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

Research Data Computing Day at IST
Penny Beile, PhD
Associate Director, Information Services and Scholarly Communication
University of Central Florida Libraries
February 11, 2014
The Research Lifecycle at Univ of Central FL

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

Planning Cycle

- Impact Measures & Prestige
- Ideas
- Global Scholarly Community
- Research Planning
- Grants Search
- Grant Planning
- Proposal Development
- Research Concept

Project Cycle

- Experiment/Project
- Funding
- Compliance
- Grant Management
- Data/Output
- Conclusions
- Grant Conclusion
- Final Reports
- Publication/Presentation
- Peer Review
- Comments/Revisions
- Draft Work
- High Performance Computing

21st Century Digital Scholarship Cycle

- Disseminate
- Preserve
- Communication
- Not yet supported
- Library
- Office of Research and Commercialization
- Institute for Simulation & Training
- Research Data Management Infrastructure
- Faculty Center for Teaching and Learning
- Research Data Management

Legend

All content is available under the Creative Commons Attribution-ShareAlike license. http://creativecommons.org/licenses/by-sa/3.0/us/

Design Inspiration by OpenWetWare
Acknowledgements

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 – Erica England, Joel Lavoie, Bobby Ciullo

University of Florida – Mark Sullivan and Laurie Taylor
Survey sections

- Demographics
- Data collection
- Data storage and preservation
- Data recording and analysis
- Data sharing
- Needs
Sample and response rate

- ARGIS database plus workshop attendees
- 549 invitations sent
- 15 inactive for a total of 534 researchers
- 110 (20.6%) opened, but 13 did not select responses and were removed
- 97 people participated; 18.2% response rate
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
What type(s) of data do you generate? Please indicate an approximate percentage.

- Numerical data, e.g. ocean temperatures (%)
- Text, e.g. historical records and literature (%)
- Still images (%)
- Audio files (%)
- Video files (%)
- Medical data, e.g. patient health information (%)
- Biochemical data, e.g. raw and processed “omic” data (%)
- Tabulated data (%)
- Other (%)
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.
Some takeaways...

- **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
And now, those of you who *didn’t* respond to the survey??