. Retrieved from: <http://www.ariadne.ac.uk/issue25/app-profiles/> .
- [12] Harper, C., Greenberg, J., Robertson, W. D., & Leadem, E. (2002). Abstraction versus implementation: issues in formalizing the NIEHS application profile. Paper presented at the DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, Italy, October 13-17.
- [13] Nielsen, J. (1994) *Usability Engineering*. San Francisco: Morgan Kaufmann.
- [14] Preece, J., Rogers, Y., & Sharp, H. (2002) *Interaction Design: Beyond Human Computer Interaction*. New York: John Wiley & Sons.

Appendix

Figure 1, NIEHS Metadata Generation Application input form, Version 1.0

Document's URL: <input type="text"/>	
Title <input type="text"/>	Document's Language English <input type="text"/> Format text/html <input type="text"/>
Author/Contributor(s) (<i>last name comma first name</i> ; one author per line) <input type="text"/>	Date Created (YYYY-MM-DD, or YYYY-MM, or YYYY, or decade, e.g., 1980s) <input type="text"/> Date Modified (as above) <input type="text"/>
Subject(s) (one key word or phrase per line) <input type="text"/>	Audience (ctrl-click to select more than one) General Public <input type="text"/> Researchers <input type="text"/> NIEHS Employees <input type="text"/> Teachers <input type="text"/> Students <input type="text"/> Kids <input type="text"/> Type (ctrl-click to select more than one) Text <input type="text"/> Sound <input type="text"/> Software <input type="text"/> Service <input type="text"/> Interactive Resource <input type="text"/> Image <input type="text"/>
Alternative Title(s) (one per line) <input type="text"/>	NIEHS Project or Event Number: <input type="text"/> Other ID (e.g. ISBN) <input type="text"/>
Geographic Coverage (e.g., town, county, state, and/or country for which this resource most <i>narrowly</i> applies) <input type="text"/>	Time Coverage (e.g., <i>time span</i> 1999-2001, <i>era</i> Vietnam War or 19th century, or a <i>specific date</i>) <input type="text"/>
Description (enter text) <input type="text"/>	
Related URL(s) (one per line) <input type="text"/>	Original Source(s) (one per line) <input type="text"/>
<input type="button" value="Submit"/> <input type="button" value="Reset"/>	

Figure 2, NIEHS Metadata Generation Application input form, Version 2.0

<p>Document's URL: <input type="text"/></p> <p>Title <input type="text"/></p>			
<p>Author/Contributor(s) (<i>last name comma first name; one author per line</i>)</p> <input type="text"/>	<p>Document's Language (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> English ▲ Spanish ▲ French ▲ Chinese ▼ 		
<p>Keywords (one key word or phrase per line)</p> <input type="text"/>	<p>Date Created (YYYY-MM-DD, or YYYY-MM, or YYYY, or decade, e.g., 1980s)</p> <input type="text"/> <p>Date Modified (as above)</p> <input type="text"/>		
<p>Description (enter text)</p> <input type="text"/>	<table border="0"> <tr> <td> <p>Audience (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> General Public ▲ Researchers ▲ NIEHS Employees ▲ Teachers ▲ Students ▲ Kids ▼ </td> <td> <p>Type (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> Text ▲ Sound ▲ Software ▲ Service ▲ Interactive Resource ▲ Image ▼ </td> </tr> </table>	<p>Audience (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> General Public ▲ Researchers ▲ NIEHS Employees ▲ Teachers ▲ Students ▲ Kids ▼ 	<p>Type (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> Text ▲ Sound ▲ Software ▲ Service ▲ Interactive Resource ▲ Image ▼
<p>Audience (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> General Public ▲ Researchers ▲ NIEHS Employees ▲ Teachers ▲ Students ▲ Kids ▼ 	<p>Type (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> Text ▲ Sound ▲ Software ▲ Service ▲ Interactive Resource ▲ Image ▼ 		
<p>Alternative Title(s) (one per line)</p> <input type="text"/>	<p>NIEHS Project or Grant Number:</p> <input type="text"/> <p>Other ID (e.g. ISBN) <input type="text"/></p>		
<p>Related Resource(s) (URL, standard number, or title; one per line)</p> <input type="text"/>	<table border="0"> <tr> <td> <p>Geographic Coverage (e.g., town, county, state, and/or country for which this resource most <i>narrowly</i> applies)</p> <input type="text"/> </td> <td> <p>Time Coverage (e.g., <i>time span</i> 1999-2001, <i>era</i> Vietnam War or 19th century, or a <i>specific date</i>)</p> <input type="text"/></td> </tr> </table>	<p>Geographic Coverage (e.g., town, county, state, and/or country for which this resource most <i>narrowly</i> applies)</p> <input type="text"/>	<p>Time Coverage (e.g., <i>time span</i> 1999-2001, <i>era</i> Vietnam War or 19th century, or a <i>specific date</i>)</p> <input type="text"/>
<p>Geographic Coverage (e.g., town, county, state, and/or country for which this resource most <i>narrowly</i> applies)</p> <input type="text"/>	<p>Time Coverage (e.g., <i>time span</i> 1999-2001, <i>era</i> Vietnam War or 19th century, or a <i>specific date</i>)</p> <input type="text"/>		
<p>Submit Reset</p>			

Figure 3: Mandatory fields error page (catch page), Version 2.0

You must fill in the following items before your metadata can be processed and stored.

<p>Subject (one key word or phrase per line)</p> <input type="text"/>
<p>Date Modified (YYYY-MM-DD) (YYYY-MM-DD, or YYYY-MM, or YYYY, or decade, e.g., 1980s)</p> <input type="text"/>
<p>Type (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> Text Sound Software Service Interactive Resource Image
<p>Date Created (YYYY-MM-DD, or YYYY-MM, or YYYY, or decade, e.g., 1980s)</p> <input type="text"/>
<p>Audience (ctrl-click to select more than one)</p> <ul style="list-style-type: none"> General Public Researchers NIEHS Employees Teachers Students Kids
<p>Document's URL: <input type="text"/></p>
<p>Authors/Contributors (<i>last name comma first name</i>; one author per line)</p> <input type="text"/>
<p>Description (enter text)</p> <input type="text"/>
<p>Title</p> <input type="text"/>

If you have more specific information about the Related Resources, please fill it in below

URL, Standard Number, or Title for the resource	Type of relation
http://www.foo.com/	Relation
The Title of Something	Relation
ISBN 8-304882784-39	Relation
<input type="button" value="Submit"/>	