Digging into Metadata: Enhancing Social Science and Humanities Research

Michael Khoo, Xia Lin, Jae-wook Ahn: Drexel University, U.S.A.
Doug Tudhope, Ceri Binding: University of South Wales, U.K.
Diana Massam, Hilary Jones, Dan Needham: MIMAS/Manchester University, U.K.

Overview of goals and objectives of your DiD project

The *Digging Into Metadata* project is developing a novel method for automatically augmenting existing metadata sets with Dewey Decimal classes. The method generates DDC 'tags' for digital resources by digging into metadata records and performing analyses of the metadata in each record. The analysis generates terms that are matched with subject categories in the DDC system. The technique is automatic, scalable, and permits resource discovery across multiple collections without first having to crosswalk metadata to a standard format.

Challenges and lessons learned from international collaboration across disciplines and domains

Challenges in an international project often center about project coordination and communication. *Digging Into Metadata* has relatively successful in these areas. Some reflections on the factors that contributed to this success include:

- *Networks and contacts.* Some key members of the project team already knew each other, either through previous collaboration, or through online communication.
- *Regular meetings.* Regular project telephone conferences were scheduled.
- *Formal and consistent meeting documentation.* Early on, the leaders of the US and the MIMAS (UK) components of the project, agreed to create detailed agendas for meetings, and to keep detailed minutes, including tasks and responsibilities.
- *Face-to-face interaction.* A face-to-face meeting was also held in May of Year 2. In addition to dealing with project issues, this was also beneficial for team morale and confidence. A further face-to-face meeting is planned for December 2012. Project members also met at conferences.

Overall project communication used a document-based approach, where issues were not 'discussed' solely in email, but rather in telephone and email discussions that addressed circulated documents. The documents laid out and explained the issues involved, including previous issue history. Document revision was common. All current documents were also uploaded to a private document wiki.

Digital humanities, social sciences and computational based research methods in the context of big data projects

Big data projects in the digital humanities and social science often depend on common metadata aggregated from multiple sources. The sharing of resources and metadata within and between research communities raises a number of practical issues with data management; at the same time, these data management activities may be unfamiliar ones for many humanities and social science researchers. *Digging Into Metadata* addressed some of these metadata research issues in the context of mining metadata records in an attempt to create a federated backend for three digital libraries. Even with this small number of libraries, using a common metadata format, issues were encountered with metadata quality and consistency. Discussions with other aggregator projects suggest that these problems were not isolated examples, *and that addressing legacy databases and metadata provenance will continue to be an important and foundational issue for big data projects in the digital humanities and social sciences.*

Indicators of success

The project is still in development. A logic model and Gantt chart were used to help plan the proposal, and these helped to identify internal project linkages and milestones. Interim indicators of success include the successful development of functional databases to support the various stages of project workflow. The project is now developing alternate visualization interfaces for searching and browsing aggregated metadata. The first real indicators of success will be derived when these different interfaces are user-tested against each other. Further indicators will be derived with the user-testing and further development of the selected interface with representatives of the target user group.

Measuring impact

Initial impacts have yet to be measured but it is anticipated that initial assessments of impact will be obtained from quantitative and qualitative user-testing with representatives of the target user group.

Knowledge dissemination mechanism and tools

Knowledge dissemination has so far focused on the 'traditional' routes such as conference submissions, workshops, etc. The project has a public wiki, where copies of these publications are available.

Importance of working with libraries, archives and data repositories
Digging Into Metadata has worked closely with three digital libraries. Referring back to the 'big data' section above, in our experience it is necessary to work closely with those involved in data and metadata management in these organizations, as this supports the building of the underlying data/metadata architectures. Again, in our experience, the metadata of a particular organization cannot be taken at 'face value,' and has to be contextualized and described through communication with the metadata owners.

Capacity building and training (students and highly qualified personnel)
The *Digging Into Metadata* has used part of the grant funding to partly support a researcher position in metadata at Drexel University. This has proved to be a very useful decision; the researcher is able to devote significant time to the project, and also has usefully developed his metadata research program at the same time.