



Assessing Research Data Management Needs at Emory University

Jennifer Doty, Katherine G. Akers,
Jonathan Bodnar, and Lori Jahnke
Emory Libraries & Information Technology

Vincent Carter
Emory Office of Institutional Research,
Planning & Effectiveness

Southeastern Library Assessment Conference
October 21, 2013

RDM Assessment @ Emory

- Data Services @ Emory University Libraries
- Why an(other) assessment?
- Survey design and methods
- Survey results: analysis and conclusions
- Interviews: planning, enacting, reviewing
- Future Directions

Briefly covering today:

1. History of data services at Emory Libraries
2. Primary reason for our data needs assessment
3. Survey design and methods
4. Analysis of results and conclusions drawn
5. Interviews—how we planned for them, how we've conducted them, what we've learned from this type of assessment
6. And future directions for our assessment strategy

Context:

1996: Emory University Libraries' Electronic Data Center founded

- *Initial service focus on access to and analysis of data*

2012: added two new positions:

- Data Management Specialist
- E-Science Librarian

Develop services to address the rest of data life cycle

SELAC 2013: Oct 21st



3

Emory University Libraries' Electronic Data Center was founded in 1996. Services more focused on early stages of working with research data (access to and analysis of datasets).

After participating in the 2011 ARL E-Science Institute, and considering developments in data collection, storage, analysis and long-term access needs across campus, the libraries added two new positions (Data Management Specialist, E-Science Librarian) in the Summer of 2012 to develop services to address the rest of the research data life cycle.

Why an(other) assessment?



Research by neil conway on Flickr / CC BY

SELAC 2013: Oct 21st

RDM Assessment @ Emory

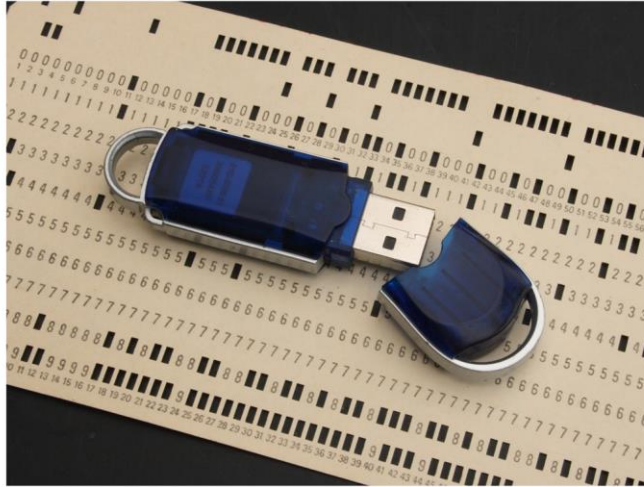
4

We first reviewed other survey and assessment examples from other institutions (GaTech, Cornell, U.Va., Cal Poly, UNC).

Ultimately decided a unique assessment of RDM needs across our campus would help to identify areas where library staff can target most beneficial, tailored assistance. And we believed it would greatly help to have hard data to support proposals for additional services and resources to address identified needs across campus.

Another incentive is that a campus-wide survey also serves as a good outreach tool—we weren't looking just to quickly obtain quantitative, institution-specific information; we also wanted to use it as a way to raise awareness among researchers about data management issues.

Survey: Design & Methods



Data storage - old and new by Ian S on Flickr / CC BY-NC-ND 2.0

Next, we met with Susan Bailey, Emory Libraries' Coordinator of Assessment, who conducts the libraries' annual survey of all faculty & students. Thanks to Susan's work, the libraries already had a strong working relationship with our Office of Institutional Research.

We then met with Vince Carter in the Emory Office of Institutional Research, Planning & Effectiveness to discuss the process for conducting a campus-wide survey.

Survey: Design & Methods

Planning:

- Survey design—software/platform to use, composition and format of questions
- Timing of survey distribution
 - avoiding conflicts with other campus surveys
 - maximizing response rate by targeting a break in faculty teaching schedules (e.g. Fall break)
- Methods for distribution
- Collection and analysis of responses

In our first meeting we discussed various aspects of the survey design, identifying the best time to run the survey to avoid conflicts with other campus surveys and maximize the response rate. We chose to initiate it during Fall break, when faculty have a brief lull in their teaching schedule and may be more inclined to complete a survey.

By conducting the survey using the campus implementation of Qualtrics, we were able to pipe in university human resources information—emails are distributed more efficiently, those who have already completed the survey won't be bothered again, and people can opt out of receiving further messages about this particular survey.

Maintaining the connection with HR data also meant that the responses collected could be analyzed using the demographics of the respondent pool. But we did have to include a disclaimer in the survey introduction so participants acknowledged that their responses would be confidential, not anonymous.

Survey: Instrument

1. Do you conduct research that generates some type of data (spreadsheets, text, images, videos, audio files, instrument files, photographs, physical samples/specimens, etc)?

#	Answer	Response	%
1	Yes	358	83%
2	No	71	17%
	Total	429	100%

Statistic	Value
Min Value	1
Max Value	2
Mean	1.17
Variance	0.14
Standard Deviation	0.37
Total Responses	429

1a. Are you planning to conduct research that generates some type of data (spreadsheets, text, images, videos, audio files, instrument files, photographs, physical samples/specimens, etc) in the future?

The questions in our survey were developed in collaboration with other members of our research data management working group. We had drafted questions using a variety of examples (e.g. the Data Asset Framework used by GT, and other examples from surveys run at other institutions).

With Vince's input and guidance, we finalized the questions and the survey was distributed using Qualtrics online survey software (Emory has a site license) to our target audience of all faculty researchers across campus.

We considered sending the survey to a broader population, but it's difficult to identify students and staff members involved in research using the university HR data, and faculty are more likely to be principal investigators guiding the data management for research projects.

Survey: Representation

College/School	Survey Respondents	Statistical Analysis	Representation
Candler School of Theology	n = 6 out of 98	Z = 0.07, p = 0.94	Representative
Goizueta Business School	n = 8 out of 106	Z = 0.66, p = 0.51	Representative
Oxford College	n = 8 out of 97	Z = 0.90, p = 0.37	Representative
Nell Hodgson Woodruff School of Nursing	n = 8 out of 210	Z = 0.83, p = 0.41	Representative
College of Arts and Sciences	n = 105 out of 975	Z = 6.50, p < 0.001	Over-representative
School of Law	n = 3 out of 225	Z = 2.87, p = 0.004	Under-representative
School of Medicine	n = 163 out of 3312	Z = 3.43, p < 0.001	Under-representative

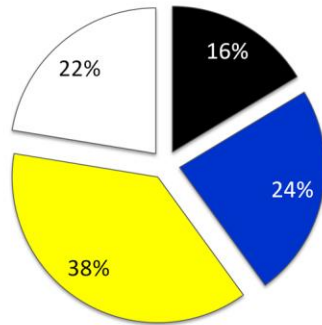
SELAC 2013: Oct 21st

RDM Assessment @ Emory

8

From a total of 5590 Emory University faculty members, 456 initiated the survey for an 8% response rate (which may sound low but is actually quite good for a campus-wide survey at our institution). Of these, 330 answered ‘yes’ to an initial question of whether they conducted research that generates some type of data and provided answers to at least one subsequent survey question. All analyses focused only on these 330 faculty members, who represented all of Emory University’s major schools and colleges. Statistical analysis was run on the response rates by college or school affiliation to determine if samples were representative, which the majority were.

Survey: Analysis



Faculty Respondents by Disciplinary Categories

- Arts & Humanities (54)
- Social Sciences (78)
- Medical Sciences (124)
- Basic Sciences (74)

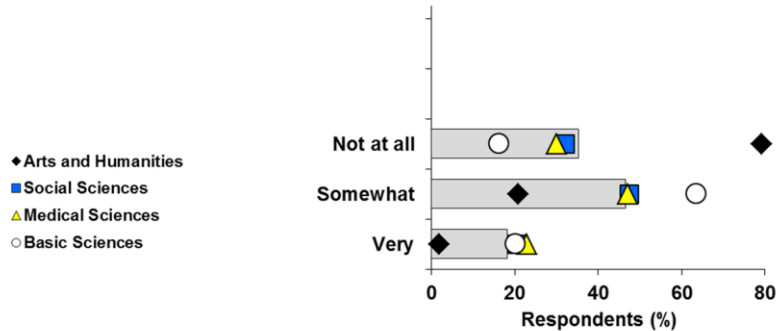
Total Respondents: 330

Once we cleaned the data for incomplete answers or those who answered “no” to any data collection in their research, we were left with 330 responses to further analyze.

Our method for grouping by major disciplinary categories: to evaluate for possible differences among fields, assigned respondents to one of 4 groups. Some were by primary departmental affiliation (e.g. all of Art History assigned to Arts/Hum); some by reviewing their specific research topics or methods (e.g. Psych folks could go to Basic Sci or Soc Sci). Medical Science—research conducted in clinical, “applied” setting. Basic Science—lab, “experimental” setting.

Survey: Analysis

How familiar are you with the requirement for data management and/or sharing plans as components of many funding applications (e.g., NSF, NIH, NEH)?



SELAC 2013: Oct 21st

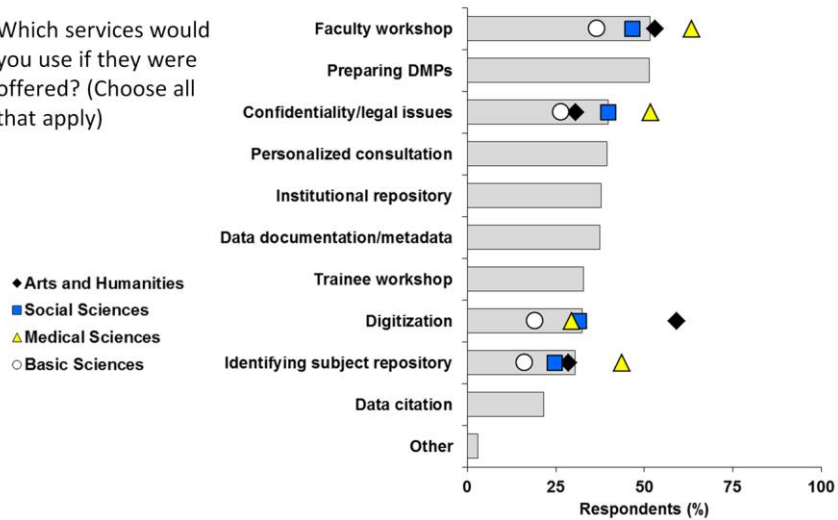
RDM Assessment @ Emory

10

We then could look at both overall trends in researchers' responses (ie. they are not at all, or only somewhat, familiar with funders' data management plan requirements), and also examine the data for significant differences based on disciplinary category—e.g. researchers in the Basic Sciences have greatest levels of familiarity with data management plans, while the majority of Arts & Humanities faculty have not yet encountered them.

Survey: Analysis

Which services would you use if they were offered? (Choose all that apply)



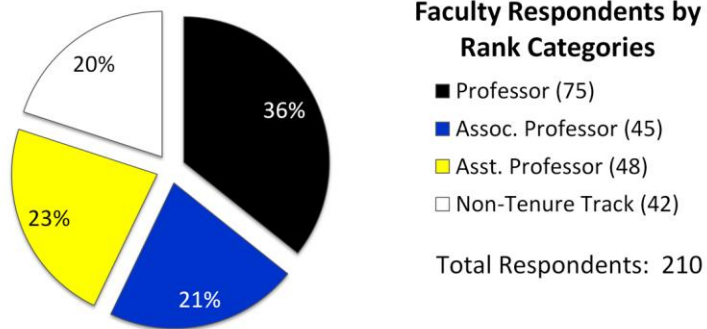
SELAC 2013: Oct 21st

RDM Assessment @ Emory

11

We ran this analysis on the responses to each of the survey questions, which led to some interesting conclusions. Surprisingly, faculty workshops on data management were the most popular service requested. This has led us to investigate possible partnerships across campus to address areas of greatest interest and concern (i.e. confidentiality) and seek to incorporate training into ongoing schedules of offerings to PIs and administrators. And the strong interest in digitization by Arts & Humanities faculty could help guide future strategy by the new center for digital scholarship in the library.

Survey: Analysis



SELAC 2013: Oct 21st

RDM Assessment @ Emory

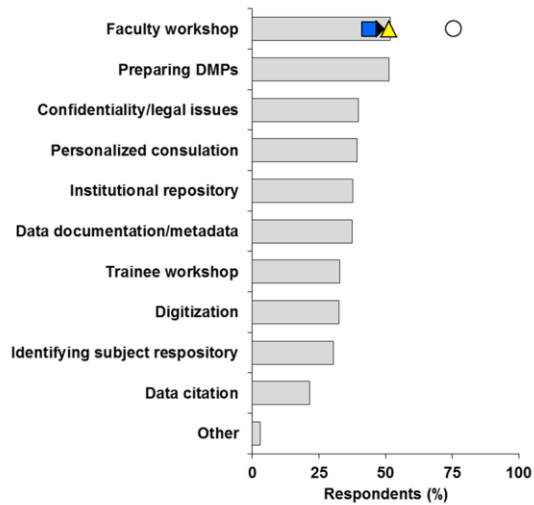
12

We ran further analysis based on rank of faculty members to evaluate for possible differences across stages of careers. Our method for grouping by major rank categories was to assign respondents to one of 4 groups (excluding the clinical/research track in medical sciences). Initially we found it interesting to note that largest slice of respondents are tenured full professors, however, in analyzing the general numbers, full professors are the largest segment of the overall faculty population as well.

Survey: Analysis

Which services would you use if they were offered? (Choose all that apply)

- ◆ Professor
- Associate Professor
- ▲ Assistant Professor
- Non-Tenure Track



SELAC 2013: Oct 21st

RDM Assessment @ Emory

13

Perhaps not surprisingly, the significant number of non-tenure track faculty interested in workshops on data management may explain its overall popularity as a possible service.

Next steps: conducting one-on-one interviews with researchers; collaborations with campus partners to develop appropriate services.

Interviews

- Planning
 - IRB approval process
- Enacting
 - Recruiting participants
 - Scheduling and conducting interviews
 - Recording/notekeeping methods
- Analysis...

Interviews—we began planning as we completed the survey, with submission of the IRB application taken care of first. Initial interviewees were recruited from survey respondents who had indicated they were willing to participate in follow-up interviews. Additional participants have been recommended by folks we interviewed, subject librarians with prior knowledge of faculty member and graduate students in their departments who conduct research using data.

Scheduling interviews has proved one of our greatest challenges. We are audio recording each interview after getting the participants' consent, but we also make sure two people can attend—one to lead the interview and one to take notes. The notes are then entered into an Access database that was modified from an existing one used by Lori Jahnke, our anthropology librarian, for a previous interview project examining researchers' data curation practices.

And analysis—the discussion for how best to analyze and report out on what we are learning from the interviews is ongoing.

Interviews: Lessons Learned

- Self-selection bias of participants
- Realistic expectations for timeframe to conduct and analyze
- Challenges of selecting appropriate analysis method(s)

People who consented to do interviews trend towards those who are already thinking about research data management issues related directly to their work. Conducting an assessment on this scale requires commitment of personnel time and resources that can be challenging to combine with ongoing projects and responsibilities. Also, the process for getting IRB approval is lengthy and requires planning adequate time to complete any mandatory training prior to submitting the application.

Future Directions



Direction by Ramkarthikblogger on Flickr / CC BY 2.0

SELAC 2013: Oct 21st

RDM Assessment @ Emory

16

19 interviews completed, but would like to conduct more to include broader representation from disciplines across campus. In process of formally chartering a working group that crosses the organization (library, IT, new center for digital scholarship) to focus on areas where there are gaps in services to support RDM and make recommendations.

References

- Akers, K.G. & Doty, J. (In press). Research data management practices and perspectives: Differences among the arts and humanities, social sciences, medical sciences, and basic sciences. *International Journal of Digital Curation*.
- Asher, A. & Miller S. (2011). A practical guide to ethnographic research in academic libraries. <http://www.erialproject.org/publications/toolkit/>
- DaMaRO project, University of Oxford Research Data Management Survey 2012: <http://blogs.oucs.ox.ac.uk/damaro/2013/01/03/university-of-oxford-research-data-management-survey-2012-the-results/>
- Jahnke, L., Asher A., & Keralis, S. D. C. (2012). The problem of data. Council on Library and Information Resources, pub 154.
- Parham, S. W., Bodnar, J., & Fuchs, S. (2012). Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech. *C&RL News*, 10–13.
- Scaramozzino, J. M., Ramírez, M. L., & McGaughey, K. J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, 73(4), 349–365.
- Westra, B. (2010). Data services for the sciences: A needs assessment. *Ariadne*, 64.

EMORY



Contact Information:

Jennifer Doty | jennifer.doty@emory.edu
Data Management Specialist
Emory Center for Digital Scholarship

[bit.ly/EmoryRDM]

Questions?



Green Question Mark by mikecogh on Flickr / CC BY

SELAC 2013: Oct 21st

19