

San Diego, CA, 11 August 2012

76th Annual Meeting of the Society of American Archivists

Infusing Consumer Data Reuse Practices into Curation and Preservation Activities

Ixchel M. Faniel, Ph. D.

OCLC Research

fanieli@oclc.org



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

DIPIR: A Project Overview

Reuse Practices of
Quantitative Social
Scientists
&
Archaeologists

Q&A



An Overview: www.dipir.org

The DIPIR Project

Research Team

ICPSR | INTER-UNIVERSITY
CONSORTIUM FOR
POLITICAL AND
SOCIAL RESEARCH
A PARTNER IN
SOCIAL SCIENCE
RESEARCH

Nancy
McGovern
ICPSR/MIT



Ixchel Faniel
OCLC
Research
(PI)



Elizabeth
Yakel
University of
Michigan
(Co-PI)

**DIPIR
Project**



William Fink
UM Museum
of Zoology

Eric Kansa
Open
Context

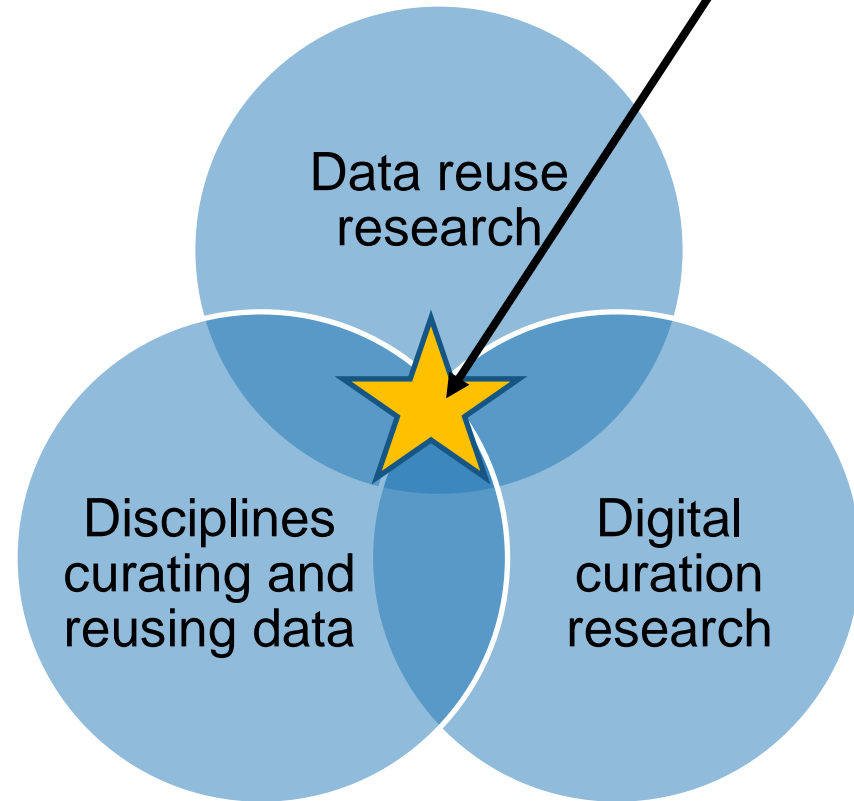


Two Major Goals

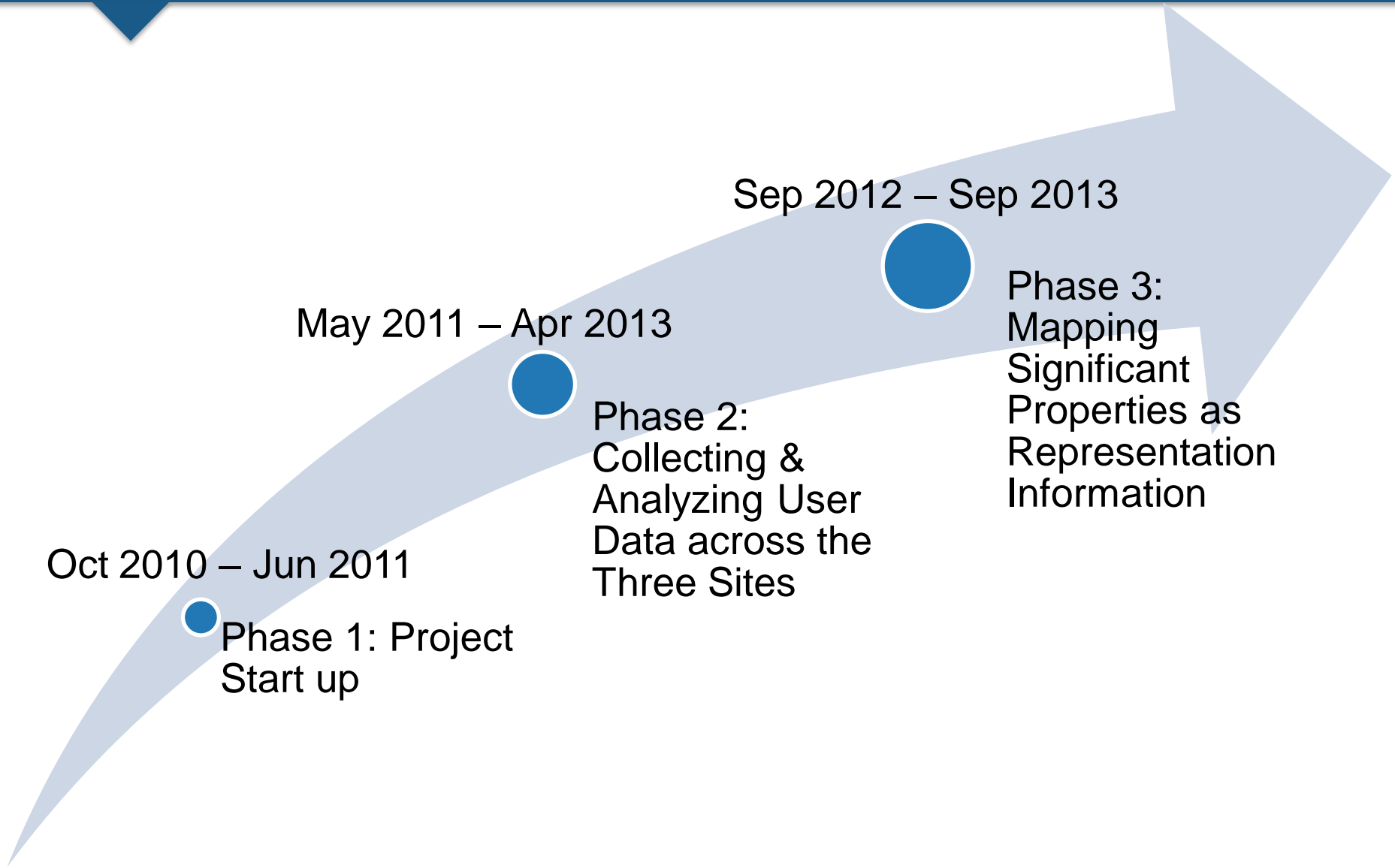
1. Bridge gap between data reuse and digital curation research
2. Determine whether reuse and curation practices can be generalized across disciplines

Faniel & Yakei (2011)

Our interest is in this overlap.



1. What are the significant properties of data that facilitate reuse in the quantitative social science, archaeology, and ecology communities?
2. How can these significant properties be expressed as representation information to ensure the preservation of meaning and enable data reuse?

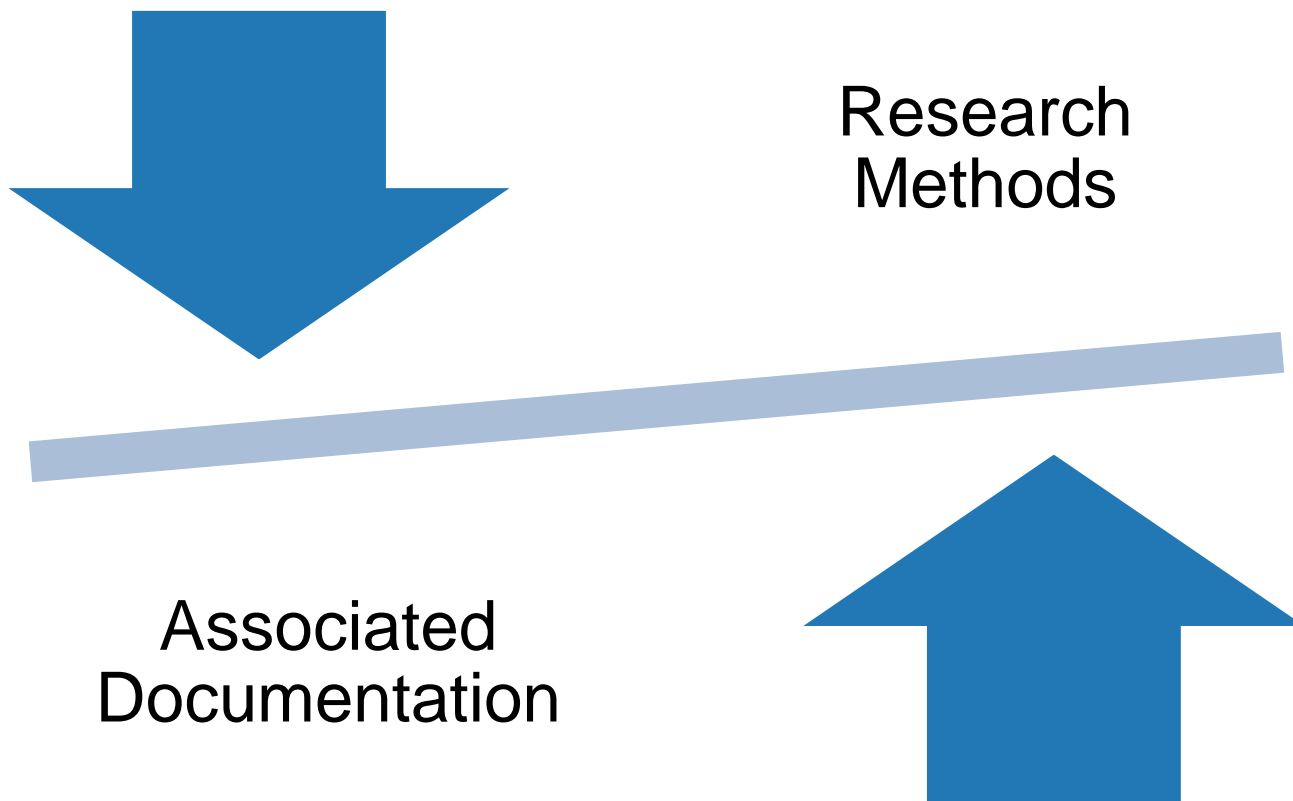


Data Collection Plan

	ICSPR	Open Context	UMMZ
Interviews staff	10 ✓ Winter 2011	4 ✓ Winter 2011	10 ✓ Spring 2011
Interviews data consumers	40 ✓ Winter 2012	22 ✓ Winter 2012	20 Spring 2013
Survey data consumers	2000 ✓ Summer 2012		
Web analytics data consumers		Ongoing	
Observations Data consumers			10 Spring 2013

Quantitative Social Scientists & Archaeologists

Consumer Data Reuse Practices



Quantitative Social Scientists & Archaeologists

Research Methods

“Archaeology is done by different people in different ways...” (CCU18)

“[Social scientists] could come up with...a lot of ways to measure it, but some of them are not really what I would be interested in.” (CBU16)

Data Collection Instruments

**“I mean, obviously, if it was public opinion, any type of public opinion looking at surveys, how you word the question really influences the responses you get. So, I think it’s really important the exact questions’ wording, which most places are pretty good about.”
(CBU23)**

Data Collection Instruments

“I have to know what the parameters are in which they gathered those coordinates. So I have to know the type of GPS in this case that with which they collected that data or they collected it from a map, the map systems, et cetera.” (CCU03)

Data Collection Procedures

“...when they rank democratic-ness...they say they do it based on...data developed from gathering newspaper articles... I’d like to know for example what years those were and what newspapers it used. And if there were those newspapers they used for example to gather data on how open it is of freedom of press..., I want to know if that was consistently used, if that was the same newspaper that was used by another...”
(CBU03)

Data Collection Procedures

“Everybody excavates in a slightly different way, and that information can be very important. For example, if you're in different sort of mesh size when you're screening your dirt for sorting archaeology can result in very, very different samples being collected, you may have large goals in one sample, you miss all the small goals.” (CCU02)

Quantitative Social Scientists & Archaeologists

Associated Documentation

Social Scientists – Codebooks

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR
POLITICAL AND SOCIAL RESEARCH

ICPSR 3680

**Los Angeles Homicides,
1830-2003**

Eric H. Monkkonen
*University of California-Los Angeles.
Departments of History and Policy Studies*

(Part 1: Annual Homicides and Related Data)

Codebook

“It was definitely the codebook was the... I mean without that I'd have no clue because it's numerical value on things that don't have numerical value.” (CBU14)

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Social Scientists - Bibliographies

- Monkkonen, Eric, "Estimating the accuracy of historic homicide rates: New York and Los Angeles." *Social Science History*. Spr 2001, 25, (1), 53 -66. DOI: 10.1215/01455532-25-1-532.
- Monkkonen, Eric, "New standards for historical violence research." *Crime, Histoire et Societes*. 2001, 5, (2), 5 - 26. DOI: 10.4000/chs.7333.
- Monkkonen, Eric H., "Homicide in New York, Los Angeles and Chicago." *Journal of Criminal Law and Criminology*. Fal 2001/Win 2002, 92, (3-4),.4.
- Monkkonon, Eric H., "Homicide in Los Angeles, 1827-2002." *Journal of Interdisciplinary History*. Fal 2005, 36, (2), 167 - 183. DOI:10.1162/0022195054741235

“Also the file with the publications. I mean I always go through that just to see what other people have done with the dataset.” (CBU02)

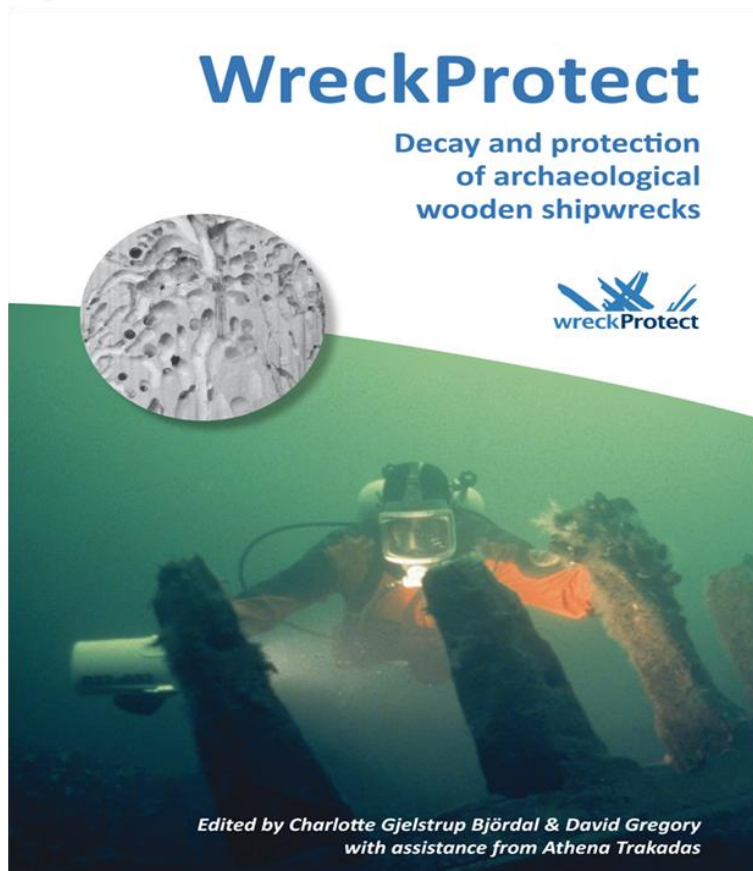
Archaeologists: accession books



“What I did is I took those accession books, I copied them, and I had them scanned and put on a CD, and then we went in and created the database to pull that information out.” (CCU09)

<http://england.prm.ox.ac.uk/englishness-PRM-Artefact-Documentation.html>

Archaeologists: monographs



“And there was a monograph that was published... So we got some information. So those were like PDF documents just for the mainstreamed published data about the site... we heavily relied on those publications about the sites.” (CCU16)

Associated Documentation

- **Typically had the context they needed**
 - “So I don’t think we could have done it without the codebook. The coding book is just indispensable.” (CBU13)
- **Often wanted more**
 - “Other things that could be provided in the codebook...If they take out questions, I think it’s helpful to know why they took out, 2 to 3 lines that would be small enough.” (CBU18)

Associated Documentation

- **Often needed additional context**
 - “...there is all sorts of metadata that you would like to have...I am quite used to not having any frankly.” (CCU17)
- **Made do with what they got**
 - “It’s not my general reaction to look at something and say it’s unusable.” (CCU18)
- **Asked data producers when possible**
 - “So I was contacting him [the excavator] for other specific information. Where was this found, what period did it date to,....” (CCU01)

Implications for Preservation & Curation

- **Consider reuse practices and the nature of the context of data production**
- **Distinguish and prioritize aspects of the data production context that is needed vs. wanted**
- **Take an integrated stakeholder view**

Acknowledgements

- Institute of Museum and Library Services
- Co-PI: Beth Yakel, Ph.D.
- Partners: Nancy McGovern, Ph.D., Eric Kansa, Ph.D., William Fink, Ph.D.
- Students: Adam Kriesberg, Julianna Barrera-Gomez, Laurinda Weisse, Jessica Schaengold, Gavin Strassel, Michele DeLia, Kathleen Fear, Morgan Daniels, Mallory Hood, Molly Haig, Annelise Doll, Monique Lowe

