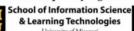
QUALITY METADATA: REINVENTING SERVICE

MISSOURI LIBRARY ASSOCIATION 2014 CONFERENCE OCTOBER 10, 2014

FELICITY DYKAS
HEATHER LEA MOULAISON
KRISTEN GALLANT

This project funded in part by a grant from the MU Wallace Foundation.









WHY DOES QUALITY MATTER TO SERVICE?

"Metadata ultimately supports retrieval and use, but if records do not contain quality metadata, the content that they describe will be hidden from end-users" (Moulaison, 2014, p.4).

→ Quality metadata IS important to Missouri libraries.

WHAT IS QUALITY?

- "Quality judgments are by definition subjective and incomplete" (Conway, 2011, p. 299)
- · Not one size fits all; based on needs
- Some characteristics (Bruce and Hillmann, 2005, p. 243) ...
 - completeness
 - accuracy
 - provenance
 - · conformance to expectations
 - · logical consistency and coherence
 - timeliness
 - accessibility

1

AGENDA

- Background and key concepts
- Study
- Method
- Results
- Implications for users
- Repositories in Missouri
- Reflections
- Next steps

DIGITAL REPOSITORIES FOR LOCAL CONTENT

- Digital Libraries (DL) → Provide storage, preservation, and access for digitized and born digital content.
- Institutional Repositories (IR) → Primarily at research institutions; hold electronic theses and dissertations (ETDs), faculty publications, etc.

5

DIGITAL REPOSITORY SOFTWARE

Depending on the type of repository, a wide selection of software products are available for data creation and sharing.

Characteristics: open access, customizable, hosted or installed locally, search engine friendly, support variety of media types

Some of the more common products being used are:

• <u>Dspace</u>

- Digital Commons / bepress
- CONTENTAM

To implicate

Internal Control Co

THE DIRECTORY OF OPEN ACCESS REPOSITORIES - OPENDOAR

Is an international directory of academic, open access repositories.

- Authoritative
- Repository lists
- Search for repositories
- Search for repository content

http://www.opendoar.org/

7

WHAT WE DID

We surveyed 50 randomly-selected U.S. digital repositories that were listed in the OpenDOAR directory. There are a total of 328 U.S. repositories included in OpenDOAR.

We asked administrators about their repositories and about practices and approaches for creating and evaluating metadata in digital repositories.

Conducted May-June 2014.

RESULTS

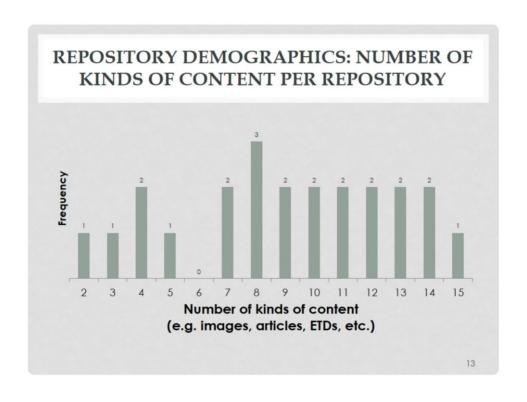
- Respondents: 23
- Repository demographics
- Metadata practices
- Metadata creation and environment
- Quality control

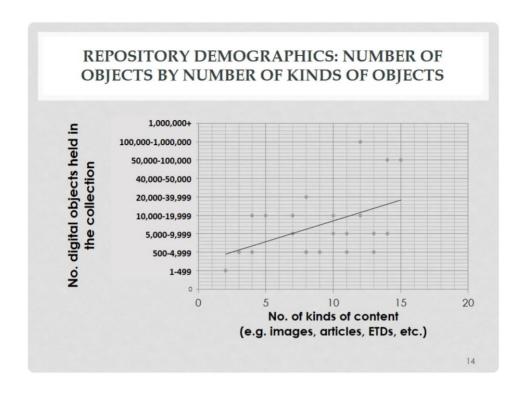
10

REPOSITORY DEMOGRAPHICS: REPOSITORY SOFTWARE

Repository/Software	Number Using (N=23)	%
DSpace	10	43%
Digital Commons /	6	26%
bepress		
Fedora	3	13%
Ex Libris DigiTool	2	9%
Hydra	2	9%
Islandora	1	4%
Omeka	1	4%
CONTENTdm	1	4%
Locally developed	2	9%
software		
Other	2	9%

Kinds of Content	Number Using (N=23)	%	I
Images	18	78%	REPOSITORY KINDS (
Individual articles	18	78%	スラ
Student projects	18	78%	
Photographs	17	74%	ZS
ETDs	15	65%	
Presentations	14	61%	OSITO
Reports	14	61%	
Digitized books	14	61%	2
Video	14	61%	F
Journals	14	61%	0 7
Newspapers	12	52%	0 5
Audio	11	48%	70
White papers	9	39%	
Research data/datasets	8	35%	
Born digital books	6	26%	
Databases	3	13%	CONTENT
Websites	1	4%	IC
Other: government documents	1	4%	S
Other: University Archive items	1	4%	
Other: collective bargaining agreements	1	4%	12





METADATA PRACTICES: CONTROLLED VOCABULARIES (6 DID NOT RESPOND)

ibrary of Congress Subject Headings (LCSH) Name authority file (NAF) Art & Architecture Thesaurus (AAT)	15 4 3	88% 24%
		24%
Art & Architecture Thesaurus (AAT)	2	
	3	18%
Medical Subject Headings (MeSH)	2	12%
hesaurus for Graphic Materials (TGM)	2	12%
ibrary of Congress Genre/Form Terms (LCGFT)	2	12%
Getty Thesaurus of Geographic Names (TGN)	1	6%
Union List of Artist Names® (ULAN)	0	0%
aceted Application of Subject Terminology (FAST)	0	0%
Other	3	18%

Encoding Schema	Responses (N=23)	%
Dublin Core	12	52%
Qualified Dublin Core	11	48%
MODS	6	26%
MARC	4	17%
PREMIS	3	13%
EAD	2	9%
ETD-MS	1	4%
MIX	1	4%
MADS	0	0%
CDWA	0	0%
PBCore	0	0%
TEI	0	0%
TextMD	0	0%
VRA	0	0%
Other	6	26%

METADATA CREATION ENVIRONMENTS					ΓS	
Team Member	Creates DESCRIPTIVE metadata	% out of 19	Creates ADMINISTRATIVE metadata	% out of 19	Reviews metadata	% out of 19
Librarian (master's level)	16	84%	14	74%	15	79%
Paraprofes- sional	10	53%	3	16%	6	32%
Administrator (outside department)	7	37%	3	16%	3	16%
Department head	3	16%	3	16%	4	21%
Subject specialist	4	21%	2	11%	3	16%
Student worker	4	21%	2	11%	0	0%
Volunteer	2	11%	1	5%	1	5%
IT	0	0%	2	11%	2	11%

METADATA CREATION ENVIRONMENT: BEST PRACTICES DOCUMENTATION

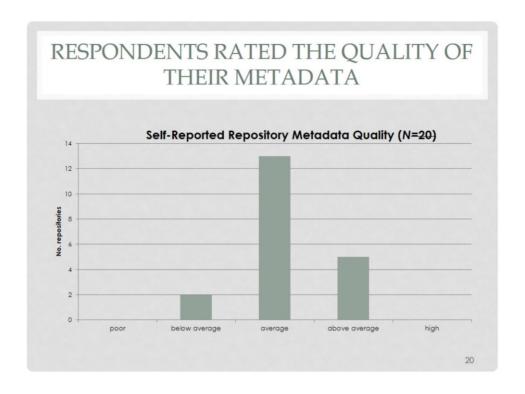
Best Practices	Number (N=18)	%
Best practices: homegrown	11	61%
Best practices: RDA	4	22%
Best practices other	1	6%
Best practices: Western States	1	6%
None mentioned	6	33%



8

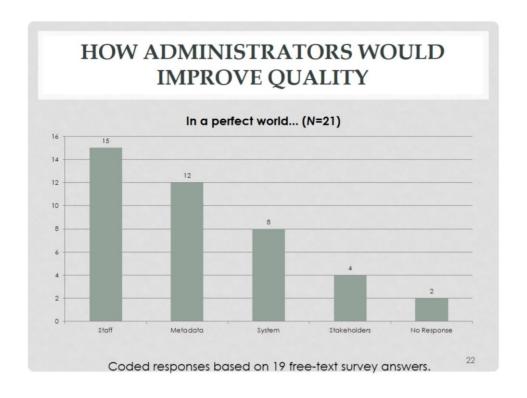
METADATA CREATION ENVIROMENT: OTHER RESOURCES USED IN METADATA CREATION

Other Resources	Number Using (N=18)	%
OCLC Connexion	8	44%
oXygen XML editor	4	22%
RDA Toolkit	4	22%
Cataloger's Desktop	3	17%
Classification Web	3	17%
MARCedit	3	17%
VIAF	3	17%
id.loc.gov	2	11%
DublinCore Generator.com	1	6%
ORCID	1	6%
None mentioned	5	28%



Q.C. IN THE REPOSITORY

- In response to a question about quality control for collections or the overall repository, the following were mentioned:
- Developing workflows
- use of an application profile or core data dictionary
- authority control
- Templates
- · use of metadata schemes.



SAMPLE RESPONSES: HOW ADMINISTRATORS WOULD IMPROVE QUALITY

Staff: Hire somebody who has time to do this instead of tacking it onto the schedule of somebody who already has a full time position (#1)

- Metadata: I would have all records described in qualified Dublin Core (#28)
- All of the concerns (staff, metadata, system, stakeholders):

1. Much more outreach about it - particularly in the context of workflows that mean something to researchers - and thus more willingness from non-metadata librarians, such as liaison librarians, to learn some metadata basics that they could consult with researchers on. / 2. More controlled vocabularies & authority standards connected to our fields, so that users could select, rather than type in, what applies to their objects. / 3. Dedicated metadata expertise for the repository service as it develops (particularly strategic thinking around metadata) (#17)

OBSTACLES TO MAKING THE "PERFECT WORLD" SCENARIO HAPPEN

Response	No. Responses (N=20)	%
Time limitations and staff hours	17	85%
Repository software limitations	12	60%
Institutional priorities	7	35%
Skill levels of staff	7	35%
Nature of legacy data	5	25%
Concerns about interoperability	3	15%
Current standards meet user needs	2	10%
Lack of necessary tools	2	10%
Other	4	20% 24

BROAD APPLICATIONS FOR QUALITY METADATA BIG PICTURE PERSPECTIVES, OR, THE VIEW FROM 30,000 FEET

USER TASKS

- In the bibliographic universe, the conceptual model Functional Requirements for Bibliographic Records (FRBR) tells us that our metadata records should allow users to:
 - Find
 - Identify
 - Select
 - Obtain

library content. Quality metadata is how this happens.

26

FROM THE USER'S POINT OF VIEW:

- •Find
- Identify
- Select
- Obtain
- Quality metadata means retrieval
 - If users can't find electronic content, they don't know it's there
 - If users can't identify and subsequently select content based on the metadata, they won't know if it meets their needs
- Quality metadata enables sharing
 - Retrieving across a number of platforms is possible when the metadata is good
 - · OpenDOAR is a great example of this
- Quality metadata means access
 - Once users find content and evaluate it based on the metadata, they will be able to proceed to access it

REINVENTING PATRON SERVICES

- Quality metadata is key to a number of automated services libraries promote
 - · Federated Search/Discovery (i.e. retrieval)
 - Library discovery layers combining repository content with ILS content for search
 - · DPLA and the new STL hub!
 - Community Engagement
 - Soliciting digital content that appeals to a variety of stakeholders – and that content's metadata
 - · Serving up content through automated means
 - · "Library as platform"

28

REFLECTIONS

- Shareability is a major goal of metadata, with interoperability a dimension of quality that can be reported
 - Whereas high quality might be context-dependent, changing from repository to repository, shareable metadata is consistently interoperable across databases and systems
 - Yet not everyone is using the same best practices documentation
 - Hillmann (2008) discusses the differences between traditional cataloging and work done in digital libraries. She observes that in digital libraries, "few communities of practice have been able to define their needs as a community" (p. 68).
 - Given that repositories are not sharing records, the need for shared best practices does not occur at the repository level and the need for a community of practice may not be perceived as important.
 - On the other hand, a large number of respondents are using controlled vocabularies with at least some of their repository material.

REFLECTIONS, CONTINUED

Size Matters

- In comparison to the traditional holdings of large academic libraries, digital repositories are still small.
- 82% of respondents answering the question indicated that their repositories held fewer than 20,000 items, and 37% reported fewer than 5,000 items.

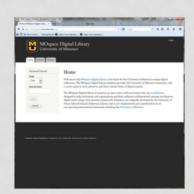
Content Is King

- The smaller repositories also reported having fewer kinds of content.
 - Stvilia and Gasser (2008) observe that large repositories may receive greater use than smaller repositories and their need for quality metadata may be greater, but, in what they call the "cycle of diminishing returns," larger repositories may have "greater difficulty in providing those metadata with limited resources as the metadata collection continues to grow and becomes increasingly diverse" (p. 67).

30

UNIVERSITY OF MISSOURI: MOSPACE





UNIVERSITY OF MISSOURI: MOSPACE

- Institutional Repository
 - Content is mostly born digital and includes electronic theses and dissertations, faculty papers and presentations, videos, podcasts, MU publications, and a few sets of research data
 - Dspace
 - · Qualified Dublin Core
 - OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)

- Digital Library
 - Content is mostly material that was digitized locally: books, maps, posters.
 Also we are starting to add born digital videos.
 - Islandora
 - MODS (Metadata Object Description Schema) and Dublin Core

32

REPOSITORY MANAGER PERSPECTIVE, PART 1

- We have great resources and want to make them available to the others outside MU
- The resources in our institutional repository are the work of faculty and students and we want to support and promote that work

REPOSITORY MANAGER PERSPECTIVE, PART 2

- Resources are tight
- What will help users to find, identify, select, and obtain resources in the MU digital repositories?
- This is a local repository. How closely do we need to adhere to shared standards?

34

IN PURSUIT OF ANSWERS ...

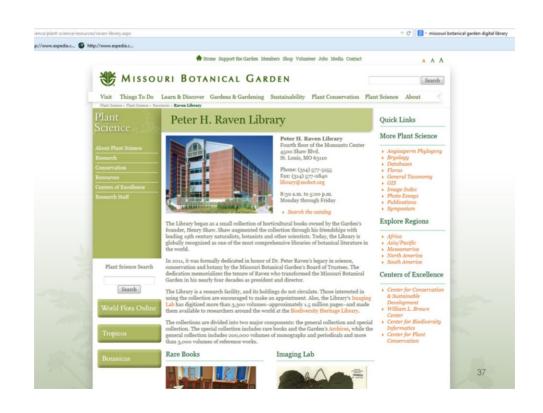
- Current research project
- Looked at other Missouri repositories
- Considered how metadata in the repositories travel and is used elsewhere
- Identified some next steps

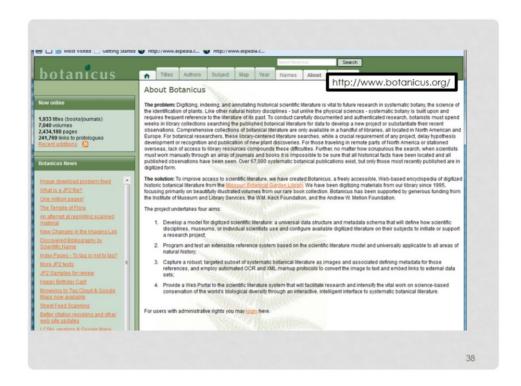
MISSOURI DIGITAL REPOSITORIES

- What metadata fields are included?
- What search options are available?
- Where else is the metadata used?

Quality criteria

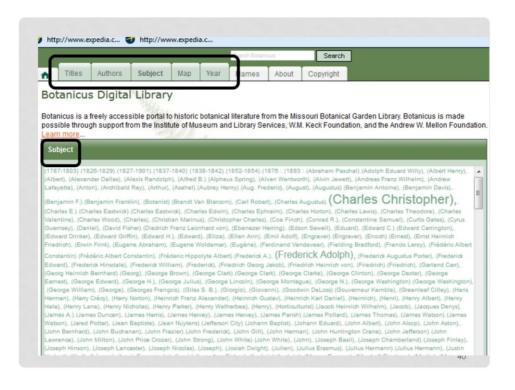
- Find
- Identify
- Select
- · Obtain





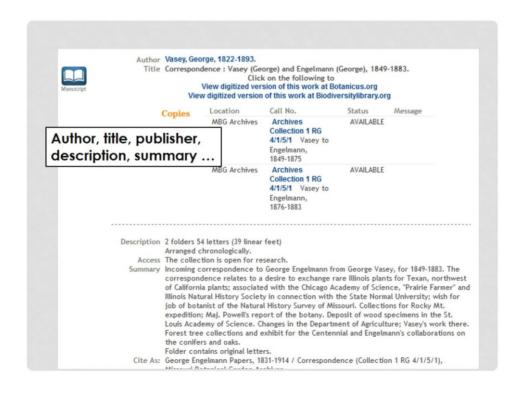
BOTANICUS: THE PROJECT UNDERTAKES FOUR AIMS:

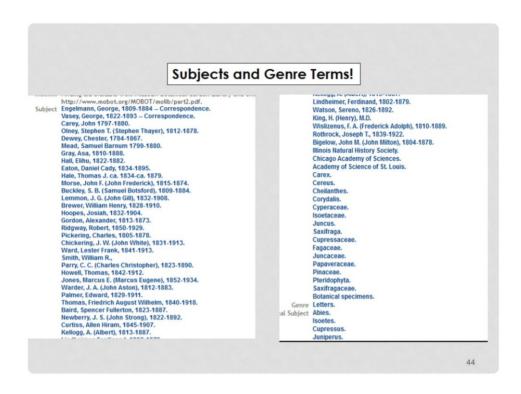
3. "Capture a **robust**, targeted subset of systematic botanical literature as images **and associated defining metadata** for those references, and employ automated OCR and XML markup protocols to convert the image to text and embed links to external data sets."

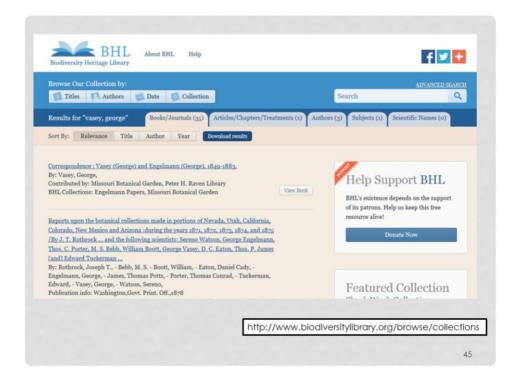


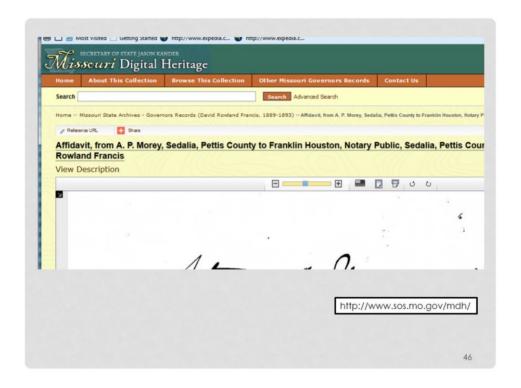


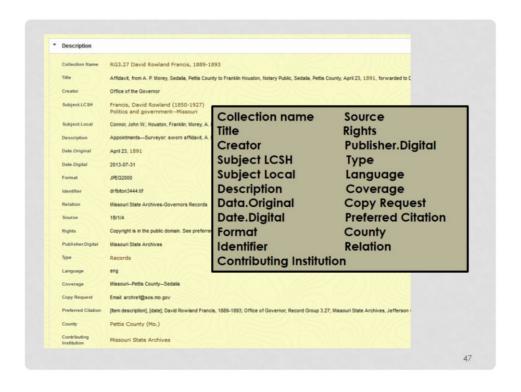




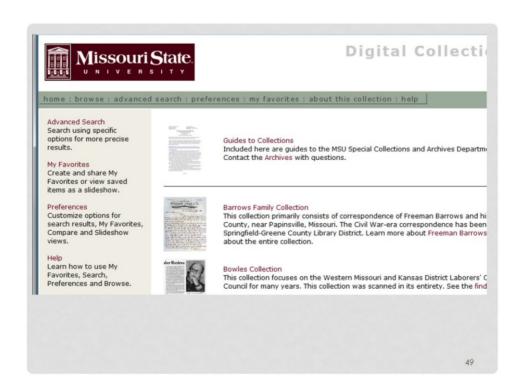


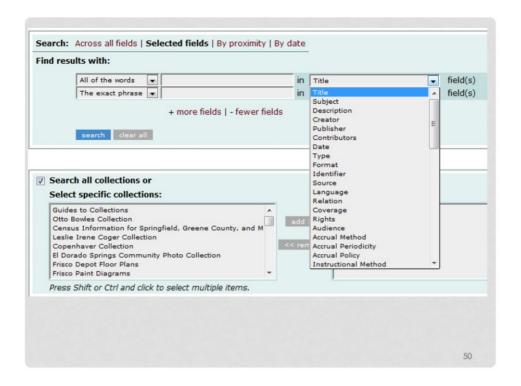


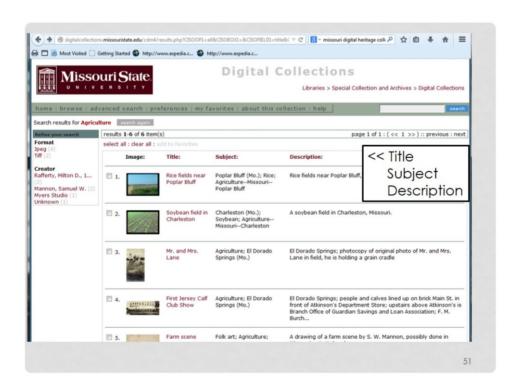


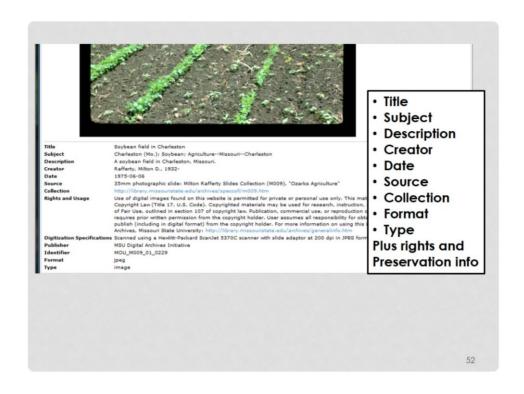


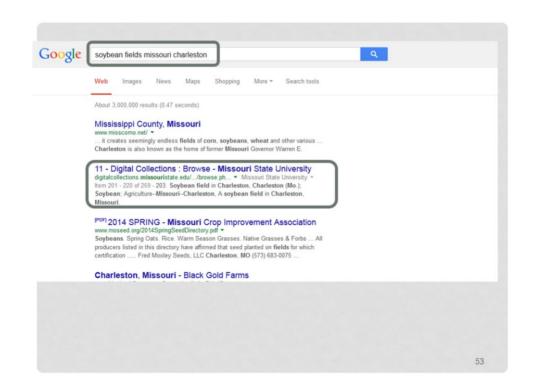


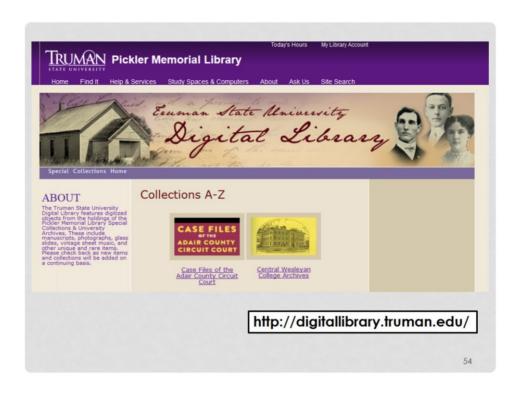












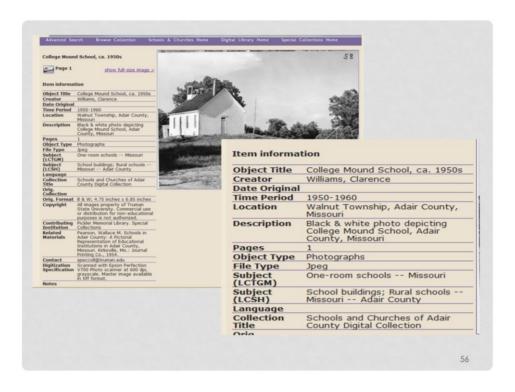
TRUMAN STATE DIGITAL LIBRARY

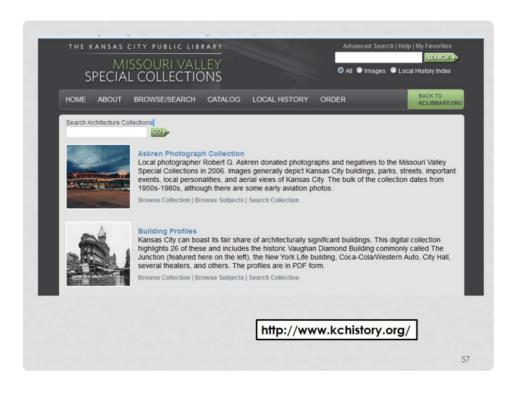
Browse options (varies by collection)

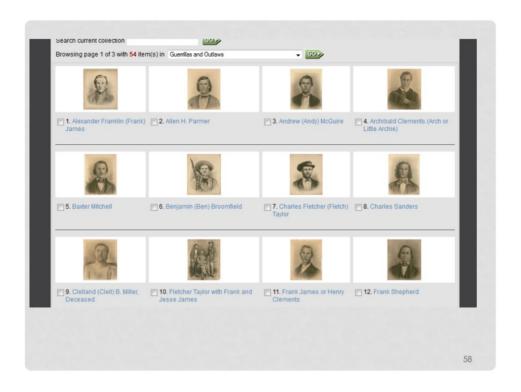
- Collection
- Title
- Author
- · Date
- Location
- Time period
- Category
- Composer
- Instrument

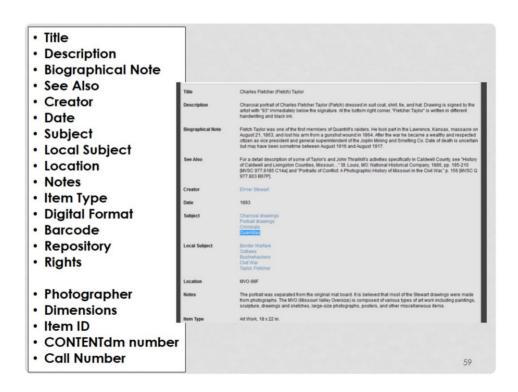
Advanced search

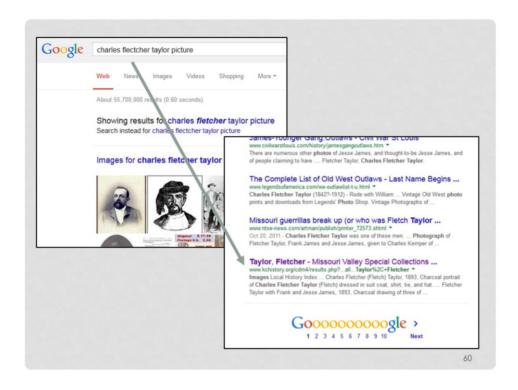
 Keyword for some or all words entered in the search box

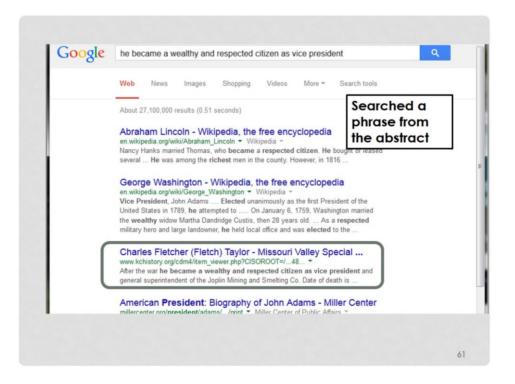




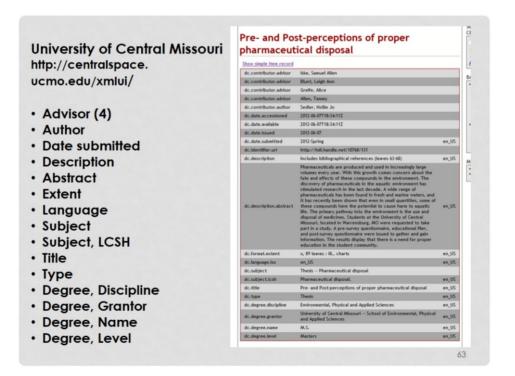


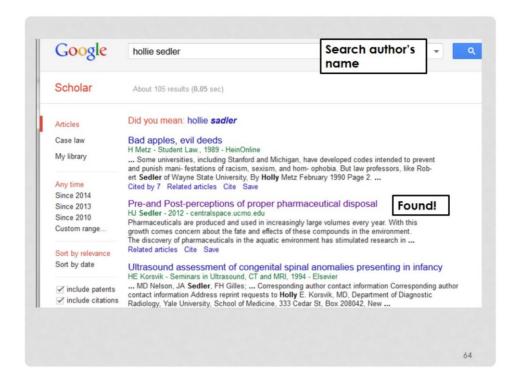


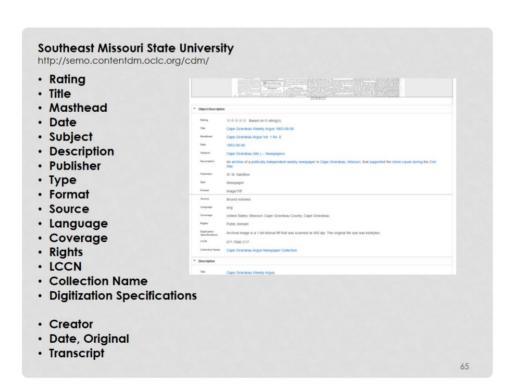


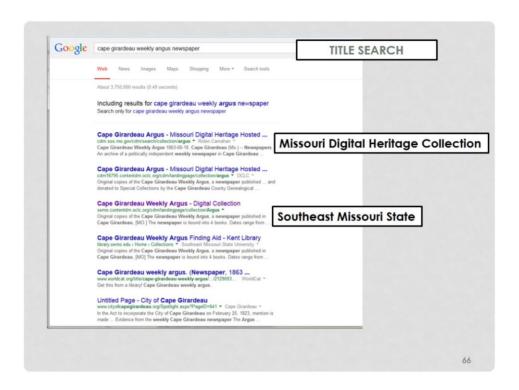


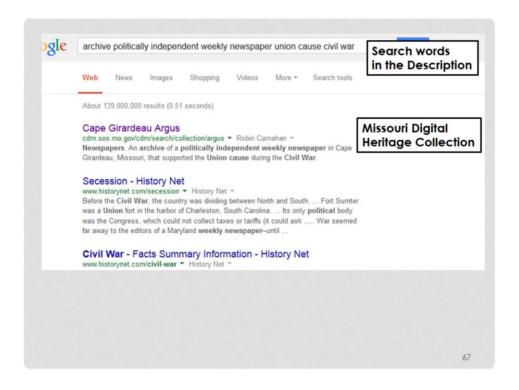


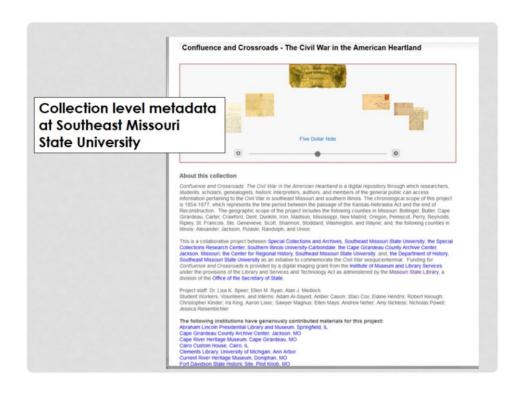


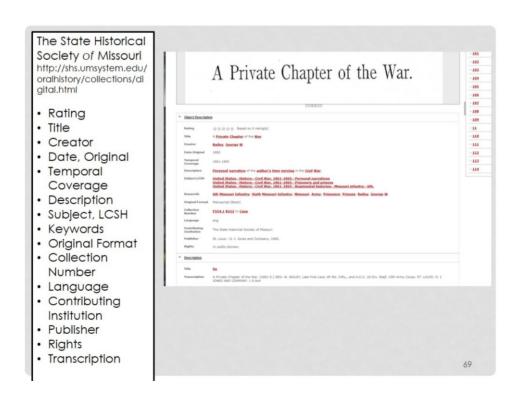


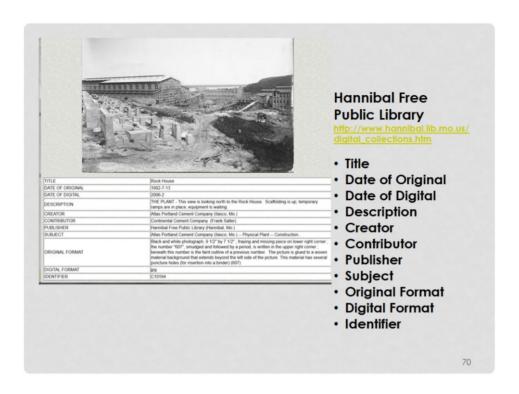


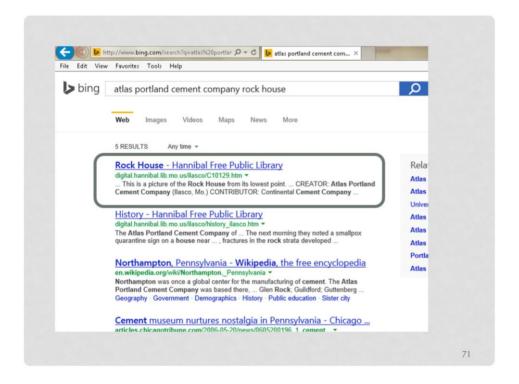


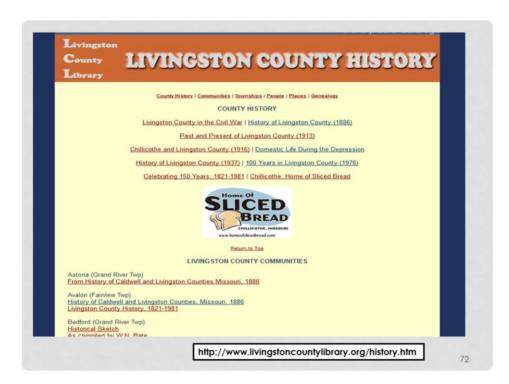




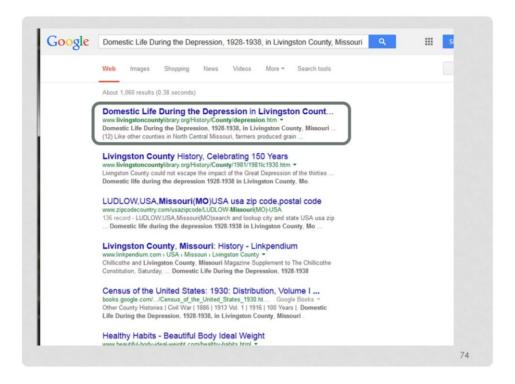








Domestic Life During the Depression, 1928-1938, in Livingston County, Missouri by James R. Rashan, 1974 A Research Paper Presented to The Faculty of the School of Arts and Sciences Central Missouri State University In Partial Fulfillment of the Requirements for the Depree Specialist, History by James R. Nashan, May, 1974 The Great Depression of the 1930's which created unemployment, business and bank failures, farm bankruptcy and relief lines was a failure of our economic system. September 3, 1929 has been used by some as the start of the depression, because this was the day the Big Bull Market reached its peak. From this high point, the market tumbled, (1) and America moved into a period of enomous economic and political transformation. (2) President Hoover took several steps to change the financial crisis. The Reconstruction Finance Corporation was established to aid banks and business in trouble with the intent of safeguarding sarings and securing worker's jobs. In the seconomy. Several of President Hoover's proposals were defeated by the Democratic Congress, but those that were passed were inadequate to resolve the financial crisis. (3) Franklin Roosevelt, campaigning in 1932, promised the nation a New Deal. Once he was elected, he began a process of experimentation to attempt to improve conditions in the nation. Hundreds of new policies were designed to pump money into the economy. Among these policies, some of the major ones were: The Affacultural Adjustment Act which provided help to farmers. The National Recovery Act for help in business. The Federal Emergency Relet Administration which provided help to farmers. The National Recovery Act for help in business. The Federal Emergency Relet Administration which provided help to farmers. The National Recovery Act for help in business. The Federal Emergency Relet Administration which provided help to the server end. The Vitoris Progress Administration, Cell Words Administration which provided place to the unemployed (4) The conomic set back created by th



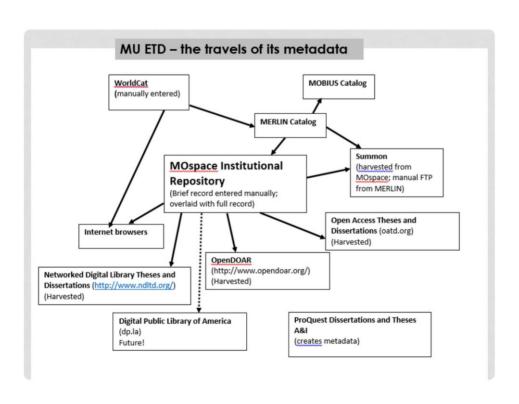
SUMMARY

Missouri cultural institutions are doing a great job cataloging material in their digital repositories!

- A variety of search options used with good metadata help users find their material
- Authors, titles, dates, descriptions, etc., help users identify whether the resource is the item for which they are looking
- Summaries and subject headings are very useful in selecting relevant material
- Controlled vocabularies are as being used extensively in local

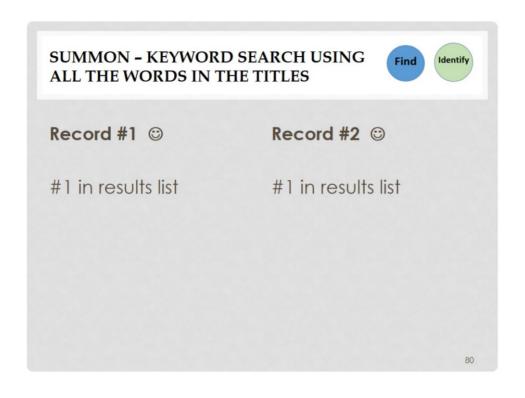
SUMMARY (CONTINUED)

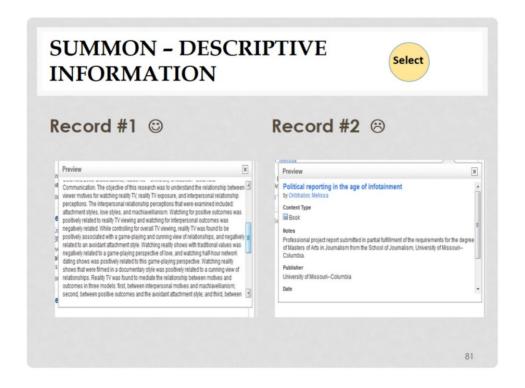
Even if you have a small collection and may not need the detailed metadata in-house, the addition of it will make your digital material findable by a broader audience via internet searches



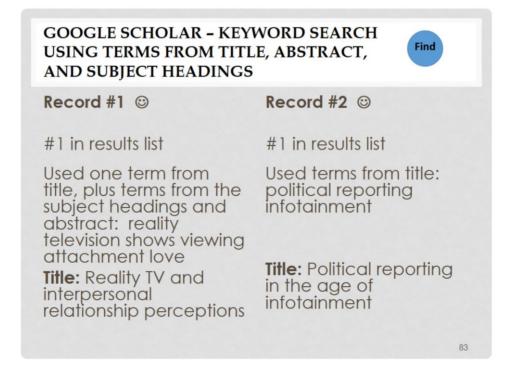


METADATA Record #1 Record #2 Abstract Advisor Author Author Collection Collection Date issued Date submitted Date issued Description Identifier, OCLC Date submitted Publisher Language Language code Subject, LCSH Subject, Keywords Publisher Title Type Degree department Degree grantor Degree level Degree name





GOOGLE SCHOLAR - KEYWORD SEARCH USING ALL WORDS IN TITLES Record #1 ③ Record #2 ④ Direct hit!!! Really TV and releparational relationship perceptions Really TV and releparational relationship perceptions Record #2 ⑤ Did not appear in the first four pages of the search for each results Stocking ratio fit valiety to you come of instance of the continued by these cores makes by the search for each fit of the search fit of the search for each fit of the search fi



OPENDOAR - KEYWORD SEARCH USING TERMS FROM TITLE, ABSTRACT, AND SUBJECT HEADINGS



Record #1 ©

Record #2 ©

#1 in results list

Used one term from title, plus terms from the subject headings and abstract: reality television shows viewing attachment love

Title: Reality TV and interpersonal

relationship perceptions

#1 in results list

Used terms from title: political reporting infotainment

Title: Political reporting in the age of infotainment

OPENDOAR - BROWSE DISPLAY





Record #1

Reality TV and interpersonal relationship perceptions

Identify: 😊

https://mospace.umsystem.edu/xmlui/handle/10355/5532

Select: (2)

Watching for positive outcomes was positively related to reality TV viewing and ... view of relationships, and negatively related to an avoidant attachment style. ... of love, and watching half-hour network dating shows was positively related to ...

Record #2

Political reporting in the age of infotainment

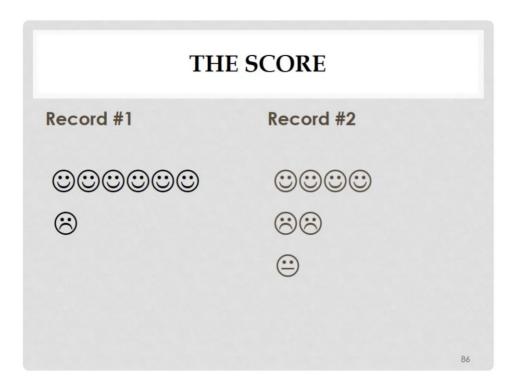
https://mospace.umsystem.edu/xmlui/handle/10355/43855

Identify: ©

Title: Political reporting in the age of infotainment. Author: Oribhabor, Melissa. Date: 2014.

Select: 😑

Publisher: University of Missouri--Columbia.



SUMMARY (CONTINUED)

- Detailed metadata is effectively used by search engines and other sites
- Detailed metadata can effectively be used by people to find, identify, and select resources
- Metadata standards in local digital repositories are important

REINVENTING METADATA, OR, APPLYING WHAT CATALOGERS KNOW

What is important?

- Standards
- Cross-compatibility; looking at how metadata is used in other databases and searches
- Using the FRBR user tasks as guidelines for ALL our metadata work
 - Find
 - Identify
 - Select
 - · Obtain
- We should catalog, not just for our local repositories, but also for the use of metadata by search engines and other repositories
- There is value in cataloging collections AND the items in the collections

88

NEXT STEPS

MU Digital Repositories

- Renewed commitment to full cataloging
- Explore various options, including FAST (Faceted Application of Subject Terminology)
- Become more aware of how metadata works outside MU

Research

- Continue to explore the need for shared standards for digital repositories
- Explore ways to create or get quality metadata in an environment with limited resources (systems, templates, etc.)
- Determine how to priorities metadata elements for the greatest benefit

QUESTIONS AND DISCUSSIONS

- Felicity Dykas
 - MU Libraries
- Heather Lea Moulaison
 - · MU iSchool
- Kristen Gallant
 - MU iSchool Graduate Student

REFERENCES

Bruce, T. R., & Hillmann, D. I. (2005). The continuum of metadata quality: Defining, expressing, exploiting. In D. I. Hillmann & E. L. Westbrooks (Eds.), Metadata in Practice (pp. 238-254). Chicago: American Library Association.

Conway, P. (2011). Archival quality and long-term preservation: A research framework for validating the usefulness of digital surrogates. Archival Science, 11 (3-4), 293-

Functional Requirements for Bibliographic Records (FRBR): Final report (1997 2009 rev) International Federation of Library Associations and Institutions

Hillmann, D. I. (2008). Metadata quality: From evaluation to augmentation. Cataloging & Classification Quarterly, 46(1), 65-80. DOI: 10.1080/01639370802183008

Moulaison, H. L. (unpublished manuscript). The expansion of the personal name authority record under RDA: Current status and quality considerations.

Stvilia, B., & Gasser, L. (2008). Value-based metadata quality assessment. Library & Information Science Research, 30(1), 67-74.

Contents of this presentation based in part on:

Moulaison, H. L., Dykas, F., & Gallant, K. (accepted: 2015, March). OpenDOAR repositories and metadata practices. *D-Lib*.

View a PDF of the survey instrument through the MOspace repository: