

Data Management/Curation Task Force¹

Wed., Dec. 11, 2013, 1-2pm; HSC Library C2-41

Members : Hannah Norton, Laurie Taylor, Rolando Garcia-Milian, Denise Bennett, Val Minson, Joe Aufmuth, David Schwieder, Blake Landor, Mark Sullivan, Sara Russell Gonzalez, Erik Deumens, Robert Ferl, and Cecilia Botero; *Invited:* Matt Gitzendanner and Aaron Gardner

Draft Agenda

Updates and Discussion:

- Discussion/updates from work and activities of interest at UF and external
- 12/11 meeting: Mark Sullivan presenting on the IR@UF data support now, and possible for futures
 - Question/discussion topic: Tool/portal to provision dinky databases and data websites (similar to what REDCAP does for clinical trial surveys)
 - Reef data: <http://ufdc.ufl.edu/ltrstudy/datasets>
- Summer school opportunity: introduction/data science translator course
- Best days/times to meet for 2014 (Wednesdays at 1pm are bad)
 - Rotating locations, does this work? Should this just be one place? Should other locations be added?

Work towards Year One Report, strategic recommendations, specific problems/projects, etc.

- Review materials thus far
- Other questions and ideas for Year Two and report
 - Possibly sharing recommended readings or creating a reading/working group? If so, what types of readings? Are these of interest:
 - Article on BIBFrame: <http://www.inthelibrarywiththeleadpipe.org/2013/charles-a-cutter-and-edward-tufte-coming-to-a-library-near-you-via-bibframe/>
 - Chapters/all of *Big Data: A revolution that will transform how we live, work, and think*
- Other new resources in support of Year One Report:
 - Data Services, Text for Outreach and Promotion: <http://ufdc.ufl.edu/l/AA00019190/00001/pdf>
 - Fact Sheet / Overview on Data Management Support from the UF Libraries with the IR@UF & More: <http://ufdc.ufl.edu/l/AA00017119/00018/pdf>
 - IR@UF: Loading Large Files & Data Sets: <http://ufdc.ufl.edu/l/AA00017119/00016/pdf>
 - Draft text for requesting information on dinky databases: <http://ufdc.ufl.edu/AA00014835/00022/pdf>
 - Research Computing Vision; note radical collaboration: <http://ufdc.ufl.edu/l/AA00014835/00023>
 - IR@UF :: Theses and Dissertations (includes section on supplemental data): <http://ufdc.ufl.edu/l/AA00017119/00002/pdf>

Upcoming events scheduled and to be discussed/planned

- Zotero workshops (citation management software for data in bibliographic databases and connects to many tools for text/data mining)
- DMPTool, scheduling hands-on training
- Workshop for outreach for HiPerGator

Resources

¹ Data Management: <http://www.uflib.ufl.edu/datamgmt> & DMCTF resources: <http://ufdc.ufl.edu/AA00014835/>

- Meetings: alt. Wed.; HSC Library C2-41, Library West 429, Marston Science Library L107

Ongoing

- Planning and supporting different informational, training, and outreach activities and events on data and related resources like HiPerGator
- Workshops (types for different groups: researchers, and data service providers); known needs:
 - DMP Tool for Librarians (and other Data Liaisons/Supporters to be identified)
 - DMP Tool and creating a plan
 - Possible workshop: Primer on Data Management, 2 hour version, expanded primer within 2 day workshop, co-taught with teaching faculty in-field; expanded primer within lab-style courses as with research and methods courses, etc.

Deadlines/Events

- November:
 - Presenting to libraries; work on survey result analysis; RC Day; GIS Day
 - Work towards larger Year One report and strategic directions/recommendations
 - Quarterly report due for July-September
- 2014 January:
 - Quarterly report due for October-December
 - Year One Report, draft due to group²
- 2014 February:
 - Year One Report due to Deans of the Libraries
 - Future surveys/data gathering for feedback on data needs with possible questions³

² See charge and notes: Draft proposed recommendations as whitepapers for review/approval/implementation to include: Recommendations for the Libraries' campus-level role in support of data management and curation; proposing a corresponding framework and resources for library support of the data life cycle; recommending the role of the institutional repository and research computing in storing, finding, and accessing working and final data, and linking publications to supporting data; and, recommending a framework for liaisons and subject specialists to incorporate data instruction and consultation into their workflows. Outline with detailed plan for training and other supports based on information gathered during Focus Groups, survey, and other activities; plan for ideal (more resources) and for conservative (current resources); Outline with detailed information on how the IR fits in the overall supports for data; and same for other applicable resources that can be used/leveraged as is now, and detailed information on how to enhance or make best fit

³ Possible questions:

-- How would you like authenticated users to be able to interact with the data on-line, if you were to make it available? [Download only; Search on site, no download; Run statistical analysis across my data; etc.]

-- What type of data visualizations would you like authenticated users to have access to regarding my data on-line? [A, B, C, D, etc., write-in]

-- If you (or other authenticated users) could add individual records through a form on the online system, would you transfer the data to the system and rely on it for working access and long-term preservation?

Initial Draft for Discussion

The initial draft notes below are towards a possible course to aid in translation competency with data (for working with Data Scientists, no prereqs, not necessarily heavily technical, etc.). The course could draw on theories of the database age, procedural rhetoric, data provenance for reproducible research, and help frame questions and learning for changes in working, thinking, and doing scholarship and research overall in the Data Age. Readings could include Manovich, Bogost (*Persuasive Games*: “practice of authoring arguments through processes [...], through the authorship of rules of behavior, the construction of dynamic models” (29)).

Introductory Concepts in Research Computing

3 Credits

Fall/Spring, or Summer A/B compressed course

Undergraduate/Graduate sections possible (at what level?)

Purpose

Working with “Big Data”, or large numbers of digitized texts, images, sounds, and other information sets enables students and researchers to ask new and exciting questions in their fields. This research is termed ‘computational’ because it involves harnessing computer power to examine more sources than is possible by any individual or team. This course is an introduction to the basic concepts that will enable students to collaborate with computer scientists to develop or support computational research projects in different fields. The primary goals of the course are to help researchers to determine what types of data modeling tools to use for their research, and to provide an introduction to associated computing concepts. This course will not teach or involve computer programming.

Prerequisites: None. Anyone interested in using computers for research is encouraged to attend.

Format

Classes will be part lecture, discussion, and guided inquiry with hands-on examples to work through different concepts and learn different programs. Students will produce a computational research proposal at the end of the course.

Course Content

Overview of Research Computing, Common Uses and Tools

- What are Data, and Where Do They Come From?

- Computer Simulations

- The Monte Carlo Method

- GIS

- Data Mining and OCR

- Visualizations and Everything Else

Unit Operations

- Procedural Rhetoric

- Grounded Theory Approaches to Analysis (Functional and non-Functional Requirements)

- Introduction to *nix and Shell Scripting

- Other Systems Operations

Overview of Applications, Programming Languages, and Libraries used in Research

- Commercial Software Examples

- Open Source Software

- Package Managers

- Scripting Languages

- High Performance Compiled Languages

Brief Overview of [Parallel Computing](#) Techniques and Resources

- GPUs and Moving Data

- On-Campus resources: HiPerGator

Data Management

- Data Storage and Curation

- Ethics of Big Data