

**The AMeGA Project:
Recommended Functionalities for Automatic Metadata Generation Applications
(for Section 4.2, LC Bibliographic Control Action Plan)**

Status Report Summary

Work on the AMeGA project has followed the Activity Plan during our 3rd quarter (Activity plan and Gantt chart are found at: http://ils.unc.edu/mrc/amega_activity.htm). A summary list of accomplishments and important developments follows:

1. Metadata creator survey

The survey was launched June 25, 2004, and ran through July 31, 2004. The survey gathered data from metadata professionals and other persons interested in metadata generation. The focus was on functionalities desired in automatic metadata generation applications.

Participants included administrators/executives (from both educational and corporate environments), catalogers/metadata librarians, digital resource librarians and librarians in other roles, Web architects, systems personnel, researchers, archivists, and professors/lecturers. Two-hundred and seventeen (217) survey participants provided responses useful for data analysis. The initial goal was to recruit at least 100 participants; the high response rate is encouraging. It should be noted that a total of 320 people actually started the survey, although approximately 1/3rd of these participants did not complete it mainly because they found it was beyond the scope of their work experience. (Further inquiry confirmed that often a large percentage of people who start Web surveys, disseminated freely via listservs, fail to complete them for this very reason.)

Survey data is currently being analyzed and both qualitative and quantitative results will be included in the final report. Below are some initial findings:

- A large majority (over three-fourths) of participants were in favor of automatic metadata creation tools. Most participants communicated that automatic tools need to allow for human intervention or editing during the metadata creation sequence.
- Asked about current use of metadata applications, about half of the participants (51.1%, or 94 of the 184 participants who answered the question about this topic) reported that one

application is being used for metadata creation in their organization. The other near-half of the 184 participants (48.9%) noted that persons involved in metadata creation in their organization are using two or more applications. One participant reported the use of 7 different applications, which was the upper limit. The use of multiple applications suggests an inadequacy in the overall functionality of current applications, although further research is required to confirm this conclusion.

- Over a third of the 184 participants (38.5%) noted directly above are using applications developed in-house. The high use of homegrown systems may also support conclusions related to the inadequacy of current application functionality, although other factors, such as the cost of developing an in-house application compared to the cost of purchasing an off-the-shelf system need to be considered.
- Asked about quality control (QC), slightly over half of the participants (53.1%, or 85 of the 160 participants who answered the question about QC) reported that their organization had a formal QC process to evaluate metadata. Among the procedures that could be clearly determined, most QC methods appeared to be fully dependent on humans, although a few participants reported on the use of automatic or semi-automatic QC methods.
- Participants were generally in favor of automatic metadata generation for metadata elements that can be easily and fairly accurately extracted or harvested from document content or source code (e. g., “identifier” and “date modified” metadata), but were less favorable about the use of fully automatic techniques for metadata that is more intellectually demanding to create (e.g., “subject/keyword” metadata).

[All survey data, including the results reported on here, are being further analyzed, and will be detailed in the final report.]

2. Metadata creation application functionalities

Documentation associated with the Web sites from both Content DM (www.contentdm.com) and Infomine’s iVia (<http://infomine.ucr.edu/>) was assessed to assist with the identification of recommended functionalities for automatic metadata generation applications. CONTENTdm was found to be among one of the most frequently used metadata creation applications in the metadata creation survey, and Infomine’s iVia has been promoted as an application with a sophisticated automatic subject indexing algorithm. The documentation associated with these two applications was assessed in order to identify

functionalities that were not apparent in other components of the AMeGA research project. Due to time and funding limitations, a further analysis of the usability and effectiveness of these applications could not be part of the AMeGA project. Future research calls for a further analysis of both applications.

3. Document presentation analysis

The analysis of metadata generated by document presentation software (e.g., software for word processing, Web pages, etc.) was completed. A case study approach was used for this part of the AMeGA project. Metadata creation facilitated by templates or generated automatically at the time a document is created or updated was tracked. Additional functionalities were also tracked, such as automatic harvesting capabilities. The final report will include an analysis and discussion on this aspect of the AMeGA project.

4. American Library Association (ALA) Annual Conference

The AMeGA project was reported on in the ALCTS CCS Cataloging & Classification Research Discussion Group (CCRDG) and the LC Action Plan Forum, at this year's annual ALA conference. The presentation for both sessions was entitled: Optimizing Metadata Generation Practices.

AMeGA reporting was combined with a report on the Metadata Generation Research (MGR) project—a project that preceded and has informed the AMeGA project (MGR project Web site: http://ils.unc.edu/mrc/mgr_index.htm). AMeGA reporting focused the planning and launching of the survey, and general project progress. Both sessions provided an opportunity to announce the survey launch and recruit participants. Fruitful discussions followed the reporting in both sessions.

The CCRDG discussion session was recorded by AMeGA task force member, Amanda Wilson, with the assistance of former student team member, Michelle Mascaro. These notes were distributed to AMeGA task force members and staff and to cataloging and metadata-related listservs that requested the report. Session notes are also publicly available via the AMeGA Web site: http://ils.unc.edu/mrc/amega_ccrd.htm and on the ALA/CCS/CCRDG Web site:

<http://www.ala.org/ala/alctscontent/catalogingsection/catdisc/catalogingresearch/ResDGan04.doc>.

MGTF member and LC Contact, John Byrum was present at the CCRDG meeting, and MGTF members Robin Wendler and David Williamson (also a LC contact) were present at the LC Action Plan Forum. It was great to have task force members present at both sessions.

5. Project Web site

The project's Web has been updated and a new section labeled "Project Documentation" (http://ils.unc.edu/mrc/amega_docs.htm) has been added to the left-hand side-bar. This new section links to the project's activity plan, ALA/CCRDG discussion notes, and completed quarterly reports.

6. Project dissemination

The *International Journal of Metadata, Semantics and Ontologies* (IJMSO) (<https://www.inderscience.com/browse/index.php?journalID=152>), sponsored by Inderscience Publishers, is a new metadata journal that is being launched this fall. The editorial board for this new journal includes a mix of people from library science, information science, and computer science.

As UNC's AMeGA project team members prepare the final project report, we are also working on an article to submit to this new journal inline with the first public call for submissions. This is an excellent opportunity to disseminate project findings and gain feedback from the larger metadata community.

6. Next steps

Next steps include:

- Complete a draft final report and article draft for submission to *IJMSO*.
- Have the MGTF review, provide feedback, and endorse the final report.
- Incorporate comments of the MGTF and produce the final, endorsed report.
- Distribute to final report to the library / bibliographic control community, and other communities actively involved in metadata generation.

Respectfully Submitted by Jane Greenberg.

//signed jg

CC: MGTF members and Project Staff