

Preserving State Government Digital Information Minnesota Partners Meeting



Minnesota Historical Society

Friday, February 22, 2008
Pillsbury Conference Room, Minnesota History Center
St. Paul, Minnesota

As the result of a recent LNET meeting (Legislative Networking Group, who are information and systems staff from the legislature and who govern legislative Web content), the Minnesota Historical Society hosted a second partner meeting for its NDIIPP-sponsored project, Preserving State Government Digital Information. Nancy Hoffman, Bob Horton, Jennifer Jones, Charles Rodgers, Shawn Rounds, and Christopher Welter represented the Minnesota Historical Society (MHS). Isaac Holmlund (Web developer), Tim Orr, and Michele Timmons represented the Revisor's Office (RO). Robbie LaFleur and Elizabeth Lincoln represented the Legislative Reference Library (LRL).

The following account is paraphrased.

Bob Horton began by summarizing the project's background and by highlighting recent updates, including the project Web site.¹ He was seeking a consensus on how the project can best move forward quickly. The NDIIPP grant money realistically allows for the development of tangible products with 1 or 2 states. Based on well-established working relationships from the previous E-Legislature project,² Minnesota and California are the two likely candidates.

Meanwhile, the MHS project team will work with its other state partners—Illinois, Kansas, Mississippi, Tennessee and Vermont—in these areas: initial meetings to discuss project possibilities, ongoing communication and education, appraisal of existing government records, potential business cases for record preservation (an initial point of interest among the state partners), and benchmarks to allow for fuller participation down the road (gap analysis). The National Conference of State Legislatures (NCSL) will also be consulted in these matters.

Tim Orr: We had an impromptu brainstorming session after a recent LNET meeting. I understand that one talking point from the January 28 meeting was a possible online research tool. But it would be of limited use, as states are saving data in different formats, and technology is changing so rapidly.

What about preservation? Long-term access (i.e., archiving) should be considered very early on in an information technology project, as we learned from the E-legislature project. For

¹ <http://www.mnhs.org/preserve/records/legislativerecords>

² <http://www.mnhs.org/preserve/records/elegislature/elegislature.htm>

example, considering a standard, minimum XML schema for legislative records. As there's no national standard, each state is developing its own. If this project could develop a minimum standard for the 80% of common content across states—bills, statutes, session laws, administrative rules—that would be a tangible and beneficial outcome.

What about West³ as a potential partner? They're already using information from several different states. Are they already using some standard schema, or a type of translator? I suspect that they are.

Bob: Is there anyone specifically we can talk to?

Tim: Mark Bergen [651-687-7000, 610 Opperman Drive, Eagan, MN 55123-1340]. I understand, second hand, that West is revising its XML schema.

Robbie: What about NCSL? Didn't they attempt this?

Michele: Yes, there were two attempts, including just a common database (since outsourced to a vendor). But there were so many differences among state legislatures that NCSL couldn't do it. If we could come up with some sort of standard tag set, then NCSL's project—or anyone else's—is much more likely to be useful.

Tim: Projects with data/records are much easier if standardized. It facilitates work across states.

Bob: How can we better NCSL's attempt? Perhaps partnering with West. Also, by promoting a schema with minimal data sets—something akin to the DublinCore model—but one that's extensible.

Michele: The last major attempt was several years ago. Many states are much better equipped now than back then. We could bring together states with well-developed XML systems (e.g., Minnesota, Kansas, and Illinois) to come up with a common set of tags.

Robbie: It was a premature attempt back then.

Tim: As a new technology, states wanted to adapt XML to their unique situations.

Robbie: It might be more apparent to discern what's common now, as states have made inroads, whereas this was not possible several years ago.

Tim: I've been trying to entice West with this idea because standard data capture would presumably make their work more efficient and cost effective.

Charlie: Could you explain what data you're talking about?

Tim: XML-formatted bills, session laws, statutes, administrative rules. The whole process, from idea to law.

Robbie: What about other important policy documents like legislator backgrounds and studies/reports? For people who do legislative research (e.g., House Research), what's in XML right now isn't the be-all and end-all of content. They're interested in tying in other types of

³ <http://west.thomson.com/overview/>

information. There's interest in connecting to other data sets that illuminate the laws themselves.

Bob: There are 3 possibilities: 1) map it to DublinCore or the Minnesota Record Keeping Metadata Standard to facilitate linkages; 2) pair with a Google-like discovery tool to uncover non-cataloged items; 3) allow for annotation and social tagging. These 3 options could work together to provide access to/use of other important items, and would position us to do more in the future.

Robbie: Reports are already saved as e-docs [*electronic documents*] with a certain level of metadata.

Bob: Yes, but report creators themselves won't add the metadata. LRL and MHS catalog items, but others won't. IDOL, on the other hand, can do a brute force search on a term rather than metadata tags.

Isaac: Data silos currently exist, and we want these silos to have relationships with each other. The big buzzword right now is the Semantic Web. [*According to the W3C, the Semantic Web "provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries. It is a collaborative effort led by W3C with participation from a large number of researchers and industrial partners. It is based on the Resource Description Framework (RDF)."*⁴]. Think of the Web as a database rather than a book [*i.e., dynamic rather than static*].

Bob: So what's our process?

Shawn: 1) Contact West, and perhaps Lexis-Nexus if West doesn't pan out; 2) get state teams together; 3) vet through NCSL to determine commonalities; 4) provide demonstration model using Minnesota documents.

Charlie: Can we find out what didn't work with NCSL, to develop an XML schema for legislative content?

Bob: I think what we've discussed here today has already pinpointed the problems.

Michele: Tim and I have tried but there just isn't much documentation from NCSL. [*In Spring 1999, the National Association of Legislative Information Technology (NALIT) conducted a survey of bill-drafting systems. The end result was a 50-state comparison tool.*⁵]

Tim: I'm thinking of what other states use XML currently—Texas, Kansas, Maryland, Idaho, California, Minnesota, Florida. Who else? Devan Shepherd is the Minnesota consultant. [*Devan Shepherd is the CEO and CTO of XMaLpha Technologies, in Shoreview, Minnesota*]. When XML pilot projects start, often on the state level, the process gets bogged down when it comes time to agree upon specific tags.

Bob: If we can create a demonstration model, it would show a definite value and use.

Tim: RO can offer experience with schemas, and Isaac has XSLT experience [*eXtensible Stylesheet Language Transformations, a language for transforming XML documents into*

⁴ <http://www.w3.org/2001/sw/>

⁵ <http://www.ncsl.org/programs/lis/nalit/billtools.htm>

XHTML documents or to other XML documents]. He could transform documents into schema and develop programs to bring documents together. And LRL could perhaps offer input as to what the interface might look like.

Robbie: The idea is not just another static Web page, but a portal that can make connections among these different XML data sets.

Tim: LRL knows what it wants so it should work with whoever develops the Web site.

Isaac: There are many more ways to utilize XML data, like SOA [*Service-Oriented Architecture*], where people can create their own product rather than relying upon just the book and Web site [*i.e., static*] product we currently offer now.

Robbie: Maybe we focus on technical stuff for the next year, and about a year from now, perhaps talk to academics, non-profits, businesses or others about what they might want at an aggregate level?

Bob: We should also discuss the revenue issues, as Michele has brought up before, then there's the matter of authentication.

Tim: By discussing this with law school faculty, perhaps the schools themselves will be interested in building an interface.

Bob: I've been invited by Temple University to a May meeting [30-31st] on legislative records [*Eighth Annual State Politics and Policy Conference*]⁶. They're looking at grabbing legislative information in batch form.

Elizabeth: LRL has also been contacted about our legislator database [*Minnesota Legislators Past and Present*]⁷. It has proven very popular among genealogists, of course, but other audiences have found their way to it, too, for study purposes, including election studies. Information ought to be served up in a way that it can be sucked up wholesale.

Nancy: Illinois has done this by dumping legislative information into an FTP site so that it can be readily harvested—one method of providing access to data.

Charlie: At the MHS library, we get lawyers and paralegals coming in for legislative intent research.

Robbie: At the LRL, legislative history researchers want context above and beyond just the straight statute or law. They want associated records, documents, audio/video content, information on legislators themselves.

Michele: If we do have a presentation or focus group, we don't want to exclude lawyers. They're a big part of the customer base.

Charlie: What did you mean about revenue and copyright issues that you mentioned earlier?

⁶ <http://www.temple.edu/ipa/events/2008%20State%20Politics%20and%20Policy%20Conference--Conference%20Details.pdf>

⁷ <http://www.leg.state.mn.us/legdb/index.asp>

Michele: Prior to my becoming the Revisor, the RO copyrighted bills, session laws, statutes, and administrative rules, partly because there's been value added by RO staff. We sell copies of the session laws, statutes, and administrative rules, and the revenue goes to the state's general revenue fund. We also sell our databases to West and LexisNexis. Putting information directly online undercuts these current customers. I'm also concerned about authentication, moving from paper to electronic. How can the RO say to the public that what you're seeing online is in fact trustworthy? Revenue is not the most important consideration, but it's not unimportant [*Michele estimates a \$250,000 annual revenue for both hardcopies and databases*].

Bob: To some extent, this is happening already. Some of this is an educative process, adjusting to the new electronic medium. The Association of American Law Libraries (AALL) asserts that an electronic version is inherently untrustworthy, whereas case law and UETA assumes the opposite—that it's trustworthy unless proven otherwise. [*"The Uniform Electronic Transactions Act (UETA) provides a legal framework for electronic transactions. It gives electronic signatures and records the same validity and enforceability as manual signatures and paper-based transactions. This model act was adopted by the National Conference of Commissioners on Uniform State Laws (NCCUSL) in 1999."*⁸]

Michele: Some of this will be discussed by the Uniform Law Commission, which I'm working with. [*The ULC includes the Committee of Online State Legal Materials, which will study authentication of said materials* <http://www.nccusl.org/Update/CommitteeSearchResults.aspx?committee=321>>.]

Isaac: We'll need a strategy to disallow frequent, overwhelming hits by large services/crawlers hitting the data and sucking bandwidth, scraping sites.

Nancy: Illinois said that their FTP site has cut down on Web scraping.

This leads to an aside by Bob, who explains a recent occasion where Ancestry.com purchased copies of MHS microfilm, digitized them, and mounted them on their Web site for resale, including to MHS and its patronage.

Michele: As for the Uniform Law Commission, each state has a different process by which it adds people to the commission. Once laws are approved by the commission, then each state can consider adapting it to its own level.

Charlie: RO gives only certain access to legislative content? Are the drafts in XTEND non-public?

Tim: Yes, files are assigned different statuses, separating public data from non-public data.

Bob: In summation, there are 3 tracks: 1) development of minimum standard XML schema for bill-drafting system; 2) development of access pilot for XML documents; 3) preservation and distribution of batches/aggregation, including examination of authentication, revenue, and copyright issues.

Next steps:

- 1) Michele Timmons will contact Ellen Gillespie at West.

⁸ <http://www.ncsl.org/programs/lis/CIP/ueta.htm>

- 2) Choose which states with whom to work (Chris will compile which states are using XML bill drafting systems currently).
- 3) Bob will coordinate all this with NCSL.

Michele: Are we going to move ahead simultaneously with content capture [*as mentioned in the January meeting*]?

Robbie: If we want to discuss what information silos we want to link, then we need a brainstorming session with legislators and staff to identify what is the totality of “perfect” legislative research.

Bob: Let’s identify a group and use the LRL’s legislative history report as a starting point. [*In January 2008, a LRL team issued a 2-page series of recommendations for improving legislative history research, many of which address Web navigation and usability of online legislative content.*]

Michele: Let’s start with LNET and others, like Peter Wattson [*Senate Counsel & Research*].

Robbie: We can help with setting up a session.

Michele: Bob, are you envisioning we can do all these suggestions, or might we have to prioritize down the line?

Bob: The Library of Congress wants content captured, so that’s the project’s first priority. The XML schema has potential if we look at it as incremental steps. I think we can do the minimal amount that we’ve been talking about because we can’t mandate this, but only pass it off to NCSL or someone else. Fundamentally, this project offers guidelines, but other states will have to prove the guidelines’ worth.

Michele: As I have mentioned before, I have concerns about staff time/cost, given the several ideas floated by this project.

Elizabeth: Has anyone looked at West and Lexis to compare/contrast how and what they’re sucking up, data-wise?

Robbie: They must have figured out what they wanted or needed.

As the meeting concluded, Bob thanked all present for their time and ideas.