



THE UNIVERSITY of TENNESSEE 
School of Information Sciences

Data Help Desk:

*(Informing Science Data Help Desk Staffing through
Transaction Analysis)*

Bradley Wade Bishop, PhD
Rose Borden,
Graduate Research Assistant
January 31, 2019

Executive Summary
Informing Science Data Help Desk Staffing through Transaction Analysis

Purpose:

The purposes of this project were to directly observe and assess the question transactions at the Data Reference Desk at two conferences and inform future research data management trainings and staffing of information services.

The project’s objectives include: (1) create a data question typology of from those asked at the Data Help Desk; (2) assess the transactions using reference guidelines; and (3) inform future research data management trainings and staffing of information services.

As science data questions grow in number and complexity the optimal staffing of information services will be a key to the success of data access and use. Ultimately, better science will result in more informed scientists accessing and re-using data more fit for their purposes. For ESIP, the impact of a better understanding of the current Data Help Desk will improve and expand the current quality of assistance. In addition, more knowledge of scientists research data management needs may inform the curation and assessment efforts related to the Data Management Training Clearinghouse.

Description of key project steps and timeline:

Project Timeline	
Aug 2018	<ul style="list-style-type: none"> • Met with Data Help Desk organizers and volunteers to discuss project objectives • Hired graduate student researcher • Submitted IRB for approval of human subjects research • Attend Ecological Society of America Annual Meeting 2018 • Conducted direct observation of Data Reference Desk (Aug 6-9)
Sep-Nov 2018	<ul style="list-style-type: none"> • Reviewed questions asked and developed data question typology • Assessed user responses to exit survey • Discussed findings with Data Help Desk organizers and volunteers, Knowledge Motifs, and DataONE to inform future staffing of the service • Modified methodology for next direct observation of interviews using reference guidelines was problematic in a conference space
Dec 2018-Jan 2019	<ul style="list-style-type: none"> • Attended American Geophysical Union Fall Meeting 2018 • Conducted direct observation of Data Reference Desk (10-14) • Reviewed questions asked and modified the data question typology • Assessed user responses to exit survey • Determined how these questions could inform future research data management trainings needs of scientists • Created a model for staffing of information services of science data
Ongoing	<ul style="list-style-type: none"> • Dissemination of findings to data-related communities through webinars and conferences (e.g., ESIP Summer Meeting, Data Management Working Group calls, and so forth)

Current Model of Data Help Desk and Data Collection Process:

This project seamlessly integrated its methodology with the existing Data Help Desk staffing model. At ESA 2018 and AGU 2018, the Data Help Desk was set-up as a booth within exhibit hall space. Conference attendees could approach the desk or be encouraged by those staffing the Data Help Desk to ask a question. Although many were drawn more by chocolate, stickers, and other tchotchkes, this informal approach resulted in a considerable amount of traffic albeit passive.

Those staffing the Data Help Desk used technology at both conferences to scan badges and capture attendees' information for follow-up regardless of if they asked a question or not. At ESA, Kevin Love from iDigBio was meticulous to gather this information through an iPad and shared this later with the group; however that data is not included in this report. At AGU, a scanner was intermittently used to capture badges even for attendees that did not ask a question.

When attendees asked the staff a more substantive question than “can I take a sticker” and it was clear that their transaction concluded, each attendee was read the recruitment and informed consent language prior to completing a survey about their experience. The standard Institutional Review Board (IRB) informed consent language states there is no penalty for not participating in the project. The proposed method included direct observation for each transaction; however, informed consent in an informal conference setting would have been awkward and impacted the actual service and was jettisoned. Still, this researcher could generalize from past experiences when studying reference transactions and less systematically highlight the tactics used by Data Help Desk staffers in answering questions. The survey questions used were informed through the Data Help Desk staffers to evaluate this information service. The data from the survey is presented in this report.

Questions:

1. What was your question for the Data Help Desk?
2. What is your highest level of education? PhD; Master's; Bachelor's
3. What is your current job title?
4. In your job, what are your top three research data management needs?
5. If you have received training, what types of data research management training did you receive?
6. If you have not received training, what types of data research management training would you be most interested in?
7. What delivery formats for training do you prefer (e.g., half day workshop, webinar, graduate coursework)?
8. What was your level of satisfaction with your service at the Data Help Desk?

(Likert scale of satisfaction used)

Findings:

The most common questions asked at the Data Help Desk had to do with the various organizations represented by ESIP, as well as questions related to general data management, data sharing, and data storage (Table 1). Scientists also wanted to know about data analysis, finding and accessing data, and what software is available to help with data management.

Topic	ESA	AGU	Combined
Organizations	10	7	17
Data Management	9	0	9
Sharing data	6	4	10
Data storage	6	5	11
Data analysis	4	1	5
Finding data	3	5	8
Other	3	9	12
Accessing data	2	4	6
Software	0	3	3
Total	43	38	81
Others	General info; Can I help out by hosting a data workshop here?; Grasped diversity of initiatives	What are you doing here; If you, have search engine; What is data? What counts as data?; How to get involved	

Table 1: Question typology.

The most common education level at each conference was a PhD, followed by a Bachelor's and then Master's (Table 2). The most common job title was student, then research scientist and professor. There were a few postdocs at ESA, but none at AGU, or at least none who identified as such (Table 3).

The findings for top data management needs are a bit more complicated. The most common data management need for both ESA and AGU was data storage. Data accessibility and data sharing were also in the top 5 for both conferences. General data management techniques were higher at ESA, while scientists at AGU more commonly mentioned interoperability. Data analysis was also a concern of scientists at ESA. Needs with fewer responses at ESA were reproducibility, cleaning data, metadata, resources, quality control, visualization, and privacy. There were also a number of responses that were not easily categorized, listed under Other in Table 4. Needs with fewer responses at AGU were data organization, finding data, metadata, databases, collecting data, data processing, data availability, visualization, and restricted data. There are also a number of other responses that did not easily fit into a category and are listed under Other in Table 4.

About one third of the respondents had received no prior data management training (Table 5). Three were uncertain whether they had or not. Of those who had, the most common format was as a workshop (Table 6). Another common format mentioned was seminars at AGU or in grad school, as well as being self-taught. The only general topics of these trainings mentioned more than once each were data carpentry, R, and general data management.

When asked what format they would like for future trainings, the most common response at AGU was webinar, followed by workshop (Table 7). At ESA the most common responses were workshop and online. When asked about the preferred topics of future trainings, the most common responses at AGU were data archiving, database management, data organization, and documentation/nomenclature. The

most common responses from ESA were data management, data archiving, DataONE, data analysis and databases.

To provide a content analysis of these question and answer sessions, the Reference and User Services Association's (RUSA) *Behavioral Performance of Reference and Information Service Providers* guidelines will be used for coding the transactions (<http://www.ala.org/rusa/resources/guidelines/guidelinesbehavioral>). Other transaction metrics will be captured (e.g., time) to inform a future staffing model of information services for science data. At the conclusion of the service, users will be given a short web-based exit survey to complete. The exit survey will assess their experience and perceived needs of further research data management training and information services.

Overall 88% of respondents said they were highly satisfied with the help they received and another 8% were somewhat satisfied (Table 8). One respondent was neither satisfied nor dissatisfied and two respondents were somewhat dissatisfied.

Education Level	ESA	AGU	Combined
PhD	20	16	36
Bachelor's	11	15	26
Master's	9	6	15
Total	40	37	77

Table 2: What is your highest level of education?

Title	ESA	AGU	Combined
Student	12	15	27
Research scientist	11	11	22
Postdoc	6	0	6
Professor	6	4	10
Other	5	7	12
Total	40	37	77
Others	Florida DEQ	Head of R&D	
	Biological Science Technician	Manager	
	Program Coordinator	Public Affairs Specialist	
	President of First People's Council	Program Officer	
	Project coordinator for CWMA	Policy Fellow	
		Principal systems engineer	

		Manager of Departmental Computing	
--	--	---	--

Table 3: What is your current job title?

ESA					AGU				
Category	#1	#2	#3	Weighted	Category	#1	#2	#3	Weighted
Data storage	8	10	4	16.0	Data storage	11	2	0	12.3
How to do data management	7	5	4	11.7	Data accessibility	5	6	0	9.0
Data access	6	2	1	7.7	Interoperability (readability, standardization)	2	3	4	5.3
Data sharing	5	1	2	6.3	Sharing data	1	5	3	5.3
How to do data analysis	2	3	5	5.7	Data organization	0	1	1	1.0
Reproducibility	0	3	1	2.3	Finding data	3	0	0	3.0
Cleaning data	2	0	0	2.0	Metadata	1	2	1	2.7
Metadata	0	2	1	1.7	Databases	1	0	1	1.3
Resources	1	1	2	2.3	Collecting data	1	0	0	1.0
Quality control	1	0	0	1.0	Data processing	0	1	0	0.7
Visualization	0	0	1	0.3	Data availability	0	1	0	0.7
Privacy	0	0	1	0.3	Visualization	0	0	1	0.3
					Restricted data	0	0	1	0.3
Total	32	27	22		Total	25	21	12	
Others	4	4	5		Others	6	7	2	

Table 4: In your job, what are your top three research data management needs?

Answer	ESA	AGU	Combined
No	24	30	54
Yes	15	7	22
Not sure	2	1	3
Total	41	38	79

Table 5: Prior to coming to the Data Help Desk, have you received any research data management training?

ESA					AGU
Type of training	Count	Topic of training	Count		
Workshop	7	Data carpentry	4		Seminar through doc program, programming
Self-taught	1	R	2		Image data time series; spatial data
Grad seminar	1	Data lifecycle	1		Self taught, nothing formal
Grad course	1	Python	1		Research group on the job at JPL
		Modeling	1		Seminars and town halls at AGU
		iDigBio	1		Data management basics Science Base
		Reproducibility	1		Unidata committees
		Metadata	1		
		Data management	1		
		Statistics	1		
		GitHub	1		
		Provenance	1		
		Automating scripts	1		
		Software	1		

Table 6: If you have received training, what types of data research management training did you receive?

Format	ESA	AGU	Combined
Webinar	0	13	13
Workshop	5	6	11
Online	5	2	7
Face to face	0	2	2
Refresher Course	1	0	1
Book	1	0	1
Feedback from organizations	1	0	1
Written tutorial	0	1	1
Grad course	0	1	1
Total	13	25	38

Topic	ESA	AGU	Combined
Data archiving	4	5	9
Data management	6	2	8
Database management	0	5	5
DataONE	4	0	4
Unsure	0	4	4
Data analysis	3	0	3
Databases	3	0	3
Sharing data	1	2	3
Data organization	0	3	3
Documentation/nomenclature	0	3	3
Data processing	0	2	2
Background research	1	0	1
Coding	1	0	1
Data cleaning	1	0	1
Python	1	0	1
Version Control	1	0	1
Available services	0	1	1
Data access	0	1	1
Big data	0	1	1
Finding data	0	1	1
Science data communities	0	1	1
Total	26	31	57

Table 7: If you have not received training, what types of data research management training would you be most interested in?

Answer	ESA	AGU	Combined
Extremely satisfied	37	32	69
Somewhat satisfied	2	4	6
Neither satisfied nor dissatisfied	1	0	1
Somewhat dissatisfied	1	1	2
Extremely dissatisfied	0	0	0
Total	41	37	78

Table 8: What was your level of satisfaction with your service at the Data Help Desk?

Discussion

- (1) A data question typology will inform data products, tools, and services. A better understanding of the questions asked and how they are answered will improve future information services, but also anticipate users' needs and assist with usability design of other data product and tools.
- (2) An assessment of the question and answer sessions will review the current service and provide productive feedback on how to improve upon the existing success. The transaction analysis will provide organizers and volunteers a reflective opportunity to better understand the relational and content aspects of questions and answering. Successful reference encounters include not only aspects of accuracy of information and correct responses, but attitudinal aspects of interpersonal communication. Also, user feedback will give perceived needs of further research data management training and information services.
- (3) All the data collection and analyses will inform a staffing model for future information services of science data. This model will likely include the Data Help Desk conference service as well as a consortia of virtual reference network for assistance beyond conferences.

Findings will be disseminated to data-related communities through webinars and conferences (e.g., ESIP Summer Meeting, Data Management Working Group calls, and so forth). This will include a final project report to ESIP and likely several journal articles.

Recommendations for Next Steps

One area where this service could be improved upon would be to have some handouts describing exactly what data management is, as that is one topic that not all scientists have the same awareness of yet and was one of the most commonly asked questions by attendees who stopped by the booth at AGU.

Another suggestion is to have a service year-round where scientists and other data users can ask questions of data management experts, similar to "Ask a Librarian" chat and e-mail services common in academic libraries. Some attendees at conferences like this may not have specific data management questions in mind while at the conference or may need information ask their question that is back on their office computers. However, one of the main considerations is whether this would be a funded service with dedicated paid experts or whether it would rely on volunteers to sign up for shifts. An alternative option could be to have volunteers send their contact info and areas of expertise and to keep a database of experts, or at least experienced data users, that less experienced users could use to ask data management questions of outside of conference help desks.

References

American Library Association. Reference and Users Services Association. Guidelines for Behavioral Performance of Reference and Information Service Providers.

<http://www.ala.org/rusa/resources/guidelines/guidelinesbehavioral>