Getting in on the Ground Floor: Developing a Data Management Portal for Graduate Students

by Andi Ogier (alop@vt.edu) and Laura Miller (lam2c@virginia.edu), University of Virginia Library

Problem Statement

Data Management is a well-known and oft-cited problem for researchers. Because almost no advance training is available, researchers tend to learn data management practices on the job in an ad hoc manner. The pressure to publish provides little incentive to learn about and adhere to data management practices, such as creating appropriate metadata and planning for format obsolescence, which often create more work for a researcher at a later stage in their research process. However, education on and adoption of good data management practices early in the career of a researcher can do much to later alleviate the necessary work of data management.

Project Vision & Partnership

A partnership between the UVA Library's Data Management Consulting Group (formerly SciDaC) and the Vice President of Research's Graduate Studies Office was formed to create a scalable and sustainable Data Management Portal aimed at graduate students. This portal would deliver best practices, institutional services, and curated web resources to students at their time of need.

Benefits:
- Combines data management expertise with graduate research experience
- Provides different levels of information (beginner, advanced, etc.), mapped to the graduate career trajectory
- Uses Grad Studies website and listserv as a communication channel to reach students where they are - with no need to seek out help in the library

Challenges:
- Determining how to structure, brand, and pitch information so that it will meet the needs of both partners and scholars of all disciplines
- Devising strategies for keeping content dynamic, current, and relevant

Market Research

We interviewed 13 Graduate Students in Biomedical Engineering, Educational Psychology, Physics, Chemistry, Economics, Politics, Spanish, Italian and Portuguese.

Common Needs:
- Standards and tools for file organization, storage, and sharing
- Better documentation of procedures, data transformations, and data itself (especially in labs with inherited on-going research)
- Tips for effective and transparent spreadsheet design
- Early management of coursework and readings
- Mechanism to clarify (long and short-term) ownership and responsibility for data

Other findings:
- Information streamed through departmental rather than administrative channels was most likely to be read
- Best source of information was other grad students, not advisors

Site Architecture

In order to facilitate graduate student access to immediately useful content, we created three different entrances into the site: by stage in the Graduate Student Lifecycle; by topical Data Need; and by research Subject Area. Each content item was tagged with a category from at least two of these three entry points.

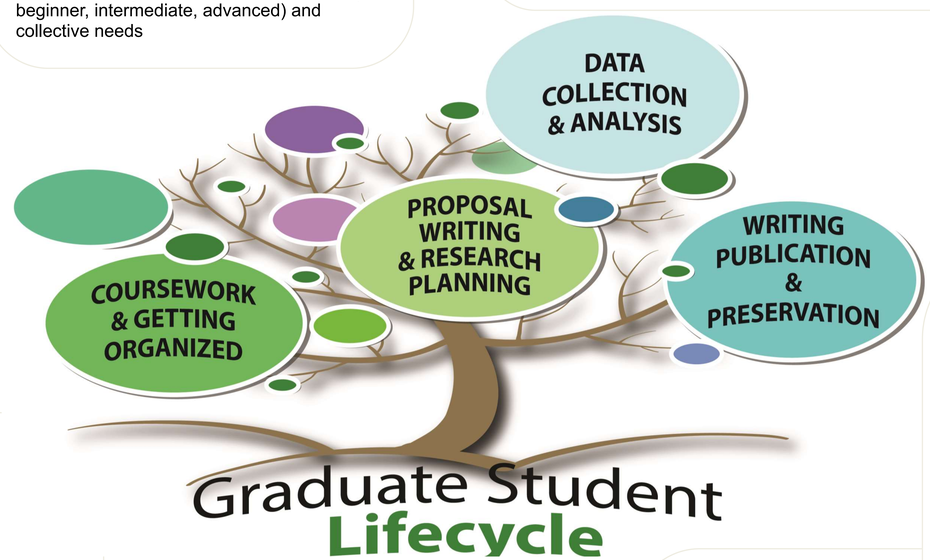
Benefits:
- Creating three distinct entry points through which to offer content allows us to tailor the presentation of that content to fit the needs of a specific graduate student at a particular time.

Challenges:
- Classifying content according to the models (e.g., a article on the effective use of spreadsheets could be classified according to multiple Data Needs or stages within the Graduate Student Lifecycle. In the end, we categorized and tagged content with an eye towards its immediate use, and allowed important content in the portal to show up within multiple contexts.

Graduate Student Lifecycle

The Graduate Student Lifecycle (GSLC) model was developed to deliver immediately useful content to students in a way that integrates directly into their individual graduate school experience.

The GSLC model:
- Connects topical content to curriculum benchmarks as a way to estimate user skill level (i.e. beginner, intermediate, advanced) and collective needs



Delivery Stream

We used a locally-hosted instance of WordPress to build the Data Management Portal.

Benefits:
- Ability to aggregate content posts through categorization and display them on a static page (invaluable in creating a dynamic, scalable, self-referential, and easily sustainable site). The WordPress platform also allowed easy integration into the VPR's website.

Challenges:
- While a WordPress platform can be very customizable via Themes and Plugins, the learning curve for creating a site can be very high.
- Without a programmer to dress it up, it retains that scrolling, blog-like look.

Data Needs & Subject Areas

- The Data Needs model was developed to deliver topically useful information to grad students at different levels of depth. It offers Graduate Students access to content based on suggested actions that could be performed on data.
- However, our market research revealed that Graduate Students often prefer to encounter content through the filter of their school or department. Coding discipline-specific resources with subject tags was a simple way to respond to their requests.

Content Types

- Videos
- How-to's
- Brief topical introductions
- In-depth lectures
- Step-by-step written guides
- Screencasts
- Best practices
- Recommended tools
  - Institutional (for UVA users)
  - External
- Self-guided instruction modules
- Programming tutorials
- Links to software and data forums
- Discipline-specific resources
- Links to funding agencies and guidelines

Data Needs Chart

Launch & Reception

The Portal was soft launched in the Fall of 2012 at and has been the basis for informational sessions since. See it for yourself at http://dmconsult.library.virginia.edu/.

- Reviews have been positive, but students still seek a more personalized method for triaging and fulfilling their specific data needs.
- The new Graduate Studies website launch was delayed a full semester, so we have not yet seen the increased traffic we were anticipating.

Next Steps

- We are in the process of hiring a programmer to bring the portal's design more closely in line with the VPR's and the Library's sites.
- Building on continued conversations with grad students, we have hosted a UVA/VT Data Management Bootcamp, and the first of several Software Carpentry training sessions to address gaps in DM and programming knowledge.

Special thanks to: Thea Atwood, Brian Craig, Andrew Sallans, Phil Trella, the DMConsult staff, and all the students we were lucky enough to interview.

University of Virginia Library

acrl 2013 peer revered