



12-27-2016

Understanding scientific data sharing outside of the academy

Danielle Elaine Pollock

University of Tennessee - Knoxville, dpolloc2@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_dataone

 Part of the [Information Literacy Commons](#), [Other Social and Behavioral Sciences Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Pollock, D. (2016). Understanding Scientific Data Sharing Outside of the Academy. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-5. doi:10.1002/pr2.2016.14505301144

This Conference Proceeding is brought to you for free and open access by the Communication and Information at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in DataONE Sociocultural and Usability & Assessment Working Groups by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Published as Pollock, D. (2016). Understanding scientific data sharing outside of the academy. *Proceedings of the 79th ASIS&T Annual Meeting, Vol. 53: Creating Knowledge, Enhancing Lives through Information & Technology*, 53(1). doi:10.1002/pr2.2016.14505301144

Understanding Scientific Data Sharing Outside of the Academy

Danielle Pollock

University of Tennessee, Knoxville, USA
dpolloc2@vols.utk.edu

ABSTRACT

Sharing and reuse of scientific data, which can enhance the transparency and reproducibility of research and lead to the creation of new knowledge from existing data, is both a growing scholarly communication practice and an expanding area of interest in information science. However, much of the literature to date has focused on the data practices of scientists working in academic environments, with less research done on understanding the practices of scientists working in other types of environments, such as government or industry. This poster presents the results of a study in which data from a worldwide survey of scientists were analyzed to determine if differences in data practices, perceptions, and access to resources for data sharing existed between scientists who reported their primary work sector as academic and those who reported a non-academic primary work sector. Researchers' perceptions of data sharing and reuse were generally positive and did not differ significantly by work sector. However, differences were found in actual reported data sharing practices, even when controlling for researchers' age, geographic location, and subject discipline. Researchers outside of academia had lesser odds of reporting sharing all their data. Differences were also found in reported barriers to data sharing, as well as in reported access to and use of data sharing resources, suggesting that data sharing challenges faced by scientists working outside of academia may differ from those faced by their academic peers. Implications for the adoption of data sharing practices and technologies, as well as for

knowledge sharing and creation across work sectors, are discussed, and suggestions are offered for further research.

Keywords

Data sharing, data reuse, scientific data.

INTRODUCTION

The sharing and reuse of research data—increasingly mandated by governments, publishers, and funding agencies—can enhance the transparency and reproducibility of research and lead to the creation of new knowledge from existing data; however, while data sharing is increasing, the practice has not yet been widely diffused throughout the entire community of practicing research scientists (Kim & Zhang, 2015; Schmidt, Gemeinholzer, & Treloar, 2015; Tenopir et al., 2015). Previous research has found that data sharing practices and norms can vary by discipline, geographic location, and age of researcher (Borgman, 2015; Poole, 2015; Tenopir et al. 2011; Tenopir et al., 2015).

The adoption of innovative practices such as data sharing may also be influenced by factors related to the specific social systems within which a researcher works. These include community norms, the existence of policies related to data sharing, compatibility with existing work and communication practices, and access to needed resources (Kim & Zhang, 2015; Fecher, Friesike, & Hebing, 2015; Rogers, 2003). Just as social factors may vary by discipline and academic region, they may also vary by specific work environment or work sector. Yet the majority of studies of scientists' data sharing behaviors have focused on scientists working in academic environments, while fewer studies have examined scientists working in other environments, such as government and industry (Douglass, Allard, Tenopir, Wu, & Frame, 2014; Poole, 2015; Stvilia et al., 2014).

Better understanding the data sharing practices and perceptions of scientists working outside of academia is not

{This is the space reserved for copyright notices.}

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.

[Author Retains Copyright. Insert personal or institutional copyright notice here.]

only essential for developing systems and services to better meet the needs of these scientists, but may also help enable essential sharing of data across work sectors for research into complex problems in areas such the environmental and health sciences (Collins & Varmus, 2015; Downs, Duerr, Hills, & Ramapriyan, 2015; Thessan, McGinnis, & North, 2015).

In this study, secondary analysis was performed using data from a worldwide survey of scientists' data sharing practices and perceptions (Tenopir et al., 2015) to discover if work sector differences exist in scientists' data sharing behaviors, perceptions of data sharing and reuse, and access to resources for data sharing.

METHOD

Data for this study come from a survey of scientists' data sharing practices and perceptions conducted by the DataONE Usability and Assessment working group from October 17, 2013 to March 19, 2014 (Tenopir et al., 2015).

The survey received a total of 1,015 valid responses. Respondents were free to skip any question. Analysis was done on using the 1,004 responses from researchers who specified their primary work sector. A dummy variable was constructed based on whether the respondent indicated their primary work sector as academic or as non-academic (government, industry, non-profit, or other).

Dependent variables in this study were respondents' answers to questions related to:

- Current data sharing practices. Example: *I share my data with others.*
- Perceptions and attitudes towards sharing and reusing data. Example: *Lack of access to data generated by other researchers or institutions is a major impediment to progress in science.*
- Access to resources, including technologies and services, for data management and sharing. Example: *My organization or project provides the necessary tools and technical support for data management beyond the life of the project (long-term).*

Two models were constructed for use with each dependent variable:

Model 1 included only researchers' work sector

Model 2 included researchers' work sector and all demographic variables for which differences in data sharing practices were found to exist in prior studies. These control variables included age, geographic location, and discipline.

Depending on the nature of the survey question, ordered logit, logistic, or robust regression analysis techniques were used with both models to determine if a statistically significant relationship existed between the dependent

variables and researchers' reported work sector, and if this relationship remained statistically significant when controlling for other variables.

RESULTS

Data practices

Researchers outside of academia had a lesser odds ratio of reporting sharing all their data (see Table 1).

	Model 1 Bivariate	Model 2 Control Variables
Non-Academic Work Sector	.443** (.149)	.342* (.159)
Pseudo R-squared	0.004	0.047
n	831	804
LR Chi Square	8.89**	92.37***

Table 1. Ordered Logit Regression of Work Sector and Other Variables on "How Much of Your Data Do You Make Available to Others?"^{1,2}

The impact of specific barriers to data sharing was found to vary by work sector. Among those who did not make all their data available to others, logistic regression showed that for researchers indicating a non-academic work sector, the odds of reporting failing to make data available due to a need to publish first decreased by 50% in model 1 ($p < .001$) and 49% in model 2 ($p < .001$). For these researchers, tenure and promotion practices tied to publication may not be as large a factor in their communication decisions as for researchers in academic environments. Somewhat more surprisingly, working in a non-academic environment was associated with a 46% decrease in odds of reporting failing to share data due to lack of standards in model 1 ($p < .05$) and a 45% decrease in odds when age, geographic location, and subject discipline were controlled for in model 2 ($p < .05$).

Researchers in non-academic work environments may experience more barriers related to not having the ability to make research data public. The odds of these researchers reporting not having the rights to make their data public increased by 61% over academic researchers in model 1, and 72% when control variables were introduced in model 2. For non-academic researchers, when controlling for other variables, model 2 showed a 90% increase in odds of reporting that their data should not be available.

¹ Table abbreviated for space. Coefficient and standard error for control variables in Model 2 not shown. *** $p = .001$ ** $p = .01$ * $p = .05$

² Coefficients reported as odds ratios. Standard error in parentheses.

Lack of rights to make the data resulting from their research available may be a factor in why robust regression showed non-academic researchers had lower levels of agreement with the statement *Others need my permission to access my data* and lesser odds of requiring legal permission as a condition for data reuse (see Tables 2 and 3), as these permissions may not be within the researchers' ability to grant.

	Model 1	Model 2
Non-Academic Work Sector	-.444*** (.118)	-.369** (.124)
Constant	3.790	3.23
n	725	701
F	14.14***	3.23***

Table 2. Robust Regression of Work Sector and Other Variables on Level of Agreement with “Others Need My Permission to Access My Data”^{1,3}

	Model 1	Model 2
Non-Academic Work Sector	.544** (.114)	.582* (.138)
Pseudo R-squared	0.012	0.124
n	545	526
LR Chi Square	8.80**	89.09***

Table 3. Logistic Regression of Work Sector and Other Variables on Agreement with Obtaining Legal Permission as a Condition for Data Reuse^{1,2}

Perceptions and attitudes toward data sharing and reuse

Overall, scientists' perceptions and attitudes toward data sharing and reuse in Tenopir et al. (2015) were generally positive, and were not found to vary significantly by work sector in the current study.

The exception was that researchers outside of academia reported higher levels of agreement with the perception that *Lack of access to data generated by other researchers or institutions has restricted my ability to answer scientific questions* (see Table 4).

	Model 1	Model 2
--	---------	---------

³ Coefficients represent change in level of agreement on five-point scale from 1 = disagree strongly to 5 = agree strongly. Standard error in parentheses.

Non-Academic Work Sector	.294* (.124)	.322* (.127)
Constant	3.324	3.702
n	624	603
F	5.79*	2.87***

Table 4. Robust Regression of Work Sector and Other Variables on Level of Agreement with “Lack of Access to Data Generated by Other Researchers or Institutions Has Restricted My Ability to Answer Scientific Questions”^{1,3}

Resources for data sharing

The availability of many resources for data sharing and reuse, including formal processes for managing and storing data, training on data management practices, and assistance with data management activities was not found to differ significantly by work sector. Nor were levels of satisfaction with processes and tools for collecting, storing, searching and analyzing data, and for preparing metadata and data documentation.

Non-academic researchers did have higher levels of agreement that their organizations or projects provided funding to support data management both during and beyond the life of the project (see Table 5).

	“...during the life of the project (short-term).”	
	Model 1	Model 2
Non-Academic Work Sector	.452** (.153)	.410* (.163)
Constant	2.70	2.451
n	589	572
F	8.78**	1.84**
	“...beyond the life of the project (long-term).”	
	Model 1	Model 2
Non-Academic Work Sector	.386* (.154)	.356* (.147)
Constant	2.20	2.188
n	562	545
F	6.26*	1.20

Table 5. Robust Regression of Work Sector and Other Variables on Level of Agreement with “My Organization or Project Provides the Funds to Support Data Management...”^{1,3}

These researchers also differed from their academic peers in where they reported storing their data. Researchers in work sectors outside of academia had greater odds of reporting storing their data via their own institution's server or repository, and lesser odds of making use of discipline-specific repositories, of storing data on personal computers, or on paper in their own offices (Table 6).

“On my institution’s server”		
	Model 1	Model 2
Non-Academic Work Sector	2.087*** (.429)	2.192*** (.492)
Pseudo R-squared	0.012	0.058
n	715	691
LR Chi Square	13.87***	50.09***
“In my institution’s repository”		
	Model 1	Model 2
Non-Academic Work Sector	1.989*** (.392)	1.913** (.417)
Pseudo R-squared	0.017	0.067
n	575	548
LR Chi Square	12.06***	46.45**
“On my personal computer”		
	Model 1	Model 2
Non-Academic Work Sector	.339*** (.096)	.320*** (.104)
Pseudo R-squared	0.034	0.119
n	752	664
LR Chi Square	13.39***	42.32**
“On paper in my office”		
	Model 1	Model 2
Non-Academic Work Sector	.545** (.103)	.591* (.124)
Pseudo R-squared	0.012	0.076
n	636	610
LR Chi Square	10.18**	60.65**
“In a discipline-based repository (e.g. LTER or NEON)”		
	Model 1	Model 2
Non-Academic Work Sector	.533** (.128)	.496** (.130)
Pseudo R-squared	0.011	0.094
n	587	562
LR Chi Square	7.40**	62.57**

Table 6. Logistic Regression of Work Sector and Other Variables on Storing At Least Some Data...^{1,2}

CONCLUSION

The results of this research indicate that scientific data sharing practices are still yet unevenly diffused across work sectors, and that while attitudes toward data sharing and reuse do not differ significantly, the actual data sharing

practices of scientists working in government, industry, non-profit, and other environments may differ from those of their academic peers.

Specifically, scientists in non-academic work sectors had lesser odds of sharing all of their data and greater odds of reporting limitations related to lack of ability to make their own research data available to others. These restrictions mean that individual scientists’ attitudes towards data sharing may not ultimately impact their data sharing practices. Much in the way that differences in disciplinary cultures impact data sharing practices, differences in the cultures and policies of organizations may impact data sharing (Borgman, 2015; Douglass et al. 2013). Further study is needed to examine what role institutional policies, legal requirements, the nature of data, and other potential restrictions play in limiting these scientists’ ability to share data.

Resources available for data sharing also differ by work sector. Scientists outside of academia were more likely to agree their project or organization provided funds for data management activities and had greater odds of using an available institutional repository or server for data storage. While these findings suggest that these scientists are being given support for data management within their own organizations, it is unclear whether this support extends to sharing data with others, even in cases where data could be made available outside the institution.

In addition to restrictions on sharing their own data, scientists outside of academia may also have less access to others’ data when needed. Further research into data sharing across institutions and work sectors could help uncover the specific barriers to data access faced by these researchers and to what extent these barriers represent limitations to knowledge sharing and creation.

Limitations of this study include the fact that the size of the data set meant useful comparisons could not be made across types of non-academic work environments. Government, industry, non-profit and other organizations each have their own unique cultures, and may have different social factors that impact data sharing behaviors. Respondents were free to skip any question, which further limited the number of responses.

Additional mandates for data management and sharing, some of which have particular impact on government-funded scientific research, have been enacted since the survey data was collected (Tenopir et al., 2015). A follow-up study could examine how these mandates have changed data sharing practices and perceptions across work sectors, including examining the specific impact of these mandates on the data sharing practices of non-academic researchers and overall work sector differences in data sharing and access.

Additional focused research is needed to understand the data practices and uncover the specific data sharing

challenges faced by scientists working outside of academia. Research focused on scientists in government, industry, and other work environments will lead to greater insight into institutional cultures and practices surrounding data sharing in these environments, and help information professionals identify technical and cultural barriers to data sharing and access, and better design systems and services to meet the needs of these researchers.

ACKNOWLEDGMENTS

Data analyzed in this research was created by the Data Observation Network for Earth (DataONE) Usability & Assessment Working Group (Bill Michener, principal investigator, NSF award #0830944).

REFERENCES

- Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.
- Collins, F. S. & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, 372, 793-795. doi:10.1056/NEJMp1500523
- Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among full-time government employees. *Journal of the American Society for Information Science and Technology*, 65, 251-262. doi:10.1002/asi.22988
- Downs, R. R., Duerr, R., Hills, D., & Ramapriyan, H. K. (2015). Data stewardship in the earth sciences. *D-Lib Magazine*, 21(7). Retrieved from <http://www.dlib.org/dlib/july15/downs/07downs.html>
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing? *PLoS ONE*, 10(2). doi:10.1371/journal.pone.0118053
- Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, 37(3), 189-200. doi:10.1016/j.lisr.2015.04.006
- Poole, A.H. (2015). How has your science data grown?: Digital curation and the human factor: A critical literature review. *Archival Science*, 15, 101-139. doi:10.1007/s10502-014-9236-y
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Schmidt, B., Gemeinholzer, B., & Treloar, A. (2016). Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. *PLoS one*, 11(1). doi:10.1371/journal.pone.0146695
- Stvilia, B., Hinnant, C. C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., . . . Marty, P. F. (2015). Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *Journal of the Association for Information Science and Technology*, 66, 246-263. doi:10.1002/asi.23177
- 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(6). doi:10.1371/journal.pone.0021101
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., Dalton, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, 10(8). doi:10.1371/journal.pone.0134826
- Thessen, A. E., McGinnis, S., & North, E. W. (2016). Lessons learned while building the Deepwater Horizon Database: Toward improved data sharing in coastal science. *Computers & Geosciences*, 87, 84-90. doi:10.1016/j.cageo.2015.12.001