This meeting was cancelled

Data Management/Curation Task Force

Wed., Nov. 27, 2013, 1-2pm; Marston Science Library L107


Draft Agenda

- Meeting Cancelled.
- Next Meeting is Dec. 11, 2013.

Resources


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 group
- 2014 February:
  - Year One Report due to Deans of the Libraries

1 Data Management: http://www.uflib.ufl.edu/datamgmt & DMCTF resources: http://ufdc.ufl.edu/AA00014835/
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
Future surveys/data gathering for feedback on data needs with possible questions

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