

Differences among Faculty Ranks in Views on Research Data Management

by Katherine G. Akers¹ and Jennifer Doty²

RDM@
EMORY

Abstract

Academic libraries are increasingly engaging in data curation by providing infrastructure and services to support the management of research data on their campus. Efforts to develop these resources can benefit from a greater understanding of the social factors that affect how researchers manage their data during and after their research projects. In particular, the age or amount of experience of researchers is often thought to be an important factor influencing their viewpoints on research data sharing and preservation. In this study, we categorized faculty members who responded to our campus-wide survey on research data management into four ranks—professor, associate professor, assistant professor, and non-tenure track—and analyzed differences in their patterns of survey responses. We found statistically significant differences among faculty ranks in familiarity with funding agency requirements for data management plans, reasons that might prevent data sharing, and interest in potential research data services. These findings reveal key distinctions among different ranks of faculty members in their outlook toward research data management, which can help guide academic librarians and data curation professionals to develop research data services that are tailored to the unique needs of specific populations of researchers.

Keywords: Data curation, research data management, data sharing, researchers, faculty rank, library

Introduction

Academic libraries are increasingly providing support for the management and dissemination of research data by offering infrastructure (e.g., insitutional

repositories), services (e.g., consultation on data management plans), and education (e.g., best practices in data management) to campus researchers (ACRL Research Planning and Review Committee, 2012; Fearon, et al., 2013; Heidorn, 2011; Monastersky, 2013). To build research data management support systems that are both effective and desirable by researchers, academic librarians and other information professionals have conducted surveys and interviews with researchers to further understand how they manage data throughout the research lifecycle and their opinions on issues such as data sharing (e.g., Bardyn, 2012; Jahnke & Asher, 2012; Scaramozzino et al., 2012; Wells Parham et al. 2010; Westra, 2010; Witt et al., 2009). These investigations indicate that researchers exhibit a myriad of approaches to managing their data depending on their discipline, research topic and methodology, source of funding, data privacy concerns, and collaborative networks.

The age or amount of experience of researchers is another factor that may influence data management actions and attitudes. Within conversations among information professionals, two assumptions are often made: (1) younger researchers have been raised in a culture of greater openness of information and therefore are more willing to share their data, and (2) researchers nearing retirement are concerned about their research legacy and therefore are more eager to preserve their data. These assumptions, however, are primarily based on anecdotal evidence or small numbers of researcher interviews (e.g., Office of Policy and Analysis, 2011). Moreover, the few formal investigations on this topic have yielded contradictory results. For instance, Kuipers & van der Hoeven (2009)

found that less experienced researchers (< 10 years of experience) were more willing to deposit their data into a disciplinary repository than more experienced researchers (> 20 years of experience). Likewise, Piwowar (2011) found evidence suggesting that younger researchers are more likely to share their data than older researchers. By contrast, Tenopir et al. (2011) found that younger scientists (< 50 years of age) were less likely to make their data available to others without restrictions than older scientists (> 50 years of age), and Andreoli-Versbach & Mueller-Langer (2013) found that junior-ranking economy professors were less likely to share their data than full economy professors. Therefore, the nature of the relationship between researcher age/experience and tendency to preserve or share their research data remains in question.

To further explore how the age or amount of experience of researchers is related to their views on managing data throughout the research lifecycle, we took faculty rank into account when analyzing the results of our campus-wide survey of researchers' practices and perspectives on research data management. Specifically, we categorized faculty member respondents into four different ranks—professor, associate professor, assistant professor, and non-tenure track—and searched for differences in their patterns of survey responses.

Methods

In the fall of 2012, Emory University Libraries, in cooperation with the Office of Institutional Research, Planning, and Effectiveness, administered an online, 13-question survey on research data management practices and perspectives using Qualtrics software. A link to the survey was sent via email to all Emory University employees with faculty status according to Human Resource records (N = 5,590). The survey was open for 4 weeks, and three email reminders were sent at 1-week intervals. The survey was initiated by 456 faculty members (~8% response rate).

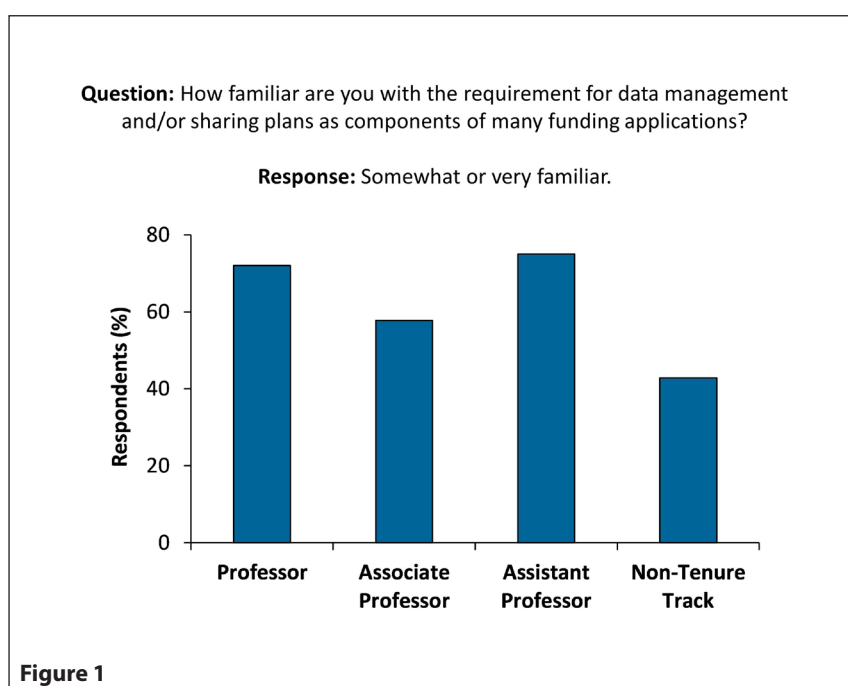
Our analysis focused on respondents who answered 'yes' to an initial question of whether they conducted research that generated some type of data (e.g., spreadsheets, text, images, videos, audio files, instrument files, photographs, physical samples/specimens, etc.; n = 330). Due to difficulties in equating rank among tenure, clinical, and research tracks, faculty members with clinical or research track designations were not included in the analysis. The remaining faculty members (n = 210) were divided into four groups based on Human Resource records: professor (professor, professor emeritus, or dean), associate professor, assistant professor, or non-tenure track (instructor, lecturer, visiting scholar, or adjunct professor). These faculty members were predominately based in the Emory College of Arts and Sciences (50%), with others from the School of Medicine (25%), Rollins School of Public Health (10%), Goizueta Business School (4%), Oxford College (4%), Candler School of Theology (4%), Nell Hodgson Woodruff School of Nursing (3%), and School of Law (1%).

Differences in survey responses among the four ranks of faculty members were evaluated using chi-square (χ^2) tests. Statistical significance was set at $p < 0.05$. Data are shown only for survey responses for which there were statistically significant differences among ranks. Complete survey results, including differences among arts & humanities, social science, basic science, and medical science domains, were previously reported (Akers & Doty, in press).

Results

Data Management Planning

We found no significant variations among different ranks of faculty members in the amount of research data they were storing or their methods for data storage and back-up (e.g., computer hard drive, external hard drive, instrument hard drive, university server, internet-based storage, lab notebooks, discs/tapes).



However, we did find variations among faculty ranks in their familiarity with federal funding agency requirements (e.g., National Science Foundation (NSF), National Institutes of Health (NIH), National Endowment for the Humanities (NEH)) for data management or data sharing plans as components of grant applications ($\chi^2 (3, n = 210) = 13.5, p = 0.004$; **Figure 1**). The majority of full and assistant professors stated that they were either somewhat or very familiar with data management plans, and over half of associate professors also expressed familiarity with these requirements. By contrast, most non-tenure track faculty members were not familiar with data management plan requirements, which may reflect a greater focus on teaching and less reliance on research grants.

Data Sharing

Faculty rank did not predict faculty members' willingness to share their research data with other people (e.g., researchers working on project, researchers outside of project, funders, instructors, general public) or their method of sharing research data (e.g., email upon request, supplementary material linked to journal article, data repository, university or personal website).

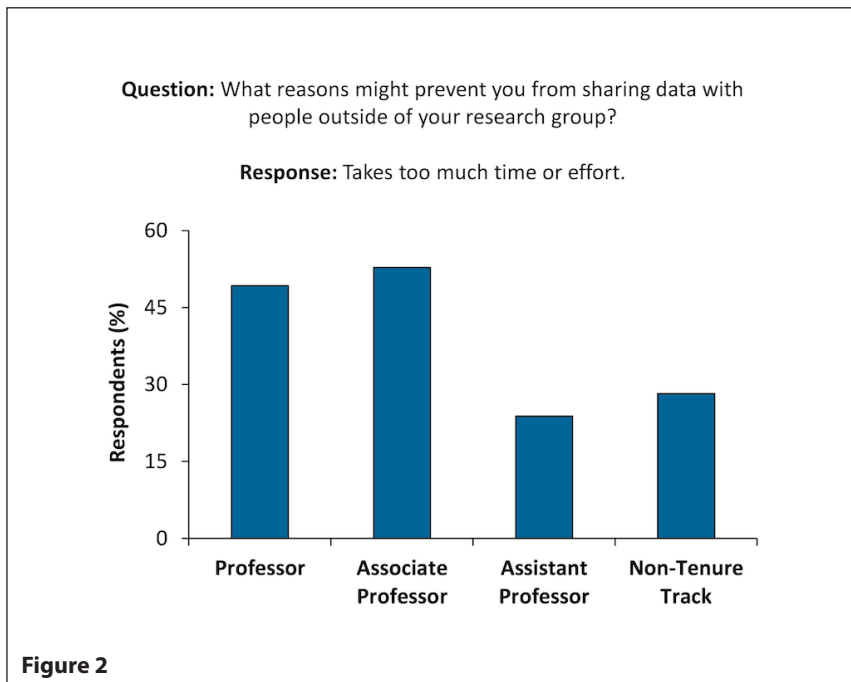


Figure 2

However, different ranks of faculty members expressed different opinions on why they might not share their data. Specifically, full and associate professors were more likely than assistant professors and non-tenure track faculty members to state that it takes too much time or effort to share their research data ($\chi^2(3, n = 199) = 10.1, p = 0.018$; **Figure 2**). This finding may reflect that senior-ranking faculty members may simply feel they are too busy to organize, document, and compile their data into shareable data packages that can be understood and used by others. Alternately, junior-ranking faculty members may feel that sharing their data with others is an expected part of the research process and thus may not perceive preparation for data sharing as an imposition on their time.

There were no differences among faculty ranks in other reasons that might prevent data sharing, including having data that contain private or patentable information, having data that require restricted access, fear of not getting credit for their data, fear of possible misinterpretation or misuse of their data, or belief that their data are of little use to others. Different ranks of faculty were also equally likely to deposit their data in data repositories or express familiarity with data documentation and metadata.

Interest in Data Services

In the final survey question, we offered a list of ten potential research data services and asked faculty to select which services they would use if available. The service garnering the most interest was faculty workshops on general data management. This service was desired by non-tenure track faculty members more than by assistant, associate, or full professors ($\chi^2(3, n = 191) = 11.6, p = 0.009$; **Figure 3**).

There were no rank-related differences in interest for the other potential services, including assistance preparing data

management plans, consultation on data confidentiality and/or legal issues, personalized consultation on research data management for specific researchers or research groups, an institutional repository for research data, assistance with data documentation or metadata creation, research data management workshops for trainees (i.e., graduate students or postdocs), digitization of physical research materials, assistance identifying appropriate disciplinary data repositories, or methods for data citation.

Discussion

It is often assumed that younger researchers are more supportive of open data and therefore more likely to share their research data with others via websites or data repositories/archives (e.g., Johnson, 2008; Lin, 2013; Boulton, 2013). However, empirical studies have not consistently provided support for this assumption. Although evidence from Kuipers & van der Hoeven (2009) and Piwowar

(2011) suggests that younger researchers are indeed more willing to share or archive their data than older researchers, our survey failed to find differences among ranks of faculty members in their willingness to share research data or their preferred method of data sharing. Moreover, Tenopir et al. (2011) and Andreoli-Versbach & Mueller-Langer (2013) found that younger researchers were less likely to share their data than older researchers. Therefore, younger age or less research experience may not always be a predictor of data sharing.

Rather than a simple correlation between researcher age and willingness to share data, findings by Tenopir et al. (2011) suggest that the situation is more complex. Their survey revealed that younger scientists were less likely than older scientists to place their data in a central repository without restrictions. However, younger scientists were slightly more likely than older scientists to make their data available if they could place conditions on data

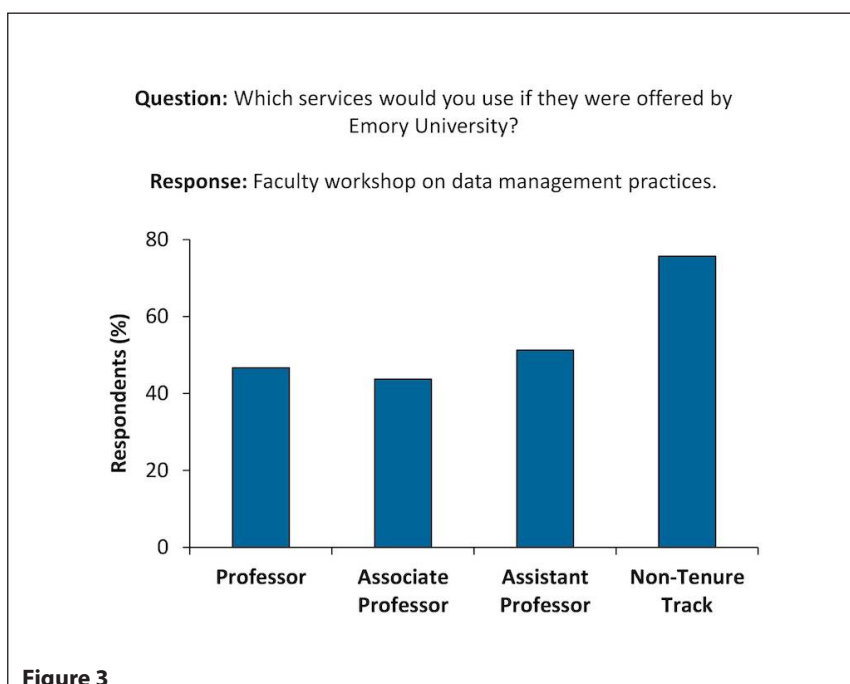


Figure 3

re-use, such as requiring legal permission for re-use of their data or the receipt of a complete list of products that make use of their data. Andreoli-Versbach & Mueller-Langer (2013) speculate that young researchers might be hesitant to openly share their data because this could enable other researchers to use their data before they can be fully exploited for additional publications. In other words, young faculty members who have not yet secured tenure may act more competitively than their tenured counterparts, choosing to withhold their research data or impose restrictions on data re-use. As we found no rank-related differences in faculty members' tendency to state that they might not share their data due to fear of not getting credit, our results do not directly support this possibility. Nevertheless, junior-ranking faculty members may be the ideal target population for outreach on ways of turning datasets into citeable outputs of scholarly research to increase personal research impact, including assigning digital object identifiers (DOIs) to datasets, depositing data into disciplinary or institutional repositories, or publishing data papers. Younger researchers might also be particularly receptive to evidence indicating that openly sharing research data increases the citation rate of associated journal articles (Bueno de Mesquita et al., 2003; Dorch, 2012; Henneken & Accomazzi, 2011; Ioannidis et al., 2009; Piwowar et al., 2007; Piwowar & Vision, 2013; Sears, 2011).

Our survey did not contain questions about data re-use, but previous studies indicate that the age of researchers may be an important indicator of their likelihood to re-use or re-purpose other people's data. Kuipers & van der Hoeven (2009) found that less experienced researchers are more eager to re-use data from other disciplines than more experienced researchers. Similarly, Tenopir et al., (2011) found that younger scientists are more likely to consider lack of access to data as a barrier to scientific progress that has restricted their ability to answer research questions. These findings underscore our suggestion that junior-ranking faculty members, in particular, could benefit from learning about ways to publically disseminate and thereby open their research data for re-use. Also, younger researchers may be more interested in receiving assistance with discovering and accessing pre-existing datasets.

Although we observed no differences among faculty ranks in willingness to share data, we found that senior-ranking faculty members were more likely to state that they might not share their research data due to the amount of time and effort involved. Indeed, Tenopir et al. (2011) found that insufficient time was the top reason that scientists did not make their data available to others, and others have also recognized this potential barrier to data sharing (Cragin et al., 2010; Peters & Riley Dryden, 2011; Williams, 2013). However, Vickers (2006) makes the case that a fundamental responsibility of data producers is to create clean, accurate, and well-annotated datasets, after which it should take little time to delete extraneous variables, remove personal identifiers, convert files into accessible formats, and briefly describe the dataset contents. Therefore, objections to data sharing on the premise that it takes too much time suggests that researchers do not always develop clean and well-annotated datasets (Savage & Vickers, 2009). We speculate that this is because researchers often have no clear incentive to do so; instead, they may be motivated only to develop datasets that are sufficient to support their own analyses for a particular project and not to invest time to make their datasets understandable to other researchers or to themselves at future dates. Therefore, our results suggest that senior-ranking faculty members, in particular, may benefit from increased university investment in the management and

dissemination of research data, including the creation of electronic research data management systems that could automatically organize data, generate metadata and documentation files, and push data packages into open repositories after project completion.

Finally, we found that non-tenure track faculty members were least familiar with funding agency requirements for data management plans but most interested in taking advantage of faculty workshops on general data management practices. At our university, faculty members are hired onto one of three different tracks: tenure track, research track, or clinical track. Due to difficulties in equating rank status across tracks, however, we removed research and clinical track faculty respondents from our analysis, meaning that the professional focus of the remaining non-tenure track faculty members was more likely to be teaching than other types of scholarly activities such as research. Therefore, their unfamiliarity with data management plans may be a result of their lack of dependence on external funding for research projects. Nevertheless, these faculty members also answered 'yes' to a question of whether they performed research that generated some type of data, indicating that they were indeed engaged in research to at least some extent. As such, our finding that non-tenure track faculty members expressed the highest desire for learning about "best practices" in research data management is very interesting and suggests that this faculty contingent, which is continuing to increase in size across academia (Curtis & Thornton, 2013), may be an overlooked population of researchers that might welcome greater outreach from academic librarians and other data curation professionals.

Acknowledgments

We thank Vincent Carter for his critical role in creating and administering the survey. We also thank members of the Emory University Libraries Research Data Management Team for their feedback on survey design and analysis

References

- ACRL Research Planning and Review Committee. (2012). 2012 top ten trends in academic libraries: A review of the trends and issues affecting academic libraries in higher education. *College & Research Libraries News*, 73, 311-320. Retrieved from <http://crln.acrl.org/content/73/6/311.full>
- 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*.
- Andreoli Versbach, P. & Mueller-Langer, F. (2013). Open access to data: An ideal professed but not practiced. *RatSWD Working Paper Series No. 215*; Max Planck Institute for Intellectual Property & Competition Law Research Paper No. 13-07. Retrieved from the Social Science Research Network: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2224146
- Bardyn, T.P., Resnick, T., & Camina, S.K. (2012). Translational researchers' perceptions of data management practices and data curation needs: Findings from a focus group in an academic health sciences library. *Journal of Web Librarianship*, 6, 274-287. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/19322909.2012.730375?journalCode=wjwl20#.UYPaHLXCaSo>
- Boulton, G. (2013). Open data: why and how it matters to the future. *Knowledge Exchange Berlin*, April 2013. Retrieved from <http://webcache.googleusercontent.com/search?q=cache:vHC50Qd1u8CJ:www.knowledge-exchange.info/>

- Admin/Public/DWSDownload.aspx%3FFile%3D%252FFiles%252FFile%252Fdownloads%252FPrimary%2BResearch%2BData%252FMaking%2BData%2BCount%2Bworkshop%252FBoulton_Presentation%2B-%2BBerlin.pdf+&cd=2&hl=en&ct=clnk&gl=us
- Bueno de Mesquita, B., Gleditsch, N.P., James, P., King, G., Metelits, C., Ray, J.L., Russett, B., Strand, H., & Valeriano, B. (2003). Symposium on replication in international studies research. *International Studies Perspective*, 4, 72-107. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/1528-3577.04105/full#>
- Curtis, J.W. & Thornton, S. (2013). Here's the news: The annual report on the economic status of the profession, 2013-2013. American Association of University Professors. Retrieved from <http://www.aaup.org/report/heres-news-annual-report-economic-status-profession-2012-13>
- Cragin, M.H., Palmer, C.L., Carlson, J.R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A*, 368, 4023-4038. Retrieved from <http://rsta.royalsocietypublishing.org/content/368/1926/4023>
- Dorch, B. (2012). On the citation advantage of linking to data. *hprints & Humanities*, hprints-00714715, version 2. Retrieved from <http://hprints.org/hprints-00714715/>
- Fearon, D., Gunia, B., Sherry, L., Pralle, B.E., & Sallans, A.L. (2013). SPEC Kit 334: Research data management services. Association of Research Libraries. Retrieved from <http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/>
- Heidorn, P.B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51, 662-672; Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01930826.2011.601269#.UYvnL7WG2Hs>
- Henneken, E.A. & Accomazzi, A. (2011). Linking to data – effect on citation rates in astronomy. arXiv:1111.3618. Retrieved from <http://arxiv.org/abs/1111.3618>
- Ioannidis, J.P., Allison, D.B., Ball, C.A., Coulibaly, I., Cui, X., Cuthane, A.C., Falchi, M., Furlanello, C., Game, L., Jurman, G., Mangion, J., Mehta, T., Nitzberg, M., Page, G.P., Petretto, E., & van Noort, V. (2009). Repeatability of published microarray gene expression analyses. *Nature Genetics*, 41, 149-155. Retrieved from <http://www.nature.com/ng/journal/v41/n2/full/ng.295.html>
- Jahnke, L.M. & Asher, A. (2012). The problem of data: data management and curation practices among university researchers. Council on Library and Information Resources. Retrieved from: <http://www.clir.org/pubs/reports/pub154/problem-of-data>
- Johnson, C.Y. (2008) Out in the open: some scientists sharing results. *Boston Globe*, August 21, 2008. Retrieved from http://www.boston.com/news/local/articles/2008/08/21/out_in_the_open_some_scientists_sharing_results/
- Kuipers, T. & van der Hoeven, J. (2009). INSIGHT into issues of permanent access to the records of science in Europe. PARSE. Insight. Retrieved from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Lin, T. (2013) Imagining data without division. *Quanta Magazine*, September 30, 2013. Retrieved from: <https://www.simonsfoundation.org/quanta/20130930-imagining-data-without-division/>
- Monastersky, R. (2013). Publishing frontiers: The library reboot. *Nature*, 495, 430-432. Retrieved from <http://www.nature.com/news/publishing-frontiers-the-library-reboot-1.12664>
- Office of Policy and Analysis. (2011). Sharing Smithsonian digital scientific research data from biology. Smithsonian Institution. Retrieved from <http://www.si.edu/content/opanda/docs/Rpts2011/11.03.DataSharing.Final.pdf>
- O'Reilly, K., Johnson, J., Sanborn, G. (2012). Improving university research value: A case study. *SAGE Open*, 2, 1-13. Retrieved from <http://sgo.sagepub.com/content/2/3/2158244012452576.full.pdf+html>
- Peters, C. & Riley Dryden, A. (2011). Assessing the academic library's role in campus-wide research data management: a first step at the University of Houston. *Science & Technology Libraries*, 30, 387-403. Retrieved from <http://www.tandfonline.com/doi/pdf/10.1080/019462X.2011.626340>.
- Piwovar, H.A., Day, R.S., & Fridsma, D.B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, 2, e308. Retrieved from <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0000308>
- Piwovar, H. & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ PrePrints*, 1, e1v1. Retrieved from <http://dx.doi.org/10.7287/peerj.preprints.1v1>
- Piwovar, H.A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE*, 6, e18657. Retrieved from <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657>
- Scaramozzino, J.M., Ramirez, M.L., & McGaughey, K.J. (2012). A study of faculty data curation behaviors and attitudes at a teaching-centered university. *College & Research Libraries*, 73, 349-365. Retrieved from <http://crl.acrl.org/content/73/4/349.full.pdf+html>
- Savage, C.J. & Vickers, A.J. (2009). Empirical study of data sharing by authors publishing in PLoS Journals. *PLoS ONE*, 4, e7078. Retrieved from <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0007078>.
- Sears, J.R. (2011). Data sharing effect on article citation rate in paleoceanography. *American Geophysical Union*, Abstract# IN53B-1628. Retrieved from <http://adsabs.harvard.edu/abs/2011AGUFMIN53B1628S>.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE*, 6, E21101; Retrieved from <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0021101>
- Vickers, A.J. (2006). Whose data set is it anyway? Sharing raw data from randomized trials. *Trials*, 7, 15. Retrieved from <http://www.trialsjournal.com/content/7/1/15>.
- Wells Parham, S., Bodnar, J., & Fuchs, S. (2010). Supporting tomorrow's research: Assessing faculty data curation needs at Georgia Tech. *College & Research Libraries News*, 73, 10-13. Retrieved from <http://crln.acrl.org/content/73/1/10.full>
- Westra, B. (2010). Data services for the sciences: A needs assessment. *Ariadne*, 64. Retrieved from <http://www.ariadne.ac.uk/issue64/westra>
- Williams, S. (2013). Data sharing interviews with crop sciences faculty: why they share data and how the library can help. *Issues in Science and Technology Librarianship*, Spring 2013. Retrieved from <http://www.istl.org/13-spring/refereed2.html#8>
- Witt, M., Carlson, J., Brandt, D.S., & Cragin, M.H. (2009). Constructing data curation profiles. *International Journal of Digital Curation*, 4, 93-101. Retrieved from <http://www.ijdc.net/index.php/ijdc/article/view/137>

Notes

1. Katherine G. Akers was the eScience Librarian and a Council on Library and Information Resources (CLIR) Postdoctoral Fellow at Emory University Libraries. She is now the eScience Librarian and a CLIR Postdoctoral Fellow at the University of Michigan Libraries. She can be reached by email: kgakers@umich.edu.
2. Jennifer Doty is the Data Management Specialist at Emory University Libraries and can be reached by email: jennifer.doty@emory.edu.