Knowledge Domain

Paul Courant
University Librarian and Dean of Libraries
Harold T. Shapiro Professor of Public Policy
Arthur F. Thurnau Professor, Department of Economics and School of Information

June 3, 2011
Knowledge Domain Description

• Enable accessible, preserved and reusable knowledge

• The knowledge domain contains all plausibly catalogued items
  • Specimen collections in the herbarium
  • Millions of fish, feathers and furs in our various museums
  • Museum holdings
  • Library holdings
  • Audio, video, objects of various kinds
  • Data (research data we want to use, share and reuse)

• Questions of cataloguing and holding.
  • Findable, usable, reliably reusable
  • Rules for coming in, rules for exiting
Two Distinct Areas of Focus

Libraries, Archives and Museums

Research Data
Knowledge Domain

LIBRARIES, ARCHIVES AND MUSEUMS
Libraries, Archives and Museums Participants

Fran Blouin*
  • Bentley Historical Library
Karen Sikkenga
  • Botanical Gardens
Corey Seeman
  • Business Library
Kevin Graffagnino*
Shneen Coldiron
  • Clements Library
Tim Richards
  • Dearborn Library
Elaine Didier*
  • Ford Library
Sharon Herbert*
  • Kelsey Museum
Margaret Leary
  • Law Library

Rick Francis
  • LSA Dean’s Office
Carla Sinopoli*
  • Museum of Anthropology
Amy Harris*
  • Ruthven Museum
Joe Rosa*
  • UM Museum of Art
Ken Fischer*
John Kennard,
John Peckham,
  • University Musical Society
Bryan Skib,
Becky Dunkle
Maria Bonn
John Wilkin
Paul Courant*
  • University Library

*Members of Public Goods Council
Four primary focuses

• Teaching and Learning Support
  • Direct curriculum support provided by libraries and academic museums.
  • Librarians work with specific colleges, departments, faculty members
  • Librarians part of Ross MBA MAP teams

• Research Support
  • Support U-M faculty and graduate students research needs
  • Non U-M research encouraged and supported
  • Provide some mechanisms for cost sharing requirements

• Enrichment/Outreach
  • The Botanical Gardens, Museum of Art, Natural History Museum and University Musical Society have substantial public enrichment and outreach function
  • K-12 outreach programs

• Preservation
  • Fundamental mission of most libraries, museums, archives
• **Publishing shift from paper to electronic**
  - Current copyright laws assume physical asset and are not able to accommodate the new realities of electronic publishing
  - In aggregate, U.S. academic libraries spend over a billion dollars each year to make information available — but even then access is limited by the scope of the licenses involved
  - Long-standing ecosystem of information in scholarly publishing
  - Today’s technologies allow for free and open access at essentially zero marginal cost but the legal arrangements generally do not permit this
Libraries, Archives and Museums **Strengths**

- Leader in digital library space (HathiTrust)
- Rich portfolio of collections
- Significant institutional financial support
  - Including Art Museum expansion, large display museum renovation and expanded museum storage facility
- Meeting changing needs of research/faculty staff/students in a robust way
  - Closely aligned with teaching, learning and research missions of the university, schools and campuses
- Reliable, easy-to-use, authenticated access to cultural/scholarly record
- Deeply embedded in University fabric
- Nationally recognized academic libraries and museums
- Strong service orientation
- Outreach activities to K-12, community, and other scholars
- Free services
- Positive recruiting impact from cultural and academic assets
- Heavily used by students
Libraries, Archives and Museums Weaknesses

- Uneven collection management
  - Units use various methods and systems to manage their collections
- Many objects not cataloged
- Insufficient funding for digitization of collections
- Undervalued free services
- Under-marketed services
- Limited and fragmented IT support across public good organizations
- Sub-optimal collaboration
  - Don’t know who’s doing what
- Inadequate ticketing system for UMS
  - New system not fully implemented because of PCI compliance and technical issues
Libraries, Archives and Museums Opportunities

- Leverage facilities
  - Library space conversion and performance venue rentals
- Remove physical constraints to access library and museum collections
  - May not apply to Art Museum
- Teaching and learning expansion
- Target marketing
- Expansion of current publishing capability
- Shared collaboration
- Making sense of fragmented information
- Create the scholarship and teaching of the digital age
- Resource sharing
  - Very broad: staff, systems, services
- Increase user support
- Expand fund raising to cover collection digitization

“Transform the landscape of scholarship”
  - New methods and new ways of using collections are emerging
Libraries, Archives and Museums Threats

- General locking of IP
- Unstable funding model for publishing
- Fractured marketplace for information
  - Scholarly prices are increasing
- Current publishing system works quite well for most faculty at Michigan and other elites
  - The advantages of open access to much publishing are therefore hard to articulate. This is costly, but the costs are diffuse
- Lack of space
  - Collections are outgrowing available space
- Reorganization of IT on campus
- Shifts in national / legal / scholarly publication environment
- Information fragmentation
- Library as destination
  - Why come to library?
- Pressure on outreach programs from poor economy and decreasing funds for education
The future vision for the U-M libraries, archives and museums is to make all collections accessible and available to anyone irrespective of location or time.

Ronald Reagan famously invoked the image of a —shining city upon a hill for our country, as did John F. Kennedy years before him. In terms of research, we have sought this ideal for the world. We now have the opportunity to assure that the doors to that city’s library and museum are always open, for the benefit of all. We will not and cannot know which visit to that library or museum, which hit on a database, which download from a repository, or which subsequent published work will solve a pressing social, medical, or other problem, so we must first set our sights on accessibility, knowing that use and progress will follow.
• **Provide reliable, easy-to-use and authenticated access to and use of the cultural and scholarly record.**
  • Increase the quantity of electronic “collections”
  • Downsize physical collections in libraries
  • Ensure compliance with law (change the law?)
  • Effectively manage collections
  • Increase sharing with other institutions
  • Optimize, design and deliver content to meet scholars’ needs
Libraries, Archives and Museums Priorities

• **Expand use of shared spaces**
  - Libraries and museums sit in the center of each campus and have always been a place you can walk to to get access to things you can’t individually afford to buy
  - Increase use of library labs and museum displays in teaching and learning.
  - Become locus of incubation expertise with respect to the development and deployment of learning technology
  - Provide access to scholarly gadgets, e.g. 3-D labs and 3-D printers
  - Increase student collaboration spaces
Libraries, Archives and Museums Priorities

• Become publisher of scholarly content
  • Transfer funds from purchasing scholarly content to subsidizing the scholarly publications of our own faculty
  • Provide advice, a suite of nonexclusive publication venues on the Web, and modest portable subventions for faculty to use in getting published where ever they choose.
  • Work to change copyright laws and establish new business model for publishing
Knowledge Domain

RESEARCH DATA
Research Data Blue Ribbon Panel Participants

- H. V. Jagadish, Chair, EECS, database systems
- Ken Pienta, Medicine, Health system information inventory
- Drago Radev, SI, text mining
- Margaret Hedstrom, SI, long-term access
- Richard Rood, AOSS, climate research data systems
- Elaine Brock, DRDA, data policy
- Shawn McKee, Physics, big science data needs
- Paul Millis, University Audits, compliance
- Paul Killey, Operational data systems
- John Wilkin, Digital library systems
- Frank Manion, CIO Cancer Cntr., clinical research information systems
- Margaret (Maggie) Levenstein, ISR, Bus. Ad., social science data
- Mary Simoni, Music, performing arts
- Chris Miller, Astronomy, data mining
- Ron Barbret, Financial Operations, funding data systems
- Joanna Mirecki-Millunchick, Material Sci., engineering research data
- Heather Carlson, Pharmacy & Chemistry, database mining
- Cindy Wells, ITS, Sr. Staff Support
Research Data Culture

- Wide range of research activities at UM with corresponding spectrum of data needs.

- Large unmet need to support research data management
  - Researchers are “on their own” to find solutions for their data needs

- Metadata is difficult to create and maintain
  - Ontologies and standards are different across disciplines

- Privacy and regulatory requirements limit how some data can be used and merged with other data

- Highly variable norms about data curation and sharing
Data storage needs are growing exponentially
- DNA Core generates approximately 0.6TB daily and currently need 300TB per year for long term storage of data that should be stored forever
- LHC collider at CERN generating PBs of data for analysis in High Energy Physics

Data requires not just storage, but moving it for use with necessary computational resources, analytical tools, consistent with access restrictions
• 81% of research groups use less than 2TB
• Only 4% use about 10TB annually
• 40% of researchers share their data with outside of U-M
• Nearly half of PIs prefer affordable and centralized back-up services over increased centralized or localized storage
• For nearly half of PIs, it is either required or recommended by their funding agency to provide continued access to their data after the grant has ended
Research Data External Scan

- Peer universities are providing basic data services for their researchers above what U-M is providing
  - Indiana provides unlimited data storage for researchers
  - Northwestern provides 500 MB for each researcher and graduate student

- Federal research funding agencies have begun to require data sharing plans
  - NIH is vigorously starting to enforce in the last year
  - NSF began requiring as part of grant applications in January 2011
Research Data Strengths

- ICSPR has large multi-year data sets that are leveraged across the world for research
- DRDA has service to UM researchers to assist in meeting requirements for data management and sharing plans
- Deep Blue public archiving system
Research Data Weaknesses

• Limited central data storage as part of public good – 10 MB

• Highly-fragmented or non-existent services

• No public good HIPAA compliant storage solution available for researchers dealing with Personal Health Information (PHI)

• Most research data is stored locally by researcher, either on laptop/desktop or on servers under their desks, including data with high security or regulatory requirements

• Deep Blue is primarily geared towards research publications rather than research data

• Difficult to de-identify data – no de-identified data set available for health care researchers
Research Data Opportunities

- Leapfrog peer institutions through investment in infrastructure to support research data
- Reduce research project startup time
- Increase dissemination of research findings throughout the world and increase the prestige of the university and U-M researchers
- Establish Architecture and Data Governance using framework such as TOGAF or similar industry framework
Research Data Threats

• Increased competition for faculty and graduate students by peer universities who provide better research data infrastructure

• Aggressive enforcement of data sharing plans by federal funding agencies

• Loss of research funding because of lack of research data services for U-M researchers

• Regulatory fines or actions because of non-compliance

• Security breaches through theft of computer equipment or penetration into unsecured/unpatched locally managed devices
The cost of preservation, access, and dissemination of research articles would be modest as a portion of the overall cost of the research itself and would be provided as a common good by U-M.

U-M provides a specified level of data management services to each funded research project without charge to the project or the researchers involved. This is to accelerate research by removing the barriers that research teams face, allowing them to quickly focus on the research rather than on creating a data management infrastructure for each research project.

In addition to multiple data management services, expert consulting also needs to be provided to guide researchers.
Research Data Core Public Good Priorities

• **Data Storage Services**
  • Provide up to 1 TB per year per funded research project in public good
  • Include tools/facility/encouragement for researchers to define and create appropriate metadata for their data
  • Provide a few additional “flavors” of data management services to meet widely needed regulatory requirements, e.g. HIPAA.

• **Compliance Service**
  • Expand current DRDA offering
  • Focal point to educate users about other service offerings and support them in developing a strong data management plan rather than one that is minimally compliant.

• **Expert Consultation**
  • Security and compliance
  • Use of data management interfaces
  • Metadata recording and preparation of data for archiving
  • Choice of data representation
  • Other task, such as data recording or data analysis
Research Data Core Public Good Priorities

• Metadata Recording Services
  • Any storage service for research data should include some sort of free tools/facility/encouragement for researchers to define and create appropriate metadata concerning their data.
  • Do not set up any special indexing system for its archive of research data. Rather, follow lead from inter-institutional collaborations and emerging best practices, some of which maybe discipline-specific.

• Data Publication Services
  • Provide support for publishing data sets on the web.
  • Combine web hosting service with the data storage service to provide a unified service with natural functionality.

• Data Search Service
  • Provide minimum data search service as an extension to the data publication service.
Research Data Toll Good Priorities

- **Software Licensing Services**
  - For situations in which there is a desire to license or sell rights to some data, or digital creation.
  - Create digital licensing service that would allow a small fee to be charged per access to some digital artifact.

- **Data recording service**
  - Expert consulting is likely to be valuable, particularly in metadata capture.
  - Even though much of this consulting may be customized to the specific project, since metadata capture benefits the research enterprise more than the individual researcher, we recommend that the recharge for this expert consulting exclude the fraction related to metadata capture and preparation for archiving.
Research Data Toll Good Priorities

• **Data Analysis Service**
  • Provide a set of “lower-level” tools supporting common tasks can be offered as a data analysis creation service; including methods for data selection, data access and data visualization
  • Extend CSCAR facilities and work with the data storage service

• **Data with Computation Service**
  • Extend FLUX to include some sort of “data with computation” plan
  • May require increased network capacity and high speed storage
  • Design funding model to support long term storage of large data sets – different than current FLUX funding model
QUESTIONS?

COMMENTS?