Exploring the Data Management and Curation Practices of Scientists in Research Labs within a Research University via an Adapted Data Asset Framework (DAF) Survey – Phase 1

Plato L. Smith II, Doctoral Candidate, Florida State University - School of Information
International Association of Social Science Information Services & Technology Presentation
Toronto, Canada - June 4, 2014
Research Purpose

- Identify origin and types of data assets
- Investigate how data assets are stored, managed, & preserved across research labs
- Articulate implications
- Explore multiple DMC practices & perspectives

Data Asset Framework (DAF)

Research Labs

1. Center for Advanced Power Systems
2. Antarctic Marine Geology Research Facility
3. Center for Ocean-Atmospheric Predication Studies
4. Geophysical Fluid Dynamics
5. Marine and Coastal Laboratory
6. National High Magnetic Field Laboratory (NHMFL)
7. National Science Foundation (NSF) EarthCube

<table>
<thead>
<tr>
<th>Role</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior Researcher</td>
<td>23</td>
<td>23%</td>
</tr>
<tr>
<td>Principal Investigator</td>
<td>29</td>
<td>29%</td>
</tr>
<tr>
<td>Research Assistant</td>
<td>26</td>
<td>26%</td>
</tr>
<tr>
<td>Research Technician</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>Research Support</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>Research Student</td>
<td>10</td>
<td>10%</td>
</tr>
<tr>
<td>*Other</td>
<td>7</td>
<td>7%</td>
</tr>
<tr>
<td>Total</td>
<td>101</td>
<td>1.01</td>
</tr>
</tbody>
</table>

*Other:
- IT Support
- Postdoctoral research associate
- Research associate
- Operation project manager
- Data management
- Postdoctoral research associate
- Postdoc
Research Design & Methodology

- Sequential Mixed-Methods Explanatory Research Design (Creswell & Plano Clark, 2011)
- Quantitative/qualitative DAF survey
- Qualitative semi-structured DAF interview
- Metatriangulation (Lewis & Grimes, 1999)
- Adapted Conceptual Framework (Burrell & Morgan, 1979; Morgan & Smircich, 1980; Morgan, 1983; Solem 1993; Smith II, 2013)
- Three Perspectives (Martin, 1992)

Findings

<table>
<thead>
<tr>
<th>DAF</th>
<th>Surveys</th>
<th>Interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starts</td>
<td>129</td>
<td>7</td>
</tr>
<tr>
<td>Completes</td>
<td>107</td>
<td>6</td>
</tr>
<tr>
<td>Completion Rate</td>
<td>83%</td>
<td>86%</td>
</tr>
</tbody>
</table>

What is your primary disciplinary domain?

- Multidisciplinary: 52%
- Interdisciplinary: 25%
- Other: 23%
- Multidisciplinary
- Interdisciplinary
- Other
Research Design & Methodology

- Sequential Mixed-Methods Explanatory Research Design (Creswell & Plano Clark, 2011)

- Quantitative/qualitative DAF survey

- Qualitative semi-structured DAF interview

- Metatriangulation (Lewis & Grimes, 1999)

- Adapted Conceptual Framework (Burrell & Morgan, 1979; Morgan & Smircich, 1980; Morgan, 1983; Solem 1993; Smith II, 2013)

- Three Perspectives (Martin, 1992)
## Findings

<table>
<thead>
<tr>
<th>DAF</th>
<th>Surveys</th>
<th>Interviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starts</td>
<td>129</td>
<td>7</td>
</tr>
<tr>
<td>Completes</td>
<td>107</td>
<td>6</td>
</tr>
<tr>
<td>Completion Rate</td>
<td>83%</td>
<td>86%</td>
</tr>
</tbody>
</table>

**What is your primary disciplinary domain?**

- Multidisciplinary: 52%
- Interdisciplinary: 25%
- Other: 23%
Findings - Primary Data Types

- 78% (74) Experimental (scientific experiments and computational results)
- 61% (58) Derived data (processing or combining 'raw' or other data)
- 51% (48) Computer code (model & simulation code)
- 44% (42) Observational (scientific phenomena at a specific time or location)
- 27% (26) Reference (ex. gene sequences, chemical structures or literary texts)
- 3% (3) do not hold any primary data
- 2% (2) Other (videos, images, audio files; project funding, cost & budget analysis)
Who is responsible for managing your research data (select all that apply)?

- Other: 6
- You: 70
- National data center: 8
- Research groups: 16
- Research assistant: 17
- Project manager: 20

Findings - Research Data Management Responsibility
Findings - Standards, Best Practices, and Guidelines
Findings - Barriers
Research Implications

A. Good DMC practices stimulate organized research data management awareness;

B. Organized research data management awareness allows stakeholders, institutions, and users to increase ROI;

C. Data management education exposure across multiple disciplines and departments raise data management cognition.
Practical Implications

1. Adherence to best practices, standards, and guidelines foster cogent data policies, promote good DMC practice, and enable new research built on accessible & existing data;

2. Data standards improve departmental and institutional level data management accountability;

3. Good data policies support funding agencies data management plan requirements.

Social Implications

- Proper data lifecycle management increases data access, discovery, use/reuse;

- Metadata standards provide the origin, nature of research data, and extend the usefulness of data to science, research, and education;

- The current and future use of data allows users and the research learning communities to study, duplicate, and/or advance existing research thus creating new and/or derivative research.
Practical Implications

1. Adherence to best practices, standards, and guidelines foster cogent data policies, promote good DMC practice, and enable new research built on accessible & existing data;

2. Data standards improve departmental and institutional level data management accountability;

3. Good data policies support funding agencies data management plan requirements.
Social Implications

- Proper data lifecycle management increases data access, discovery, use/reuse;

- Metadata standards provide the origin, nature of research data, and extend the usefulness of data to science, research, and education;

- The current and future use of data allows users and the research learning communities to study, duplicate, and/or advance existing research thus creating new and/or derivative research.
Conclusions

1. Scientists from research labs with a DMP and/or part of a consortium with DMP exhibited RDM Integration Perspectives (Martin, 1992).

2. Scientists from research labs without a DMP exhibited RDM Differentiation and Fragmentation Perspectives (Martin, 1992).

3. There are opportunities for Universities and research labs to collaborate on developing RDM & DMP education, guidelines, strategies, and policies across disciplines, domains, & projects.

4. It is recommended to develop a research gateway that bridges the divide between ‘big science’ and ‘little science' (i.e. RCUK GtR).
Exploring the Data Management and Curation Practices of Scientists in Research Labs within a Research University via an Adapted Data Asset Framework (DAF) Survey – Phase 1

Plato L. Smith II, Doctoral Candidate, Florida State University - School of Information
International Association of Social Science Information Services & Technology Presentation
Toronto, Canada - June 4, 2014