Proc. Int. Conf. on Dublin Core and Metadata for e-Communities 2002: 45-52

Semantic Web Construction: An Inquiry of Authors' Views on Collaborative Metadata Generation

Jane Greenberg
School of Information and Library Sciences
University of North Carolina at Chapel Hill
janeg@ils.unc.edu

W. Davenport Robertson
National Institute of Environmental Health Sciences
robert11@niehs.nih.gov

Abstract

Increasing the amount and quality of metadata is essential for realizing the Semantic Web. The research reported on in this article addresses this topic by investigating how resource authors might best collaborate with metadata experts to expedite and improve metadata production. Resource authors, working as scientists at the National Institute of Environmental Health Sciences (NIEHS), were surveyed about collaborating with metadata experts (catalogers) during the metadata creation process. The majority of authors surveyed recognized cataloger expertise is important for organizing and indexing web resources and support the development of a collaborative metadata production operation. Authors discovered that, as creators of web resource intellectual content, they too have knowledge valuable for cataloging. This paper presents the study's framework and results, and discusses the value of collaborative metadata generation for realizing the Semantic Web.
Keywords: *Semantic Web, Collaborative Metadata Generation, Human generated metadata, Dublin Core, National Institute of Environmental Health Sciences (NIEHS), Government Agencies.*

1. Introduction

Envisioned by Tim Berners-Lee, the inventor of the World Wide Web (web), and further defined by a number of key researchers and visionaries, the Semantic Web aims to bring structure to the web's meaningful content. The goal, as Lassila et al. (2001) explain, is to build a structured environment where software agents roam and carry out sophisticated tasks, such as arranging *all* the components of a summer vacation, from air travel and hotel to a night on the town. Structured knowledge representation underlying the Semantic Web needs to be built upon trusted metadata—that is accurate, consistent, sufficient, and thus reliable metadata.

Although researchers agree creating trusted metadata is fundamental to realizing the Semantic Web, examining partnerships among persons involved in metadata creation does appear to be a major focus. A probable reason for this predicament is the need to *first* clarify the Semantic Web's conceptual design, an undertaking being documented via numerous theoretical and practical discussions (see links from: (www.w3.org/2001/sw/). Another possible factor is the need to invent and test Semantic Web languages, or what may be thought of as infrastructure technologies (e.g., Resource Description Framework (RDF) (www.w3.org/RDF/), DAML (DARPA Agent Metadata Language) + OIL (Ontology Inference Layer) Reference Description (www.w3.org/TR/daml +oil-reference), and now OWL (Ontology Working Group Language) (Patel-Schneider, 2002). This second focus is evident by research presented at the recent Semantic Web Conference (ISWC2002), Sardinia, Italia (see conference proceedings at: link.springer.de/link/service/series/0558/tocs/t2342.htm). These research emphases are critical to the Semantic Web's development, although they do not specifically address the fact that a vast majority of web content is not semantically encoded with the metadata required for agent roaming and automatic processing activities.

If the amount and quality of web content metadata is to increase, Semantic Web efforts need to also prioritize metadata generation research. Important foundation work designing metadata schemas (e.g., Dempsey et al. 1997) and developing metadata tools

(For example, see Dublin Core tools at: www.dublincore.org/tools/) is in progress. Paramount now is the need to discover the best means for efficiently producing good quality metadata, drawing from both human and automatic processes.

2. Human-Metadata Generation

Human-metadata generation, the focus of this study, takes place when a person is responsible for the identification and assignment or recording of resource metadata. Human-metadata generation is often explained by distinguishing it from *automatic-metadata generation*. In the first case a person intellectually manages in the metadata generation, whereas in the latter case a machine-based algorithm automatically extracts metadata from the resource content. Both methods have strengths and weaknesses, and experts, particularly in the area of subject indexing, agree that the most effective results can be achieved through methodological integration (e.g., Schwartz 2000, p. 152). Although empirical evidence is limited, it appears that the variety of persons involved in metadata generation also exhibit different strengths and weaknesses, making it likely that the best results will be achieved through skill integration.

This research focuses on potential collaboration between metadata experts (here after referred to as experts) and resource authors (hereafter referred to as authors). These classes of persons have been selected as they are among two of the most active producers of descriptive metadata. A discussion of the persons involved in metadata generation is found in Greenberg (2002). Descriptive metadata includes elements, such as "title", "author/contributor", and "subject"; these elements provide surrogates for information resources and facilitate discovery. *Information resources* are objects housed in digital and physical libraries, museums, archives, and like information centers.

Experts include catalogers, indexers, and other persons having formal education, earning an advanced degree in information or library science. They are often preferred metadata creators because their ability "to make sophisticated interpretative metadata decisions and work with classificatory systems" (Greenberg 2002) aids in the production of high quality metadata (Weinheimer 2000). Experts' skills are, however, insufficient when addressing common web resource problems stemming from the absence of "title pages" and other standard bibliographic features, which are heavily relied on in cataloging. As third-party metadata creators, experts may not be privy to resource details needed for creating descriptive metadata.

Authors include persons who produced the intellectual content of the resource being cataloged. They are intimate with their creations and have knowledge of *unrecorded* information valuable for producing descriptive metadata. An example is "date of creation" metadata. A scientist/author may know when a report was originally published, although the web version may not show this information. Exploratory research demonstrates to some degree that authors can produce acceptable metadata (Barrueco & Krichel 2000, Greenberg et al. 2001). Further evidence is found in that commercial information databases (e.g., *Dissertation Abstracts*) index resources with abstracts, keywords, and other author-generated metadata. In fact, many publishers of scientific journals require authors to submit subject "keywords" with their manuscripts. A limitation with authors as metadata creators is that they may lack knowledge of indexing and cataloging principles and practices, and are more likely to generate insufficient and poor quality metadata that may hamper resource discovery (Milstead & Feldman 1999, Thomas & Griffin 1999, Weinheimer 2000).

The field of information and library science has a substantial body of research studying automatic and human indexing, a recent summary of which is found in Anderson and Pérez-Carballo (2001). Additionally, metadata generation tools experiment with the integration of automatic and human processes e.g., www.lub.lu.se/tk/metadata/dctoollist.html. In efforts to realize the Semantic Web, it makes sense to further extend comparisons and integration activities to collaboration among different classes of persons generating metadata—the goal of this research.

3. Towards A Collaborative Metadata Generation Framework

Collaborative metadata generation, as defined in this study, is the *joint production of web resource metadata*. While official collaborative metadata generation programs appear scarce, collaboration research, together with long-standing indexing practices and recent web-based initiatives, provide a framework for developing such an operation. Several of these developments are highlighted below:

"Collaboration" research. A growing area of research focuses on collaboration between "system designers" and "potential users" during information system design activities (e.g., Sonnenwald and Lievrouw 1996). This work examines social and behavioral issues that arise when "technical experts" (system designers) and "clients" (persons for whom a system is being designed and who have intimate discipline knowledge) collaborate. Results provide insight into issues that may arise when experts (e.g., catalogers) and authors, who are domain experts with respect to their creations, collaborate during metadata creation.

De-facto collaborative metadata generation. As highlighted before, scientists and scholars generated "abstracts", "keywords" and other metadata for their

publications. Commercial databases adopt and enhance this metadata for access. Frequently a metadata expert conducts authority control work to standardize subject and name-headings. This framework is one of *economy*, allowing metadata experts (generally indexers) to take advantage of author knowledge and devote their valuable and more costly time to metadata activities requiring professional training. The partnership may be viewed as a *de-facto* collaboration rather than an *active* collaboration because of the absence of real time communication between author and professional.

Dublin Core metadata. The Dublin Core Metadata Initiative (DMCI) has facilitated the development and use of the Dublin Core Metadata Element Set (1997), a schema comprised of 15 elements deemed essential for resource discovery (Weibel 1995, Duval et al. 2002). An underlying principle is that this schema is simple enough for nonprofessional use. Authors can create metadata working with simple templates or editors, and experts can subsequently enhance this metadata following a more complex schema or by performing authority control work. OCLC's Cooperative Online Resource Catalog (CORC) (www.oclc.org/corc) project provides framework for this type of collaboration.

Open Archives Initiative. The Open Archives Initiative (OAI) (www.openarchives.org/) promotes interoperability standards that facilitate efficient access to web content and other forms of electronic documentation. The OAI has adopted the Dublin Core metadata schema. OAI projects use a variety of metadata generation techniques, including metadata produced by experts or authors. Metadata from any OAI compliant initiative can be harvested and placed in *single service*, and may result in a collection of metadata generated by experts for some resources and by authors for other resources. Integrating expert and author produced metadata records, post-metadata creation, may be viewed as a partnership in that both parties (experts and authors) are contributing to a larger pool of resource representation—generally for a particular domain. It's likely that some OAI projects carry out collaborative metadata generation during the initial metadata production stage, although documentation is limited.

Metadata tutorials. Metadata initiatives associated with the web expand well beyond the traditional library environment to other information communities (e.g., commerce, health science, and geo-science). As part of this development, experts have been called upon to write schema specifications and design tutorials instructing authors and other persons about metadata creation. Additionally, many HTML guides instruct web developers and resource authors about the creation of meta tags, often highlighting the "keyword," and "description" tag (e.g., Dr Clue's HTML/CGI Guide (http://www.drclue.net/F1.cgi/HTML/META/META.html). Expert designed tutorials providing metadata creation guidance to authors and other non-metadata experts provides a form of collective metadata generation that may have implications for collaborative activities.

The developments reviewed here provided a framework for this paper's examination of authors' attitudes about collaborative metadata generation involving experts.

4. Research Goals

This study was conducted to gain insight into authors' perceptions about collaborative metadata generation. The study was conducted as part of a larger ongoing study that is examining human and automatic metadata generation methods. An underlying goal of this study is to assist the National Institute of Environmental Health Sciences (NIEHS) metadata project. A broader objective is to share these results with similar organizations aiming to increase the amount and quality of metadata, while contributing to the Semantic Web's construction. Questions guiding the study were:

- Do authors think expert assistance would be useful during the metadata creation process?
- What communication methods do authors prefer in a collaborative metadata generation operation?
- What types of metadata are authors most likely to seek expert help generating in a collaborative environment?

5. Method

The survey method was used to gather data on author's views about collaborative metadata generation. This survey was supplemented by data gathered via a participant profile questionnaire and a post-metadata creation questionnaire implemented in a larger ongoing metadata generation study, which will be reported on in a future paper.

The test domain was the National Institute of Environment Health Sciences (NIEHS), an Institute of the National Institutes of Health (NIH), which is a component of the U.S. Department of Health and Human Services (DHHS). Participants were NIEHS scientists who had created Dublin Core metadata records in the larger metadata generation study for at least one of the web resources they had authored. Thirty-four scientists were each sent a printed copy of the collaboration survey and printed copies of the metadata records they had created in the larger study. The metadata records were reproduced on yellow paper to distinguish them from the collaboration survey and remind participants that, approximately three-months earlier, they had produced at least one Dublin Core metadata record for their web resource. The survey materials were sent via NIEHS inter-departmental mail with the assistance of library staff and student interns, who are active members of the

NIEHS metadata team. Printed survey materials were used instead of electronic materials because library staff indicated that this would most likely result in the highest return rate.

The survey was brief and included a series of questions asking participants if they thought cataloger assistance would be useful during metadata generation, if so—through what methods would they prefer to communicate with a cataloger, and for what types of metadata generation might they seek expert help. An open question at the end asked participants to describe other ways they envision scientists collaborating with catalogers to generate metadata. Participation in the study was optional.

The survey was efficiently designed on a single page that could be folded in-half upon completion to display the library director's return address. The design made it possible for participants to answer the survey and easily place it in inter-departmental mail without the complications of finding an envelope. A period of two weeks was given for survey completion, during which time the library director sent out two e-mails encouraging scientists to respond.

6. Results

Nineteen NIEHS scientists responded to the collaborative metadata generation survey. As indicated under methodology, the collaboration survey was sent to 34 scientists participating in the larger study. Of the 19 responses received, 18 (52.9% of the 34 originally distributed) were useful for data analysis. One returned survey was eliminated from data analysis due to a failure to answer any of the questions.

Results of data analyzed for this study fall into two categories: 1) Participant background and metadata knowledge, and 2) Collaborative metadata generation survey results.

6.1 Participant background and metadata knowledge

Participant assessment was based on data gathered via a participant profile questionnaire and the post-metadata creation questionnaire implemented in the larger study noted above. This data was culled to provide contextual information for the current study's data analysis and discussion. Of the 18 participants partaking in the collaborative metadata generation survey, nine (50%) search the web daily, six (33.3%) search weekly, and three (16.7 %) search monthly or less than once a month. These results indicate a fairly good comfort level with public and consumer-oriented web technologies (e.g., search engines and consumer web sites, such as Amazon.com). Participants' understanding of the word "metadata" appeared limited with only four (22.2%) of the 18 indicating they had heard the word metadata prior to the NIEHS metadata research. Three of these participants attempted to define metadata, with one response being accurate, giving the definition of "data about information". Limited metadata knowledge was further evidenced by the fact that only one participant had created web resource metadata prior to participating in the NIEHS metadata generation study, although six (33.3%) participants had experience creating HTML (hypertext markup language) documents.

All 18 participants had created at least one metadata record in the NIEHS metadata generation study. This task was completed by inputting metadata into the NIEHS metadata form, which is a simple template based on the Dublin Core metadata schema. Post-questionnaire data gathered after this activity provided insight into participants' views on the value of metadata and metadata creation. A semantic differential scale, on which "1" indicated "with difficulty" and "5" indicated "easily" gathered participants' opinions about the difficulty of the metadata creation task. The majority of participants indicated that it was an *average* to *easy* task, with 16 (88.9%) selecting a "3"or above. A semantic differential scale where "1" indicated "never" and "5" indicated "always" gathered data about participants' views on the need to create web resources metadata. Fifteen participants (83.3%) selected a "3" or above indicating an *average* to *always* support for web resource metadata. A final question asked participants who should create metadata. A check list included the following: "No one," "Authors," "Departmental heads," "Librarians," "Web Masters," "Secretaries" and "Other." Participants were encouraged to select as many options as they would like. Ten participants (55.6%), selected author, whereas 6 participants (33.3%) selected librarians. The results of the profile and post-metadata generation questionnaires show that authors value metadata and believe they should create metadata for their works.

6.2 Collaborative metadata generation survey results

The collaborative metadata generation survey gathered data on authors' views about collaborating with a "cataloger" during metadata generation. Data gathered on establishing a collaborative metadata generation program was fairly, although not unanimously positive with 12 of the 18 participants (66.7%) indicating that assistance from a professional cataloger would have been useful during the metadata creation process. Reasons given noted the ability of catalogers to be consistent and make resources accessible to potential users. A few replies also revealed a slight insecurity among participants in terms of their own ability to produced quality metadata. Examples of responses include the following:

- It [cataloger assistance] would ensure that consistent terms were being used across the various programs/initiatives.

- I can't do it adequately without assistance.
- Professional cataloger will/should be up to date and understand communication via this vehicle and enable a broader audience.
- I'm not sure of the best word to use to 'hit' a particular audience.

Six participants (33.3%) indicated that professional cataloger assistance would not have been useful during the metadata creation process. Only one participant provided a textual response, which was related to the fact that he was cataloging a "project website," not a "publication." The response was, "I'm open to ideas, but I think the only webpage that might fit would be publications." Three participants (16.7%) (two supporting cataloger assistance and one not in favor of cataloger assistance) provided textual responses indicating they were confused by the word "metadata" and its meaning. The NIEHS library staff surmised that the confusion stemmed from the survey's use of the words "metadata" and "metadata record", and several responses suggest participants may have actually equated the word "metadata" with HTML encoding required for a webpage.

Despite the noted confusion, participants were clearly cognizant of the fact that they created a "metadata record". Thirteen participants (81.3%) indicated "yes" they wanted to be notified if their metadata record was to be expanded or modified by a professional cataloger in *any way*, while three participants (18.7%) replied "no", they did not want to be notified. (Percentages, based on the population of 18 participants. are given here and in the remainder of this paper). A cross-tabulation was performed to see if there was a correspondence between participants supporting cataloger collaboration and wanting to be notified about metadata records enhanced or modified by a cataloger. The majority of participants (8 of 12, 66.7%) supporting cataloger collaboration wanted to be notified of changes to their metadata record. Five participants (27.8%) supporting cataloger collaboration did not want to be notified of changes; and two participants (11.1%) not supporting cataloger collaboration, wanted to be notified. Three participants did not answer this question. Results for the correspondence analysis indicate that participants' supporting collaboration were enthusiastic about communicating with catalogers even after they created their initial metadata record. These participants may have had a sense of metadata record ownership and/or a desire to learn more about metadata creation from a cataloger. It's likely that participants who did not want cataloger assistance, but wanted to be notified of any changes, had a sense of metadata record ownership for their work (the metadata record they created in the larger experiment and which was given to them on yellow paper for this study). More research is needed to verify these observations, and well as why participants supporting collaboration did not want to be notified of cataloger changes.

Participants were asked about preferred method of communication in a collaborative metadata generation operation. Among options given were: "Face to Face (a personal contact with a cataloger), "E-mail", "Phone", "Web-form" and "Other". Seventeen participants responded to this question. Percentages based on the entire population (18 participants) are given below in Table 1.

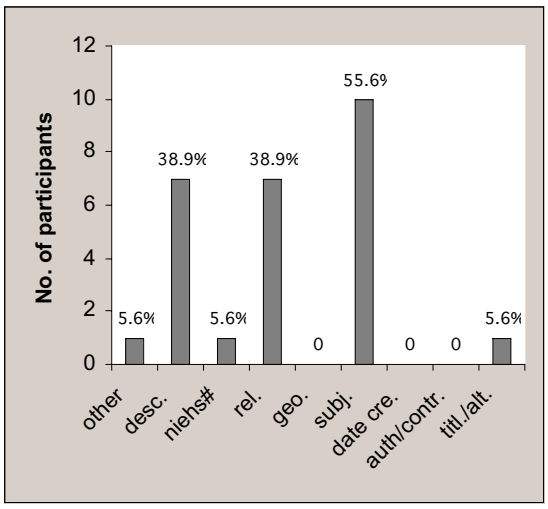
Table 1. Communication Methods Preferred for Collaborative Metadata Generation

Communication Method	Valid Percent selected
Face to face	7 (38.9%)
Email	6 (33.3%)
Phone	2 (11.1%)
Web form	2 (11.1%)
Other	0 (0.0%)

The preferred communication methods were "face to face" (personal contact) and "electronic mail". No relationship was found between preferred communication methods and desire for (or not for) cataloger assistance.

The last segment of the collaboration survey examined metadata elements that participants would like help generating. The NIEHS metadata schema is an application profile based on the Dublin Core metadata schema, the GEM (Gateway to Educational Materials) (www.geminfo.org) schema, and the NIEHS namespace (Robertson et al. 2001). The NIEHS application profile underlying this study is reported on in this conference's proceedings (see: poster session, Harper et al. 2002). A checklist consisting of descriptive labels for eight NIEHS Dublin Core metadata elements was given to participants, with the option to add "other" elements as desired. (The check-list included "Title/Alternative title", "Author/Contributor", "Date created", "Subject keywords", "Geographic coverage", "Relationship w/other resources", "NIEHS number", and "Writing a description"). Participants were asked to each select three or fewer elements from the checklist. It should be noted that the checklist was purposefully not inclusive but served to prompt participants thinking about their metadata generation experience. To facilitate this process, participants were also encouraged to view the metadata records they produced prior to the collaboration survey, which was reproduced on yellow paper.

Figure 1 graphically compares the selection of each individual metadata element by participants. Results show that more than half of the participants favored cataloger help for "subject keyword" metadata (10 of 18 participants, 55.6% selected this element). Participants also favored cataloger help for "relationship" and "description" metadata to a fair degree, as both of these elements were selected by 7 of the 18



*percentages given past on participant sample of 18 per element.

Figure 1. Author's Selection of NIEHS Dublin Core Elements for Cataloger Assistance

participants (an 38.9% selection rate). "Other", "NIEHS number", and "Title/Alternative title" were each selected once. Among the elements not selected by any of the participants for assistance were "Author/Contributor", "Date created", and "Geographic coverage". "Author/Contributor" and "Date created" are first-hand knowledge for authors and not very complicated to generate, so these results make sense. Most NIEHS web pages do not concern a geographic focus, so leaving this element blank may have been an obvious choice.

One-third (6 of 18, 33.3%) of the participants did not select any elements. Four of these participants where four of the six (66.7%) who indicated that they did not think cataloger assistance would be useful during metadata generation. The other two participants who originally indicated that they did not think cataloger assistance would be useful each selected one metadata element where they thought cataloger guidance would be useful: One of these participants selected "other" and identified "new publications" and the other participant selected "description".

A final open-ended question asked participants how they envisioned collaborating with a cataloger. Responses indicated participants' awareness and concern about metadata currency. For example, one participant said that, "as new pages are developed or old pages are modified, program staff would meet with web designer and cataloger to ensure the site is easily accessed by the appropriate audience(s)". Another example is offered by a participant who asked, "how often will this [metadata] be upgraded? (PI [principal investigator] leaves in a few days and new PI arrives in mid-June)". This second example reveals the participants concern for the PI represented in the metadata, not being up-to-date. Several participants commented on preferred communication methods and interest in collaboration, while three participants noted their confusion about the word metadata.

7. Discussion of Results

The results of this study provide insight into authors' views on collaborative metadata generation, preferred communication methods for collaborative metadata generation, and types of metadata authors are most likely to seek expert help generating. Moreover, they provide clues about how to expedite the creation of quality metadata through a collaborative model to help build the Semantic Web.

Authors' views on collaborative metadata generation. The majority of participants in this study support collaborative metadata generation. They recognized that catalogers have special knowledge and skills—particularly in working with standards. Furthermore, participants recognized that, as resource authors, they too have knowledge valuable to cataloging. Scientists are in the business of generating data. They are avid users of commercial database and often depend on data sets created by other researchers. Data integrity is critical to scientific work; good data furthers research efforts, leads to new discoveries, and advances knowledge. Given these scenarios, it makes sense that the scientists participating in this study demonstrated an understanding of the importance of producing quality metadata in an efficient manner, and supported collaborative metadata generation.

Preferred communication methods for collaborative metadata generation. Communication methods are key to any collaborative operation. Participants in this study were equally in favor of both personal (face to face) communication with catalogers and using electronic communication protocols supported by e-mail. The web offers glorious new communication capabilities for disseminating and accessing resources. For example, it's fairly easy to video-conference with colleagues across the globe from the comfort of your own office. Although the results of this study indicated two preferences, it's very likely that these results will change over time, particularly with the introduction of new technologies. Likewise, preferences will change as collaborative partnerships grow, and partners (authors, catalogers, etc.) comfort levels are established. Additionally, different collaborative partnerships and partners will prefer different communication protocols, or work with a combination of methods (for example, "e-mail" and "fact to face" meetings or "telephone"). The researchers in this study advocate that collaborative metadata operations remain open to and test new technologies and various combinations of methods on both a team and

individual level. Related to this is the need to explore human computer interaction (HCI) questions and the design of web forms and tutorials developed to help authors in a collaborative metadata project.

Types of metadata authors most likely to seek expert help generating. Participants selected "subject" metadata for cataloger assistance more than any other element, indicating that this element might be the most problematic for authors. Participants' selection of "subject" metadata was further supported by their high selection of the "description" metadata element, which includes *abstracts*, *summaries*, *content notes* and other descriptions with an *intellectual aboutness* (subject) component. Based on these results, subject metadata is an area where experts might focus their attention. Experts could help educate authors through interactive sessions. Experts might also provide tutorials for using the wide-variety of subject tools available on the web, many lack user friendly introductions about the principles of subject analysis. The larger metadata generation experiment referred to above included a metadata tutorial with slides illustrating how to achieve subject *specificity* and *exhaustivity* when creating metadata. The results of the author generated metadata needs to be analyzed and may provide further insight into this issues.

What is perhaps most significant about subject metadata is its relationship to ontology construction and use and the goals of the semantic web. Achieving semantic interoperability and sharing ontological structures are critical for building the Semantic Web (Heflin & Hendler 2000). (Example, see also homepage for ECAI-02 Workshop on Ontologies and Semantic Interoperability: www.informatik.uni-bremen.de/~heiner/ECAI-02-WS/). Underlying this goal is the need to accurately employ metadata schemas and assign ontological terminology, particularly *subject terminology*. Explained another way, without semantic interoperability, there can be no Semantic Web, because agents have no intelligent knowledge structure to roam, make linkages, and complete tasks. Subject metadata is at the core of many ontologies that are being constructed and needs to be studied from a number of different venues.

8. Conclusion

This study provides insight into aspects of collaborative metadata generation that may be useful for achieving the Semantic Web. Although the sample size was limited, the results are fairly consistent and provide data for comparing results gathered from additional research in this area. Another contribution of this work is that the research design provides a framework for future studies examining collaboration among authors and experts during metadata creation, as well as for other classes of persons (e.g., professionals and para-professionals).

The goal of the Semantic Web is to support sophisticated tasks, such as planning a vacation. Web agents are essentially problem solvers, in that a person seeks assistance from a web agents, which roams the web to complete a task or provide an answer. One of the major limitations to this simple idea is that there is not nearly enough web content metadata to facilitate sophisticated web agent roaming and task activities. Examining the potential for collaborative metadata generation by drawing upon the expertise of different classes of persons is one way to contribute to remedying this problem—herein is the topic underlying this paper.

Scientific, government, and educational institutions are among leading users of the web technology. Their information is non-proprietary and produced for betterment of society. These agencies have a vested interested in their resources being used for problem solving and in seeing the realization of the Semantic Web. Research needs to further explore options whereby authors and experts in these institutions may effectively collaborate to efficiently generate good quality metadata and contribute to a foundation for the Semantic Web. The results presented in this paper indicate that the authors' surveyed are supportive of a collaborative metadata operation, at least in a governmental institution.

In conclusion, 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. The Dublin Core's application in a wide-variety of metadata initiatives and its use by many different classes of persons (experts, authors, web developers, etc.) provides opportunity for collaborative metadata generation involving different classes of persons and in different environments. If such partnerships are well planned and evaluated, they could make a significant contribution to achieving the Semantic Web.

Acknowledgements

This research was funded in part by OCLC Online Computer Library Center Inc. and Microsoft Research.

The authors would like to acknowledge the following members of the NIEHS metadata team and School of Information and Library Science, University of North Carolina (SILS/UNC) for their help with the collaborative metadata survey research: Ellen M. Leadem (NIEHS Library), Corey Harper (University of Oregon Libraries), Oknam Park (SILS/UNC), Bijan Parsia (UNC/University of Maryland Semantic Web Research Team), Cristina

Pattuelli (SILS/UNC), Nikki Warren (SILS/UNC and NIEHS intern), and Wayne Wiegand (SILS/UNC and NIEHS intern).

References

Anderson, J.D. & Pérez-Carballo, J. 2001. The nature of indexing : how humans and machines analyze messages and tests for retrieval. Part I. Research, and the nature of human indexing. Information Processing & Management, 37: 231-254.

Dublin Core metadata element set, version 1.1: reference description. 1997. Available at: http://purl.org/DC/documents/rec-dces-19990702.htm.

Duval, E., Hodgins, W., Sutton, & Weibel, S. 2002. Metadata principles and practicalities. D-Lib Magazine, 8(4): http://www.dlib.org/dlib/april02/weibel/04weibel.html.

Dempsey, L., Heery, R., Hamilton, M., Hiom, D., Knight, J., Koch, T., Peereboom, M. & Powell, A. 1997. DESIRE - Development of a European Service for Information on Research and Education (RE 1004): http://www.ukoln.ac.uk/metadata/desire/overview/.

Greenberg, J. 2002. Metadata and the World Wide Web. Encyclopedia of Library and Information Science. New York: Marcel Dekker, Inc. 72: 344-261.

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/.

Harper, C., Greenberg, J., Robertson, W.D. & Leadem, E. (2002). Abstraction versus implementation: issues in formalizing the NIEHS application profile. In DC-2002: Metadata for e-Communities: Supporting Diversity and Convergence, Florence, October, 13-17, 2002.

Heflin, J. & Hendler, J. 2000. Semantic Interoperabilty on the Web. Extreme Markup Languages, 2000. http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf

Lassila, O., Berners-Lee, T. & Hendler, J. 2001. The Semantic Web
A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, (May, 2001): http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2

Milstead, J. & Feldman, S. 1999. Metadata: Cataloging by any other name. Online, (January/February): 25-31. http://www.onlineinc.com/onlinemag/OL1999/milstead1.html.

Patel-Schneider, P.F., Horrocks, I. & van Harmelen, F. 2002. Feature Synopsis of the OWL Ontology Language: http://lists.w3.org/Archives/Public/www-archive/2002May/att-0021/01-_owl.html

Roberston, D., Leadem, E., Dude, J. & Greenberg, J. 2001. Design and Implementation of the National Institute of Environmental Health Sciences Dublin Core Metadata Schema. Proceedings for International Conference on Dublin Core and Metadata Applications 2001, October 22-26, 2001, National Institute of Informatics, Tokyo, Japan, pp. 193-199.

Schwartz, C. 2002. Sorting Out the Web: Approaches to Subject Access. Westport, Connecticut: Ablex Publishing. Part of the Contemporary Studies in Information Management, Policies, and Services series by Hernon, P. (Ed.).

Thomas C. & Griffin, L. 1999. Who will create the Metadata for the Internet? First Monday: Peer Reviewed Journal of the Internet: 131.193.153.231/issues/issue3_12/Thomas/index.html.

Weinheimer, J. 2000. How to Keep the Practice of Librarianship Relevant in the Age of the Internet. Vine (Special issue on Metadata, part 1), 116: 14-27.