



Is Quality Metadata Shareable Metadata?

Sarah Shreeves

Ellen Knutson

University of Illinois at Urbana-Champaign

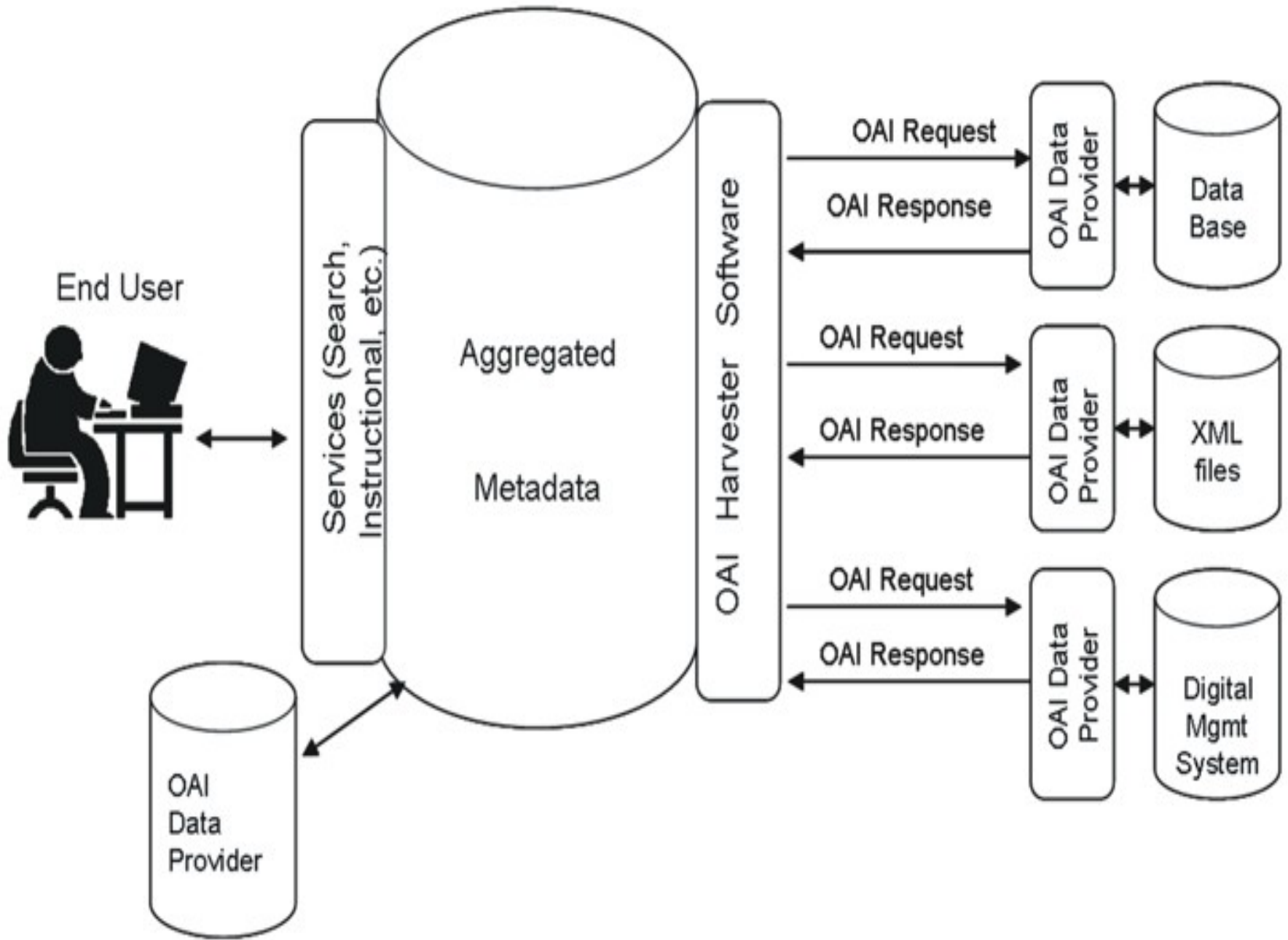
ACRL 12th National Conference

Minneapolis, MN • 2005 April 9

IMLS Digital Collections and Content

- Collection description and registry for IMLS National Leadership Grant projects with associated digital content
 - Enhance discoverability; all registry fields searchable
- Item level metadata repository for content via OAI-PMH
 - Demonstrate potential of metadata for interoperability
 - Serve as testbed for IMLS projects interested in OAI-PMH
 - Facilitate reuse of information resources
- Research question

How can resource developers best represent collections and items to meet the needs of service providers and end users?



Research Question:

What do information quality metrics and local practice help us understand about the quality of metadata at the aggregated level?

Methods:

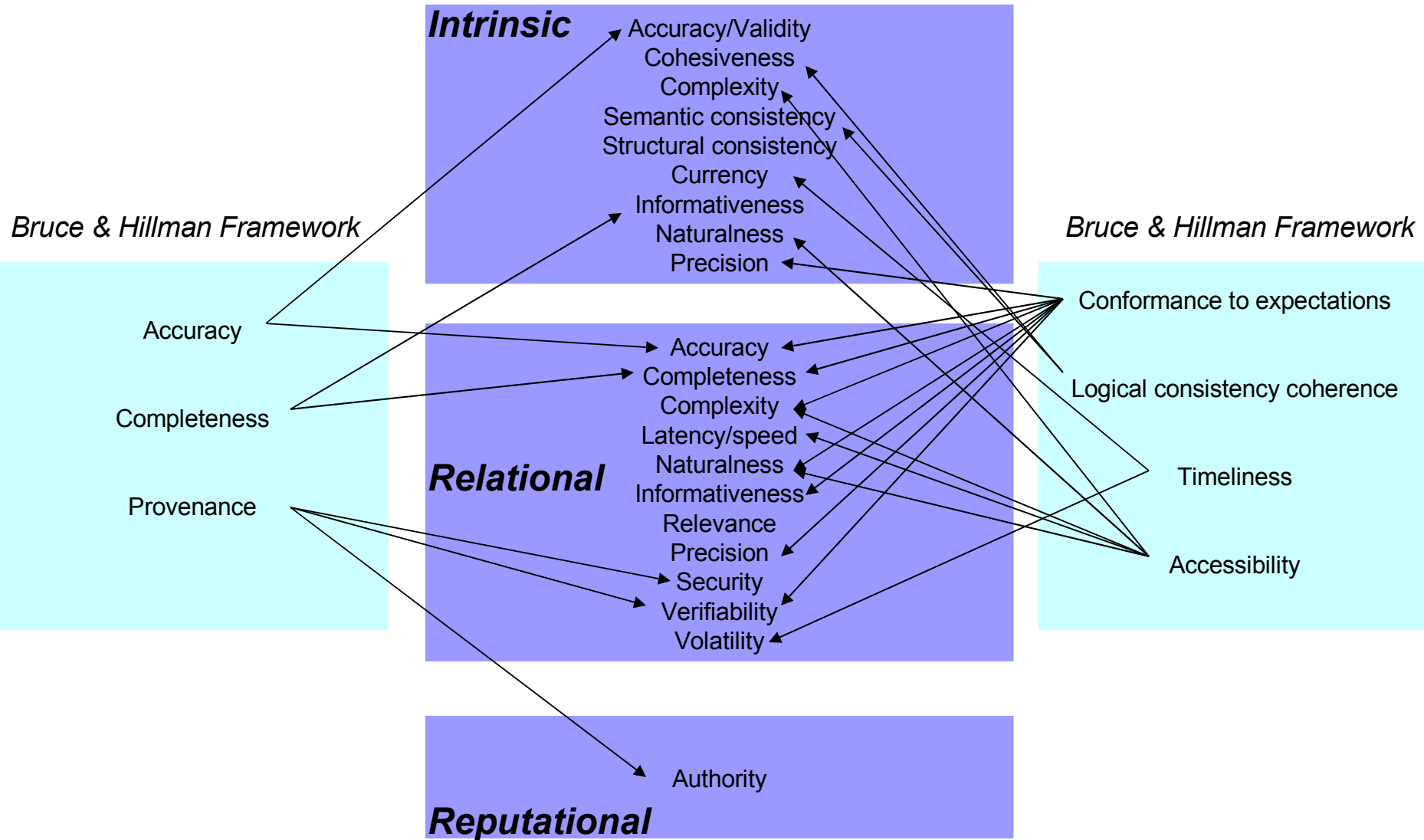
- Combination of quantitative and qualitative data.
- Various statistical analyses of the harvested metadata records from four digitization projects
- 13 open ended interviews

Characteristics of the four analyzed collections

	Collection 1	Collection 2	Collection 3	Collection 4
Total # of records	27,444	14,425	1,599	35
Type of institution	Large collaborative digitization project	Large academic library	Small academic library and public library collaboration	Small academic library
Type of resources described	Photographs, artifacts, text.	Photographs	Legal documents, letters, government documents, maps	Texts
Metadata mapped to simple Dublin Core from other metadata format?	Yes; variation of Qualified Dublin Core in use.	Yes; local metadata format in use.	No; variation of simple Dublin Core in use, but only Dublin Core elements exported.	Yes; local metadata format similar to qualified Dublin Core.
Notes about 35 record sample	Represents metadata from 12 institutions	None	Contains 14 nearly empty records exported by the content management system.	Represents entire collection.

Information Quality Frameworks

Gasser & Stvilia Framework





Aggregated Environment

Aggregation Activities

Local Environment

Content Creation Activities

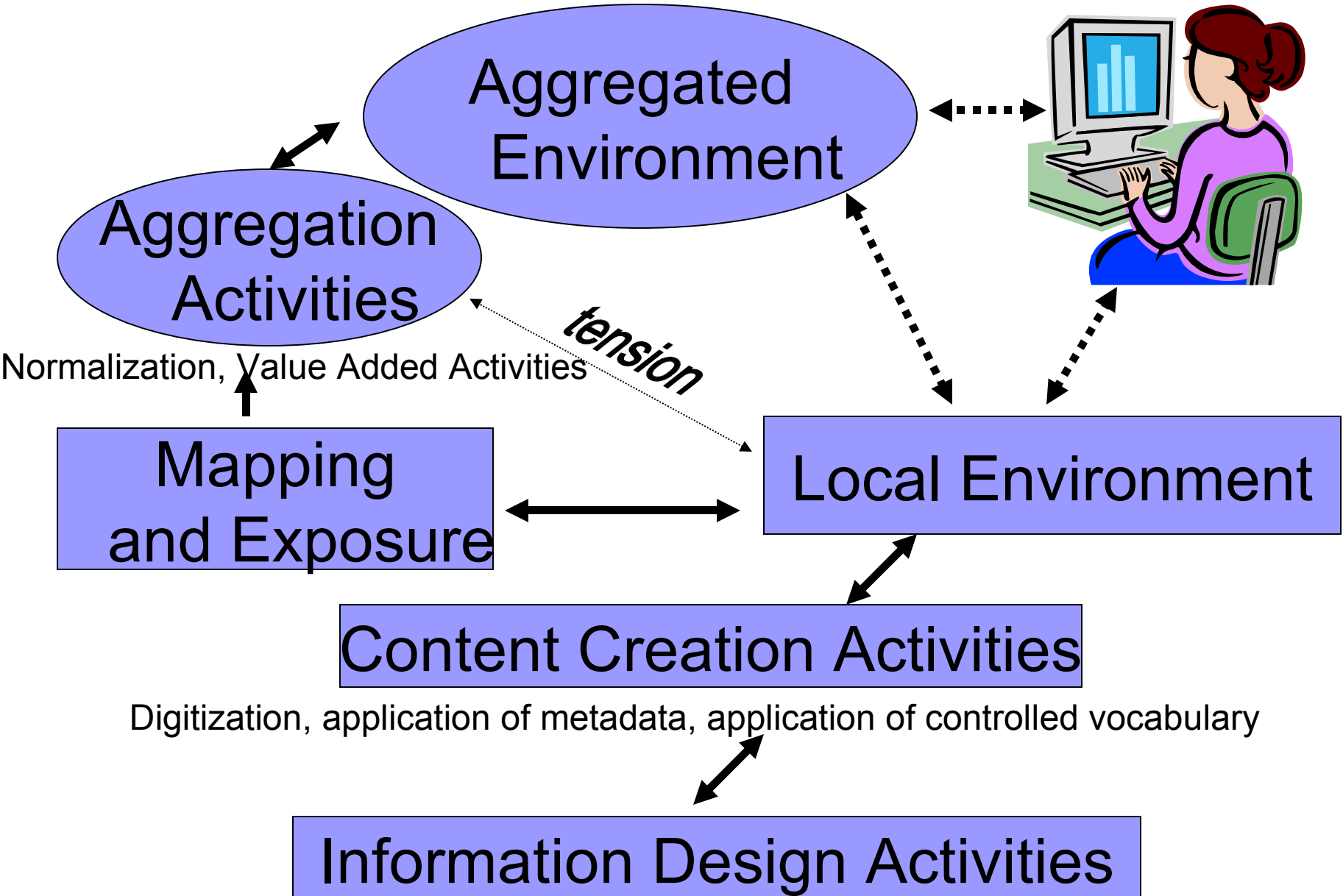
Information Design Activities

Normalization, Value Added Activities

tension

Digitization, application of metadata, application of controlled vocabulary

Collection decisions, Metadata scheme and controlled vocabulary selection





Tensions and Trade-offs

- Tensions between interoperability and local practice
 - Participants aware, but local practice takes priority
- Barrier to participation in digitization projects



What is sharable metadata?

- Attention to certain quality measures helps make metadata more sharable
 - Consistency
 - Completeness
 - Ambiguity

Example: Structural Inconsistency

10/1/1991	ca. June 19, 1901
ca. 1920.	(ca). June 19, 1901)
(ca). 1920)	1929 June 6
2001.06.08 by CAD	[between 1904 and 1908]
Unknown	[ca. 1967]
1853	1918?
c1875	191-?
c1908 November 19	1870 December, c1871
[2001 or 2002]	1920, 1921, 1922, 1923, 1924, 1925, 1926, 1927, 1928, 1929
[1919?]	20 th century

Example: Semantic Inconsistency

Date information included in:	Collection 1	Collection 2	Collection 3	Collection 4
<date> element (used once)	9 (26%)	35 (100%)	20 (57%)	0
<date> element (used at least twice)	20 (57%)	0	0	0
<coverage> element (used once)	0	0	17 (49%)	0
Date in other element	0	0	21 (60%) <title>	35 (100%) At end of <source> string
Not recorded	6 (17%)	0	14 (40%) (nearly empty records)	0

Example: Completeness

	Collection 1	Collection 2	Collection 3	Collection 4
% incomplete records	69%	71%	0%	100%

Description: 1 file : 1,131,754 bytes ; 600 dpi.

Description: Description: **Stereopair** showing the Mammoth and Ocean Bath hot spring establishments in Idaho Springs. In December 1858 when George Jackson decided to explore along the Vasquez (now Clear Creek), he observed the vapors rising from the hot springs along Soda Creek. A few weeks later, Jackson discovered gold in nearby Chicago Creek, and by June 1859 around 400 prospectors had settled along Chicago Creek. In the early years the settlement was called Jackson's Diggings, Jackson's Bar, Sacramento City or Idaho, but in 1866 the name was changed to Idaho Springs. A Cornish miner, James Jack, and his son actually tried to mine the hot springs, but their mine shaft filled with hot water. After this early misguided attempt to mine the hot springs, they were used occasionally by the miners for bathing, but Dr. E.M. Cummings was the first to commercialize the hot springs. Cummings purchased the property in 1863 and built a flimsy bath house over James Jack's mine shaft and charged people for bathing. Harrison Montague came to Idaho Springs in 1866 from New York, and he purchased a merchantile business and Charles Churchill's stage line, which ran from Central City to Georgetown via Idaho Springs. In 1871, Montague purchased the hot springs from Cummings, and he charged the miners twenty-five cents for a towel and bath. But Montague wanted to make his bathing establishment into a tourist attraction with the tourists traveling to Idaho Springs on his stage line. So he built a new bath house, Ocean Bath, which had a 24 X 40 X 4 ft. swimming bath, with separate sections for the men and women. There were also hot and cold showers, and Montague built a small hotel next to Ocean Bath for health seekers and travelers. Montague had a chemist, Dr. I.G. Pohle, determine the medicinal properties of the spring water, and Pohle's analysis was used in advertising Ocean Bath, which proved to be very popular with Denverites. Montague's success resulted in a competitor building the nearby Mammoth Bath Company, which had a larger 45 X 65 X 5 ft. bath. In 1890, the Big Five Federation purchased both the Ocean Bath and Mammoth Bath Company establishments, and Montague received \$60,000 for his business. The new owners built the Hot Springs Hotel, which had baths in the basement and an elevator to take people in wheelchairs down to the baths. They also added tennis courts and a casino. Brochures from Hot Springs featured a shapely Indian maiden at the springs, and the endorsements of Chief Idaho (probably a fictitious person), who was quoted as saying "Happy the rheumatic that takes Radium baths and is benefited, but more happy is the one that takes Radium bath every month and never has rheumatism." The establishment is still in business today and is called the Indian Springs Resort.

Date: 2000.

Date: [between 1870 and 1890].

Identifier: <http://www.digital.library.muni/SC155.gif>

Language: eng

Publisher: [redacted]

Relation: [Mining and mineral industries in the US : photographic perspectives] ;

Relation: [redacted]

Publisher: [redacted]

Rights: Rights management statement available at [redacted]

Source: 1 **stereopair** : b&w ; 10 x 18 cm.

Source: Held in the [redacted]

Source: SC155

Subject: Clear Creek County (Colo.)

Subject: Idaho Springs (Colo.)

Subject: Mineral waters -- Colorado -- Clear Creek County.

Title: [Mammoth and Ocean Baths, Idaho Springs]

Title: [Mining and mineral industries in the US : photographic perspectives] ; SC155

Format: GIF (Image file)

Title: Ragged London in 1861

Creator: John Hollingshead

Type: text

Source: Ragged London in 1861. John Hollingshead. Smith

Language: eng

Identifier: <http://www.digital.library.muni/cgi-bin/ptext?>

Example: Ambiguity

Description: Tiff master image: 300 dpi Photoshop LE 5.5 scanning software; Epson 836XL flatbed scanner; scanned by [redacted] Access jpeg image: 150 dpi. Thumbnail jpeg image: 72 dpi.

Creator: Catherine and Tom Cleveland

Description: 42 x 50"

Description: wool

Description: Classic "Teec" type partial piece still on **loom**, complete with all the tools involved in the spinning and weaving process. Mrs. Cleveland lives near Two Grey Hills and weaves only partial pieces for **loom** displays while her husband constructs the **looms**.

Description: Navajo, 1985

Date: 5/8/2001

Identifier: [http://\[redacted\]/images/F021NN/F021DCNN92Page.htm](http://[redacted]/images/F021NN/F021DCNN92Page.htm)

Relation: [redacted]

Relation: [redacted]

Publisher: [redacted]

Rights: [http://\[redacted\]images/Rights.htm](http://[redacted]images/Rights.htm)

Source: DC-NN-92

Source: 1999:12003

Subject: Navajo textile fabrics

Title: Navajo loom

Type: image

Format: image/jpeg



	Collection 1	Collection 2	Collection 3	Collection 4
% of records that describe at least 2 manifestations of a resource	86%	100%	100%	69%



Conclusions

- Semantic and structural consistency
- Minimize ambiguity
- Include documentation
- Exposure of richer metadata schemes?
- Establish best practices for 'shareable metadata' (DLF and NSDL effort)



Aggregated Environment

Aggregation Activities

Mapping and Exposure

Local Environment

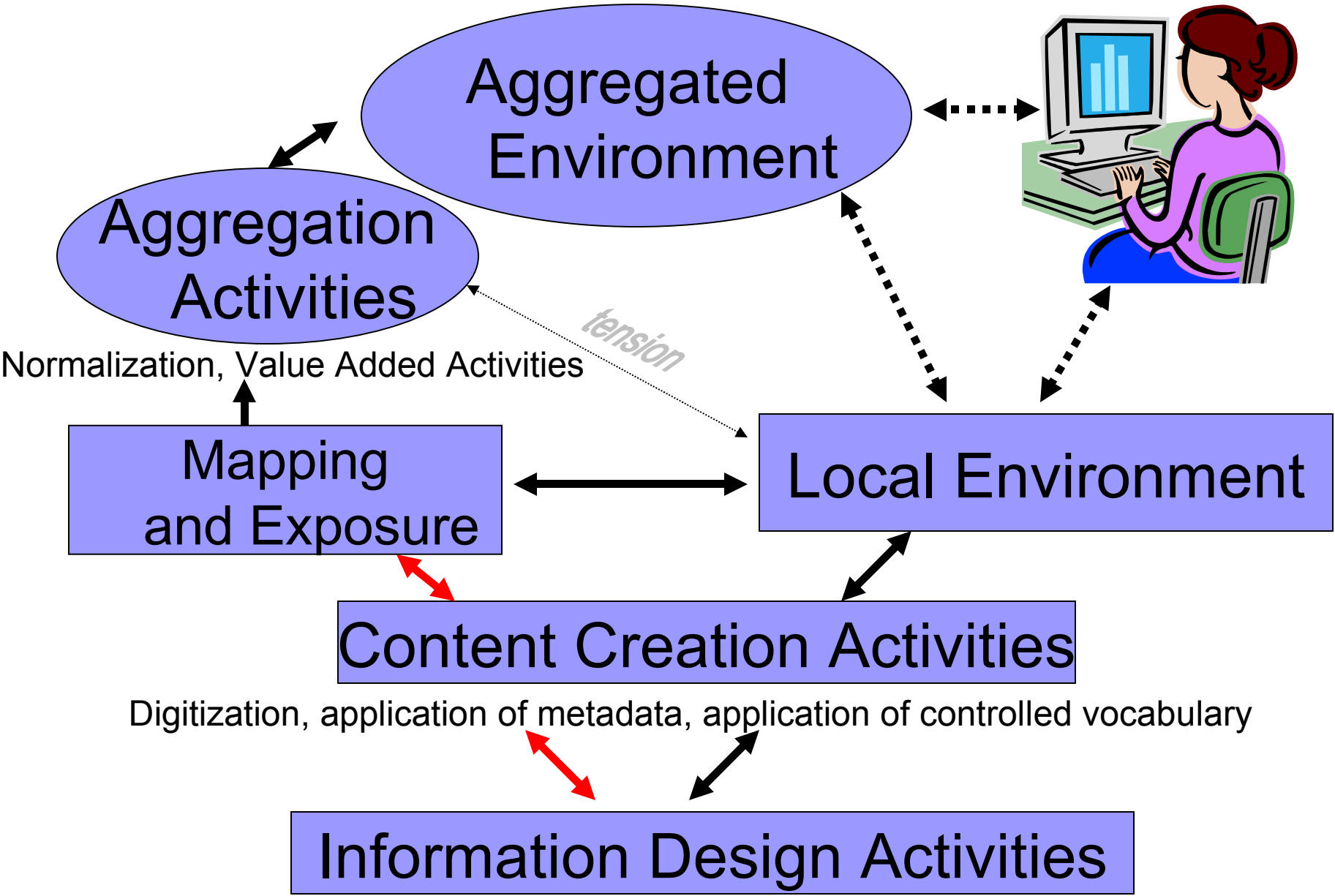
Content Creation Activities

Information Design Activities

Normalization, Value Added Activities

Digitization, application of metadata, application of controlled vocabulary

Collection decisions, Metadata scheme and controlled vocabulary selection



Questions / Comments Welcome

Sarah Shreeves sshreeve@uiuc.edu

Ellen Knutson eknutson@uiuc.edu

Acknowledgements:

Our collaborators from the IMLS DCC project team:

Timothy W. Cole – Principal Investigator
Carole L. Palmer – Co-Principal Investigator
Michael Twidale – Co-Principal Investigator
Besiki Stivila – Research Assistant

This research was funded by an IMLS National Leadership Grant

Record in Local and Aggregated Environments

Title: Saturday afternoon - first of Spring. Spreckles lake in Golden Gate Park

Title: Toy sail boats on **Spreckels** Lake - Golden Gate Park

Creator: Cushman, Charles Weever, 1896-1972

Subject: Trees

Subject: Lakes & ponds

Subject: People

Subject: Model sailboats

Subject: Cycling

Subject: Parks

Subject: Spring

Subject: Golden Gate Park (San Francisco, Calif.)

Subject: **Spreckels** Lake (Calif.)

Publisher: [REDACTED]

Publisher: [REDACTED]

Date: 1958-03-22

Type: Landscape photographs

Type: StillImage

Type: Image

Format: image/jpeg

Format: 35mm slide

Identifier: Cushman number: 258.14

Identifier: [REDACTED]

Identifier: [http://\[REDACTED\]/archives/cushman/P09709](http://[REDACTED]/archives/cushman/P09709)

Source: [REDACTED]

Relation: isPartOf [http://\[REDACTED\]/collections/cushman/](http://[REDACTED]/collections/cushman/)

Coverage: San Francisco, California, United States (San Francisco county)

Rights: [http://\[REDACTED\]/archives/cushman/rights](http://[REDACTED]/archives/cushman/rights)

Cushman ID: 258.14

Archives ID: P09709

Date: Mar. 22, 1958

Description (Slide): Saturday afternoon - first of Spring. Spreckles lake in Golden Gate Park

Description (Notebook): Toy sail boats on Spreckels Lake - Golden Gate Park

Street Address: (None)

Location: [San Francisco, California, United States \(San Francisco county\)](#)

Camera Settings: 100 f 6.3 tl

Roll Number: [2-58](#)

Frame Number: 14

Personal Names: (None)

Corporate Names: (None)

Other Names: (None)

Genre Headings: [Landscape photographs](#)

Subject Headings: [Trees](#)
[Lakes & ponds](#)
[People](#)
[Model sailboats](#)
[Cycling](#)
[Parks](#)
[Spring](#)

Other Subject Headings: Golden Gate Park (San Francisco, Calif.); Spreckels Lake (Calif.)