

Comparison of BTAA Data Services

2024-02-08

Motivation

This is a brief summary comparison of **Data Services** provided at the libraries of public BTAA institutions, based on the raw data collected and summarized here. The intention is to provide some benchmarking for Rutgers Data Services, where the public BTAAAs are the most relevant peer group, being composed of similar large, research-intensive institutions that are also dependent on state-funding and seek to serve their state populations in addition to broader research and educational goals. The BTAA also seeks to collaborate around collections and other library initiatives. Northwestern, as a private BTAA, is excluded, as are USC and UCLA, which are new to Big Ten athletics. At this point it is unclear on how these new institutions will be integrated into the academic collaborations of the Big Ten. Of the BTAA institutions included, Nebraska is substantially smaller in student population, faculty numbers, and research activity. However, the remaining institutions are relatively close matches to Rutgers across a number of dimensions.

Methodology

For each institution, I looked through the org charts and directories of public, non-Rutgers BTAA Libraries (with the help of these links) for individuals providing Public Data Services (**PDS**), and for those providing Research Data Management (**RDM**) services. Loosely speaking, we can consider that Public Data Services (PDS) help researchers and students locate and use existing data sources and data software, while Research Data Management (RDM) services focus on helping researchers manage, organize, and share their own data created from their research. Alternatively, we can call **PDS** help with discovery and use of data, and **RDM** as help with the creation and publication of data. This is combined with observations on the presence of other services such as a data repository and how institutions organize these services. When staffing numbers are reported, these are the numbers of people who have some role in either PDS or RDM, but does not necessarily represent FTE employees. However, it is clear from other institutions' organizational structure that most have more staff with full time roles relating to data, rather than wearing multiple hats across the organization. The raw data reflects some notes on the titles found at the institutions, but does not systematically distinguish between librarians and staff, status or rank among those providing services.

Staffing Numbers

The number of staff range from 1 to 17 across all institutions. Those with low numbers may carry a caveat, as are those with very high numbers. The **average staffing is 7** (84/12), with a median of 6.5

Illinois is unusual in that no library staff align with Public Data Services, but there is a distinct separate unit run out of a non-library center that does provide robust public data services, with 14 staff. If these are included, the average rises to 9 (98/12), with a median of 8.

PDS staffing has a mean of 4 (47/12), median of 3, which would rise to a mean of 5 if we were to count the larger Illinois variant. **RDM staffing has a mean of 3** (37/12), median of 2.5.

Staffing Patterns

Maryland is the low-end anomaly, with only one staff member clearly identifiable as providing Public Data Services, and none in RDM. However, they have repository services, a “*GIS and Data Service Center*”, and a GIS lab. Services are well-described, but use generic e-mail addresses as contacts. It may be that graduate students are actually delivering most of their announced services. One Data Services position is also currently vacant.

At the other extreme, Wisconsin lists very well-staffed services in both RDM and PDS. Research Data Services consists of **nine** individuals on the RDS Team from both Libraries and IT, and **eight** individuals have PDS responsibilities, some through The Data & Information Services Center. Another notable position is the **Data, Science & Engineering Diversity Resident Librarian** at the Science Library, a position that fills many needed criteria.

In the median, most institutions have some form of Data Librarian, GIS Services, Research Data Management Librarian, and one or two additional areas covered like Data Visualization or Data Curation. As mentioned above, it appears that at other BTAA’s, the “multiple hats” problem is not as severe, so their staffing delivers more effective service than the Rutgers’ model of multiple responsibilities beyond data, where we have no full-time staff completely focused on data. Further details are in the notes on each institution

Data Repositories

Five institutions (Illinois, Michigan, Minnesota, Nebraska, Purdue) have dedicated data repositories that are separate from their institutional repositories. In most cases, these have clearly identified RDM staff, such as at Illinois’ Research Data Service, with 3 staff and one graduate assistant. When dedicated data repositories exist, they tend to accrete RDM staff and services towards themselves.

Six institutions have clear statements to indicate that data (up to certain size limits) is accepted by their general institutional repository.

Only one institution, Michigan State, does not discuss data in connection to its repository. (Rutgers also has no public information about the data that is accepted into SOAR)

Presentation of Services

It appears to be a matter of taste and of aligning with the organization's staffing patterns whether or not to separate RDM and PDS when presenting these services to users, but the best practices among peers provide clear home pages and launch points to the various data services offered.

Of these, perhaps Ohio State's is the most user friendly via their Research Commons.

Others include:

- Michigan Data Services
- Michigan State Data Scholarship Lab
- Penn State Research Informations and Publishing

Discussion of Rutgers

Staffing At Rutgers, there are 3 individuals in the NBL Research Data Services group, one focused on RDM, one focused on PDS, and one in digital humanities (which is not typically included in peers' data services). Looking across the system, we can identify a (half-time) GIS Specialist and a librarian providing data services in Newark (in addition to other primary responsibilities), providing a total of five staff under a generous interpretation. As is typically the case at Rutgers, none of these staff are completely focused on these tasks, but rather have multiple additional responsibilities which would be tedious and time-consuming to enumerate. While the current staffing is a rational evolution and response to the conditions at Rutgers and makes sense for the staff members involved, future hiring in the data services area would be more effective at bolstering data services if those hires were focused on data services instead of split among "multiple hats". Generous accounting would be to consider these 5 individuals devoting 50% of their name to data services, implying a maximum of 2.5 FTE at Rutgers. The One Library approach may help us to leverage positions across campuses, but we are currently still far short of the large majority of our peers.

Repository We have incompletely developed data repository resources, and our "data deposit" is not publicly advertised or acknowledge, but is only available to those in the know, like *In and Out Burger's* secret menu, which represents a form of solution to our limited capacity, but one that is less than ideal. Moving to a situation where we could publicly describe the data deposit services available at Rutgers, whether through the Libraries or elsewhere, would be a huge bonus to the increasing percentage of researchers who deal with data deposit requirements for their grants and publications.

Presentation We lack a coherent and comprehensive presentation of our data services, which is partially due to the fragmentation of the Libraries site by campus, and partially due to the inability to create a more complete view of the services we do offer.

Recommendations

Staffing Personnel, whether library faculty or staff, at the level of *4-5 FTE who are fully dedicated to data services* would position Rutgers at the lower end of the middle range of our BTAA peers. This would require hiring at least 2 full-time staff or library faculty with dedicated data roles across RUL, but there are compelling cases to be made for expanding RUL's services into areas that we cover incompletely at present, such as Diversity in Data, Data Visualization, Bioinformatics, AI, and more. Hiring up to the 4-5 FTE level should be an immediate aspirational goal.

Repository Whether within the Libraries, or at another Rutgers unit, a better solution for data deposit and advice and consultation around data deposit to meet funder and publisher mandates should be developed.

Presentation We lack a coherent and comprehensive presentation of our data services, which is partially due to the fragmentation of the Libraries site by campus, and partially due to the inability to create a more complete view of the services we do offer which are organizationally divided. We should study the best practices of peers to develop a one-stop solution for data services from the Libraries that will then route users to the appropriate places for more specialized help.

These are simplified, summary recommendations, but studying the peer data and looking at trends in library data services can provide more nuanced recommendations that will evolve with changing circumstances. Campus collaborations such as with OARC and the Office of Research can provide even stronger services and more effective results for Rutgers in terms of data services.