

The Position of Library-Based Research Data Services: What Funding Data Can Tell Us

Karl Nilsen
Robin Dasler
Trevor Muñoz
Sarah Hovde

University of Maryland, College Park
April 5, 2013

Context

As academic research libraries develop services to support data management and curation, understanding the demand from researchers for new services and establishing parameters for pilot projects are key challenges for managers.¹

Data about proposals and awards for research funding provide evidence about the potential scale, scope, and institutional location of research and data production. Information obtained from funding data can complement and contextualize insights obtained directly from individual researchers about their data management needs.

Data Sources

To understand the composition and distribution of research funding at the University of Maryland, College Park (UMD), the authors examined data about proposals and awards retrieved from:

- University of Maryland Office of Research Administration
- NSF Awards Database
- NIH RePORTER
- Research.gov (for NASA)

Objectives

Librarians at other institutions have used funding data to support planning and outreach, typically identifying potential candidates for interviews or participants for training and instruction.² In contrast, because research data services at UMD are in start-up phase, the authors aimed to discover what funding data can tell librarians about the demand for data management support and the potential challenges for library-based services. The authors also sought to understand the limitations of funding data as a source of information. Findings from this investigation will help librarians at UMD allocate resources, develop services, and design outreach strategies.

Contact

lib-research-data@reflectors.mail.umd.edu

Acknowledgements

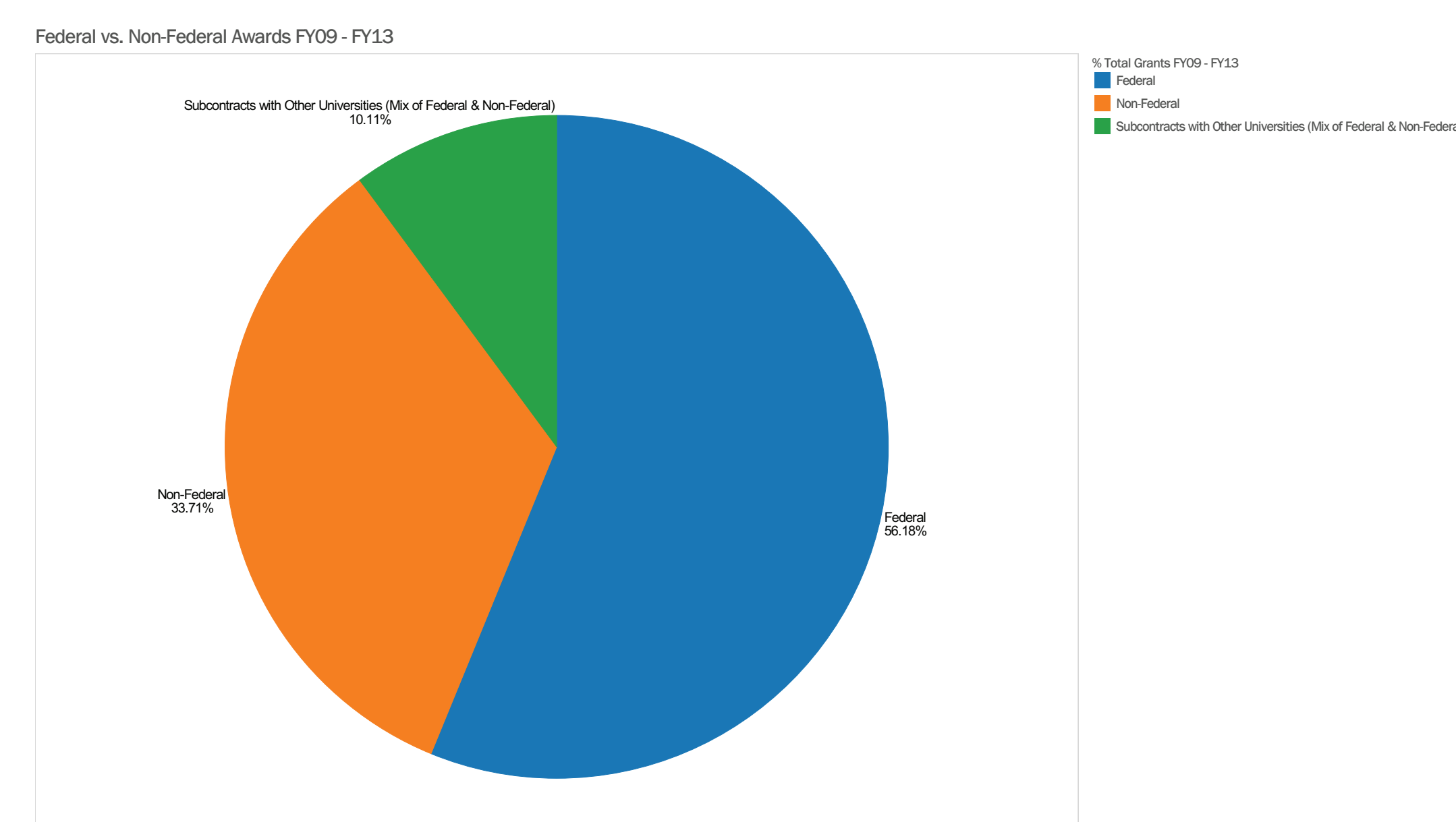
The authors are grateful to Office of Research Administration at the University of Maryland, College Park, for their assistance with this project.

References

- Fons, Ted, Mike Furlough, Elizabeth Kirk, Judy Luther, and Michele Reid. *Fit for Purpose: Developing Business Cases for New Services in Research Libraries*. Council on Library and Information Resources, 2012. <http://mediacommons.futureofthebook.org/mcpress/businesscases/>. p. 3, 17-18.
- Steinhart, Gail, Eric Chen, Florio Arguillas, Dianne Dietrich, and Stefan Kramer. "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements." *Journal of eScience Librarianship* 1, no. 2 (2012). <http://escholarship.umassmed.edu/jeslib/vol1/iss2/1/>; Johnston, Lisa, Meghan Lafferty, and Beth Petsan. "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach." *Journal of eScience Librarianship* 1, no. 2 (2012). <http://escholarship.umassmed.edu/jeslib/vol1/iss2/2/>; Peters, Christie, and Anita Riley Dryden. "Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston." *Science & Technology Libraries* 30, no. 4 (2011): 387-403. doi:10.1080/0194262X.2011.626340. pp. 389-90.
- Office of Science and Technology Policy. "Expanding Public Access to the Results of Federally Funded Research." WhiteHouse.gov, 2013. <http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>.
- Rodrigo, Allen, Susan Alberts, Karen Cranston, Joel Kingsolver, Hilmar Lapp, Craig McClain, Robin Smith, Todd Vision, Jory Weintraub, and Brian Wiegmann. "Science Incubators: Synthesis Centers and Their Role in the Research Ecosystem." *PLoS Biology* 11, no. 1 (2013): e1001468. doi:10.1371/journal.pbio.1001468.

Notable Findings

1 Federal policies put additional pressure on traditional service models



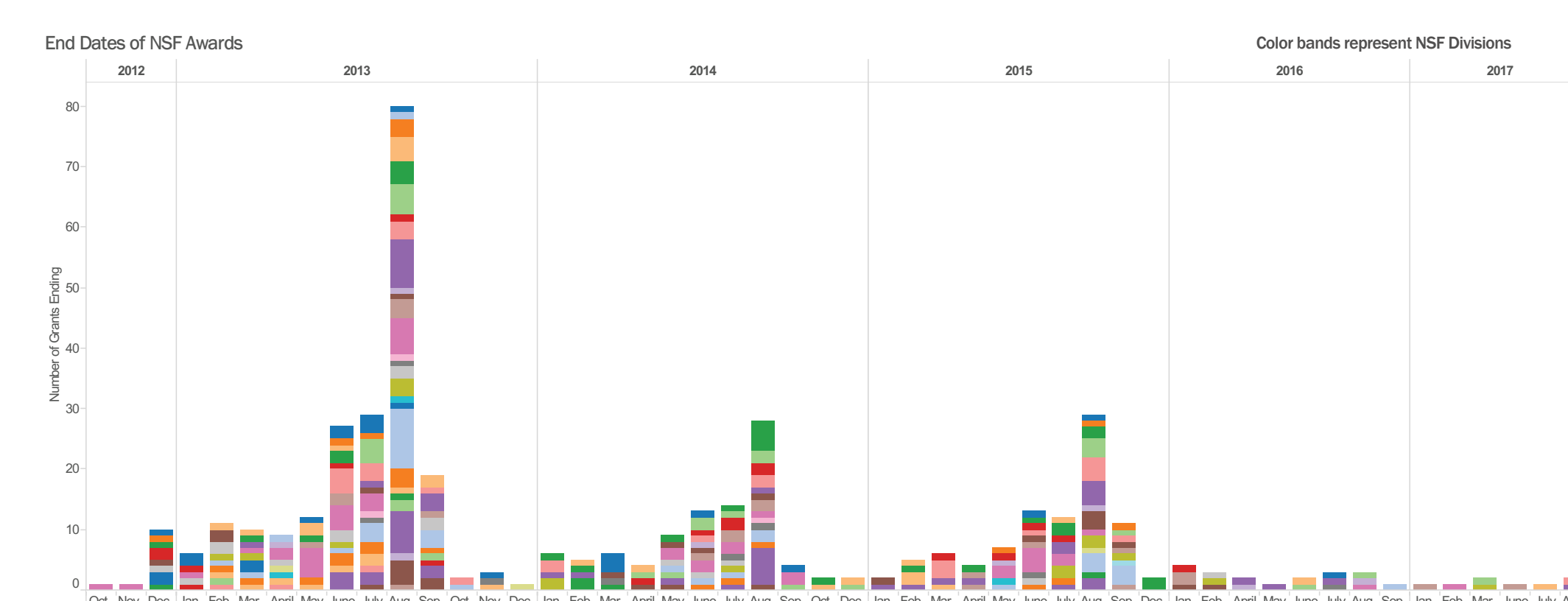
Federal sponsorship accounted for almost two-thirds of awards and supported close to 1,222 distinct investigators at UMD in FY09-13.

Given the importance of federal sponsorship to research at UMD, U.S. government science and technology policy will have a massive influence on data management support services. The Office of Science and Technology Policy memorandum directing federal agencies with over \$100 million in annual research and development expenditures to support public access to data will likely compel many UMD researchers to pay greater attention to data management.³

Implications:

As federal policies transform more and more researchers into potential clients for data management support services, it becomes difficult for libraries to provide personalized consultations or embedded support to every researcher. Unlike library services designed to deliver uniform support across the campus, research data services may be forced to allocate resources to a limited number of projects. At UMD, the authors are considering a selection process that will allocate resources to researchers whose projects match well with the priorities of the university, the relevant college, and the Libraries.

4 Award end dates signal outreach opportunities

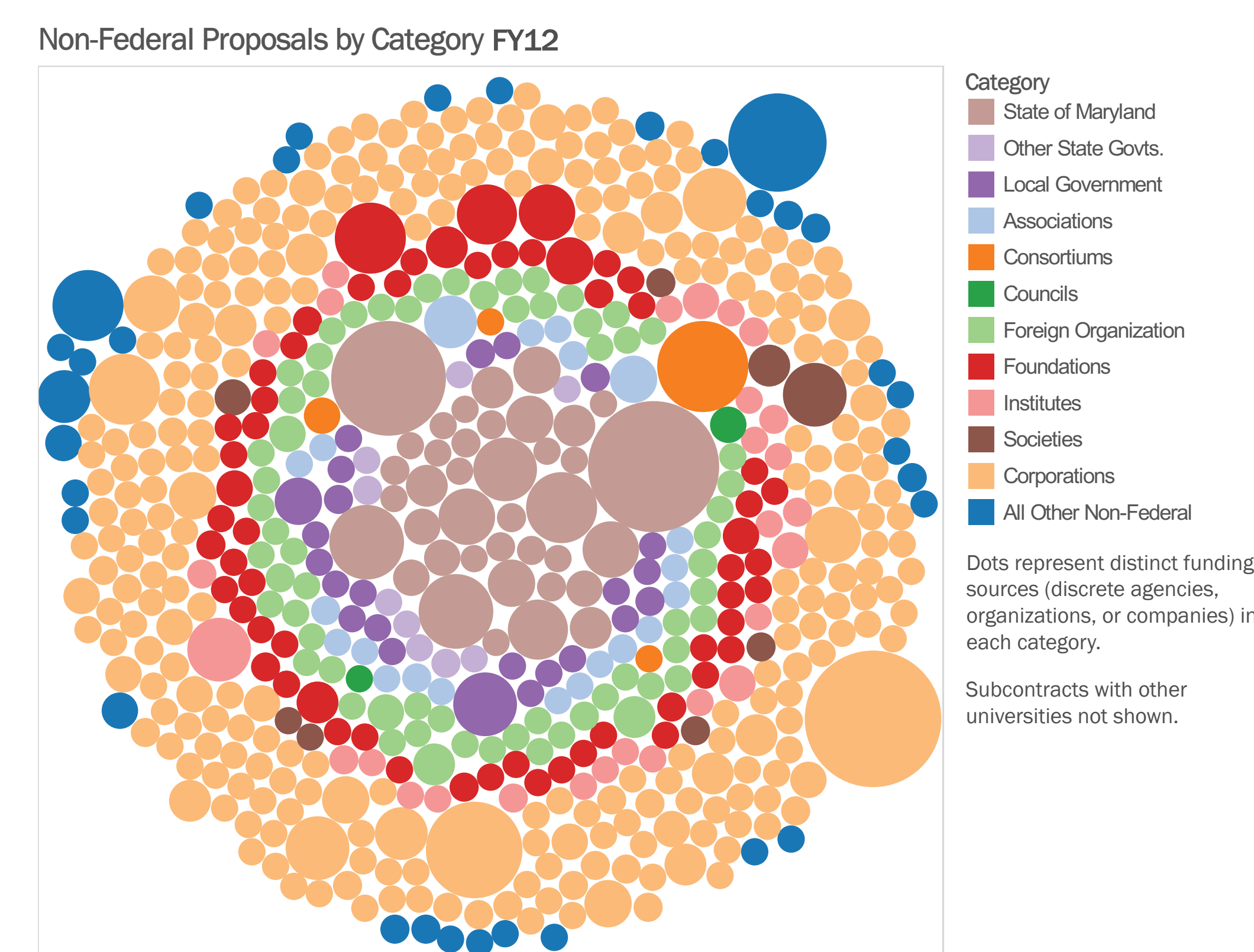


Award data from the NSF database, NIH RePORTER, and Research.gov contain end dates for individual awards. In some cases, researchers may renew an award, but, in other cases, their project may be complete and their research products available for curation and preservation. Shown here, there is a spike in NSF end dates at UMD in late summer.

Implications:

The authors intend to use upcoming end dates to identify researchers who may be interested to learn about options for curating and preserving their data. By aligning outreach efforts with an individual researcher's project lifecycle, we may be more successful at intercepting data before it is lost.

2 The long tail of non-federal grants is long and diverse



UMD researchers submitted proposals to more than 567 distinct non-federal funding sources in FY12.

While federal sources account for a substantial portion of research funding, there is a long tail of non-federal and non-government sources that may or may not impose data management or sharing requirements on researchers. In the absence of requirements, data and documentation are potentially at higher risk of being deleted, damaged, or left to languish on old media.

Implications:

A large number of research projects may not have to comply with data management requirements or submit data management plans, neutralizing a basic engagement strategy for librarians. Similarly, the intellectual property issues associated with corporate sponsorship may frustrate engagement efforts that focus on public data sharing. To build relationships with researchers in these situations, the authors intend to position data management services as activities that support research efficiency, innovation, and impact, rather than primarily compliance.

5 Some limitations of funding data

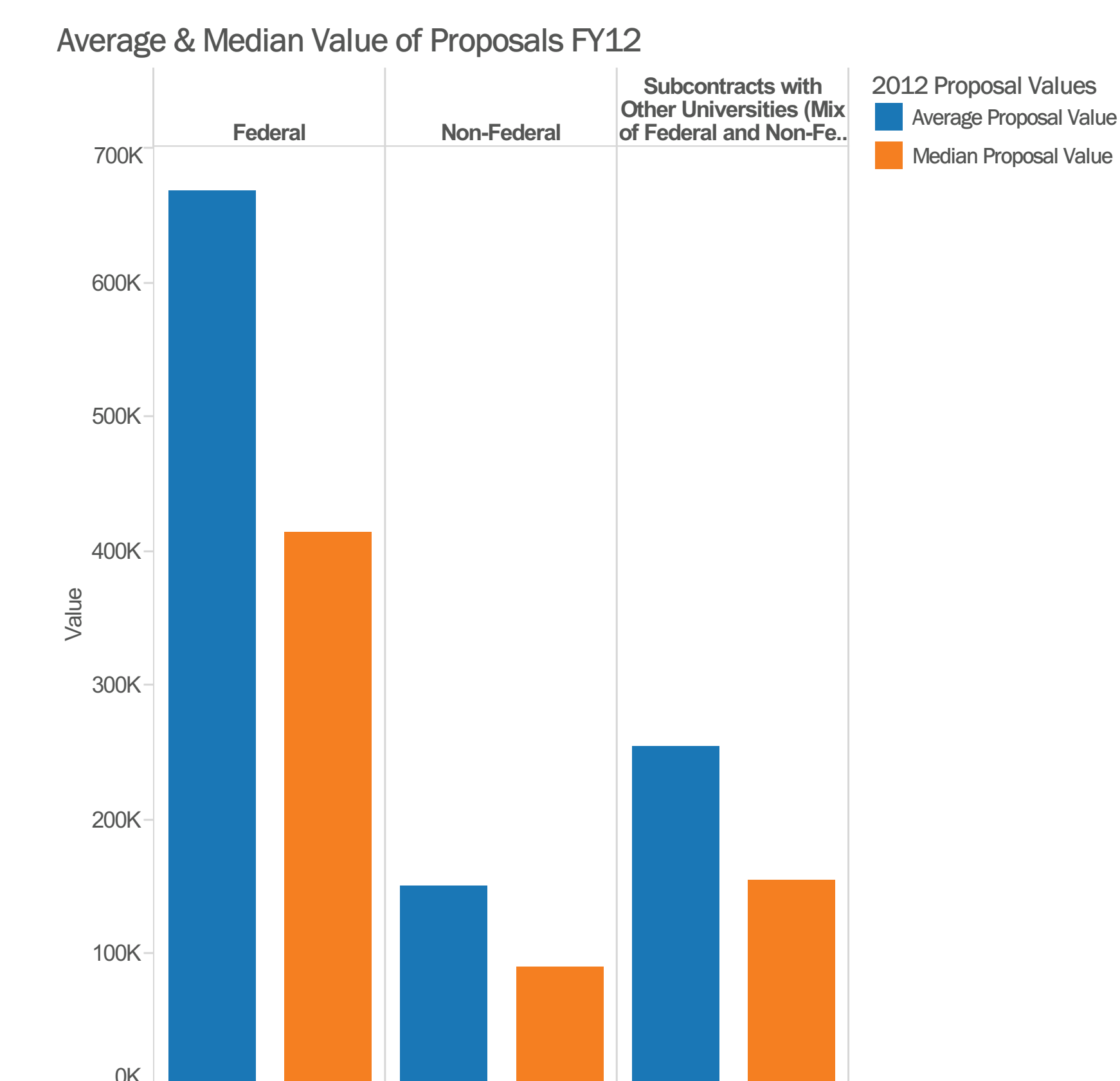
Integrating datasets

The authors sought to associate NSF Directorates and Divisions with departments, centers, and institutes at UMD in order to target assistance to particular academic units and individual researchers. We found that these data were contained in separate datasets that could not be automatically integrated. As a result, we proceeded to manually associate Directorates and Divisions with units. The results are being used to design outreach strategies, but the process was not efficient.

Funding data is an incomplete picture

Funding data can provide useful insights into the potential demand for data management services and the parameters of pilot projects, but they are not a perfect proxy for data production. Some funded research produces relatively little data, and researchers with little or no funding may generate large quantities of data.

3 Non-federal grants are less likely to support fee-based curation services



The average value of a UMD funding proposal to federal sources in FY12 was 341% greater than the average value of a proposal to non-federal sources. The median was 355% greater.*

Implications:

Recipients of non-federal awards (and low-value federal awards) may be reluctant to budget for curation and preservation. Institutions that plan to fund data curation from research awards will have to account for the many researchers who may not be able to justify allocating funds to fee-based curation services. In addition, we will have to accommodate researchers whose funding varies from project to project while the amounts of data generated may not vary significantly.

* Excluding subcontracts with other universities (mix of federal and non-federal).

Key Conclusions and Future Directions

Personalized data management consultations and embedded services will not scale to support every researcher.

- We may have to allocate resources on a selective basis that reflects the research priorities of our institution. Funding data can aid in this process.

- The subject-liaison system may not be the best model for research data services. Alternatives may come from outside the traditional library organizational model, such as the cross-disciplinary synthesis centers sponsored by the NSF's Biological Sciences Directorate,⁴ digital humanities centers, or data curation institutes.

An outreach and engagement strategy positioned around data management requirements and DMP compliance will not be relevant to all researchers.

- We need to re-position data management support services from compliance to research efficiency, innovation, and impact.

Demand for services from researchers who have no external funding, or funding from unusual sources, remains underexplored.

- Additional research is necessary to develop outreach and engagement strategies. Funding data can play a role in identifying potential participants.