Aligning Data Support Services to Researchers’ Needs

Danielle Cooper and Dylan Ruediger
Ithaka S+R

March 28, 2022
Today’s Agenda

1. Overview of the data support landscape
2. What support needs do researchers have?
3. Aligning support services to researchers’ needs.
4. Discussion
What does the research data support landscape look like and how can it be optimized to meet the needs of researchers?
Overview of the current data support landscape
S+R’s research on data support services and needs

Recent and In-Progress Projects:
- Research Data Services (2020)
- Big Data Report (2021)
- What is a Research Core? (2021)
- NSF Data Communities (2022)
- Teaching with Data in the Social Sciences (2022)
- Data Sharing Practices in the Humanities (2022)

Foundational Work:
- National faculty survey (2000-present)
- Research and Teaching Support Service projects (2012-present)
- “Data Communities” (2019-present)
- Research enterprise/SRO (2020-present)
By the numbers

1585 interviews with researchers

107 from universities

346 collaborating librarians
Mixed-method research on data support services

**Big Data Infrastructure at the Crossroads (2021)**

Cohort-based project conducted in collaboration with 23 U.S. university libraries, resulting in over 200 semi-structured interviews with researchers from c. 100 departments.

**Data Services Inventory (2020)**

Systematic web-based inventory of 120 U.S. colleges and universities from across Carnegie Classifications yielding quantitative data about number and type of data services.
Key takeaways from the 2020 Inventory

Clustered at Research Institutions

R1 institutions exceed R2s and SLACs by more than double in the number of research data services offered. The average R1 offers 7.6 research data services, while the average R2 offers 2.6 and the average SLAC offers 1.8.

Libraries are essential

Libraries are the most active providers of research data services across institution type, especially for generalist consultations and training workshops. IT/HPC units are the next most active providers of these services.

Decentralized Offerings

A wide variety of research data services, most focused on specific disciplines or methodologies, can be found in academic departments, independent research centers and facilities, and professional schools. These services most commonly focus on statistics and bioinformatics.
What support needs do researchers have and how do they access support services?
Finding Areas

- Disciplinarity and Interdisciplinarity
- Managing Complex Data
- Managing Complex Workflows
- Sharing Knowledge
- Support and Training
Disciplinarity & Interdisciplinarity

Big data research is an interdisciplinary enterprise conducted by practitioners working in settings organized around disciplines.

Main Points

- Discipline-specific incentive structures, culture, and different levels of access to funding, can hinder or encourage disciplinary participation in big data research projects and exacerbate existing academic inequities.
- The widespread adoption of methodologies drawn from computer and data sciences can create tension between researchers and raise questions about the relative value and status of disciplinary perspectives.
Managing Complex Data

In an era of data abundance, few researchers struggle to access data. In contrast, organizing those data are a major challenge.

Main Points

● Data are abundant, but discovery is decentralized and often difficult.
● Many big data researchers rely heavily on secondary data, which are generally cheaper to acquire than primary data.
● Acquiring, cleaning, processing, and organizing data are the most labor-intensive aspects of most big data projects.
Managing Complex Workflows

Big data research is almost always a collective endeavor involving students, faculty, staff, and colleagues, clients, and collaborators from in and beyond higher education.

Main Points

- The lab is the core unit for much big data research: within those labs, students (both undergraduate and graduate) play major roles in the research process.
- Many researchers view existing campus storage and computing resources as inadequate or insufficiently specialized and prefer local, lab-based, storage and computing solutions.
Sharing Knowledge

Researchers are broadly committed to the open sharing of research outputs, including data and code. However, sharing practices reflect a spectrum that extends well beyond formal, FAIR, sharing.

Main Points

- Sharing should be conceptualized as a spectrum of activities, ranging from formal public archiving of data to informal sharing with colleagues.
- Many researchers are skeptical that much of their data is worth the trouble of sharing, seeing it as either derivative, low quality, or gathered from sources that make sharing ethically complex.
Support and Training

Researchers often rely on informal modes of education as a primary mode of training. While these DIY methods work well for solving immediate problems, the need for “end to end” competencies is likely to grow.

**Main Points**

- Instructors value web resources because they can provide immediate answers to specific problems, especially those involving software and coding. Most view these resources as sufficient for many of their training needs.
- Some researchers fear falling prey to “black box” issues involving conceptual and foundational knowledge of coding, data science, and content knowledge.
How can universities align support offerings to researchers’ needs?
Core challenges for data service providers

**Need for diversified, individualized services**

Data-intensive researchers use a wide range of methodologies and tools and have a wide range of competencies. Many also perceive their needs as idiosyncratic to specific projects or disciplinary ways of knowing, resulting in a wide range of user needs.

**Decentralization and Silos**

Data support services are widely scattered across campus units, including libraries, HPCs, academic departments and research centers. This makes coherent offerings difficult to coordinate and hampers communication and marketing of existing services.

**Antiquated perception of what libraries offer**

Libraries are the single most important provider of data support services, yet researchers often perceive libraries as little more than providers of books and journal articles.
Alignment strategies

Mapping & Coordinating Resources

Universities need to assess and coordinate data services and personnel to maximize their value and minimize duplication. Trainings and funding opportunities should be targeted to disciplines and audiences, such as graduate students, who are most likely to use and benefit from them.

Fostering Learning

Big data research can be facilitated by creating spaces and forums for researchers from different disciplinary backgrounds to share and network. Two specific topics in urgent need of wider consideration are the ethical dimensions of big data research and nuanced conversations about data sharing practices.

Data Storage & Accessibility

Demands on storage and computing capacity are likely to increase considerably as big data research is normalized in more disciplines. Universities, perhaps working in consortiums, must prepare adequate storage options. Funders need to develop mechanisms for easier long-term use of cloud computing services.
Next steps
Research Data Services Assessment Cohort Project

Participating libraries will assess and map their institution’s data service support offerings and compare them to data about user needs, gathered by interviewing of current users of services across university units. Participants will use this evidence base to evolve or create their data service strategy.
Upcoming Publications

Issue Brief on Data Sharing Practices in the Humanities

Research Report on Teaching with Data in the Social Sciences
Discussion

1. What are the biggest challenges your institution is facing while developing its data support services?

2. To what extent is your institution interested in or engaged with data support services leveraging cross-institutional collaborative models?

3. What information about data service trends and user needs would be most useful for your institution?
Thank you