Learning from Success: Approaching Data Management in the Humanities with Optimism and Good Cheer

Iris Jastram, Kristin Partlo, Sarah Calhoun
Oberlin Digital Scholarship Conference
June 11, 2016
Data Management?

“In the context of research and scholarship, data management refers to the storage, access, and preservation of data produced from a given investigation.”

- Texas A&M Libraries
“Data management practices cover the entire lifespan of the data from planning to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded.”
Sharing “to the greatest extent and with the fewest constraints possible”

To “catalyze innovative breakthroughs”

“provide grist for new insights”

Goal is to develop “policies that mobilize these publications and data for reuse through preservation and broader public access” because this will “maximize the impact and accountability of the Federal research investment.”
The Spirit of the DMP

DMP = Data Management Plan

YOU GET A CAR
AND YOU GET A CAR
EVERYONE GETS A CAR

https://i.kinja-img.com/gawker-media/image/upload/s--eVIrDf03--/c_scale,fl_progressive,q_80,w_800/jcxywkck4zxdvpg3ijq.jpg
NEH Template

Office of Digital Humanities

Roles and Responsibilities
Expected Data
Period of Data Retention
Data Formats and Dissemination
Data Storage and Preservation of Access
“Data’ is defined as materials generated or collected during the course of conducting research. Examples of humanities data could include citations, software code, algorithms, digital tools, documentation, databases, geospatial coordinates (for example, from archaeological digs), reports, and articles.”
Data for a Humanist

Your own citations

https://commons.wikimedia.org/wiki/File:
Data for a Humanist

Your own citations

Metadata about metadata
Data for a Humanist

Your own citations

Metadata about metadata

Other people’s data...

- Census

<table>
<thead>
<tr>
<th>Subject</th>
<th>United States</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>WORLD REGION OF BIRTH OF FOREIGN BORN</td>
<td></td>
</tr>
<tr>
<td>Foreign-born population, excluding population born at sea</td>
<td>41,056,445</td>
</tr>
<tr>
<td>Europe</td>
<td>4,813,488</td>
</tr>
<tr>
<td>Asia</td>
<td>11,990,010</td>
</tr>
<tr>
<td>Africa</td>
<td>1,742,620</td>
</tr>
<tr>
<td>Oceanic</td>
<td>232,632</td>
</tr>
<tr>
<td>Latin America</td>
<td>21,440,938</td>
</tr>
<tr>
<td>Northern America</td>
<td>819,728</td>
</tr>
</tbody>
</table>

LANGUAGE SPOKEN AT HOME

Population 5 years and over | 294,133,373 | +/-3,657 | 294,133,373 | (X) |

English only | 232,729,203 | +/-121,279 | 78.9% | +/-0.1 |

Language other than English | 61,406,170 | +/-119,977 | 20.9% | +/-0.1 |

Speak English less than “very well” | 25,306,202 | +/-58,810 | 8.6% | +/-0.1 |

Spanish | 38,090,698 | +/-72,329 | 13.0% | +/-0.1 |

Speak English less than “very well” | 16,348,401 | +/-44,757 | 5.6% | +/-0.1 |

Other Indo-European languages | 10,800,493 | +/-40,816 | 3.7% | +/-0.1 |

Speak English less than “very well” | 3,436,072 | +/-22,879 | 1.2% | +/-0.1 |

Asian and Pacific Islander languages | 5,770,631 | +/-23,413 | 3.3% | +/-0.1 |

Speak English less than “very well” | 4,873,758 | +/-16,658 | 1.6% | +/-0.1 |

Other languages | 2,727,348 | +/-20,477 | 0.9% | +/-0.1 |

Speak English less than “very well” | 845,931 | +/-11,472 | 0.3% | +/-0.1 |

ANCESTRY

Total population | 314,107,084 | | 314,107,084 | (X) |

American | 22,305,250 | +/-53,444 | 7.1% | +/-0.1 |

Arab | 1,769,645 | +/-20,573 | 0.6% | +/-0.1 |

Czech | 1,475,506 | +/-11,823 | 0.5% | +/-0.1 |

Danish | 1,323,182 | +/-10,809 | 0.4% | +/-0.1 |

Dutch | 4,307,814 | +/-16,048 | 1.4% | +/-0.1 |

English | 25,181,291 | +/-59,031 | 8.0% | +/-0.1 |

French (except Basque) | 6,431,568 | +/-25,741 | 2.1% | +/-0.1 |

French Canadian | 2,052,979 | +/-17,133 | 0.7% | +/-0.1 |

German | 46,874,293 | +/-66,656 | 14.9% | +/-0.1 |

Greek | 1,285,866 | +/-11,557 | 0.4% | +/-0.1 |

Hungarian | 1,440,762 | +/-10,155 | 0.5% | +/-0.1 |

Irish | 33,916,058 | +/-70,417 | 10.6% | +/-0.1 |

Italian | 17,330,315 | +/-42,106 | 5.5% | +/-0.1 |

Lithuanian | 658,654 | +/-8,877 | 0.2% | +/-0.1 |

Norwegian | 4,451,908 | +/-23,841 | 1.4% | +/-0.1 |

Polish | 9,452,375 | +/-35,837 | 3.0% | +/-0.1 |

Portuguese | 1,380,904 | +/-13,688 | 0.4% | +/-0.1 |
Data for a Humanist

Your own citations

Metadata about metadata

Other people’s data...

- Census
- Primary source collections (ECCO)
1. Roles and responsibilities
2. Expected data
3. Period of data retention
4. Data formats and dissemination
5. Data storage and preservation of access

Template of a DMP
Round 1

Successful DMPs

1. Strengths/weaknesses?
2. What was vague and why?
3. What language stuck out as being in the spirit of the DMP?
4. What did you learn about DMPs?

Round 2

Services based on DMP Template

1. What questions would you ask as the faculty member starts thinking about data management?
2. Fill in a blank DMP form for either text, video, or an image.
3. How would you prepare to speak to the faculty member? Who else would you invite to this conversation? What information would you need to gather first?
4. What resources would you like to have on hand / what’s in your “toolbox”? (i.e. template, form, libguide, etc)

See: [https://dmptool.org/requirements_templates/20/basic.docx](https://dmptool.org/requirements_templates/20/basic.docx)