Infusing Consumer Data Reuse Practices into Curation and Preservation Activities

Ixcchel M. Faniel, Ph. D.  
OCLC Research  
fanieli@oclc.org

This project is possible with funding from the Institute of Museum and Library Services.
An Overview: www.dipir.org

The DIPIR Project
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 & Yakel (2011)
Research Questions

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?
Research Methodology

Phase 1: Project Start up

Phase 2: Collecting & Analyzing User Data across the Three Sites
May 2011 – Apr 2013

Phase 3: Mapping Significant Properties as Representation Information
Sep 2012 – Sep 2013
<table>
<thead>
<tr>
<th></th>
<th>ICSPR</th>
<th>Open Context</th>
<th>UMMZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviews staff</td>
<td>10</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>✓ Winter 2011</td>
<td>✓ Winter 2011</td>
<td>✓ Spring 2011</td>
</tr>
<tr>
<td>Interviews data consumers</td>
<td>40</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>✓ Winter 2012</td>
<td>✓ Winter 2012</td>
<td>Spring 2013</td>
</tr>
<tr>
<td>Survey data consumers</td>
<td>2000</td>
<td>Ongoing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>✓ Summer 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Web analytics data consumers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations Data consumers</td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spring 2013</td>
</tr>
</tbody>
</table>
Quantitative Social Scientists & Archaeologists

Consumer Data Reuse Practices
Quantitative Social Scientists & Archaeologists

Research Methods

Associated Documentation

OCLC  The world’s libraries. Connected.
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)
“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)
“...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)
“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
“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)


“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)
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