

Preserving State Government Digital Information
Illinois Partners Meeting



Minnesota Historical Society

Monday, August 10, 2009
Room 403, Gwendolyn Brooks Building, Illinois State Library
Springfield, Illinois

Illinois participants included Tim Rice from the Legislative Information System (LIS); Kathleen Bloomberg, Andrew Bullen, Anne Craig, Connie Frankenfeld, Joe Natale, Suzanne Schriar, Alyce Scott, and Cheryl Walker from the Illinois State Library (ISL); Robert Boots from the Illinois State Archives (ISA); and Nancy Hoffman and Charles Rodgers from the Minnesota Historical Society (MHS).

The following account is paraphrased.

Charles and Nancy presented a brief update about the project's progress and accomplishment during the past year, and reported on two major developments: 1) the development of a standard metadata set for legislative documents as part of an XML wrapper for the exchange and storage of XML bill files; 2) the testing of eXist, an open source XML native database for access to digital legislative documents. Besides examining practical ways to preserve digital legislative records, a primary goal of the project is to enhance the accessibility and usability of digital legislative information. The presentation by Charles and Nancy, including their notes, can be found on the NDIIPP project website at: <http://www.mnhs.org/ndiipp>

The ISL faces an ongoing concern about the cost of staff time processing electronic documents when aggregating, managing, and making accessible electronic documents at the document level. How can digital records/documents be handled on a system level? There aren't always easy and inexpensive solutions, and everyone recognized this is will continue to be a challenge, especially as limited or decreasing staff face the increase in digital content. The Minnesota NDIIPP project is very concerned about costs and benefits, and any technology model or recommendations the project puts forth will not add extra burdens or costs to state governments.

Regarding the four NDIIPP state government projects (Arizona's PeDALS, North Carolina's GeoMAPP, Washington's multi-state preservation consortium, Minnesota's digital legislative records) there is concern about the creation of separate silos of information, specifically the perceived inability of the four silos to aggregate and share data for future use. At this time it seems there is only coarse metadata the projects have in common, which may not be easily used to aggregate and access the various types of information. More information about the four NDIIPP state government projects is available at:

<http://www.digitalpreservation.gov/partners/states.html>

This led to a discussion about the value of metadata, and the need for crosswalks between different sets of information. Nancy and Charles further discussed the project's development of minimal XML schema for legislative data, and the testing of the eXist technology that will occur later this year. Tim offered to supply additional Illinois legislative data for testing the eXist technology. Illinois legislative XML documents and related metadata are in one database system, which is achieved using a customized script written by the Illinois Legislative Information System staff.

There was discussion emphasizing that the NDIIPP project should continue to advocate the need for standards and best practices in the functions involving the aggregation, management and use of digital information. A primary function of the project is to research standards and best practices about various topics, including authentication of electronic data, retrospective digitization, audio files, access options, government data mashups, and other topics. More information about the project's white papers and resources is available at:

<http://www.mnhs.org/ndiipp>

We discussed an idea for the Library of Congress NDIIPP office to consider: Explore the possibility of a public/private collaboration involving, perhaps, Google or Amazon.com, in which state government would transfer digital data with permanent value to a private company, with the understanding the company would preserve the digital data permanently. The company would provide free access to basic information, but be able to charge a fee for value added offerings like mashups or better retrieval. This led to discussion of cloud computing, ownership of data, legal issues, and costs; everyone recognized there would be a number of issues to address, and each state could only pursue a collaboration like this based on their local needs, laws, political climate, etc.

Summary of priorities or issues the Illinois partners recommend for further study:

1. Ongoing concern about the cost of staff time at the document level when aggregating, managing, and making accessible electronic documents. How can digital records/documents, particularly subject indexing, be handled on a system level?
2. Regarding the four NDIIPP projects, there is concern about the creation of separate silos of information, specifically the perceived inability of the four silos to aggregate and share data for future use.
3. Our NDIIPP project should continue to advocate the need for standards and best practices in the functions involving the aggregation, management and use of digital information.
4. An idea for the Library of Congress NDIIPP office to consider: Explore the possibility of a public/private collaboration involving, perhaps, Google or Amazon.com, in which government could give digital data with permanent value to a private company, with the understanding the company would preserve the digital data permanently, allow free access to basic information, but be able to charge a fee for other services. (Although it was not

discussed, the Library of Congress NDIIPP program is testing the use of cloud technologies to enable perpetual access to digital content.¹⁾

Connie gave a presentation about the ISL's electronic documents depository, which preserves and makes accessible Illinois state government documents published on the web. This function is mandated by Illinois law, but at times is cumbersome and expensive in time and money to maintain. The electronic documents depository provides permanent public access to electronic documents, relies on robust and accurate metadata, and has been analyzed as a Trusted Digital Depository by the Center for Research Libraries² using standards developed in partnership with OCLC. Future challenges include less staff at the ISL and Illinois state agencies, along with less funding and increasing digital content; articulating the depository's business case would be helpful in seeking additional funding to support the depository. Connie added that the NDIIPP PeDALS project being undertaken by Arizona had contacted the ISL about participating in the project, but the ISL did not apply to be a partner because the PeDALS project emphasize the use of the LOCKSS technology, and would not be compatible with the ISL's need to provide access to the materials they preserve.

¹ http://www.digitalpreservation.gov/news/2009/20090723news_article_duracloud.html

² <http://www.crl.edu/content.asp?11=13&12=58&13=142>