The UCF Research Data Management Survey: A report of faculty practices and needs

FACRL Annual Conference
October 24, 2014

Erica England
Erica.England@ucf.edu

Penny Beile, PhD
Penny.Beile@ucf.edu
The Research Lifecycle at Univ of Central FL

A library-led institutional collaboration to develop a mental model of research support and services.

Legend:
- Research Data Management
- Faculty Center for Teaching and Learning
- Libraries
- Office of Research and Commercialization
- Institute for Simulation & Training
- Not yet supported
- Institutional Repository
- High Performance Computing
- Research Data Management Infrastructure

Design Inspiration by OpenWetWire

http://creativecommons.org/licenses/by-sa/3.0/us/
Contributors

**ITR** – Joel Hartman

**IST** – Brian Goldiez

**CS&T** – Bob Yanckello, Tim Larson, Jim Ennis

**ORC** – Ivan Garibay, Josh Roney

**Libraries** – Barry Baker, Selma Jaskowski, Lee Dotson, Sai Deng

**Lib admin team** – Penny Beile, Erica England, Joel Lavoie, Bobby Ciullo

**University of Florida** – Mark Sullivan and Laurie Taylor

University of Central Florida Libraries
Population and response rate

• Worked with ORC to identify population
  • All grant seekers in last 3 years
  • Attendees at an NSF workshop
• 534 invitations sent
• 97 people participated; 18.2% response rate
Survey sections

• Demographics
• Data collection
• Data storage and preservation
• Data recording and analysis
• Data sharing
• Needs
I. Demographics

- 84% faculty, 10% administrators, 6% other
- All colleges represented except Business Administration
- 118 affiliations noted across 21 campus units
- 50 unique departments
Demographics cont’d

• 90% collaborate with external researchers
• 57% work with teams of 1 to 5 people
• 84 people identified 120 different funding agencies
• 61% indicated that the funding agency requires them to manage, store, or share data
II. Data Collection

• Wide variety of data collected, but heavily numerical, medical and tabulated
• Wide variety of formats/file extensions, but heavily spreadsheets, statistical analysis software and text
• Volume of data generated tends to be under 50 GB
III. Data Storage and Preservation

- High use of personal computer and/or the college computer network to store data; most back up using external storage
- 68% of respondents said they take measure to preserve their data; most by backing it up. Only two mentioned they migrate file formats. Only one noted an attempt to deposit in a preservation-type facility.
IV. Data Description and Sharing

• Most researchers do not add metadata to their datasets; of the 34% who do most do not use any specific standards

• Of the 69% who do or may share data, most make the data available to peers upon request, immediate collaborators and dept’l colleagues
The “typical” researcher profile

- Collaboration
  - 90% collaborate with external researchers
  - Size of teams
- Sharing
  - 61% have funding agency requirements
  - 69% share or are willing to share; need control
- Volume and types of data
  - Volume, at under 50 GBs, manageable
  - Most data are numerical and stored in spreadsheets
- Data hygiene
  - Backed up; local PCs and networks and cloud based
  - Very few preservation-type measures undertaken
... and considerations

- Volume, 65% of respondents under 50 GBs
  - 35% have data that range from 50GB to 100TB
- Types, 87% of respondents have 62% of their data as numerical, followed by medical and tabulated
  - Text, images/audio/video files, transcription, simulation
- Formats, typically stored in spreadsheets
  - .jpg, .tif, .pdf, .doc, .mpg, .mp3, .mov, LiDAR, .gis, etc.
Faculty ‘wants’...and needs

• Infrastructure
  • Networked storage
  • Ability to store a variety of data types and formats
  • Allows for sharing of data, but able to limit access
  • Backs up, migrates, preserves data for the long term

• Services, training, and support
  • Metadata and data management training
  • Easier access and more variety of analysis software
  • Data analysis support
  • Assistance with network issues, software and hardware
Data mgt? Data literacy? Informatics?  
... Datamatics??

<table>
<thead>
<tr>
<th>Grad student workshops</th>
<th>First Generation - pubs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect/create</td>
<td>Literature review &amp; WOS</td>
</tr>
<tr>
<td>Organize</td>
<td>Citation management</td>
</tr>
<tr>
<td>Describe</td>
<td>Data documentation</td>
</tr>
<tr>
<td>Save</td>
<td></td>
</tr>
<tr>
<td>Preserve</td>
<td></td>
</tr>
<tr>
<td>Disseminate</td>
<td>Where to publish/author</td>
</tr>
<tr>
<td></td>
<td>rights &amp; Presentation</td>
</tr>
<tr>
<td></td>
<td>skills</td>
</tr>
<tr>
<td>Measure</td>
<td>Citation metrics/impact</td>
</tr>
<tr>
<td>Connect</td>
<td>Researcher profiles</td>
</tr>
</tbody>
</table>
Data mgt? Data literacy? Informatics?
... Datamatics??

<table>
<thead>
<tr>
<th></th>
<th>Grad student workshops First Generation - pubs</th>
<th>Grad student workshops Second generation – add data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collect/create</td>
<td>Literature review &amp; WOS</td>
<td>Data types and stages</td>
</tr>
<tr>
<td>Organize</td>
<td>Citation management</td>
<td>File/folder structuring &amp; naming</td>
</tr>
<tr>
<td>Describe</td>
<td>Data documentation</td>
<td>Data documentation</td>
</tr>
<tr>
<td>Save</td>
<td></td>
<td>Institutional/commercial repositories</td>
</tr>
<tr>
<td>Preserve</td>
<td></td>
<td>Storage/backup/security/migration</td>
</tr>
<tr>
<td>Disseminate</td>
<td>Where to publish/author rights &amp; Presentation skills</td>
<td>Data use agreements</td>
</tr>
<tr>
<td>Measure</td>
<td>Citation metrics/impact</td>
<td>Usage metrics</td>
</tr>
<tr>
<td>Connect</td>
<td>Researcher profiles</td>
<td>Researcher profiles</td>
</tr>
</tbody>
</table>
Session handouts and related information

- “FACRL 2014 slides” and “UCF Research Data Management Survey Report”  [https://ucf.academia.edu/PennyBeile](https://ucf.academia.edu/PennyBeile)
- UCF Libraries Graduate student workshops  
- UCF Scholarly Communication and Research Lifecycle website  
- ACRL Scholarly Communication Toolkit, Data Management section  
  [http://acrl.ala.org/scholcomm/?page_id=23](http://acrl.ala.org/scholcomm/?page_id=23)
- New England Collaborative Data Management Curriculum  
  [http://library.umassmed.edu/necdmc/modules](http://library.umassmed.edu/necdmc/modules)
  [https://www.ideals.illinois.edu/handle/2142/34712](https://www.ideals.illinois.edu/handle/2142/34712)