

The Canadian Information Network for Research in the Social Sciences and Humanities
<http://www.synergiescanada.org>

Mary Westell
Libraries and Cultural Resources
University of Calgary
Netspeed Conference, October 23, 2008

Speaker Notes

1. I thought it might be interesting to give you an update on the Synergies project today for a two reasons:

First, Synergies is an innovative project that will advance the ability to make Canadian scholarly publications available worldwide. The project will use Open Source tools and a distributed approach to developing an infrastructure that will sustain a national network. Second, I thought it might be interesting for Netspeed participants to hear about a project that has a national scope – a bit different than some of the CFI projects, which focus on high performance computing, nano technology, specific subject oriented research and are institution or region specific.

Synergies is one of three national CFI projects that deal with the development of a national research network using other tools like Canarie to deliver scholarly content to researchers. The other projects are the Canadian Research Knowledge Network (CRKN) which supports licensing of digital content for libraries across the country and the national network of Research Data Centres, which is developing a strategy for storing and retrieving statistical data using international metadata standards. All these three projects fit together to solve the puzzle of sharing research information online.

While they originated with libraries, the projects depend heavily on the IT organizations in their universities and I would like to express my thanks for our own IT folks who have been very supportive as we plan and carry out this project.

2. This quote is from an excellent article by Rea Devakos and Karen Turko in the Association of Research Libraries Bulletin (June/August 2007) which describes the Synergies project.

As many of you know, the world of scholarly communications has been turned on its head because of the potential of electronic publishing. Many journals are moving to an Open Access model, where authors retain copyright of their articles and journal issues are freely available online.

Journals, especially those of research groups or scholarly organizations don't always have the means to move ahead. All of you are familiar with journals that create local web sites that aren't backed up regularly and can be abandoned or lost as editors change.

The Synergies project responds to these changes by providing an infrastructure that will assist SSH journals in particular, to move into the online world.

3. Synergies was five years in the making, led by a variety of researchers across the country who were on the cutting edge of electronic publishing and the open access movement.

As most of you know, a CFI project requires specific outcomes that address CFI criteria. As an infrastructure project, Synergies needs not only the technical champions, but

researchers that are prominent at the national and international level. Provision of the content, while critical to the success of the project, cannot stand on its own.

Without the multi-level support, this project would not have been approved. Alberta is particularly supportive of these projects and we were delighted to obtain our funding in March of this year.

4. Granting agencies increasingly are requiring that grant holders publish the results of publicly funded research in open repositories. Synergies will provide the means for integrating a wide range of digital formats into a single portal. Some of you may be familiar with DSpace, open source software developed at MIT and used widely for Institutional Repositories. Synergies will be designed to interoperate with DSpace and other institutional repository software.

This directly addressed the concerns and goals of organizations like SSHRC and CIHR (Canadian Institutes of Health Research). In fact, this year, CHIR has approved a policy requiring CIHR-funded researchers to deposit their work in Open Access repositories within 6 months of publication. SSHRC also has a requirement for the archiving of research data acquired through SSHRC funding. Until now, the ability of universities and libraries in particular to help researchers meet these criteria was limited.

5. Print models have established standards for preservation – legal deposit for permanent preservation and the ISBN/ISSN for permanent identification.

Digital objects lack this infrastructure, so Synergies is working with other agencies like Libraries and Archives Canada and the Digital Library Federation to accomplish it.

The Open Archival Information System (OAIS) reference model is one attempt, as are the definitions of trusted digital repositories developed by the centre for research libraries and the Digital Library Federation.

Persistent identifiers like the DOI (Digital Object Identifier) attach to digital objects and maintain a stable, sustainable location for long term access.

PREMIS, the North American standard for preservation metadata is another piece of the puzzle

More on preservation in a minute.

I would add, though, that Canada does not have a framework for trusted digital repositories, so Synergies is filling a real need for us here.

6. These are some of the Canadian university libraries involved in electronic publishing and digital libraries.

UNB has its Electronic Text Centre, a leader for a decade in providing digitization services along with use of text processing software for research.

UdeM is the architect of the Erudit system, which is the inspiration for Synergies. Erudit contains over 45 scholarly journals and is one of the suppliers of content to the CRKN.

University of Toronto working with the Ontario Scholars Portal has a strong track record of providing portal access and indexing of journal content. Scholars at Toronto bring their experience in the national TAPOR (Text Analysis Portal for Research) The Text Analysis Portal for Research (TAPoR) is a collaboration of research units at six Canadian universities, to build a centralized gateway to representative electronic texts and text analysis tools.

Simon Fraser, with its partner University of British Columbia, is the home of the Public Knowledge Project, whose OJS software is the platform for the regional journal nodes.

At Calgary, we will contribute our experience with large scale digitization projects and expertise in the development of preservation architecture.

All of the participants have active and well populated institutional repositories. In addition to these partners, 16 more institutions form the Synergies core, spanning the country will round out the consortium.

6. Synergies aims to cover the lifespan of scholarly work with tools for the researcher, the author, the publisher, and the librarian.

7. Here is just some brief information on our specific project in the Prairie node -

8. Adoption and use of the Public Knowledge Project software suite, developed at University of British Columbia and run for Synergies at Simon Fraser University, has continued to grow significantly during the first half of 2008. At least 1,500 online journals use OJS as a publishing platform. About 150 organizations use OCS to manage conferences online. Several consortia and groups are using the PKP metadata harvester.

Erudit, developed through a partnership of Quebec universities and hosted at the University of Montreal has just announced its new version of web platform software.

Erudit offers digital publishing services to journals and sophisticated access to over 30,000 academic articles and reviews along with a wide range of other document types.

Both of these projects have been developed using Open Source software licensing and are used by a broad and international user community.

They are relatively easy to install and support and the Open Source user community has been busy developing numerous plug-ins to integrate modules together.

9. Services for journals

- Provision of the infrastructure to move to online publication
- Developing tools to foster sustainable long term business models
- Researcher programmable interfaces
- Text and graphical results displays

10. These are some of the Canadian university libraries involved in electronic publishing and digital libraries.

UNB has its Electronic Text Centre, a leader for a decade in providing digitization services along with use of text processing software for research.

UdeM is the architect of the Erudit system, which is the inspiration for Synergies. Erudit contains over 45 scholarly journals and is one of the suppliers of content to the CRKN.

University of Toronto working with the Ontario Scholars Portal has a strong track record of providing portal access and indexing of journal content. Scholars at Toronto bring their experience in the national TAPOR (Text Analysis Portal for Research) The Text Analysis Portal for Research (TAPoR) is a collaboration of research units at six Canadian universities, to build a centralized gateway to representative electronic texts and text analysis tools.

Simon Fraser, with its partner University of British Columbia, is the home of the Public Knowledge Project, whose OJS software is the platform for the regional journal nodes.

At Calgary, we will contribute our experience with large scale digitization projects and expertise in the development of preservation architecture.

All of the participants have active and well populated institutional repositories. In addition to these partners, 16 more institutions form the Synergies core, spanning the country will round out the consortium.

10. Persistent Identification

- Excellent report from LAC:
- *“Persistent Identification of Digital Resources: Environmental Scan”*, February 2008
- No acknowledged standard in Canada yet
- Persistence is achieved through
- An identification scheme for digital documents
- Services for resolution for URLs
- Policy related infrastructure that coordinates name spaces and identifiers (Knowledge Base)

11. Article Linking (URL Resolution)

- Must:
- Use the Synergies Persistent Identifier
- Enabling links between any type of document to the full text of its bibliographical references
- Use the Identifier to bring together all indexed documents that cite a document
- Should:
- Ability for regional nodes to view and update knowledge base
- Simple text searching of the knowledge base using a citation

12. *“Digital preservation combines policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change. The goal of digital preservation is the accurate rendering of authenticated content over time.”*

13. Libraries are wrestling with this issue as we see increasing amounts of digital information being produced daily. This is especially problematic when we deal with transitory materials (for example some learning objects or web sites) which cannot be re-used or items produced with old technology. In his book Understanding Digital Libraries, Michael Lesk tells a story about the British Library, who created a contemporary digital version of the Domesday Book, only to find that its technology was obsolete before the project was finished and that the process had not ensured copyright clearance for migrating the format to newer technology.

Developing policies for digital collections is another area where we need input from our IT colleagues – both on the watch for new technologies and the strategies for keeping multiple generations of documents. Also, IT has the latest and best equipment and strategies for backing up files!

14. Again, preservation initiatives abound

LOCKSS is a set of redundant servers that can provide both accessible and “dark” archiving. With a minimum of seven identical partners, content is stored to offset potential catastrophic loss. Currently the Council of Prairie and Pacific University Libraries is doing a LOCKSS pilot project. I understand that the Ontario Scholars Portal is also investigating LOCKSS. LOCKSS is a distributed technology.

Portico is a centralized non profit “dark” archive to which libraries can subscribe.

Content is housed and brought back should a vendor go out of business or the content be otherwise lost. The Portico web site gives a sense of the scope of these initiatives: 4,103 preserved titles 7,180,979 preserved articles, 56 participating publishers, 468 participating libraries.

PREMIS metadata, supported by OCLC and the Research Libraries Group (RLG) mentioned earlier is a granular metadata standard which uses a detailed structure of identifiers and objects, with information down to the bitstream level.

Premis objectives are :

Develop a core preservation metadata set, supported by a data dictionary, with broad applicability across the digital preservation community.

Identify and evaluate alternative strategies for encoding, storing, and managing preservation metadata in digital preservation systems.

DOI technology

15. Prairie deliverables

- Develop a central resource for creators and institutions to support content creation
- Develop practices for building durable content through packaging standards like METS and metadata standards like PREMIS
- Create a network of persistent identifier servers to provide long term access handles
- Develop tools to build durable objects, capturing context and interrelationships
- Important to note that the architect of the prairie plan is Tim Au Yeung. Along with Tony Tin from Athabasca, Tim is a member of the Synergies national technical committee. That group is making tremendous strides in developing the strategy and means to achieve the Synergies goals.

16. Technology Watch System

- Will conduct migration risk analysis for formats where no risk analysis has been done and disseminate/keep a current list of risk analyses that have been done
- Will assist individual journals and institutions to conduct format, standard, and platform audits to ensure that collections are properly monitored
- Will develop a business model for long term sustaining of the technology watch system

17. Challenges

- Attracting journals and expanding our publishing program
- Working with University of Calgary Press and the publishing programs at partner institutions
- Campus Scholarly Journals Interest Group providing feedback
- Developing policies for digitization of back issues . . .
- Developing a research program around Synergies
- Working with faculty members to develop plans
- Integrating with other research proposals
- Meeting CFI guidelines for advancement of Canadian research, training of HQP, etc.
- Sustainability
- Technical support
- Software licensing issues
- Apache –vs- Open Source
- Governance
- Building the business model

18. Results

- More Canadian SSH journals online
- Variety of other scholarly output (data files,

- theses, monographs, conference
- proceedings) online
- Collaborative national technical infrastructure
- New open source software
- New collaborations within and among institutions
- Digital preservation strategy