

2012-04-20

# Scholars Portal as a Trustworthy Digital Repository

Marks, Steve

---

Marks, S. "Scholars Portal as a Trustworthy Digital Repository". Presented at Digitization and Digital Preservation: Strategies for COPPUL Libraries, Friday April 20, 2012, University of Calgary Taylor Family Digital Library.

<http://hdl.handle.net/1880/49064>

*Downloaded from PRISM Repository, University of Calgary*

# Scholars Portal as a Trustworthy Digital Repository

Steve Marks  
Digital Preservation Librarian  
Scholars Portal, Ontario Council of University Libraries

# Who are we?



# What is Scholars Portal?

- A shared IT service provider to the 21 university libraries in Ontario
- Provide services like RefWorks, RACER, Virtual Reference
- Serve as a platform for housing consortially licensed resources

# Some History

Scholars Portal was formed back in 2002 to serve as a local load point for consortially licensed materials.

Dual purposes:

- Economies of scale in making material available
- Preservation

# Fast Forward 5 Years

The project was pretty successful! All of a sudden, we realized that we had collected a whole lot of stuff owned by the 21 member libraries.

Approaches: maybe benign neglect isn't sufficient

SP started talking to experts, including CRL

# The Story Continues

OCUL Directors decided that formal certification could be a good thing. Decided to pursue the Trustworthy Digital Repository certification.

I was brought on board in September 2010 to make it happen.

# So Where did we start out?

- Very little in the way of policy
- Lots of good practice
  - Hadn't been evaluated through a long-term preservation lens
- Missing a few essential pieces



# What does our platform look like?

- MarkLogic – NoSQL XML data store
  - Same one used internally by Elsevier, Springer, others.
- Custom Java code
- Front end using XQuery

# Some Stats

- ~22 million full-text articles
  - PDF, XML
  - Supplementary materials
- 30TB of material
- High rate of ingest.
  - ~3,000 articles per day
- High use
  - 2011: 6M full-text article downloads

# TRAC/TDR/ISO 16363 Pieces

- Organizational Infrastructure
- Digital Object Management
- Infrastructure and Security Risk Management

# Self-Evaluation – Organizational Infrastructure

- Good
  - Contracts pretty solid
  - Well-defined, knowledgeable Designated Community
  - Clear Mandate
- Not as good
  - No mission statement
  - Contracts not consistent, need some work
  - Succession Planning

# Self-Evaluation – Digital Object Management

- Good
  - Good document handling practice
    - Error handling
    - Logging
    - Not overly messing with files
- Not as good
  - Not tracking some really important metadata:
    - Formats
    - Fixity values
    - Provenance

# Self-Evaluation – Infrastructure & Security Risk Management

- Good
  - Lots.
  - Robust security
  - Quality hardware and support.
- Not as good
  - Well, you can always do more.
  - Lack of risk assessment
  - Lack of formal disaster recovery plan

# Addressing the gaps – Organizational Infrastructure

- Licenses
  - Thorough review of all OCUL licenses for preservation rights
  - Creation of OCUL Model License
- Succession Planning
  - Rewrite of OCUL-UTL agreement
  - Canada could use some help
- Mission Statement
  - Nope!

# Addressing the gaps-- Digital Object Management

- We needed to track a lot of new metadata!
  - PREMIS is a natural fit for this.
  - Generate using custom code and FITS
- And we needed a way to structure it.
  - METS is hard!
  - We made a “METS-lite” standard, spPres
    - Cross-walkable to METS



# Digital Object Management, cont'd

- Place to store preservation metadata
  - Parallel database, linked by URI
- Workflows to support its creation
  - Had to stay out of the way of loading.
  - Oh, and 20 million articles worth of backfiles.

# Addressing the gaps – Infrastructure and Risk Management

- Extensive risk analysis
  - Developed a typology of risk by looking at overlap of a number of different organizations' risk planning
  - Plotted risk along two axes: likelihood and severity
- Developed new, comprehensive DR plan together with IT staff at UTL.

# Snapshot now

- All policy work is done and approved by OCUL directors.
- New preservation workflows are generating PMD and we are catching up with our backfile.
  - Currently about 3M preserved articles
    - ~10M files
  - Shooting for end of year to be caught up.

# Was it worth it?

- I think so.
- Lessons learned
  - Self-audit is great, but there's nothing like opening up the doors and inviting people in.
  - Local, large-scale approaches to digital preservation are possible!

# Our Concerns

- It was a pretty lonely process.
  - There needs to be a community to talk about this process.
- We need to own the process
  - Digital preservation practice needs to become accessible and pass into the general librarian skill set.

# We want to help!

- All\* of our documentation is freely available at:  
<http://spotdocs.scholarsportal.info/display/OAIS/>
- Policy and procedure documentation
- ISO16363 criteria responses
- Mappings of which documents answer which criteria.

\*almost all

# Thanks!

[steve@scholarsportal.info](mailto:steve@scholarsportal.info)

<http://spotdocs.scholarsportal.info/display/OAIS/>