Library Data Services: current and future state

Faculty Library Committee
October 5, 2015
Introducing: Project Open Data

Technology evolves rapidly, and it can be challenging for policy and its implementing agencies to keep up. Last week, President Obama launched the Administration’s new Open Data Policy, aimed at ensuring that data released by the government will be as open as possible, and that this tech-focused policy can keep up with the speed of innovation.

Project Open Data is an online, public repository intended to foster collaboration and the improvement of the Open Data Policy. We wanted to foster a culture change that will lead to increased collaboration and open data, and where anyone can help us make open data work better. For example, developers can take the resources and plug-and-play tools in Project Open Data and help accelerate research. The idea is that anyone, from Federal agencies to state agencies, can freely use and adapt these open source tools—and that’s exactly what we intend to see happen.

NSF Data Management Plan Requirements

Beginning January 18, 2011, proposals submitted to NSF must include a supplementary document of no more than two pages labeled "Data Management Plan" (DMP). This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. Proposals that do not include a DMP will not be able to be submitted. For more information about this new requirement, please see the Grant Proposal Guide, Chapter II.C.2.j and the Data Management and Sharing Frequently Asked Questions (FAQs).

Please note: the Engineering Directorate (ENG) has additional guidance for proposals submitted to ENG programs, [http://nsf.gov/eng/general/ENG_DMP_Policy.pdf](http://nsf.gov/eng/general/ENG_DMP_Policy.pdf). Questions and/or suggestions about this new requirement may be addressed to Dr. Maria K. Burka at mburka@nsf.gov.
Publisher mandates
Supporting Research Data Management

Consulting on Data Management Plans
- Support across disciplines: Sciences, Social Sciences, and Humanities
- Understand the data being generated/collected, explore options for sharing
- Close collaboration with A&S/SOE Research Affairs, OVPR, TTS

Programming on related issues in scholarly communications
- ELN project: best practices in research data management
- Graduate students: GREAT program, targeted workshops
- Responsible Conduct for Research
- Open workshops
Data Sharing/Data Management Plans

- "the *types of data*, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- the standards to be used for data and metadata format and *content* (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- *policies for access and sharing* including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- *policies and provisions for re-use*, re-distribution, and the production of derivatives; and
- *plans for archiving* data, samples, and other research products, and for preservation of access to them"
Data Management Planning: ELN project, etc.
Repositories: sharing and preservation

Discovery
- Scope + reach: audience and indexing vary by repository

Access
- Controls and downloads: embargo and file format options vary

Preservation
- Backup and security?
- Long-term strategy?

Tufts
- Tufts Digital Library
- Trove, MIRA
Challenges

- New research tools and how to integrate them
- New research administration tools and workflows
- Building the repository
- Compliance with emerging mandates
- Scaling and sustaining support
Supporting Data Discovery and Analysis

Teaching and research support

- Course-based workshops tailored to specific projects and assignments
- Open workshops
- Create portals for efficient access to data sources
- Course-prep/assignment design: provide insight into student experience with data and ways to integrate data into coursework
- Expand access to research technology in Tisch Library
Course-based and open workshops

Course-based:
- Methodology: exploring ways to analyze and use data for a research project
- Discovery: finding data for specific assignments
- Analysis and presentation: getting started cleaning, analyzing, and visualizing data

Open: focus on software and best practices
- Quantitative analysis: SPSS, Stata
- Quantitative analysis + scripting + visualization: R, Python
- Visualization: “do’s and don’ts of data visualization”
- Qualitative analysis: Nvivo
Data Visualization

IF BUSH TAX CUTS EXPIRE

Top Tax Rate

Now

JAN. 1, 2013

If Bush tax cuts expire...

Top tax rate

0%

10%

20%

30%

40%

35.0%

39.6%

Now

Jan. 1, 2013

TischLibrary
“I am trying to do a Wilcoxon rank sum with several if statements in Stata and am struggling. Is there any chance that you have time today to help me out with this?”

--Civil Engineering Undergraduate

“I’m really struggling with figuring out a couple things in Stata. I took Biostatistics last spring so I am familiar with Stata, but I learned how to do some very specific things with much easier data sets…Specifically, I’m struggling with obtaining percentages for some string variables, where I want to generate a new categorical variable combining all of them.”

--Biology Undergraduate

I have run a study and am writing my publishable paper for my OT doctorate. I am having difficulties with my Pearson correlation coefficients making sense, and on how to run an ANOVA…”

--Occupational Therapy, Doctoral Candidate
“I'm conducting a survey for my MA thesis and could use some guidance on how to best proceed with the statistical analysis. The survey is on the public perception of manufacturing jobs and I should wrap up data collection by next Tues. Can we set up an appointment to discuss after that?” --Urban Planning Masters Student

“My book will be published in early August -- proofs, index etc. is all in and done, but a problem came up with one graph…” --Faculty member
Data Lab

Evolution of GIS Center

- Support for spatial, numeric and textual analysis
- Integrates Tisch data discovery and analysis services
- New classroom and collaborative workstations
- An opportunity to grow services in support of data and explore models for sustaining/scaling support
Challenges

There is only one Josh…