Transformation from Digital Libraries to Digital Repositories: New Requirements for Digital Scholarship Services in Libraries

Jian Qin
School of Information Studies
Syracuse University

Advanced Digital Library Seminar, Fuzhou, China, 2017-12-04
Transformation from *digital library* to *digital repository*

- **Digital Library**: Collection of information resources

- **Institutional and data Repository**: Detailed documentation, formal publication and citation, and long term preservation for indefinite reuse

- **Digital Repository**: Specialized repositories for data, people, publications, objects, or things
The transformation is focused on

**Building blocks of knowledge**

Data, software, algorithms, visualizations, and other outputs of the research process

“…research data have arrived as first class objects of scholarly communication…”

(Clement & Schiff, 2015)
Current status: data repositories

Domain specific repositories:
- Adoption of a metadata standard for dataset description
- Detailed documentation for datasets
- Formal publication and citation of datasets
- Publications linked to datasets
- Long term preservation for indefinite reuse of datasets
Current status: Researcher Information repositories

Focus on researchers and their identity management
  ◦ Global ID for researchers
  ◦ Metadata for individuals and institutions
  ◦ Focus on improving visibility and impact of research
  ◦ Open, crowdsourced, integrated researcher information management
Current status: Code repositories

- May be open source or fee-based
- For multi-developer projects, single developers or private teams
- Support various services and tools,
  - version control
  - bug tracking
  - release management
  - mailing lists
  - wiki-based documentation
- Developers (for open source) nearly always maintain their copyright when their software is posted to the open source software hosting facilities
The larger context of research

“data intensive, multidisciplinary, collaborative, and frequently problem-oriented”

(NSF, 2011)
NSF’S BIG IDEAS

Harnessing Data for 21st Century Science and Engineering

Navigating the New Arctic

Windows on the Universe: The Era of Multi-Messenger Astrophysics

The Future of Work at the Human-Technology Frontier

The Quantum Leap: Leading the Next Quantum Revolution

Understanding the Rules of Life: Predicting Phenotype
Mid-scale Research Infrastructure

NSF 2026: Seeding Innovation

Growing Convergent Research at NSF

NSF INCLUDES (Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science): Enhancing STEM through Diversity and Inclusion
<table>
<thead>
<tr>
<th>SBE 2020 PRIORITY AREA</th>
<th>SBE COMMUNITY GOALS</th>
<th>SBE DECADAL OBJECTIVES</th>
<th>SBE ACTIVITIES (1-3 YEARS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content of science</td>
<td>Examine new approaches and pursue fundamental questions</td>
<td>Establish priorities in existing programs Consider new programs, crosscutting activities within the directorate and across the Foundation and other agencies</td>
<td>Convene workshops and expert groups to plan for next generation science in specific areas Enhance mechanisms for interdisciplinary proposal review and budgets for interdisciplinary research</td>
</tr>
<tr>
<td>Capacity building</td>
<td>Build collaborations</td>
<td>Ways to enable team building and cross-disciplinary research</td>
<td>Improve SBE’s investment in interdisciplinary training,</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>Access data, services, and program</td>
<td>Invest in data and data access; centers, enclaves, and trials</td>
<td>Build community plans for new data activities</td>
</tr>
</tbody>
</table>

Important message to academic libraries
New generation of digital repositories

- A new large scale
- Might include traditional relational data in digital form, collections of historical data, images, video recordings, administrative data/records or any other form of structured sets of data
- A usable interface that allows for the application of extant analytic tools or analytical tools

Required to support “data intensive, multidisciplinary, collaborative, and frequently problem-oriented” research

(NSF, 2017)
Bring all digital repositories together

**Research information management (RIM)** is the aggregation, curation, and utilization of information about research and is emerging as an area of increasing interest and relevance in many university libraries.

(Bryant et al., 2017)
How research information is used?

*Aggregated metadata*

*Aggregated information*

(Bryant et al., 2017)
Research information management as an emerging area in supporting digital scholarship

<table>
<thead>
<tr>
<th>TRENDS IN DIGITAL SCHOLARSHIP</th>
<th>RESPONSE FROM LIBRARIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data intensive</td>
<td>• Collect and reuse information</td>
</tr>
<tr>
<td>Multidisciplinary</td>
<td>• Showcase academic achievements</td>
</tr>
<tr>
<td>Collaborative</td>
<td>• Support Open Access initiatives</td>
</tr>
<tr>
<td>Problem-oriented</td>
<td>• Report and analyze activities</td>
</tr>
<tr>
<td></td>
<td>• Capture the evidence of Impact</td>
</tr>
<tr>
<td></td>
<td>• Simplify faculty activity reporting and assessments</td>
</tr>
</tbody>
</table>
Publications and Scholarship Expertise
- Offer help and advice on metadata
- Advise on metadata harvesting and licensing content from different sources
- Disambiguate and manage duplicate bib records
- Persistent ID management
- Standards, vocabularies, and data dictionaries

Discoverability, Access and Reputational Support
- A public discovery portal for institutional experts
- Open access to scholarly content
- Make research outputs widely available, citable, and impactful
- Provide traditional bibliometric metrics as well as altmetrics

Training and Support for End Users
- Train researchers
  - to import publications data from publication indexes and reference management tools,
  - to link their profiles with unique identifiers like ORCID, and
  - to use and understand impact indicators.
- Online guides, workshops, and individual support on scholarly communication topics

Stewardship of the Institutional Records
- Preserve and make institutional records discoverable
- Promote integration of research information management data in other systems
- Encourage broader access, availability, and reuse

Libraries in research information management
(Bryant et al., 2017)
Help researchers to:

- Collect and reuse information
- Showcase academic achievements
- Support Open Access initiatives
- Report and analyze activities
- Capture the evidence of Impact
- Simplify faculty activity reporting and assessments

Experts@Syracuse portal

Welcome to Experts@Syracuse

- Profiles: 645
- Research Units: 56
- Grants: 1564
- Research Output: 25490
- Activities: 16
- Press and Media: 6361
- More: 12
Changes brought by the transformation:

- **Content types and structures**: digital objects, digital materials, digital data, etc.
- **Emphasis on aggregation**: integration, fusion, or aggregation of data and metadata from multiple sources
- **Emphasis on interoperability and linking**: using data science methods to clean and de-duplicate data, identifier systems for disambiguation and linking
- **Management**: training to fill skill gaps, organizational change to adapt to project-oriented workflows, planning and project management
Changes in academic libraries: Cornell University Library

2012

Digital Scholarship and Preservation Services

- Conservation Services
- Digital Consulting and Production Services
- Scholarly communication services
- Preservation & archiving policies, workflows, business models.
- Web design
- Liaison to CIT & Center for Teaching Excellence
Final remarks

The new wave of digital repositories is happening now (which means not something predicted for future)

◦ How should we respond (which is reactive) to the new wave?
◦ How can we change from being reactive to proactive for future (so that we can sustain our services and innovation)?

技术不是万能的；没有技术是万万不能的
最困难的不是技术，而是相应的管理制度、灵活的组织结构、领导者的远见、以及团队的知识和技能


Questions and comments?