

Collections in the Age of E-Research; Realizing Potential through Curation and Aggregation

Carole L. Palmer

Center for Informatics Research in Science & Scholarship
Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign

Digital Library Federation Fall Forum

2 November 2010,
Palo Alto, CA



The *iSchool* AT ILLINOIS

CIRSS

What were we worried about 50+ years ago?

1948 - Royal Society Scientific Information Conference

1952 - Chicago School Symposium on Special Information

1958 - International Conference on Scientific Information

- Growth of the literature
- Complexity of formats
- Interdisciplinarity
- Connectivity among systems
- Faster circulation
- Access to pre-publications
- Document components of documents

What is fundamentally different now?

- Involvement of scientists and scholars in information system R&D
- Access to scholarly content by general public
- Modes of engagement with the universe of information
 - convenience, speed
 - inking
 - small window of access
- Higher stakes ...

*In the contemporary context of e-science, where the aim is to
re-shape scientific endeavours
and provide new infrastructures to support them, [the] goal of
studying the detail of actual practice takes on a new significance.*

(Hine, 2005)

“Reshaping” suggests professional imperatives

- Provide access to the broad landscape of information across disciplines - LIS as a metascience (Bates, 1999)
across generations, to preserve the scholarly record.
 - Foster sharing across institutions.
 - Manage scatter of information resources.
 - Coordinate in alignment with complex social structures and practices. (Shera, 1972)
 - Add value to improve current use and potential for future use. (Taylor, 1986)

Value of digital collections is in the collective

- Tremendous investment in generating content.
- Material exposed but growing base of interrelated content not exploited or broadcast.
- Scholarly and scientific potential and economies of scale obscured in the scatter.

Curated collections essential units of high-value information.

Aggregation of collections essential organizational layer for access, interpretation, and use.

Two initiatives

IMLS Digital Collections & Content (DCC)

Data Conservancy

Two large-scale digital collection efforts

Different kinds of LIS / research library R & D contributions to the emerging network of digital collections and services.

IMLS DCC - CIRSS & University Library collaboration

Carole Palmer (2007-2012)	Tim Cole (2002-2007)	PIs
Allen Renear, Miles Efron, Mike Twidale Sarah Shreeves, Bill Mischo, Tom Habing		CIRSS co-PIs Library co-PIs
Katrina Fenlon Jacob Jett, Piotr Organisciak, Sunah Suh, Richard Urban, Karen Wickett, Oksana Zavalina		Project Coordinator Research Assistants
Martin Doerr - Centre for Cultural Informatics, Greece Jeremy Frumkin - Arizona Jonathan Furner - UCLA Jean Godby - OCLC Josh Greenberg - NYPL Pete Johnston - Eduserve. UK Bill Landis - Yale Jenn Riley - Indiana Jim Scheppke - Oregon State Library Guenter Waibel - OCLC		Advisory Committee

Digital Collections and Content

Working toward interoperable digital content.



IMLS DCC - Phase 1 (2003-2007)

- Single point of access to IMLS-funded content (NLG, some LSTA).
[collection registry](#)
[item-level metadata repository](#) using OAI-PMH
- Focus on interoperability issues and trends.
[metadata sharing](#), harvesting, quality, best practices
- Collection representation and identity.
[conceptualizations of collections](#): exhibits, displays, tours ...
[cultures of description](#) in libraries, museums, archives

DCC in 2007 – 202 diverse collections, 300,000 items

The Walt Whitman Archive
Ed Folsom & Kenneth M. Price, Editors

Published Works
Leaves; Periodical Printings; Foreign Editions

Manuscripts
Transcriptions and Images; Finding Aids; More...

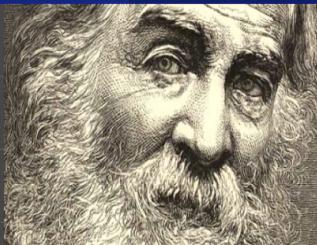
Biography & Correspondence
Life; Letters; Chronology

Criticism
Reviews; Commentary; Disciples; Bibliography

Pictures & Sound
Portraits of Whitman; Audio

Resources
Search; Teaching Materials; Tools

About the Archive



News & Updates

November 2009
Hans Reisiger's German translation of selected poetry and prose by Whitman has been added to the Archive's foreign editions section.

October 2009
Images of Whitman's Blue Book—the book that cost Whitman his government job in 1865—have recently been added to the site and are available [here](#).

January 2009
Great news! The Archive meets its fundraising goal for the NEH challenge grant. Thanks to all contributors.

Search the Archive


EXIT ART

ABOUT
PROGRAMS
EXIT UNDERGROUND
EXIT MEDIA
SUPPORT
EXIT ARCHIVE
PRESS

THE STORE
PRINT PORTFOLIOS

SEARCH EXIT ART
email address

UPCOMING EXHIBITIONS

GLOBAL / NATIONAL - The Order of Chaos
March 13 - May 1, 2010
Opening Saturday, March 13, 7-10pm



Exit Art is closed for installation and will reopen on Saturday, March 13, 7-10pm, for the opening of our Spring exhibition Global / National - The Order of Chaos. The SEA exhibition Waterpod: Autonomy and Ecology will also reopen on March 13, 7-10pm.

Autonomy and Ecology
March 13 - May 1, 2010
Opening Saturday, March 13, 7-10pm

Global / National - The Order of Chaos
March 13 - May 1, 2010
Opening Saturday, March 13, 7-10pm

Global / National - The Order of Chaos examines tensions of incommensurable forces that are dislocating our society to redefine a new global identity. The exhibition questions what is global but not national but the national is in a state of chaos. The artworks reflect how the global is national and the national has global concerns: searching inside our culture to project our global position. This exhibition tells the story of that confusion and chaos.

THE SKYSCRAPER MUSEUM

HOME

TIMELINE OF THE TALLEST TOWERS IN THE WORLD

WHAT'S UP?
Current Exhibition
Upcoming Programs
Family Programs
WTC Dossier
Exhibitions

ABOUT US
Visiting Us
History
Battery Park City Home

EXHIBITIONS
Current
Permanent
Future
Past

EDUCATION
School Visits
Resources for Educators
Cool Stuff for Kids

ONLINE ARCHIVE
Lectures & Conferences
Video Archive Index
Virtual Tours
Collections & Finding Aids
Rights & Reproductions

WEB PROJECTS
VIVA: WTC & ESB
VIVA
Bankers Trust Virtual Archive
Web Walk

HOME

Skyscraper Museum
Located in New York City, the world's first and foremost vertical metropolis. The Skyscraper Museum celebrates the City's rich architectural heritage and examines the historical forces and individuals that have shaped its successive skylines. Through exhibitions, programs and publications, the Museum explores tall buildings as objects of design, products of technology, sites of construction, investments in real estate, and places of work and residence. For a description of the gallery and photos of the spaces, please visit our Photo Slideshows page.

The Skyscraper Museum is located in lower Manhattan's Battery Park City at 39 Battery Place. Museum hours are 12-6 PM, Wednesday-Sunday. General admission is \$5, \$2.50 for students and seniors. Click here for directions to the Museum.

Call 212.968.1961 for information on upcoming lectures, exhibitions, walking tours, and membership. Subscribe to our newsletter by emailing programs@at-skyscraper.org.

Current Exhibition | China Prophecy: Shanghai
Future City 20 | 21 culminates in a close look at Shanghai, as a model for 21st century urbanism. Using architectural photography of recent towers, architectural drawings of existing and proposed towers and computer animations, documenting both the record high-rise developments and the challenges for the next generation of urban development, this installation will lead to an inquiry into the future of economics, laws, and culture shaping the present-day Chinese metropolises. Click here for more information. Click here for a Virtual Walkthrough of the exhibit.

Upcoming Programs | Book Talks
The Skyscraper Museum hosts a book talk and signing with author Andrew S. Dolkart on April 11. THE ROW HOUSE: GROWING A Neighborhood in New York City.



**VOICES
OF THE COLORADO PLATEAU**

vPlants

Welcome Chicago Region State Map County Map

HOME
About Us
Site Map

PLANTS
Fungi

Search
Help Guide
Topics
Resources
News

vPlants: A Virtual Herbarium of the Chicago Region.
The online resource for **Plants** and **Fungi** offering:

- Specimen data and images
- Species descriptions *
- Distribution by county
- Photo galleries *

* Description Pages, Features in production / Keys



Currently the site contains data for 80,000 plant specimens from three institutions with rich Chicago Region collections: the Field Museum of Natural History, The Morton Arboretum, and the Chicago Botanic Garden. Species descriptions and other content are added on a continual basis.

Specimen data from additional institutions such as the Chicago Academy of Sciences and the Illinois Natural History Survey will be added. We have also been in contact with other herbaria in the Great Lakes area such as the University of Wisconsin and University of Michigan regarding the possibility of adding or linking their Chicago Region specimen data through vPlants. This project is a first step towards building a larger online database for plants and fungi that occur in the Great Lakes area

THE MAINE MUSIC BOX



A Fogler Library, University of Maine, Bagaduce Music Lending Library and Bangor Public Library Partnership

About Maine Music Box

Contact us

Instructional Channel

© 2003 Fogler Library, University of Maine. All Rights Reserved.

Explore the Music Collections:

Search Music Collection

Browse By Collection

Browse Music Subjects

Browse or Search Sheet Music Cover Art

Search the database by Keyword

Get titles in a text list
Get titles with images



Questions driving Phase 2

What's in it, really?

Who's it for?

Limitations to opportunistic, critical-mass approaches to growth.

Need to understand whole, develop for user communities, leverage the collective.

Opening History (2007 -)

Build on extensive base of historical materials

Policy driven collection

Emphasis on research threads – metadata relationships,
emergent collections for scholarly use

Opening History - 900+ collections, over 1 million items

The screenshot shows the homepage of the *Opening History* website. At the top, there is a banner with the text "Opening History: U.S. History Resources from Libraries, Museums, and Archives". Below the banner, on the left, is a sidebar with the **INSTITUTE of Museum and Library SERVICES** logo and links for "Opening History Activities", "Sowing Culture", "IMLS DCC", "Metadata Roundtable", "Research", and "Working Groups". On the right, there is a search bar titled "Search For Items" with a "Search" button and a link to "Advanced Search for Items". Below the search bar is a section titled "Browse Collections By:" with three columns: "Subject", "Object", and "Place". The "Subject" column lists categories like Social Studies, Arts, Science, Religion, Language Arts, and Vocational Education. The "Object" column lists types like image, text, physical object, sound, moving image, and Interactive Resource. The "Place" column lists regions like Midwest U.S., Southwest U.S., Southern U.S., Pacific Coast U.S., Mountain Region U.S., and Northeastern U.S. At the bottom of the page, there are links for "Suggest a collection", "Currently 901 collection records. Last record updated On 10/20/2010.", and "Collection Development Policy (pdf)".

Opening History:
U.S. History Resources from
Libraries, Museums, and Archives

INSTITUTE of Museum and Library SERVICES

Opening History Activities:

Sowing Culture

IMLS DCC

Metadata Roundtable

Research

Working Groups

CIRSS

Search For Items

Search

[Advanced Search for Items](#)

Browse Collections By:

Subject	Object	Place
Social Studies (823) Arts (209) Science (69) Religion (37) Language Arts (24) Vocational Education (16)	image (645) text (477) physical object (119) sound (72) moving image (22) Interactive Resource (17)	Midwest U.S. (general region) (209) Southwest U.S. (general region) (193) Southern U.S. (general region) (173) Pacific Coast U.S. (general region) (120) Mountain Region U.S. (general region) (71) Northeastern U.S. (general region) (55)
View all subjects...	View all objects...	View all places...

also browse by: [Title](#) | [Hosting Institution](#)

[Search Collections Only](#)

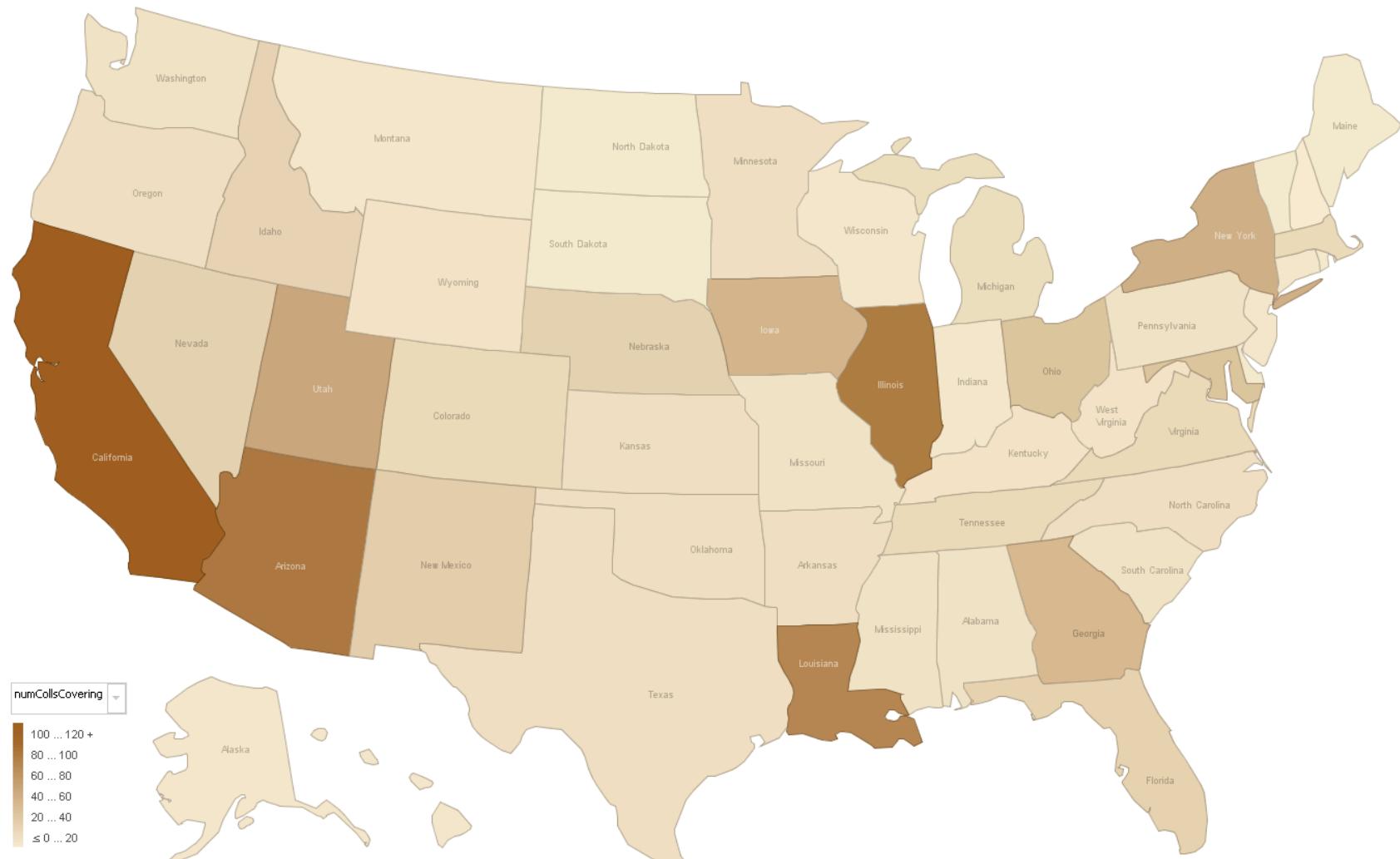
[Suggest a collection](#)

Currently 901 collection records. Last record updated On 10/20/2010.

[Collection Development Policy \(pdf\)](#)

Broad geographic coverage

Accelerated growth through engagement with state libraries



Map data © 2007 ESRI

Moving beyond size and search

- Can we improve how we collect collections by applying knowledge of:
 - how scholars build and engage with research collections.
 - research library collection development practices - conspectus-style subject evaluation.
- Can we begin to realize the collective value of the DCC aggregation by:
 - uncovering new, cohesive areas of content as they build up.
 - systematically building cohesive units of value to users.

Scholarly work with collections

special collections	unpublished, primary materials
intentionally created wholes	texts and images
dense units of content	known repositories browsing
evidence for inquiry	long term personal collections exploration and gathering
records in context	deep study over time

... Brockman et al. · Brogan · Buchanan et al. · Buckland
Case · Currall, Moss & Stuart · Duff & Johnson
Ellis & Oldman · Lee · Palmer · Tibbo ...

Guiding concepts

Contextual mass – inquiry & collecting by humanities scholars

(Brockman, Palmer, Neumann, & Tidline, 2001)

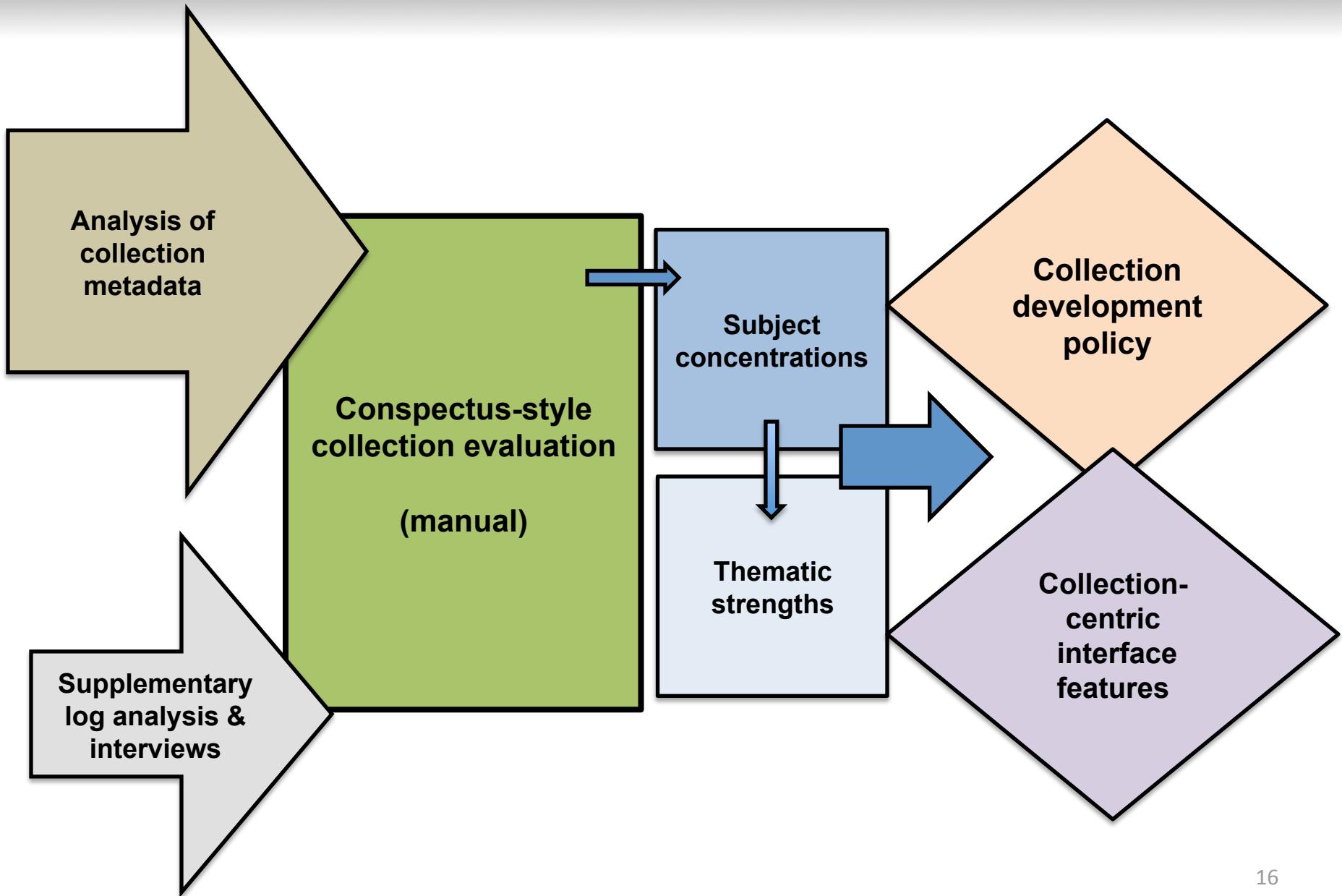
- size not a criteria for inclusion (bigger is not better)
- focus on meaningful interrelationships among distributed materials
- strive for dense, rich, cohesive groups of sources to support research

Thematic research collections – created by digital humanities scholars

(Palmer, 2004)

- primary sources and related materials sought and collected by scholars and special collections curators
- latent in the subject strengths emerging in OH

Aggregation strategy



Subject evaluations – concentrations & growth

Subject concentration	Collections with subject, April 2009	Collections with subject, April 2010	% of OH Collections with subject, April 2010
Military history	53	83	10.3%
Native American history	32	78	9.7%
Transportation history	33	48	6.0%
Asian American history	41	44	5.5%
African American history	29	33	4.1%
Mining history	17	26	3.2%
Exploration & travel history	19	23	2.9%

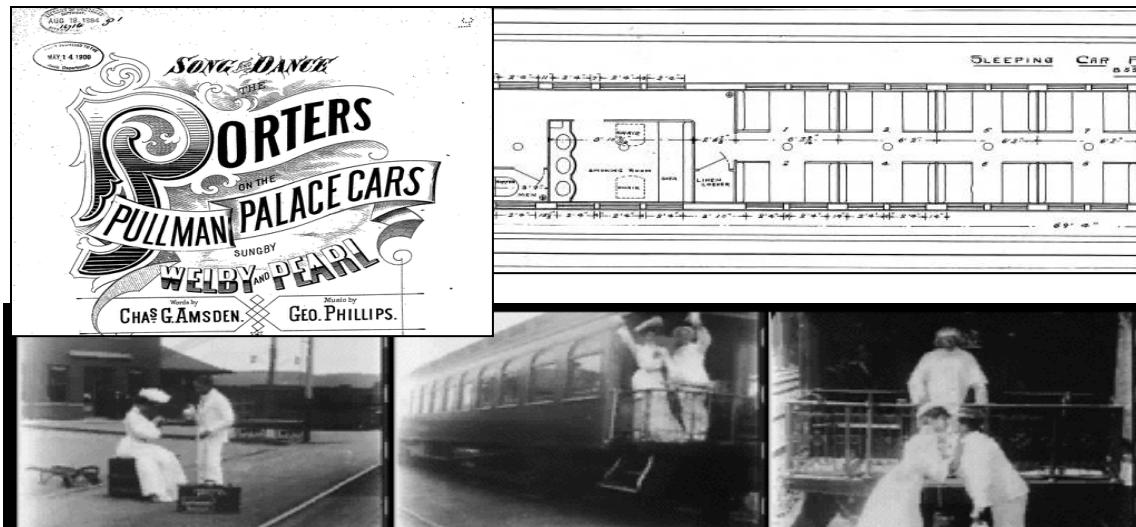
Subject concentration - contextual mass factors

Military History Strength Factors	Frequencies
Subject-focused collections	56
Items in subject-focused collections	19,388
Subject-inclusive collections	27
Items in subject-inclusive collections	48,085
Item types represented	39
U.S. states covered	24
Events covered – subject specific factor	12

Transportation thematic strength – Pullman

- Demonstrates value of small collections
 - 1500 blueprints - Newberry's Pullman Car Co. Collection
 - Scattered photos of Pullman porters from Pittsburgh, CT, WA
 - 40,000 glass negatives and prints, relevant to Pullman and era - Bain LOC,
 - Videos and songs referencing Pullman in pop culture - other LOC

1 rare book satirizing the Pullman Co from 17-item
Illinois Art and Literature digitized books collection



A STORY FROM PULLMANTOWN.

It was a quiet afternoon in October, 1893, that month of repose after fulfilled expectations, that month in which nature prepares for a health-giving rest during winter, for a calm concentration of forces, which are to burst into new life at a coming spring. Surely October, 1893, would also become a month of repose to the nations, who had exerted themselves to their utmost capacity. They had gathered together what brains had thought and hands achieved, ever since man first lifted a stone; they had built from it a world of art and beauty, of intellect and skill, till they themselves stood dazed by its splendor.

Making subject concentrations more explicit

Transportation History experimental interface
Enhanced access to a concentration with relatively low investment

Opening History: Transportation



Transportation in U.S. history and culture: a portal to thousands of history resources from libraries and museums across the country. Opening History: Transportation is part of [Opening History](#).

Browse by
[Title](#) [Subject](#) [Type](#) [Date](#)

What's Here?
122,108 Items
22 Collections
14 Institutions

Featured Content
Baltimore and Ohio Railroad
May 24, 2010 by ozavalina
The first revenue trains in the United States began service on the Baltimore and Ohio Railroad between Baltimore, Maryland and Ellicott's Mills, Maryland, 180 years ago, on May 24, 1830. This James Benney III's 1889 photograph of Baltimore and Ohio Railroad passenger platform in the downtown Pittsburgh, PA, is courtesy of Historic Pittsburgh Image Collection. [...] 

Geographic Region

Powered by Google, INET, LeadDog Consulting <[View Terms of Use](#)>

[View IMLS DCC Transportation Collection in a larger map](#)

Search
trains 

Browse by
[Title](#) [Subject](#) [Type](#) [Date](#) [Collection](#)

Item Matches **Collection Matches**

Item matches for "trains"

Bulkeley Bridge construction
Created by null
View original at University of Connect
View all matching items from Connec

[Image] Thirteenth National Ar
Created by null
View original at Library of Congress V
View all matching items from Panorai

[Image] [Men with Trains]
Created by Anderson, George Edward
View original at Brigham Young Unive
View all matching items from George

Guided Search

Time Period

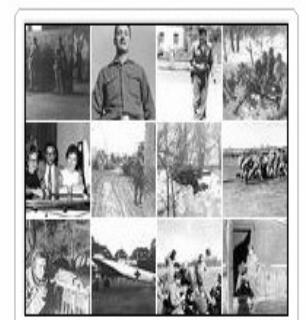
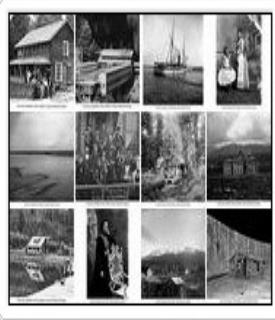
- 1910-1919 (55 items, 5 collections)
- 1900-1909 (37 items, 4 collections)
- 1920-1929 (33 items, 4 collections)
- Early 20th century (23 items, 2 collections)
- 1940-1949 (22 items, 5 collections)

+ More [Clear](#) [Filter](#)

Type

- photograph (33 items, 4 collections)
- negative (13 items, 1 collections)
- film negative (9 items, 1 collections)
- slides (6 items, 1 collections)

Extending access via Flickr



Flickr Feasibility Study

- Developing metadata, workflows, policy, technical routines
- Differences in content and representation based on type of collection and institution
- Weekly batch uploads of 100-200 (distributed across collections)
- Assessing public interaction with content
- Evaluating value of service and participation

Thematic Flickr groups using IMLS DCC images

- Cologne – Black and White
- Emmett Kelly Sr or Jr - Weary Willie
- Vintage Photos from World War II
- Vintage Monocle
- Legendary Adventures Guild
- Ford Farlaine/Torino
- Worldwide Ship Spotters (WSS)
- Hairygit's Nature Group
- Lighthouses of the World
- Clan Ross
- Too Tuff To Die! Hop Picking
- Hop Picking
- Bitterlake Seattle
- I, Magnin, & Co.
- Ticket Booth
- Seattle Dwellings
- Mail Pouch
- Bomarc Missile
- King County Creative Commons
- Roosevelt and Ravenna (Seattle)
- Snow Plows and Snow Removal Equipment
- Diamond T Trucks

For cultural heritage . . .

Principled collecting of collections valuable alternative to search and retrieval of scattered items or opportunistic aggregations.

- context of original curated collections
- links to complementary collections

Conspectus-style evaluation provides systematic process for evaluating and enhancing aggregations of research collections.

- Identify coherent bodies of emergent content
- contributions of small collections not obscured

Nationally scoped research collections can be developed out of subject concentrations through targeted recruitment.

In the scientific data arena

- Complex set of activities, dependencies, constituencies
 - Domain researchers and informaticists often leading development
 - Data and research centers with deep down-stream knowledge
 - Funding agencies and universities invested in data assets - producing competitive science; data management plans

New demands for data services requires libraries to collaborate with range of data experts and initiatives in domains, IT, data centers, etc.

We have much to contribute, and oh so much to learn.



Data Conservancy

JOHNS HOPKINS
UNIVERSITY

PI, Sayeed Choudhury,

Tim DiLauro et al., Sheridan Libraries

CoPIs and Partners	
<i>Carl Lagoze</i>	Cornell University
<i>Mary Marlino</i>	National Center for Atmospheric Research (NCAR)
<i>Carole Palmer</i>	CIRSS, GSLIS, University of Illinois at U-C
<i>Paddy Patterson</i>	Marine Biological Laboratory
<i>Chris Borgman</i>	University of California Los Angeles
<i>Ruth Duerr</i>	National Snow and Ice Data Center
<i>Mark Evans</i>	Tessella, Inc.
<i>Eileen Fenton</i>	Portico
<i>Sandy Payette</i>	DuraSpace / Fedora Commons

Aims and approach

Integrated and comprehensive **data curation strategy**

- to collect, organize, validate, and preserve data
- to address grand research challenges that face society

Network of expertise: domain and data scientists, information and computer scientists, librarians, engineers, enterprise experts.

Infrastructure builds on and connects **existing exemplar initiatives**

- deep engagement with scientists
- extensive experience with large-scale, distributed systems.

Astronomy as an exemplar community

Success in **data standards**, practices, documentation, and associated **services**

Ingest astronomy data **into preservation archive**,
connect data to existing services used by astronomers.

Demonstrate utility of hosting data in environment that supports
existing scientific capabilities in a sustainable manner.

Scope to include: life sciences
 earth sciences
 social sciences



Science and library based hubs

Marine Biological Laboratory

- Encyclopedia of Life - taxonomic organization, ontology indexing
- species identification queries for climate change analyses



National Snow & Ice Data Center

- extensive sensor network, fieldwork, aircraft and satellite data
- access node on the DC network, test bed for distributed services



National Center for Atmospheric Research

- civic decision making and climate science in megacities



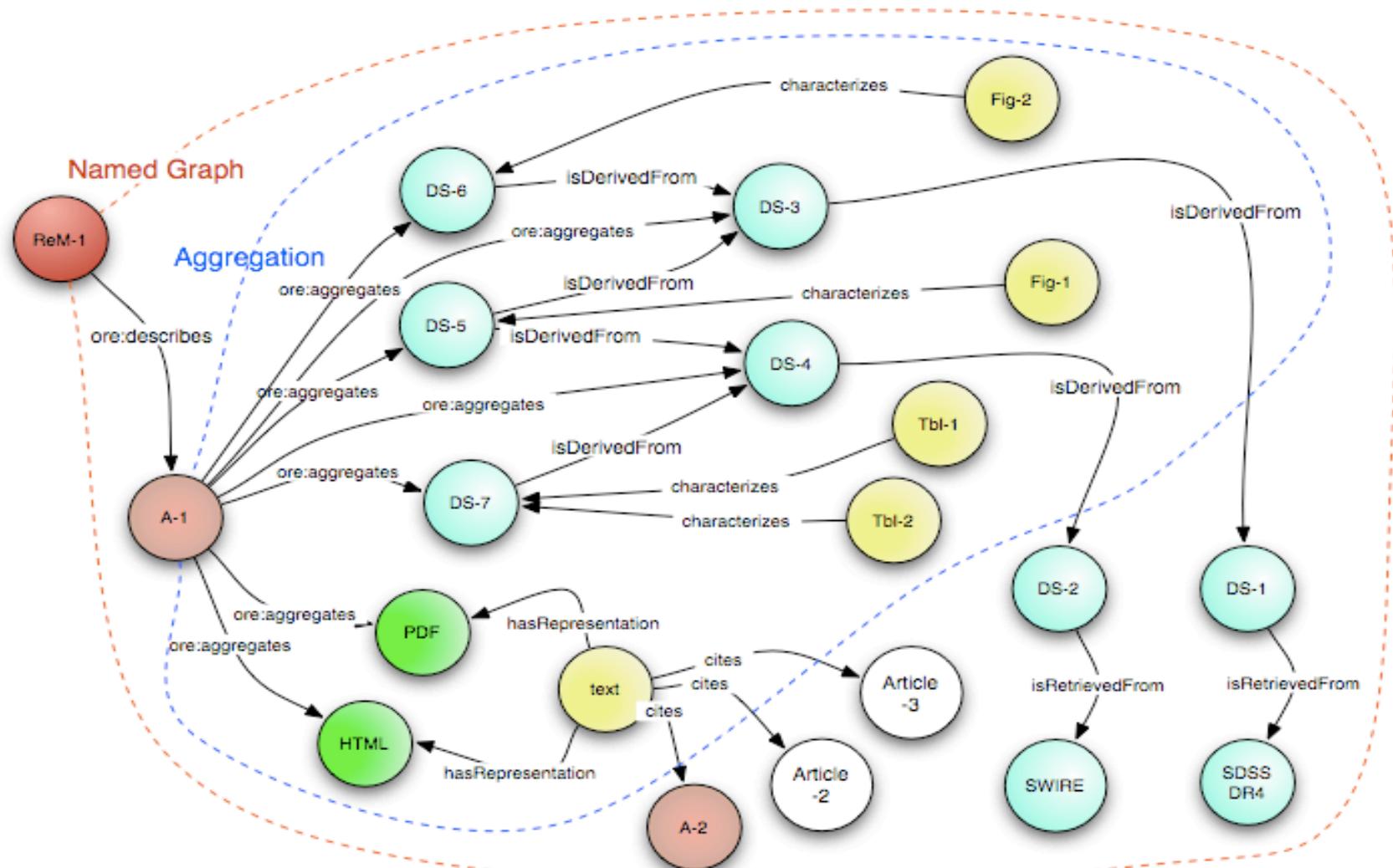
Cornell University Library

- DataStar - promotes archiving to disciplinary data centers
- arXiv eprints service capability for linking research data with publications

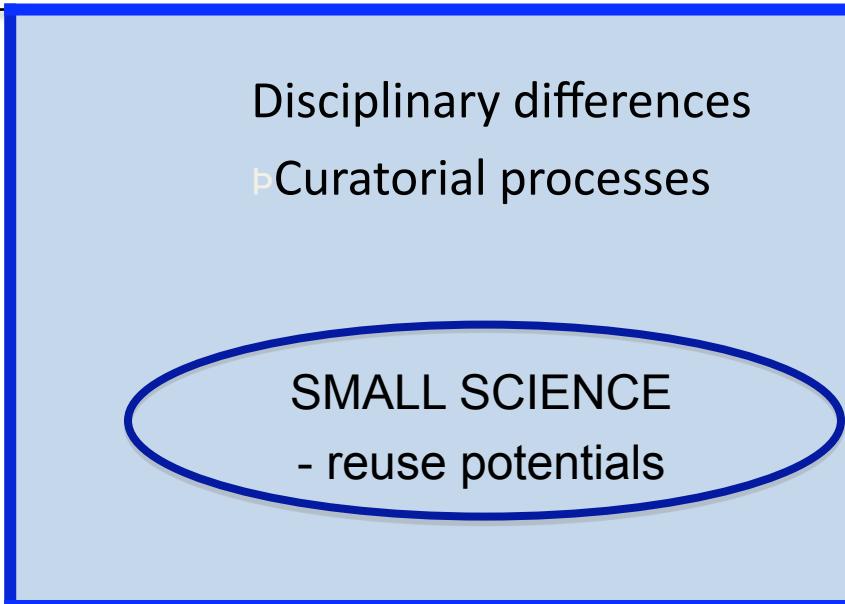


Resource map for astronomy compound object

OAI-ORE (object exchange & reuse) protocol



Use cases, data practices & curation processes

	Astronomy	Life Sciences	Earth Sciences	Social Sciences	
NCAR		Task-based design Use cases, system requirements			
UCLA	Ethnography, deep case study	Disciplinary differences Curatorial processes			ILLINOIS
		 SMALL SCIENCE - reuse potentials			

Small science is big, and poorly curated

12,025 NSF grants awarded in 2007 = \$2,865,388,605

	20%	80%
Number of Grants	2405	9621
Total Dollars	\$1,747,957,451	\$1,117,431,154
Range	\$300,000 - \$38,131,952	\$579 - \$300,000

Top 254 grants received 20% of the total awarded

(Heidorn, 2009)

Formalizing data concepts

- Conceptual precision and accuracy necessary for the reliable data curation and integration across disciplines
 - dataset, version of a dataset, part of a dataset, collection of datasets, data, granule, datum, etc.
 - representation levels (data, representation, encoding, format, etc.)
 - related problems in measurement

Research informing education



Doctoral research awards
2011 -

**Data
Curation
in Research
Centers**

Masters in bioinformatics
2006 -

**Biological
Information
Specialist**

MSLIS concentration in data curation
sciences, 2006 -
humanities, 2008 -

**Curation
In the
Humanities**

**Summer
Institutes**

In service professional
development
2008 -



Research libraries are mission-ready

- *Support research, scholarship, teaching*
- *Advance development & transmission of knowledge*
- *Preserve our intellectual heritage*
- *Contribute to the common good*
- *Promote the development of models, standards and infrastructure*
- *Develop distinctive collections, services, staff and facilities*
- *Leaders, collaborators, and innovators, to enable our communities—locally and beyond*

Many models and processes for remix

research and national libraries, consortia, CRL ...

union catalogs, bibliographies of bibliographies

collocation, just-in-time collecting ...

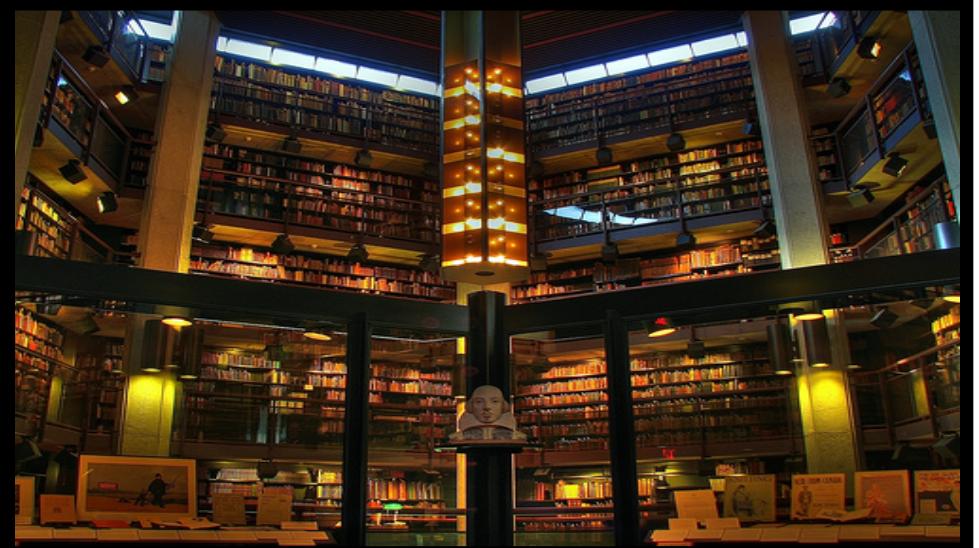
The stacks

Flickr users: stancia, rh creative commons



Special collections

flickr.com/photos/001fj/2907653323/



Profession is well positioned, but field is in its infancy

Compatible institutional and human infrastructure,
expertise, principles, and commitment.

Can leverage deep knowledge of the bibliographic universe, digital
libraries, and scholarly user communities.

But that professional knowledge began generations ago.

Studies of use of scientific literature emerged in 1960s, began in
earnest in 1980s.

Just beginning to investigate, document, and respond to
the current complex e-research collection and service environment.

Questions, comments, discussion

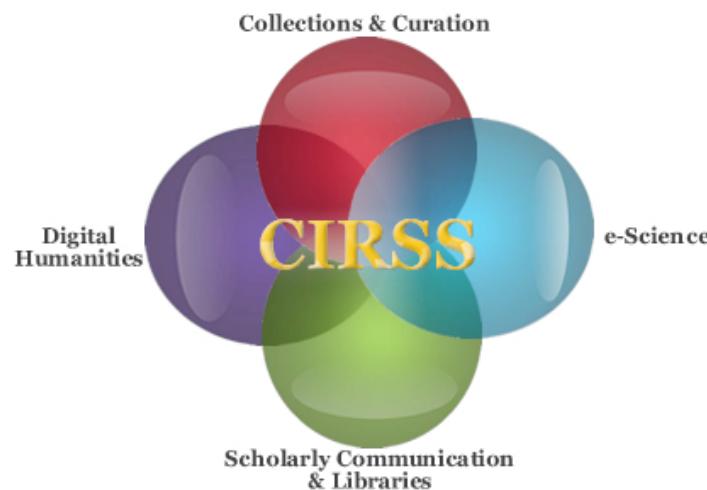
clpalmer@illinois.edu

IMLS DCC: <http://imlsdcc.grainger.Illinois.edu>

Data Conservancy: <http://DataConservancy.org>

<http://cirss.lis.uiuc.edu/>

Center for Informatics Research in Science and Scholarship



References

- Palmer, Carole L., Zavalina, Oksana, & Fenlon, Katrina (2010). Beyond size and search: Building contextual mass in digital aggregations for scholarly use. Proceedings of the ASIS&T Annual Meeting (Pittsburgh, PA, Oct. 22-27).
- Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. (2001). *Scholarly Work in the Humanities and the Evolving Information Environment*. Washington, DC: Digital Library Federation/Council on Library and Information Resources.
- Brogan, M. (2006). *Contexts and Contributions: Building the Distributed Library*. Washington, DC: Digital Library Federation/Council on Library and Information Resources.
- Buchanan, G., Cunningham, S. J., Blandford, A., Rimmer, J., & Warwick, C. (2005). Information seeking by humanities scholars. In *Proceedings of the European Conference on Digital Libraries* (Vienna, Austria, Sept. 18-23, 2005), 218-229.
- Buckland, M. K. (1999). *Library Services in Theory and Context*. (2nd ed.). New York: Pergamon Press.
- Case, D. O. (1991). The collection and use of information by some American historians: A study of motives and methods. *Library Quarterly*, 61, 61-82.
- Currall, J., Moss, M., & Stuart, S. (2004). What is a collection? *Archivaria*, 58, 131-146.
- Duff, W. M., & Johnson, C. A. (2002). Accidentally found on purpose: Information-seeking behaviors of historians in archives. *Library Quarterly*, 72(4), 472-496.
- Ellis, D., & Oldman, H. (2005). The English literature researcher in the age of the Internet. *Journal of Information Science*, 31(1), 29-36.
- Lee, H. L. (2000). What is a collection? *Journal of the American Society for Information Science*, 51 (12), 1106-1113.
- Palmer, C. L. (2004). Thematic research collections. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A Companion to Digital Humanities* (pp. 348-365). Oxford: Blackwell.
- Tibbo, H. (2003). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *The American Archivist*, 66(1), 9-50.

Image attributions

- Philips, G. (1884). The Porters on the Pullman palace cars. In ***Music for the Nation: American Sheet Music, 1820-1860, 1870-1885*** digital collection. Retrieved from: <http://hdl.loc.gov/loc.music/sm1884.16714>.
- Pullman Company (1898). Sleeping Car. Drawing. In ***Pullman Company Archives, Newberry Library*** digital collection. Retrieved from:
[http://collections.carli.illinois.edu/cdm4/results.php?
CISOOP1=any&CISOFIELD1=CISOSEARCHALL&CISOROOT=/
nby_pullman&CISOBOX1=Company](http://collections.carli.illinois.edu/cdm4/results.php?CISOOP1=any&CISOFIELD1=CISOSEARCHALL&CISOROOT=/nby_pullman&CISOBOX1=Company).
- Bech-Meyer, N. (1894). A Story from Pullmantown. C.H. Kerr. In ***Illinois Art and Literature*** digital collection. Retrieved from:
<http://hdl.handle.net/10111/UIUCOCA:storyfrompullman00bech>.
- Edison, T. A. (1903). A romance of the rail. Film. In ***Early Motion Pictures, 1897-1920*** digital collection. Retrieved from:
<http://hdl.loc.gov/loc.mbrsmi/edmp.2174>.

6th International Digital Curation Conference

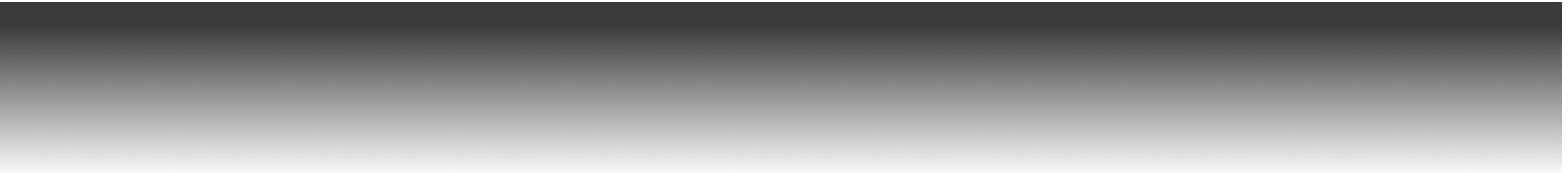
Chicago, IL
Dec. 6-8, 2010

hosted by
CIRSS / GSLIS

in partnership with
Digital Curation Centre, UK



- pre-conference Research Data Workforce Summit
 - post-conference LIS Research Summit



Data curation is . . .

The **active and on-going** management of (research) data
through its **lifecycle of interest** and usefulness
to scholarship, science, and education.

Functions

- enable discovery and retrieval
- maintain data quality
- add value
- provide for re-use over time
- archiving
- preservation

Tasks

- appraisal and selection
- representation
- authentication
- data integrity
- maintaining links
- format conversions

Convergent user data

Transaction Logs

User interactions	Page views
Viewing collection metadata records	1,760
Viewing item metadata records	368
Collection browse:	2,939
Subject browse	953
Geographic browse	533
Project browse	502
Object type browse	487
Institution browse	311
Collection title browse	153
Item browse	4,388
Collection search	880
Item search	1,860

Sessions with historians

- Collection information useful for navigation, interpretation, and assessment of content.
- OH navigated as central hub,
 - context of collection,
 - exploration of related collections at host institutions.
- Provenance information at collection level is critical for assessment of value.

Core curation content

Foundations of Data Curation

- Digital Data
- Scholarly Communication
- Lifecycles
- Collections
- Infrastructures & Repositories
- Selection and Appraisal
- Metadata
- Standards & Protocols
- Archiving & Preservation
- Intellectual Property & Legal Issues
- Workflows; Data Re-use & Value
- Policy & Cooperative Alignments
- Scientific Information Work

Assignments:

20 cases developed this semester
Critiques of data management plans

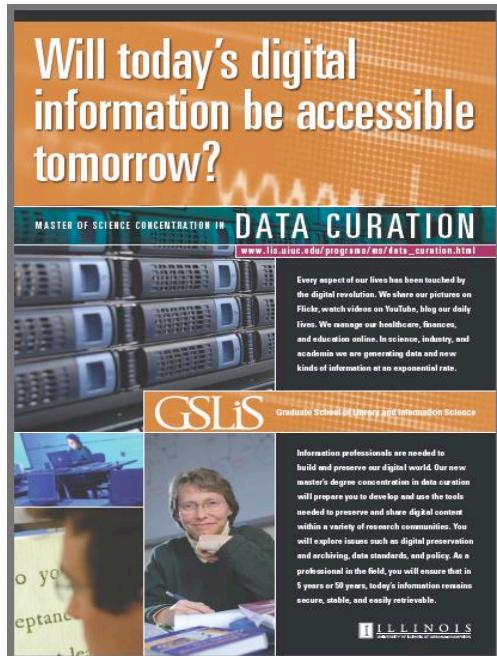
Digital Preservation

- Archival Theory & Diplomatics
- OAIS Reference Model
- Data Formats
- Digital Archival Objects
- Preservation Strategies:
- Emulation vs. Migration
- Authenticity, Integrity & Trust
- Evaluation & Value
- Digital Preservation & The Law

Assignments:

Planning Grant Application
Trusted Repository Assessment

Basic curation curriculum



Required Core Courses

Foundations of Data Curation
Digital Preservation
Systems Analysis & Management

Selected Electives – require 2, recommend 4

Metadata in Theory & Practice
Information Modeling
Ontologies in Natural Science
Foundations of Information Processing
Digital Libraries: Research & Practice
Representing & Organizing Info Resources

Internships: JHU, Smithsonian, NLM, USDA, Purdue, NSIDC . . .

MASTER OF SCIENCE CONCENTRATION IN **DATA CURATION**