

Research Forum at the 75th Annual Meeting of the Society of American Archivists
August 22, 2011 - August 27, 2011
Chicago, IL

Does trust matter?


Ixchel M. Faniel, Ph.D.
OCLC

Elizabeth Yakel, Ph.D.
Nancy McGovern, Ph.D.
Kathleen Fear
Morgan Daniels
Adam Kriesberg
University of Michigan



This project is possible with funding from the Institute of Museum and Library Services.

Agenda



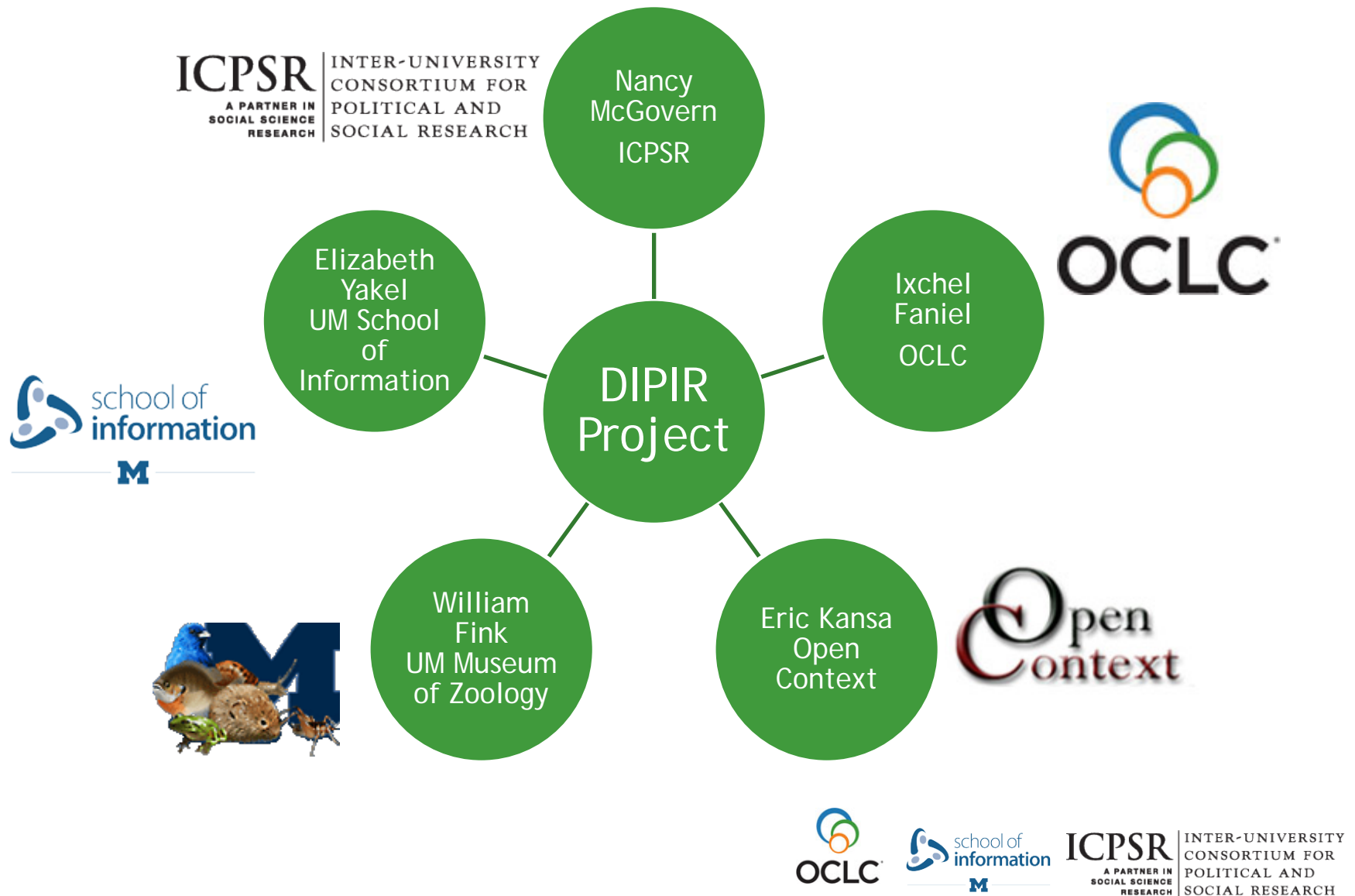
<p>Dissemination Information Packages for Information Reuse (DIPIR) Project</p>	<ul style="list-style-type: none">• Motivation• Research Questions• Research Methods
<p>Inter-university Consortium of Political and Social Research (ICPSR) Survey</p>	<ul style="list-style-type: none">• Motivation• Research Questions• Research Methods
<p>Audit and Certification of Trustworthy Digital Repositories (ISO/TRAC)</p>	<ul style="list-style-type: none">• Overview of Attributes• Example Mapping• Example Hypotheses



THE DIPIR PROJECT



Research Team

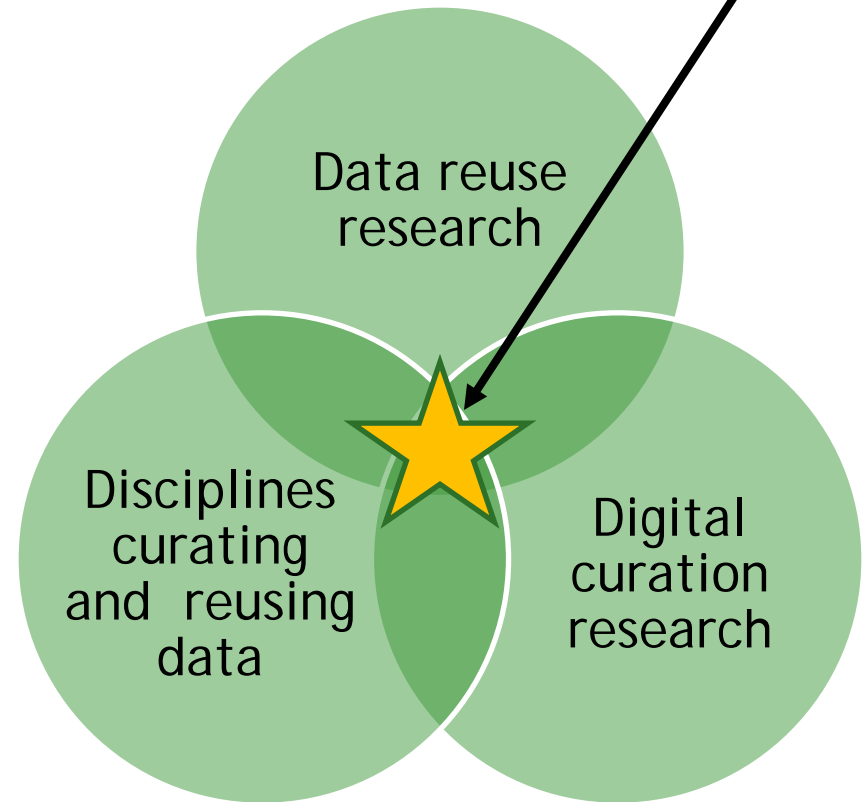


Research Motivation

Two Major Goals

1. Bridge gap between data reuse and digital curation research
2. Determine whether reuse and curation practices can be generalized across disciplines

Our interest is in this overlap.

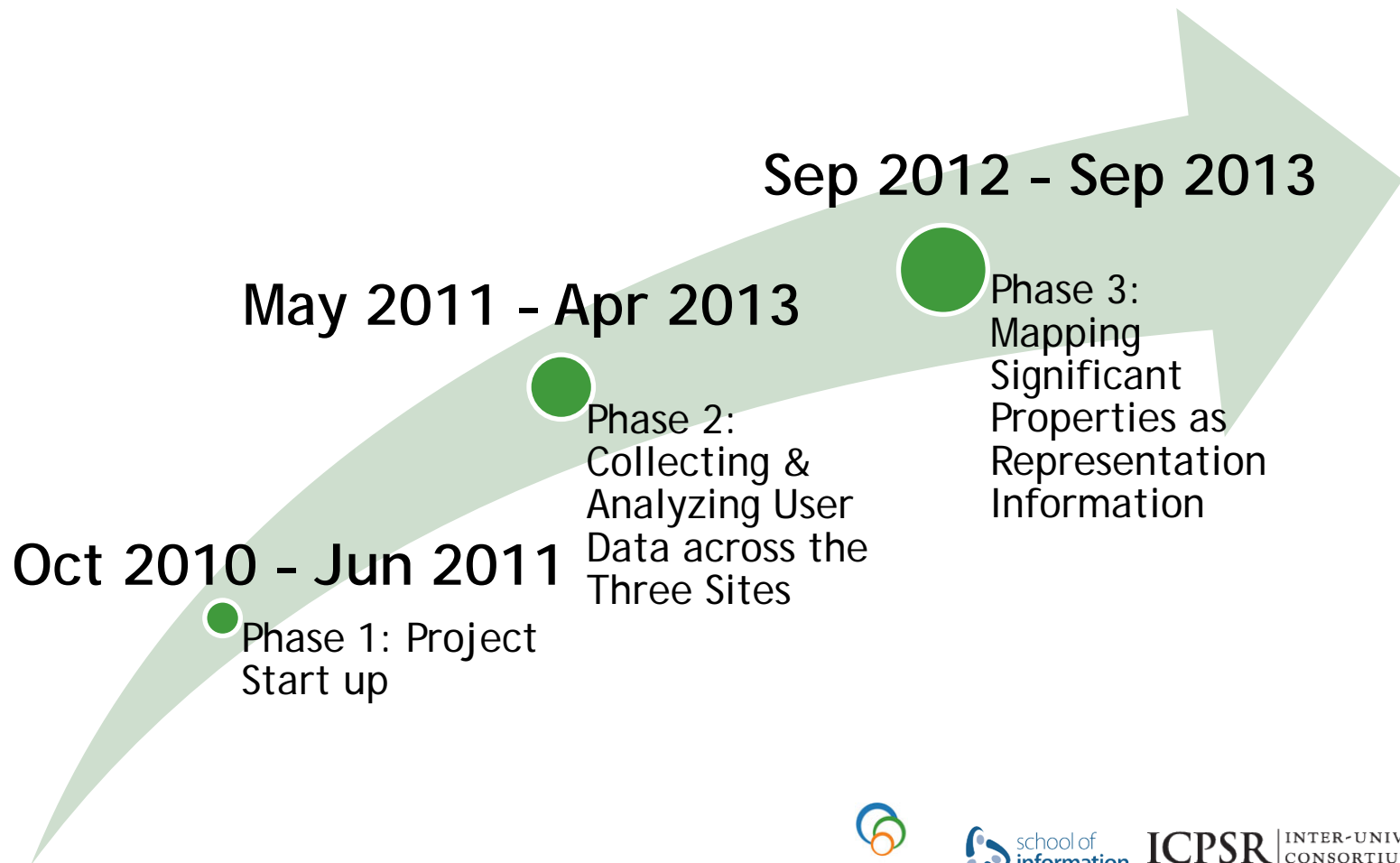


Research Questions

1. What are the significant properties of data that facilitate reuse by the designated communities at the three sites?
2. How can these significant properties be expressed as representation information to ensure the preservation of meaning and enable data reuse?



Research Methodology



ICPSR SURVEY



Findings from digital curation literature

- Suggest significant properties that provide “reliable, long-term access to managed digital resources” (Trusted Digital Repositories: Attributes and Responsibilities 2002)
 - Functionality, relationships, appearance (Coyne et al. 2007)
 - Look and feel (Hedstrom et al., 2006; Matthews et al. 2009)
 - Computing environment and usage (Morrissey 2010)
 - Purpose and use (Ashley et al. 2008)
- Interest in determining the range of significant properties a trusted repository might have to accommodate



Findings from data reuse literature

- Suggest significant properties that help users understand the data and whom and what to trust
 - Data cleaning & manipulation (Carlson & Anderson 2007)
 - Data collection methods (Faniel & Jacobsen 2010)
 - Identity of data collector (Knorr Cetina 1999; Van House 1998; 2002)
 - Selection and calibration of data collection instruments (Wallis et al. 2007)
 - Quality checks (Carlson & Anderson 2007; Zimmerman 2003)
- Interest in determining trustworthiness of data producers and an understanding their actions



Research Questions

- What are the significant properties of data that facilitate reuse by the designated communities at ICPSR?
- What differences do the attributes of a trusted digital repository (TDR) make to researchers using data from that repository?
 - What attributes of TDRs as outlined in TRAC do researchers care about?
 - How do their perceptions about repositories influence their propensity for data reuse?

Research Methods

- Outline concepts in reuse & curation literatures
- Map TDR attributes in ISO TRAC to concepts
- Develop hypotheses - *In progress*
- Operationalize concepts - *In progress*
- Administer survey to ICPSR dataset users - *Fall 2011*



Comparison of Literatures

Curation Literature	Reuse Literature	Concepts
Trusted data repository	Trusted data producer	<u>Repository context</u> perceptions of trustworthiness of source
Primarily data-centered	Primarily user-centered	<u>User context</u> characteristics of user
Identifying & maintaining data context that allows data to be rendered over the long term	Identifying & maintaining data context that allows data to be interpreted over the long term	<u>Data context</u> significant properties of data
consistently accessible and supported	accessed and supported at producer discretion	<u>Delivery context</u> ways data are accessed and supported

Both literatures agree that decisions to reuse are “not by metadata alone...”



ISO TRAC MAPPING



Overview of ISO TRAC Areas of Focus¹

- Organizational Infrastructure
 - “characteristics of the repository organization that affect performance, accountability, and sustainability.” (p. 9)
- Digital Object Management
 - “repository functions, processes, and procedures needed to ingest, manage, and provide access to digital objects for the long term.” (p. 21)
- Infrastructure & Security Risk Management
 - “adequacy of the repository’s technical infrastructure and its ability to meet object management and security demands of the repository and its digital objects.” (p. 43)

¹Definitions from Trustworthy Repository Audit and Certification: Criteria & Checklist, 2007



Summary of ISO TRAC Mapping

- 63 attributes were mapped
- 43 related to repository context
- 20 related to data context
- 11 related to delivery context
- 20 related to user context



Significant properties of data

DATA CONTEXT



Example of Data Context

ISO TRAC 4.2.5.2

ISO TRAC p. 4-11

“The repository shall have tools or methods to determine what Representation Information is necessary to make each Data Object understandable to the Designated Community.”

Quote from CBU07

“[...], like the questions on abortion, they have changed over time. And they're doing experiments to figure out better wordings, so like half of the sample got one question, and the other half of the sample got a different question. So the codebook is like your guide to all of that [...]”



Ways data are accessed and supported

DELIVERY CONTEXT



Example of Delivery Context

ISO TRAC 4.5.1

ISO TRAC p. 4-23

“The repository shall specify minimum information requirements to enable the Designated Community to discover and identify material of interest.”

Quote from CBU08

“[...] like with the religious affiliation data, I've looked and looked [...] but [...] I don't even think I found a good topical subject category [...] So it's definitely helpful when things are organized through different categories that you can search through for topics.”



Perceptions about the trustworthiness of source

REPOSITORY CONTEXT



Example of Repository Context

ISO TRAC 4.2.9

ISO TRAC p. 4-15

“The repository shall provide an independent mechanism for verifying the integrity of the repository collection/content.”

Quote from CBU03

“So there’s like a [repository that] has a bad [...] history basically. It was very biased on how they developed their methodology and such. And then they said they corrected what they did like recently or whatever, but now that history creates a history of bad reputation. I don’t know anyone that uses [the repository].”



Characteristics of user

USER CONTEXT



Example of User Context

ISO TRAC 3.3.1

ISO TRAC p. 3-5

“The repository shall have defined its Designated Community and associated knowledge base(s) and shall have these definitions appropriately accessible.”

Quote from CBU09

“Because I am so novice in these areas, I would heavily value the opinions of like professors that knew more than me [...]”



NEXT STEPS



What contributes to decisions to reuse?

- Trust in the data
 - Repository context - perceived trustworthiness of the source
- Relevance of the data
 - User context - characteristics of the user environment (e.g. task, experience level)
- Quality of the data
 - Data context - significant properties of the data that make it fit for use
- Ease of use of the data
 - Delivery context - perceived effort needed to access and get support for data



Focusing on Trust - 3 Sources

- Data producer - reputation of the person who originally collected the data
- Repository - reputation of the institution providing access to the data
- 3rd party - reputation of an independent entity endorsing reuse of the data (e.g. faculty advisor, Data Seal of Approval Board)



Example Hypotheses

- For expert users, the reputation of the data producer will be significantly more important in their decision to reuse data than the reputation of the data repository or recommendations from a trusted 3rd party.
- For novice users, recommendations from a trusted 3rd party will be significantly more important in their decision to reuse data than the reputations of the data producer or the data repository.

ISO TRAC attributes the hypotheses touch: 4.2.9, 3.3.1, 3.3.5, 3.3.6, 4.1.5, 4.2.8, 4.2.9, 4.1.4



Survey Design and Administration

- Sample
 - ICPSR's bibliography - i.e. academic literature resulting from reuse of ICPSR datasets
- Recruitment
 - To coincide with ICPSR's last major change to its delivery system
- Contact
 - Email Endorsement from ICSPR director
 - 1 email invitation to first authors with survey link
 - 2 email reminders



Acknowledgements

- Institute of Museum and Library Services  INSTITUTE of **Museum** and **Library** SERVICES
- Annelise Doll, University of Michigan, School of Information, MSI 2010
- Mallory Hood, University of Michigan, School of Information, MSI 2010
- Molly Haig, Yale University, current undergraduate



Thank you.

QUESTIONS & COMMENTS

