Preserving Principles and Transforming Practice: LIS Expertise for the Data Age

Carole L. Palmer
Information School
University of Washington

ACRL Washington & Oregon Joint Conference
Tried & True or Shiny & New?
October 19, 2017
Preparing LIS students for data intensive academia

- Data Curation in the Sciences 2006-11
- Biological Information Specialists 2006-09
- Summer Institutes in Data Curation 2008-11
- Data Curation in the Humanities 2008-12
- Data Curation in Research Centers 2010-2015
- Data Conservancy Education Initiatives 2009-12
- Preparing LIS students for data intensive academia 2016-19

- Institute of Museum and Library Services
- NSF
- ODL
- NCAR
- The National Academies of Sciences, Engineering, and Medicine
What is the contemporary meaning of your mission?

Supporting curricula and research needs

Collect and provide access to relevant and diverse academic resources

Provide access to sources of knowledge in all formats

Promote information literacy
National context - explosion of directives on data

In 2003, at least 11 national reports

Understanding Infrastructure: Dynamics, Tensions, and Design

Harnessing the Power of Digital Data for Science and Society

RCUK Review of e-Science 2009

Building the Infrastructure for Cyberscholarship
Digital data from federally funded research will be made available and useful for the public, industry, and scientific community

- Maximize impact and accountability of federal funds
- Promote entrepreneurship, enhance economic growth and job creation.

Translated into data management plan requirements for research grants

- Impact on library consultation and institutional repository services
Growth of data sharing - repositories and standards

Over 1,000 research data repositories

Now 1500

978 “databases”
703 standards
National data services & international organizations

Data Archiving and Networked Services

23 Things: Libraries for Research Data
An overview of practical, free, online resources and tools that you can begin using today to incorporate research data management into your practice of librarianship.

Data is at the heart of innovation today
Value all research products

A new funding policy by the US National Science Foundation represents a sea-change in how researchers are evaluated, says Heather Piwowar.

(Goble on Piwowar, Nature, January 10, 2013)
Scholar produced digital collections

Thematic research collections - primary [data] sources and related materials that support research on a theme.

(Palmer, 2004; Fenlon, 2017)
Digital data collections - a looming crisis?

**National Science Board (2005):** ever increasing investment in creating and maintaining collections, and the rapid multiplication of collections, with a potential for decades of curation.

**American Council of Learned Societies (2006):** Value-added ... widely shared ... collections...enabling ...interdisciplinary research ...

- Responsibility for sustainability of content and functionality is ambiguous.
- With interdisciplinary products, need to sustain paths back to disciplinary foundations to assure meaning and validity.

Transparency requires making visible both the empirical foundation and the logic of inquiry of research.  

(DA-RT, 2015)

Meaningful and valid reuse
Data analytics / science degrees and certificates

**SEATTLEU**
- MS and certificate in Business Analytics
- Bachelor of Applied Science in Data Analytics
- CS concentration in Big Data Management

**CityUniversity of Seattle**
- Seattle – Data Science Interdisciplinary MS, CS certificate
  - iSchool specialization
- Tacoma – Big Data PhD, certificate in Business Analysis

**W**
- MS or certificate in Data Analytics

---

**Two-Year College Data Science Summit**

May 10-11, 2018 (awaiting final confirmation), Washington, DC

With funding from the National Science Foundation, this workshop will bring together a diverse group of participants to make recommendations for two-year college data science programs, keeping in mind the needs of each of the three student populations:

1. Those seeking employment following an associate's degree
2. Those seeking transfer to four-year programs
3. Those seeking certificate programs and college-level courses in data science for professional development

To express interest in participating in the workshop, please fill out this Google form.
Top Trends in Academic Libraries

2012

• Communicating value
• Data curation
• Digital preservation
• Higher education
• Information technology

2014

• Data
  New initiatives / collaborative opportunities
  Cooperative roles for researchers, repositories, and journal publishers
  Partnerships for discovery & re-use
• Device neutral digital services
• Evolving openness in higher education
  open access
  open education
Top Trends in Academic Libraries

2016

- **Research data services (RDS)**
- **Data policies and data management plans**
- **Professional development for librarians providing RDS**
- **Digital scholarship**

- Collection assessment trends
- ILS and content provider/fulfillment mergers
- Evidence of learning: student success, learning analytics, credentialing
- New directions with Framework for Information Literacy for Higher Ed
  - Digital fluency in the Framework
  - Critical information literacy in the Framework

- **Altmetrics**
Definitions and distinctions

Active management of data through its life cycle of interest and usefulness to scholarship, science, and education.

**Curation:**
- managing and promoting use from point of creation
- enrichment & updating to keep **fit for purpose**
- availability for discovery and re-use

**Archiving:**
a curation activity – select and store
- logical and physical integrity
- security and authenticity

**Preservation:**
an archiving activity - specific items maintained over time
- accessed and understood through changes in technology

(JISC, 2004)

For our institutions, varying levels of service and dependencies
“Data Curation as Publishing for the Digital Humanities” (Muñoz, 2013)

Data curation = publishing work that draws directly on librarians unique skills; aligns directly with library missions and values

- making public products of scholarly work
- ensuring quality
- disseminating outputs to interested communities
### New expectations - same mission and metaphors

#### Access

| The new stacks? | (W. Tabb) |

#### Use

| The new special collections? | (S. Choudhury) |

---

Flickr users: stancia, rh creative commons

flickr.com/photos/001fj/2907653323/
The true essence of librarianship...is the maximization of the effective use of graphic records... . (Shera, 1971, p. 57).

- coordinate and integrate information in alignment with complex social structures and practices (Shera, 1972)
- add value to information to improve current use and potential for future use (Taylor, 1986)
- laying claim to the control zone (Atkinson, 1996)

LIS core of organization and access for user communities
- information behavior
- representation and retrieval of content
- collection and service development and management

(Palmer, Renear, Cragin, 2008)
<table>
<thead>
<tr>
<th>Categories</th>
<th>Types of Expertise</th>
<th>Service</th>
<th>Analytics</th>
<th>Leadership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>• Data handling</td>
<td>• Data uses &amp; users</td>
<td>• Data analysis</td>
<td>• Leadership</td>
</tr>
<tr>
<td></td>
<td>• Data landscape</td>
<td>• Data discovery</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research</td>
<td>• Research process</td>
<td>• Relationship-building</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Research instruments</td>
<td>• Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Curation</td>
<td>• Organization</td>
<td>• Data metrics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Standardization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Preservation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Data quality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Ethics</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Thompson, 2017)
Emerging new principles for practice

- Context key to meaning & validity
- Releasable ≠ reusable
- Producer datasets / consumer subsets
- Indicators of reuse value
- Primacy of method

Digital Collections and Content
Data Curation Profiles Project
Data Conservancy
Data Curation Education in Research Centers
Site-Based Data Curation at YNP
Online Data Management Course

1. Introduction to Data Management (5 min)
2. How to Inventory, Store, and Backup Your Data
3. How to Create Data that You (and Others) can Use
4. How to Navigate Rights and Ownership of your Data
5. How to Share Your Data and Ethically Reuse Data
6. How to Digitally Preserve Your Data for the Future
7. Complete Your DMP (5 min)

May, 2017

Research data management and services: Resources for novice data librarians
Sarah Barbrow, Denise Brush, Julie Goldman
Tenopir’s survey

Table 1: Frequencies and percentages for survey participants by full time equivalent (FTE) students

<table>
<thead>
<tr>
<th>FTE Students</th>
<th>Frequency (Percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1,999</td>
<td>41 (32.5%)</td>
</tr>
<tr>
<td>2,000-4,999</td>
<td>41 (32.5%)</td>
</tr>
<tr>
<td>5,000-9,999</td>
<td>18 (14.3%)</td>
</tr>
<tr>
<td>10,000-24,999</td>
<td>16 (12.7%)</td>
</tr>
<tr>
<td>25,000 or more</td>
<td>10 (7.9%)</td>
</tr>
<tr>
<td>Total</td>
<td>126 (100%)</td>
</tr>
</tbody>
</table>

The library needs to offer RDS to remain relevant to the institution

Losing data jeopardizes future scholarship

Librarians should be stewards of all types of scholarship, including data sets

<table>
<thead>
<tr>
<th>Opinions about library involvement in RDS</th>
<th>Strongly Agree</th>
<th>Strongly Disagree &amp; Disagree</th>
<th>Neither Agree nor Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The library needs to offer RDS to remain relevant to the institution</td>
<td>46.5%</td>
<td>30.2%</td>
<td>23.3%</td>
</tr>
<tr>
<td>Losing data jeopardizes future scholarship</td>
<td>59.3%</td>
<td>9.3%</td>
<td>31.4%</td>
</tr>
<tr>
<td>Librarians should be stewards of all types of scholarship, including data sets</td>
<td>75.6%</td>
<td>8.1%</td>
<td>16.3%</td>
</tr>
</tbody>
</table>

Responsible for providing research data reference consultation and instruction

<table>
<thead>
<tr>
<th>Approximate Annual External Funding</th>
<th>&lt; $50 Million (n=32)</th>
<th>$50 Million or more (n=9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual Discipline Librarian</td>
<td>84.4%</td>
<td>22.2%*</td>
</tr>
<tr>
<td>Dedicated Data Librarian</td>
<td>0.0%*</td>
<td>44.4%</td>
</tr>
<tr>
<td>Other</td>
<td>15.6%</td>
<td>33.3%</td>
</tr>
</tbody>
</table>
Workforce Trends

Trending up:
- Data Steward
- Digital repository
- Digital preservation
- Curation Science
- Digital Curator
- Data Curator

Trending down:
- Librarian

UW iSchool Educational Context

Data curation signature within iSchool MLIS with additional Data Science sequence in MSIM

Expanding curriculum
• data services and technology
• government and civic open data

Field experiences in public sector
– making open data more usable

democratizing data entrepreneurial use
When Jevin West and Carl Bergstrom quietly rolled out CallingBullshit.org in January, they hoped a few of their ideas would find their way into classrooms. At best, maybe the University of Washington would let them teach a class.

Learning objectives:
Recognize BS whenever and wherever you encounter it.
Figure out for yourself precisely why a particular bit of BS is BS.
Provide a technical explanation of why a claim is BS.
Provide an aunt or uncle with an accessible and persuasive explanation of why a claim is BS.

We will be astonished if these skills do not turn out to be among the most useful and most broadly applicable of those that you acquire during the course of your college education.
iSchool Curriculum

• INFX 551 – Fundamentals of Data Curation
• INFX 598 – Advanced Data Curation
• INFX 598 – Digital Preservation
• INFX 531 – Metadata Design
• INFX 573 Data Science I: Theoretical Foundations
• INFX 574 Data Science II: Machine Learning and Econometrics
• INFX 575 Data Science III: Scaling, Applications, and Ethics
• INFX 543 Relational Database Management Systems
• INFX 544 – Information Retrieval Systems
• INFX 561 – Visualization Design
• INFX 562 – Interactive Information Visualization
MLIS required curriculum

- 63 total credits
  - LIS 510 - History and Foundations of Libraries and Librarianship
  - LIS 520 - Information Resources, Services, and Collections
  - LIS 530 - Organization of Information and Resources
  - LIS 550 - Information and Society
  - LIS 560 - Instructional and Training Strategies
  - LIS 570 - Research, Assessment, and Design
  - LIS 580 – Management of Information Organizations
  - Your choice of one info tech core course: choose between INFX 511, 512, 542, 543, 544, 546, 547, 572, or 573
NYU Health Sciences Library
Data sharing animations

Part 1 - Request

https://www.youtube.com/watch?v=RVZbk3GEVSw
Thank you for your attention

Questions welcome
Curriculum Pathways

- Academic Librarianship
- Archiving / Special Collections / Records Management
- Data Curation
- Data Science
- Database Administration / Development
- Digital Librarianship
- Digital Youth / Children's Librarianship
- Health Information Sciences
- Information Architecture / Taxonomy
- Knowledge Organization
- Law Librarianship
- Public Librarianship
- Special / Corporate Librarianship
- User Experience
Data curation placements

**Academic**

- 40% of placements,
  - ¼ of those outside library
- Many focused on metadata and technology

**Non-academic positions**

- Data Steward Consultant
- Solutions Analyst
- Senior General Engineer
- GIS Specialist
- Director of Archive Technology
- Digital Asset Manager
- Information Architect
- Information Systems Associate
- Digital Project Coordinator
- Media Content Specialist

**Positions that (probably) didn’t exist 5 years ago**

- Research Data Management Service Design Analyst
- Data Management Consultant
- Data Science & Informatics Librarian
- Data Curator
- Assistant Dean, Digital Humanities Research
NCAR internships

Climate model metadata
Sensor data archiving
Social science data organization
Time-series temporal spatial
Analog data for digital access

metadata harvesting, standards compliance, quality
processing & file migration
cross-disciplinary data curation; subsetting
high resolution, provenance, NetCDF
50 international collections, OAIS, DOIs

Translator and Facilitator:

Understands and articulates the needs and goals of scientist
Understands and articulates data manager needs for curation
Creates guidelines to enhance communication and efficiency between scientists and data managers
Trends in national data facilities

- Some growth in positions
- Fewer specialized roles
- Value of Information Science
- Intern programs restricted

Data
- Search & retrieval
- Archiving & preservation
- Quality control
- Transformation

Policy
- Standards

Metadata
- Data structures & formats
- Measurements
- Terminology
- Communication with scientists

Domain
- Interest
- Research process

Technology
- Databases
- Programming
- Website & interface Design
- Systems analysis
- User requirements
- Storage & movement
- New technologies
Too much to lose, if we don’t get it right.

“Your analytics are only as good as your curation.”

- marshal our strengths in LIS
- leverage progress across disciplines
- build a new LIS foundation in the science of data