

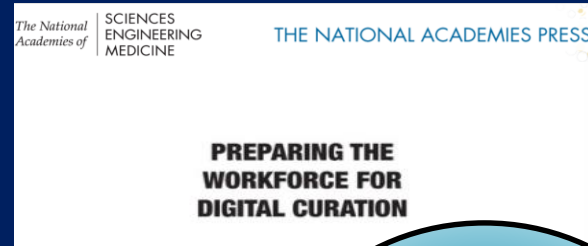
Preserving Principles and Transforming Practice: LIS Expertise for the Data Age

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



**Biological
Information
Specialists
2006-09**

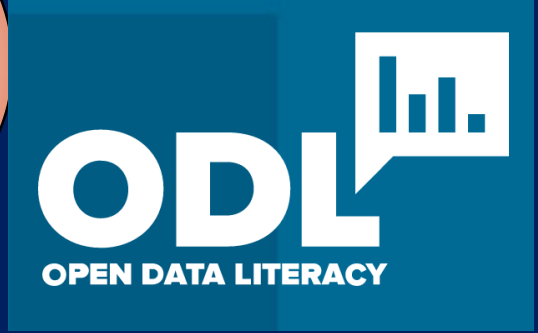
**Data Curation
in the Sciences
2006-11**

**Data Curation
in the Humanities
2008-12**

**Summer
Institutes in
Data Curation
2008-11**

**Data
Conservancy
Education
Initiatives
2009-12**

**Data Curation
in Research
Centers
2010-2015**



2016-19



What is the contemporary meaning of your mission?

Supporting curricula and research needs

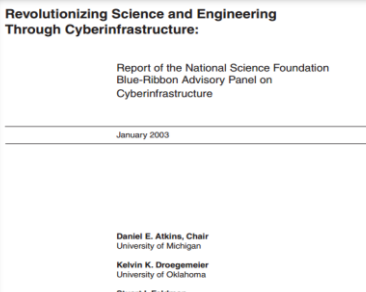
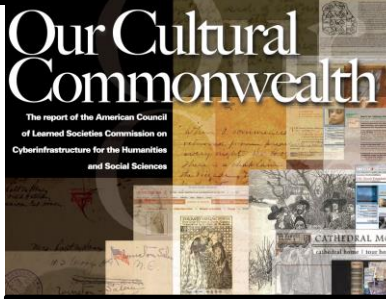
Collect and provide access to relevant and diverse academic resources

Provide access to sources of knowledge in all formats

Promote information literacy

National context - explosion of directives on data

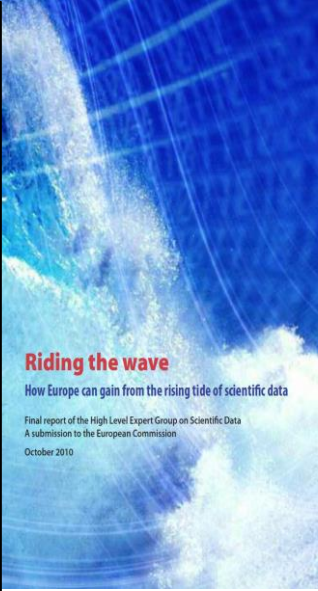
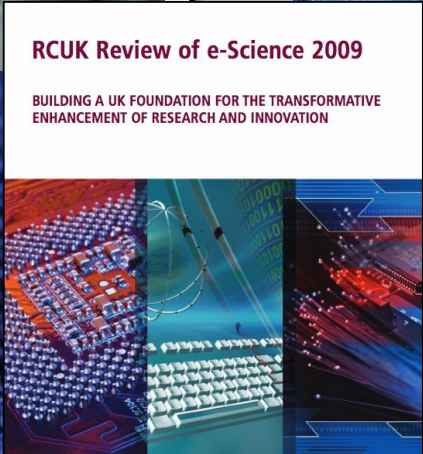
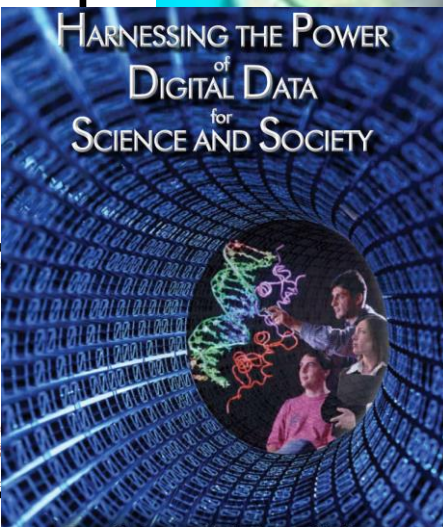
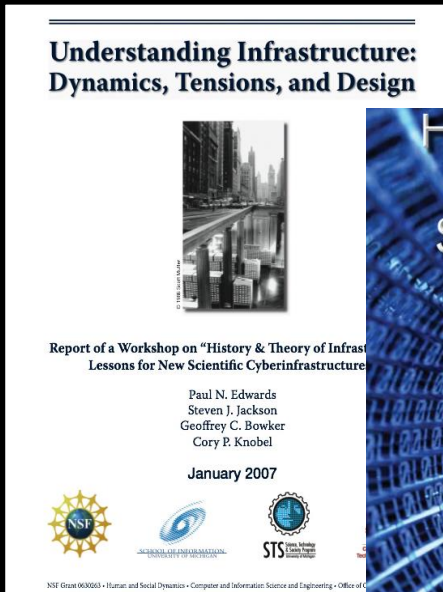
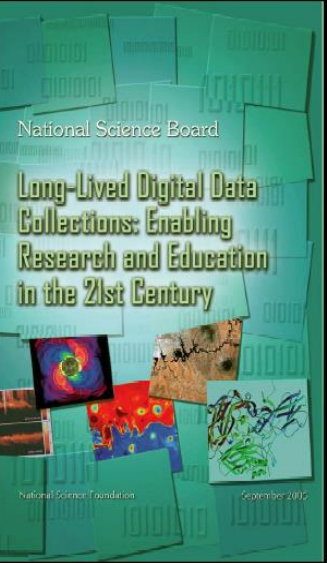
In 2003,
at least 11 national reports



Atkins

Larsen

Unsworth



10 years later, federal action

EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren 
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

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



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Home Search Browse Suggest FAQ About Schema API

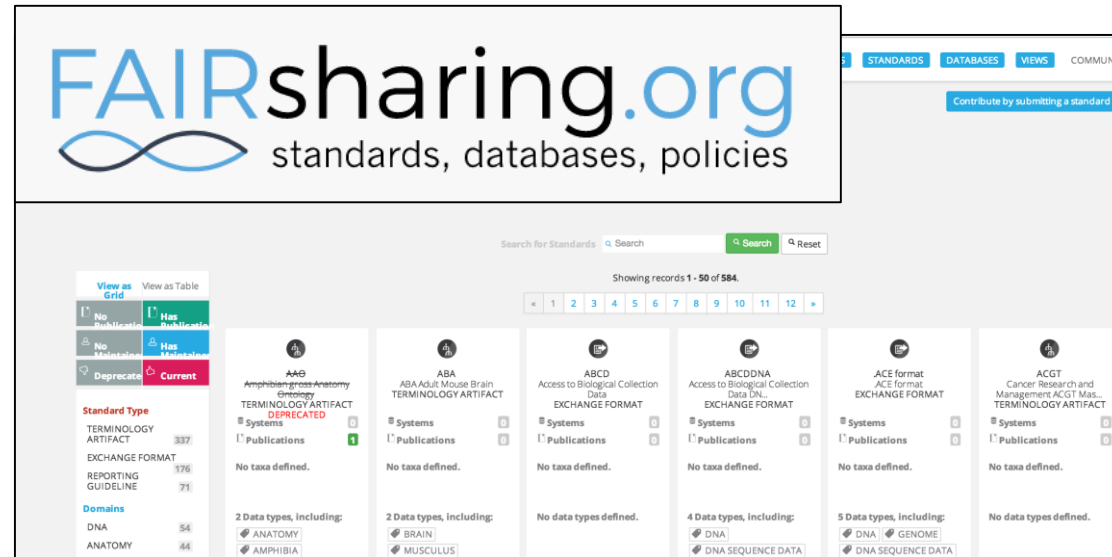
TAG ARCHIVES: DATABIB

~~Over 1,000~~ research data repositories

978 “databases”
703 standards



Now 1500



FAIRsharing.org
standards, databases, policies

STANDARDS DATABASES VIEWS COMMUNITY

Contribute by submitting a standard

Search for Standards Search Search Reset

Showing records 1 - 50 of 584

Standard Type	Count
TERMINOLOGY ARTIFACT	337
EXCHANGE FORMAT	176
REPORTING GUIDELINE	71

Domains	Count
DNA	54
ANATOMY	44

Standard	Count
Amphibian gross Anatomy Terminology Artifact (DEPRECATED)	1
ABA Adult Mouse Brain Terminology Artifact	0
ABCD Access to Biological Collection Data Exchange Format	0
ABCDNA Access to Biological Collection Data DL Exchange Format	0
ACE format ACE format EXCHANGE FORMAT	0
ACGT Cancer Research and Management ACGT Metadata Terminology Artifact	0

National data services & international organizations



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

Publishing trends

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)

required &
supplementary data

data journals

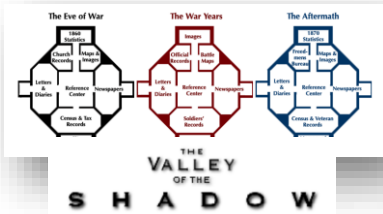


datasets
data collections
algorithms
configurations
tools and apps
codes
workflows
scripts
code libraries
services,
system software
infrastructure,
compilers
hardware

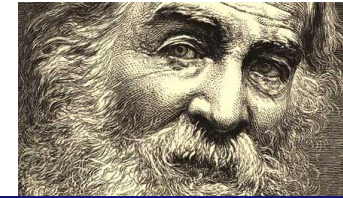
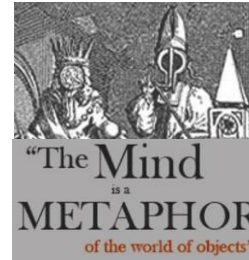
enhanced
publications

commentaries
interviews
tours
animation
archival sources

Scholar produced digital collections



THE WILLIAM BLAKE ARCHIVE



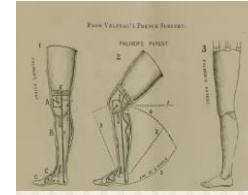
The Walt Whitman Archive



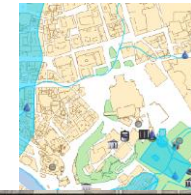
BRACERO HISTORY ARCHIVE



Journals of the Lewis & Clark Expedition



19TH CENTURY-DISABILITY CULTURES & CONTEXTS



Aquae Urbis Romae The Waters of the City of Rome



Colonial Frontier Massacres in Eastern Australia 1788-1872



THE Shelley-Godwin ARCHIVE



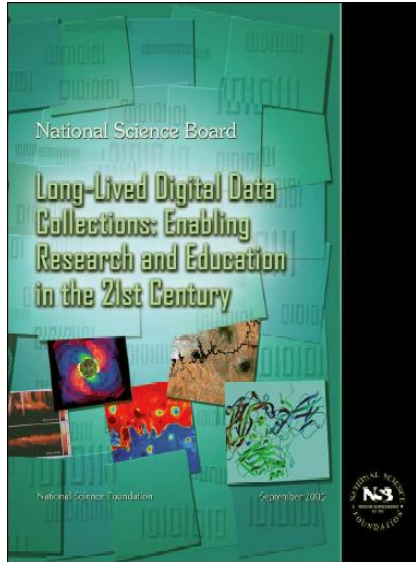
The Vault at Pfaff's



O SAY CAN YOU SEE EARLY WASHINGTON, D.C. LAW & FAMILY

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

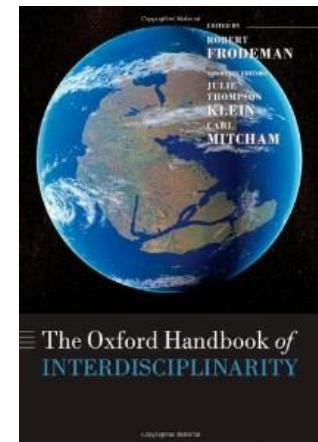
Digital data collections - a looming crisis?



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.

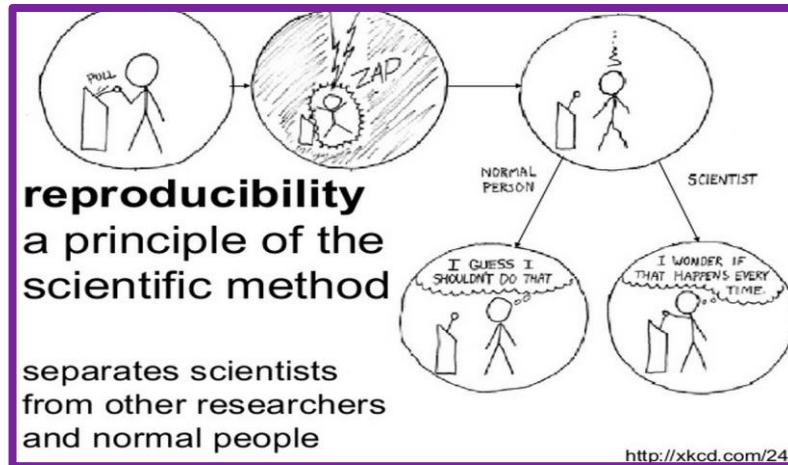
American Council of Learned Societies (2006): Value-added ... widely shared ... collections...enabling ...interdisciplinary research ...

- 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.



(Palmer, 2010, 2017)

Reproducibility, transparency, openness



replicate repeat
rerun
recompute
recreate revise
redo regenerate
restore re-examine
recycle reconstruct review
reuse repurpose

(Goble 2013)

Transparency requires making visible both the empirical foundation and the logic of inquiry of research. (DA-RT, 2015)



ANNOTATION FOR TRANSPARENT INQUIRY

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

Seattle – Data Science Interdisciplinary MS, CS certificate
iSchool specialization

Tacoma – Big Data PhD, certificate in Business Analysis

MS or certificate in Data Analytics

A screenshot of the American Statistical Association (ASA) website. The page features a blue header with the ASA logo and navigation links: 'What is Statistics?', 'Donate', 'Join', and 'Login'. Below the header is a dark blue navigation bar with links for 'ABOUT', 'MEMBERSHIP', 'EDUCATION', 'PUBLICATIONS', 'MEETINGS', 'POLICY & ADVOCACY', and 'YOUR CAREER', along with a search bar. The main content area has a white background and features the title 'Two-Year College Data Science Summit' in a large, bold, black font. Below the title is the date and location: 'May 10-11, 2018 (awaiting final confirmation), Washington, DC'. The text describes the workshop's purpose, funded by the National Science Foundation, and lists three student populations: 1. Those seeking employment following an associate's degree, 2. Those seeking transfer to four-year programs, and 3. Those seeking certificate programs and college-level courses in data science for professional development. At the bottom, it provides a link to a Google form for expressing interest.

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

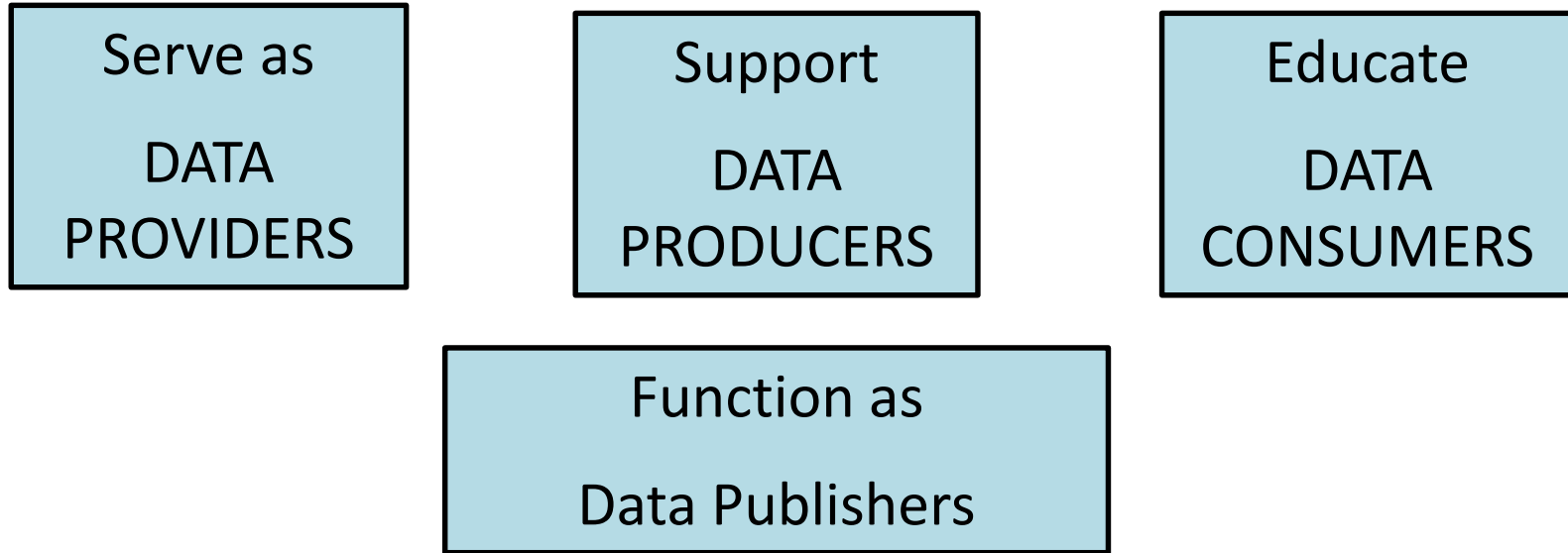
Archiving: a curation activity – select and store
logical and physical integrity
security and authenticity

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

Span of library roles



“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

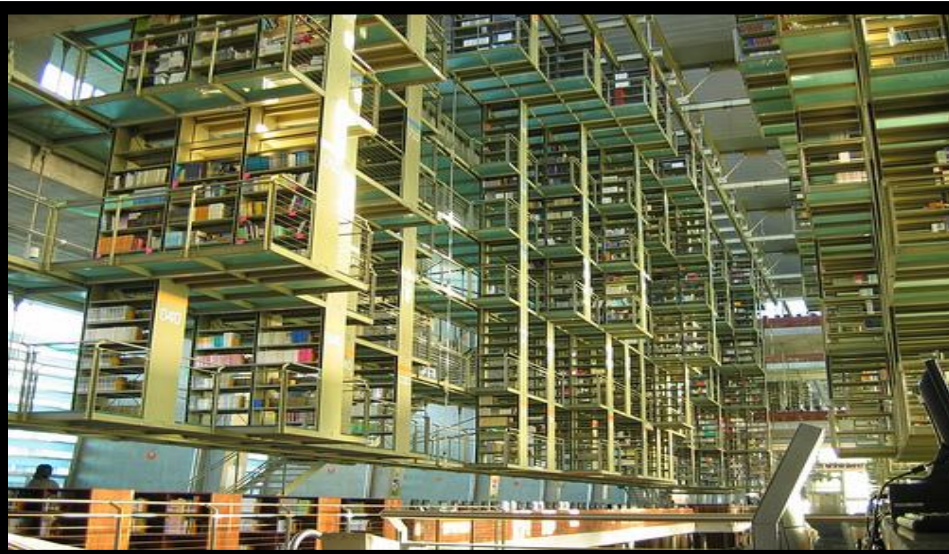
New expectations - same mission and metaphors

Access

The new stacks? (W. Tabb)

Use

The new special collections? (S. Choudhury)



Consistent core principles

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)

Risk of underestimating need for new expertise

Categories	Types of Expertise
Data	• Data handling
	• Data landscape
Research	• Research process
	• Research instruments
Curation	• Organization
	• Standardization
	• Preservation
	• Data quality
	• Ethics

Service	• Data uses & users
	• Data discovery
	• Training
	• Relationship-building
	• Collaboration
Analytics	• Data metrics
	• Data analysis
Leadership	• Leadership

(Thompson, 2017)

Emerging new principles for practice

Digital Collections and
Content

Data Curation
Profiles Project

Data Conservancy

Data Curation Education
in Research Centers

Site-Based Data
Curation at YNP

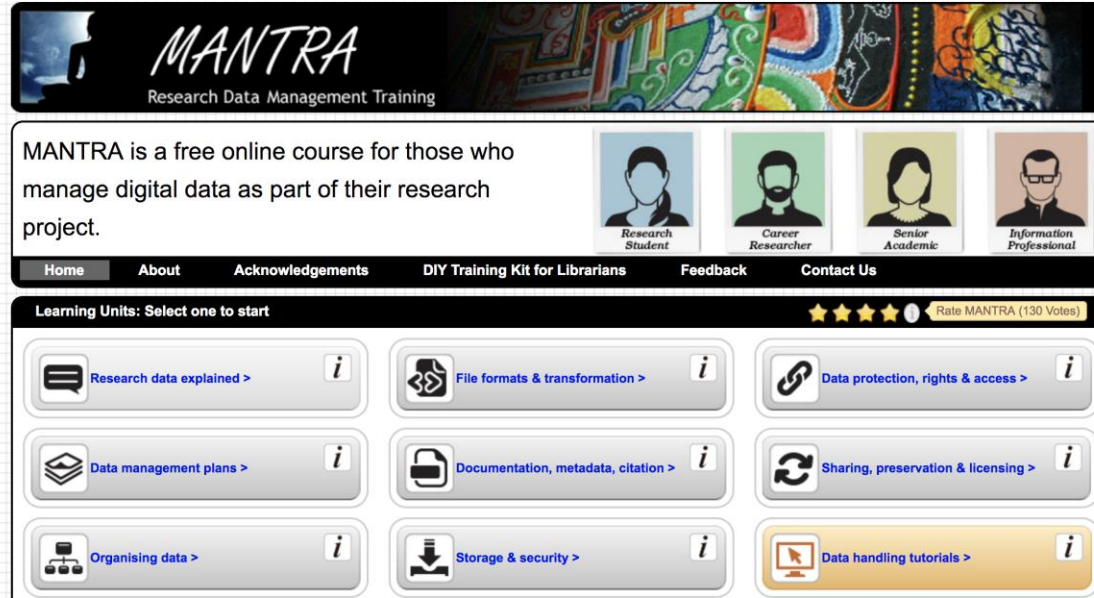
- *Context key to meaning & validity*
- *Releasable ≠ reusable*
- *Producer datasets / consumer subsets*
- *Indicators of reuse value*
- *Primacy of method*

Building capacity and expertise

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 \(5 min\)](#)
6. [How to Digitally Preserve Your Data for the Future](#)
7. [Complete Your DMP \(5 min\)](#)



The screenshot shows the MANTRA website interface. At the top, it says 'MANTRA Research Data Management Training'. Below that, a navigation bar includes 'Home', 'About', 'Acknowledgements', 'DIY Training Kit for Librarians', 'Feedback', and 'Contact Us'. A section titled 'Learning Units: Select one to start' features a grid of learning unit buttons with icons and information symbols. The units include: 'Research data explained >', 'File formats & transformation >', 'Data protection, rights & access >', 'Data management plans >', 'Documentation, metadata, citation >', 'Sharing, preservation & licensing >', 'Organising data >', 'Storage & security >', and 'Data handling tutorials >'. A 'Rate MANTRA (130 Votes)' section with five stars is also visible.



May, 2017

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

Research Data Services in Academic Libraries: Data Intensive Roles for the Future?

Tenopir's survey sample:

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

FTE Students	Frequency (Percent)
Up to 1,999	41 (32.5%)
2,000-4,999	41 (32.5%)
5,000-9,999	18 (14.3%)
10,000-24,999	16 (12.7%)
25,000 or more	10 (7.9%)
Total	126 (100%)

(Tenopir, C., et al. (2015). Journal of eScience Librarianship 4(2): e1085.
<http://dx.doi.org/10.7191/jeslib.20>)

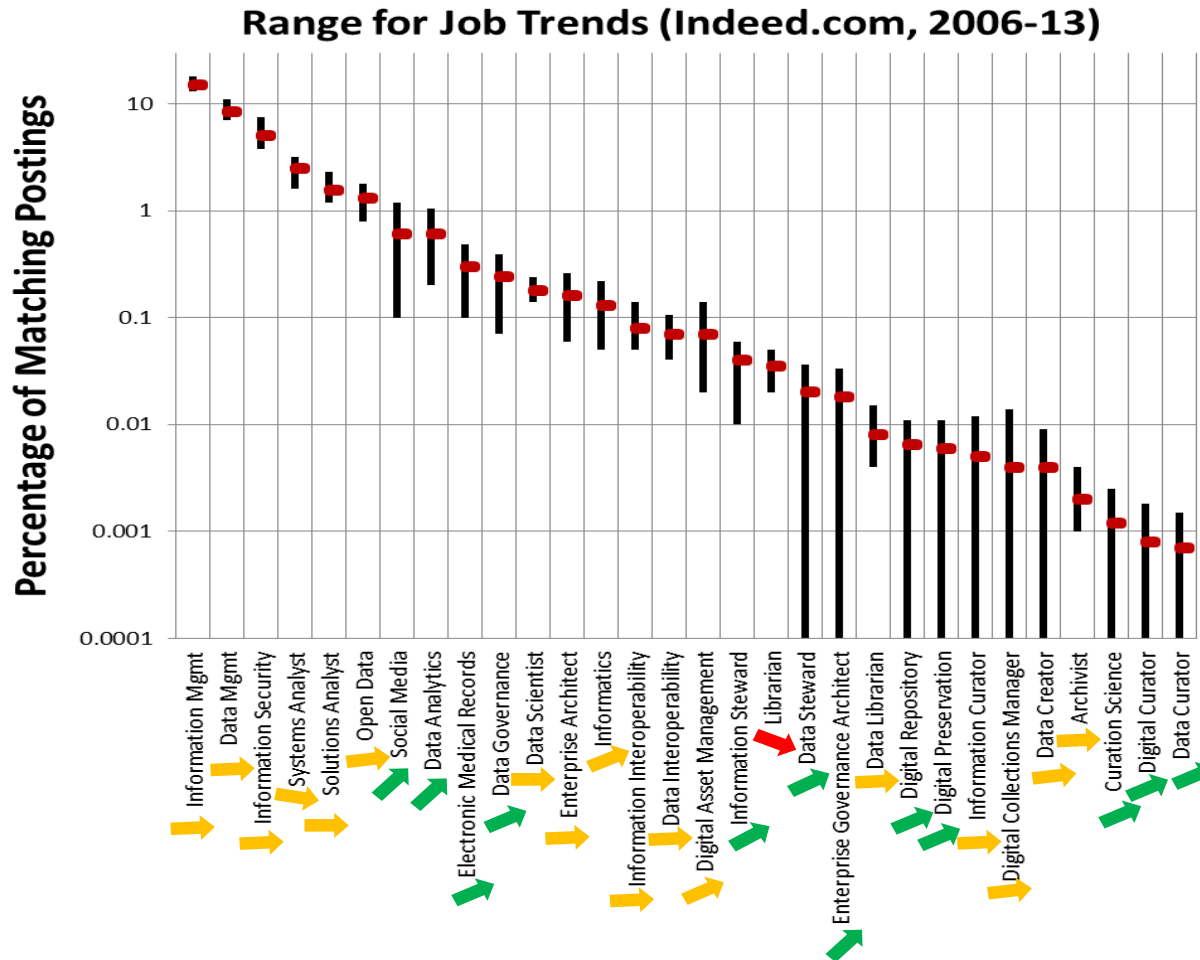
Opinions about library involvement in RDS

	Strongly Agree	Strongly Disagree & Disagree	Neither Agree nor Disagree
The library needs to offer RDS to remain relevant to the institution	46.5%	30.2%	23.3%
Losing data jeopardizes future scholarship	59.3%	9.3%	31.4%
Librarians should be stewards of all types of scholarship, including data sets	75.6%	8.1%	16.3%

Responsible for providing research data reference consultation and instruction

	Approximate Annual External Funding	
	< \$50 Million (n=32)	\$50 Million or more (n=9)
Individual Discipline Librarian	84.4%	22.2%*
Dedicated Data Librarian	0.0%*	44.4%
Other	15.6%	33.3%

Workforce Trends



Trending up:

- Data Steward
- Digital repository
- Digital preservation
- Curation Science
- Digital Curator
- Data Curator

Trending down:

Librarian

National Research Council. (2015). Board on Research Data.

Preparing the Workforce for Digital Curation. National Academies Press.

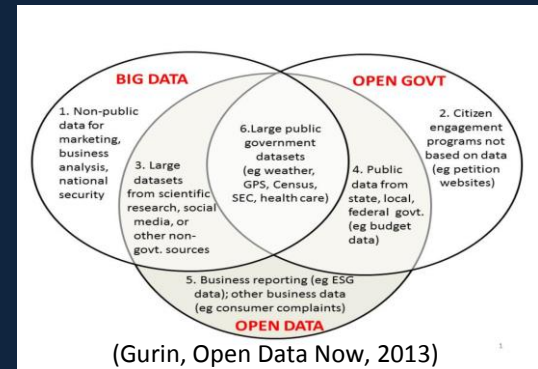
UW iSchool Educational Context

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





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

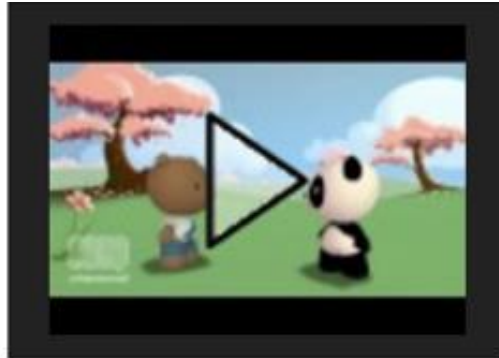
- **63 total credits**
 - LIS 510 - History and Foundations of Libraries and Librarianship
 - LIS 520 - Information Resources, Services, and Collections
 - LIS 530 - Organization of Information and Resources
 - LIS 550 - Information and Society
 - LIS 560 - Instructional and Training Strategies
 - LIS 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

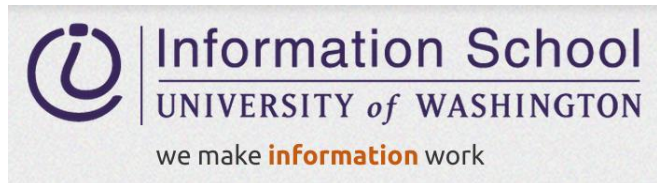
Part 1 - Request

<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,
 $\frac{1}{4}$ of those outside library
- Many focused on metadata and technology

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

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

NCAR internships

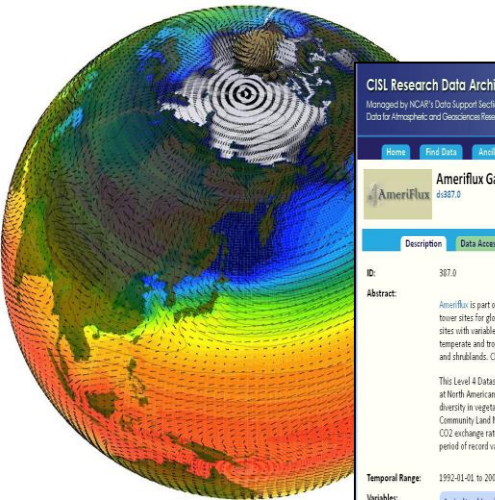
Climate model
metadata

Sensor data
archiving

Social science
data
organization

Time-series
temporal spatial

Analog data for
digital access



CISL Research Data Archive
Managed by NCAR's Data Support Section
Data for Atmospheric and Geosciences Research

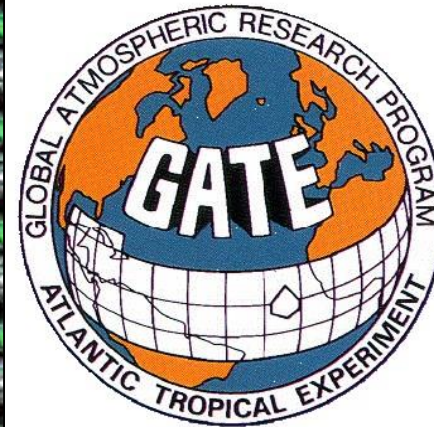
Home Find Data Ancillary Services About/Contact

AmeriFlux Gap-Filled Data from Selected Sites
49387.0 For assistance

Description	Data Access	Documentation
ID: 387.0		
Abstract: AmeriFlux is part of the larger network Fluxnet which endeavors to connect observations tower sites for global analysis. AmeriFlux towers measure the exchange of energy, water sites with variable vegetation, disturbance records, and climatic conditions. The ecosystem temperate and tropical evergreen forests, temperate deciduous forests, woodlands, grass and shrublands. Climates in these biomes vary between tundra, temperate, tropical, and		
This Level 4 Dataset accumulated from the AmeriFlux Network focuses on hourly, daily, or at North American tower sites. The selection of AmeriFlux sites were chosen based on the diversity in vegetation types, the past success of testing land surface models, and the Community Land Models. AmeriFlux Level 4 data has been gap-filled, quality-flagged and CO2 exchange rates, heat fluxes, biogeochemical pools, calculated gross productivity, in period of record varies by individual tower site, but some sites have as much as 15 years		
Temporal Range: 1992-01-01 to 2008-01-01		
Variables:	Agricultural Lands Air Temperature Alpine/Tundra Biogeochemical Cycle	
Forests Grasslands Heat Flux Humidity		
Incoming Solar Radiation Precipitation Rate Respiration Soil Moisture/Water C		
Soil Temperature Solar Radiation		

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



metadata
harvesting,
standards
compliance,
quality

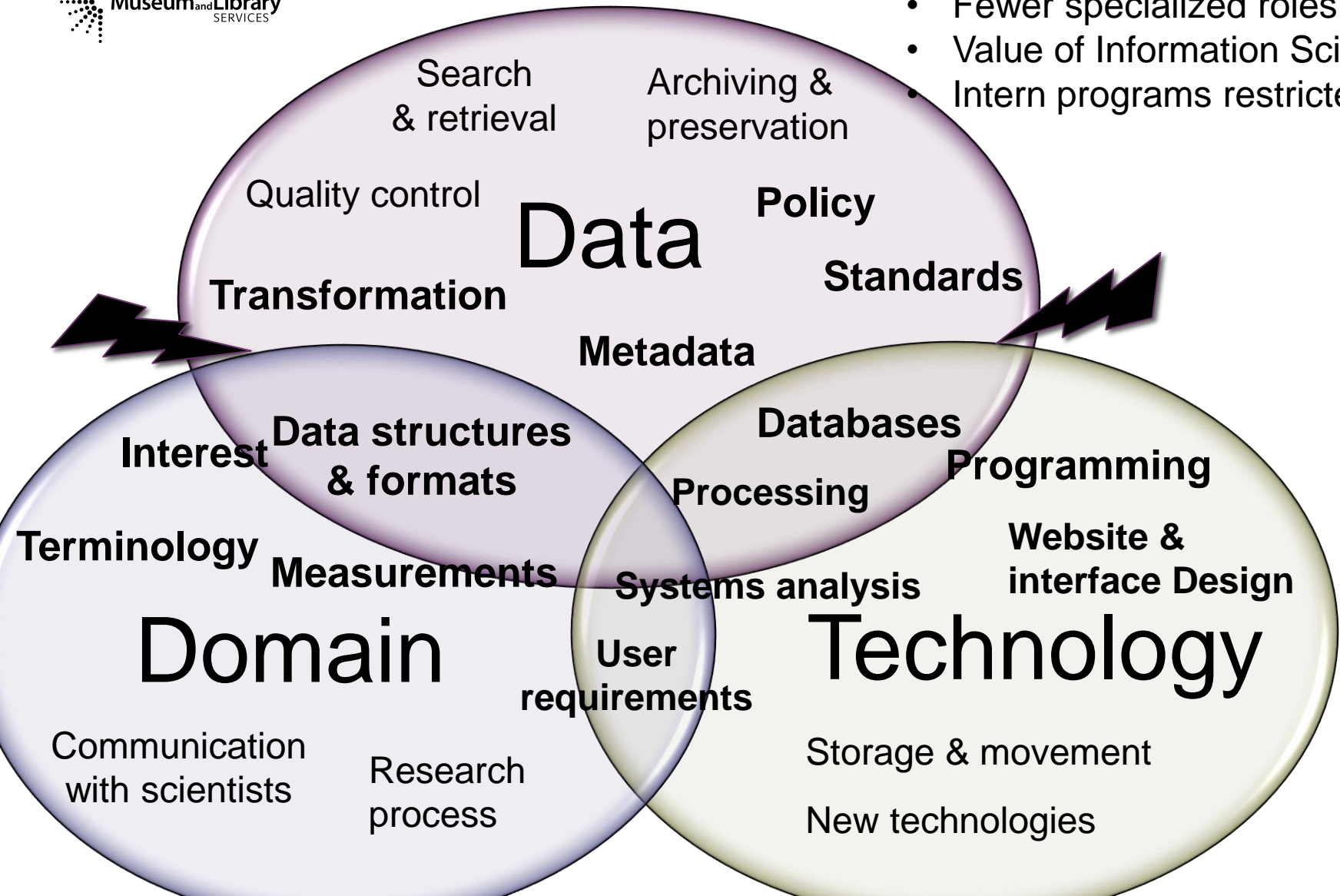
processing &
file migration

cross-disciplinary
data curation;
subsetting

high resolution,
provenance,
NetCDF

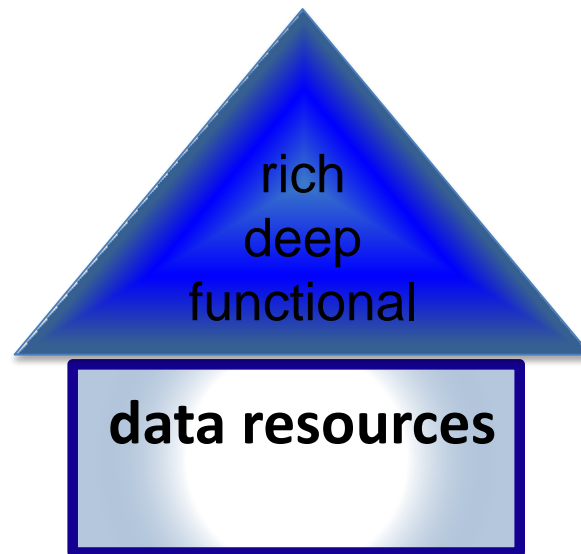
50 international
collections,
OAIS, DOIs

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



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