TDR Survey of Institutional Dataverse Liaisons

Released February 2019

Christina Chan-Park, Baylor University
Laura Waugh, Texas State University
Nerissa Lindsey, Texas A&M International University
Laura McElfresh, Texas A&M University-Galveston
Ahmet Tmava, Texas Woman's University

Table of Contents

Purpose

The Texas Data Repository (TDR) Steering Committee's Assessment Working Group (AWG) is tasked with evaluating the progress of the TDR. In Fall 2017, the AWG began an assessment to identify the needs for reporting on the TDR by addressing the following research questions:

1. Which usage and descriptive information about the TDR will be most valuable?
2. What process for gathering and distributing these metrics/information will be most useful?

The AWG administered a survey to aid in prioritization of assessment needs for the TDR Dataverse. The purpose of this survey is to identify the levels of institutional capacity at member institutions and provide recommendations on future assessment needs, metrics, and reporting moving forward. Data from this assessment is structured to anticipate potential usage of the TDR Dataverse and guide future work and needs in assessment, training, and continued growth.

Because the most valuable usage metrics should allow for comparative assessment across repositories globally, the AWG also compiled descriptions of metrics recommended by three sources which suggest best practices for tracking the impact of research data, including the Make Data Count Project's "Code of practice for research data usage metrics." The results from this compilation of best practices has been combined with the results of the survey to determine a prioritized list of metrics.

Methodology

A Google Form survey (Appendix A) was sent to all members of the TDR Steering Committee in October 2017 and collected through November 2017. Initially, there were nine respondents. As member institutions joined TDR and liaisons to the Steering Committee changed, two more respondents were contacted in Spring 2018 and one more in Fall 2018. As one of the respondents was a replacement liaison, there are a total of 12 complete responses in this analysis. We expect to continue to administer the survey to new liaisons as they join the Steering Committee. The survey consists of 34 questions covering the areas of institutional demographics; TDR staffing in member libraries; liaison experience and training; member libraries' other data initiatives; and anticipated user needs.

In addition to analyzing the results of the survey, the TDR Assessment Working Group identified the following three resources relating to metrics and reporting best practices:

- Fenner, M., Lowenberg, D., Jones, M., Needham, P., Vieglais, D., Abrams, S., Cruse, P., Chodacki, J. (2018). *Code of practice for research data usage metrics*, release 1. PeerJ Preprints 6:e26505v1 https://doi.org/10.7287/peerj.preprints.26505v1

- Bragg, M., DeRidder, J., Johnson, R., Junus, R., Kyrillidou, M., Chapman, J., & Stedfeld, E. (2017). *Best Practices for Google Analytics in Digital Libraries.* http://doi.org/10.17605/OSF.IO/CT8BS
- Lyon, C. (2017). *Texas ScholarWorks: 2017 Annual Report.* University of Texas at Austin. 20p. https://repositories.lib.utexas.edu/handle/2152/63730

For the analysis of the best practices, findings were abbreviated as “Code”, “Google”, and “TSW” respectively. The 24 compiled metrics were sorted into four categories: metrics about datasets and dataverses; metrics about users; metrics about views and access; and metrics about user sessions. Combined with the top nine metrics requested from the survey, each metric was given a high (11), medium (5), or low (8) priority.

Survey Findings

The institutional demographic information is included in Appendix B.
Based on the survey, we’ve found:

Training

The majority of respondents (73%) received training and instruction related to research data management via workshops, webinars, and on-the-job. Closely matched are online classes (64%), including MOOCs. The rest (55%) learned through conferences and committees.

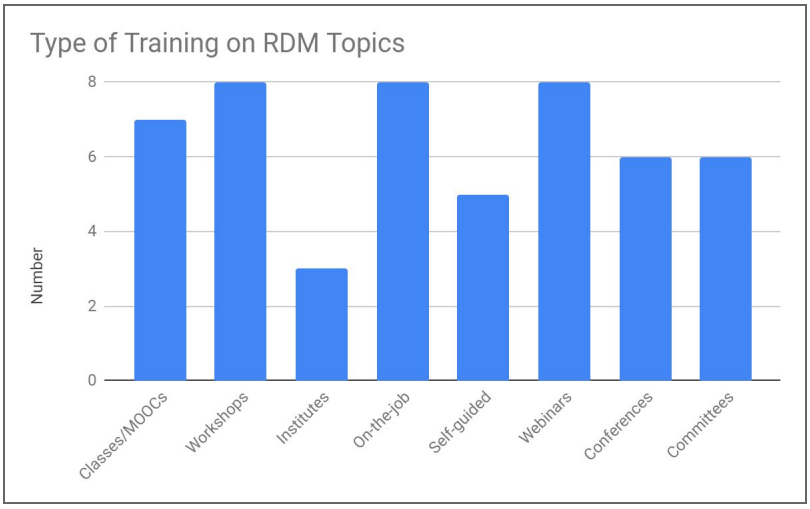

Figure 1: Type of training on research data management topics

Software Tools

Respondents indicated that in addition to institutional participation in the TDR Dataverse, the Data Management Planning Tool (DMPTool) is the most widely-used research data management tool (73%), followed by ORCID and ICPSR (64%) and DataCite/EZID (18%).

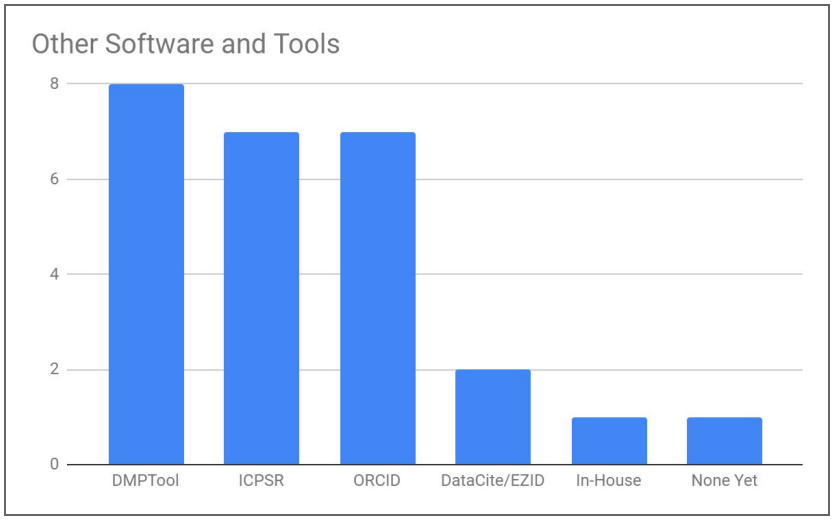

Figure 2: Other software tools for research data management

Targeted Audiences

This survey was administered during the first year of the launch of the TDR Dataverse. Respondents were asked to anticipate groups for targeted outreach and those expected as most likely to deposit their data. Faculty were the highest anticipated targeted group with 100%, followed by graduate students with 82%; research staff and postdocs with 64%; and librarians with 54%. Individual departments, colleges, research labs, and undergraduates were all under 50% as anticipated groups for targeted outreach.

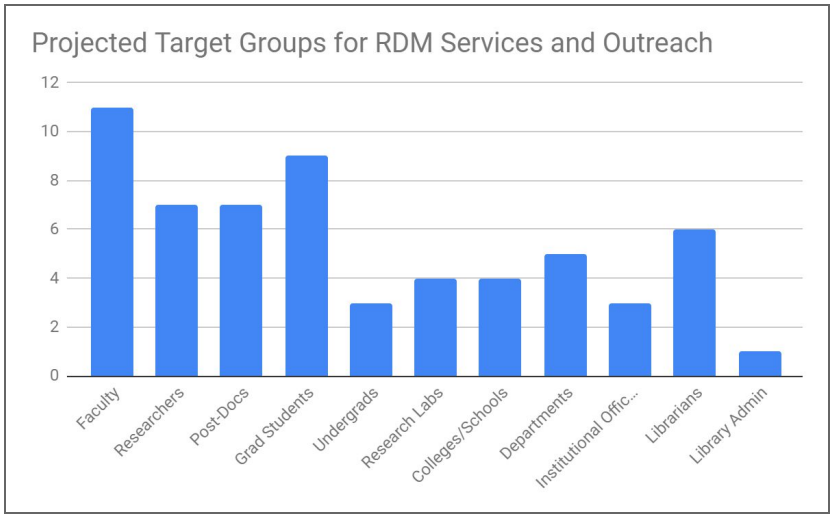

Figure 3: Projected target groups for research data management services and outreach

Projections of constituencies who would be uploading data into the TDR aligned closely with respondents' anticipated target populations for outreach services. Faculty were the most highly projected submitters of data at 92%, followed closely by research staff and graduate students at

83%. Sixty-seven percent (67%) of respondents anticipated postdocs uploading data; 50% projected participation from research labs; and only 17% expected that groups such as undergrads, departments, or institutional offices will be uploading data. Less than 1% thought that colleges or schools will upload data and 0% expected that librarians will upload data.

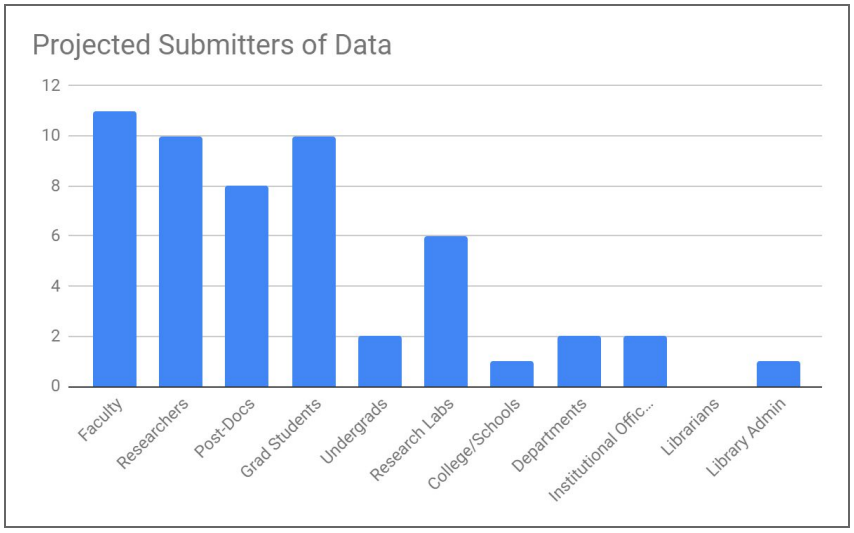

Figure 4: Projected submitters of research data

Data Management Services

In regards to services, respondents were asked to select the type of data management services they provide to their communities. Sixty-three percent (63%) of respondents indicated that their libraries provide consultations regarding data management services, followed by training (54%) and education (50%).

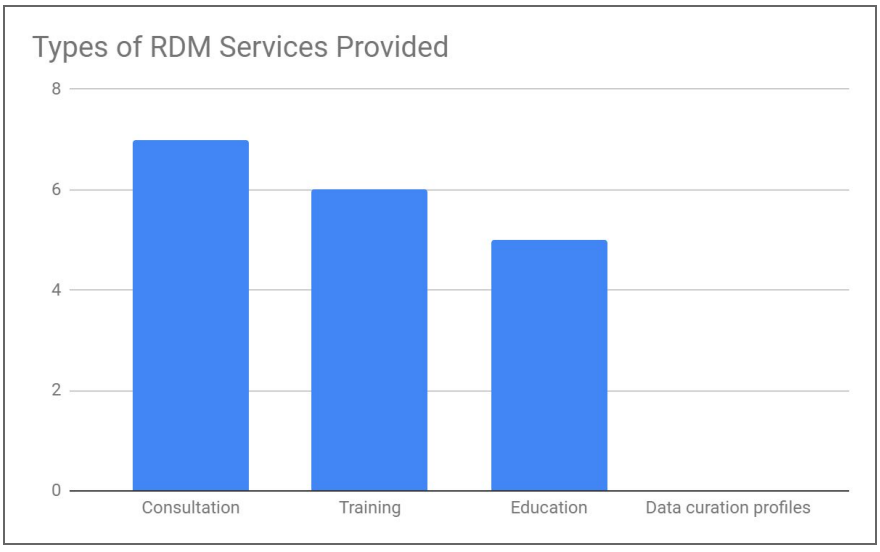

Figure 5: Types of research data management services provided

Computational Services

Respondents reported a wide variety of entities on campus that provide computational services. The largest segment (26%) indicated that their institutions' IT departments provided computational services, 19% indicated that support came from individual departments or colleges, 19% from their office of research, 15% indicated the library, and another 15% indicated they had computational services from a high performance computing center.

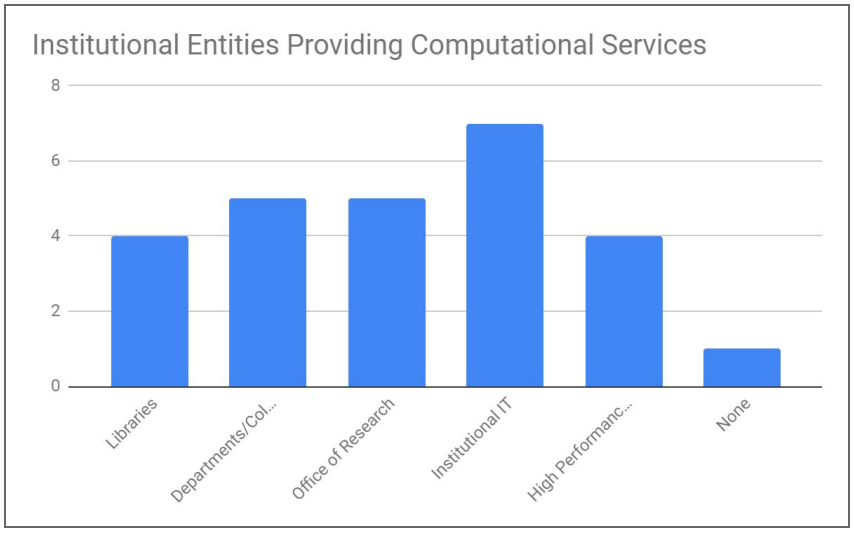

Figure 6: Institutional entities providing computational services

Thirty-eight percent (38%) of respondents indicated offering computing support; 21% offer high performance computing; and another 21% offer data storage facilities. Only 16% offer big data storage, while under 1% of responding institutions offer no type of computational service at all.

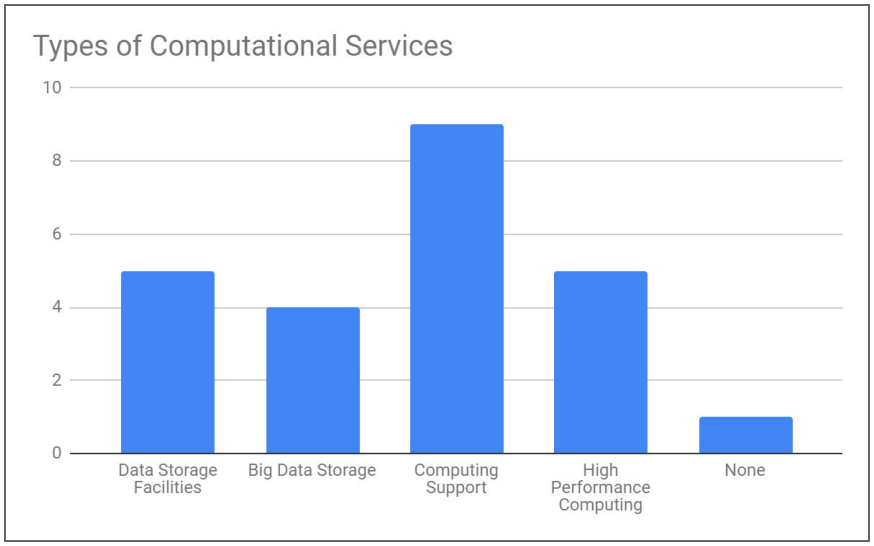

Figure 7: Types of computational services offered

Best Practices Findings

Combining findings of the survey with metrics currently offered through the TDR Dataverse and, other potential items to prioritize, and the standards outlined in three sources of research on best practices in metrics and reporting, the TDR Assessment Working Group prioritized each metric with a score of high (11), medium (5), or low (8).

	Suggested By:				Source	Priority
	Survey	Code	Google	TSW		
Metrics about datasets and collections						
Dataset download counts	X	X	X		Log File, Google	high+
Size of datasets (MB)	X				Log File	high+
Number of files within datasets	X				Log File	medium
List of dataset DOIs	X				Log File	high+
Size of collections (MB)	X				Log File	high
Hierarchy collections and datasets	X				Log File	medium
List of links to all collections and datasets	X				Log File	high+
Metrics about users						
Institutional researchers using the TDR	X				Log File	high*+
Unique page visitors	X		X	X	Google	high
Percent new users				X	Google	low
Percent users from referring site				X	Google	low
Percent users directed from search engines				X	Google	low
Location		X	X	X	Google	medium
Metrics about views and access						
Unique Dataset Investigations		X			Log File	high
Total Dataset Investigations		X			Log File	high

Number of pageviews			X	X	Google	medium
Mode of Access		X	X	X	Log File, Google	medium
Metrics about sessions						
Number of sessions			X	X	Google	high
Average session duration			X	X	Google	high
Path Through the Site			X		Google	low
Referral Traffic			X		Google	low
Site Content Reports			X		Google	low
Bounce Rate			X		Google	low
Search Terms			X		Google	low

Figure 8: Table of best practice findings for reporting

+ Metrics currently included in TDR-SC AWG reports
** Partial data about users -- includes only registered contributors of data*

Conclusion/Recommendations

Based on our findings, we recommend that the regular TDR Dataverse reports to each institution continue to include the following metrics:

- Dataset download counts
- Size of datasets (MB)
- List of dataset DOIs
- List of links to all collections and datasets

Additionally, we recommend that the following metrics be included:

- Size of Collections (dataverses)
- Unique page visitors
- Unique Dataset Investigations
- Total Dataset Investigations
- Number of sessions
- Average session duration

We also recommend expansion of the “Institutional researchers using the TDR” metric to include all registered users, rather than only those creating dataverses or datasets.

As the top priority for added services, we recommend that TDR consider size of collections; unique dataset investigations; total dataset investigations; and number of sessions.

Based on the standard best practices research, location of users and mode of access are prioritized. The TDR Assessment Working Group recommends discussing these two potential metrics for reporting with the full TDR Dataverse Steering Committee and the Texas Digital Library in order to gauge how these two metrics could be measured and reported in the future.

In light of the Steering Committee's statement of support for the principles of "Make Data Count," it is recommended that TDR contribute data to this initiative when these standards have been incorporated into the Dataverse Platform.

The TDR Assessment Working Group found wide variation in our member institutions' ability to devote FTE, financial, and other institutional resources to support TDR Dataverse services. Consequently, we should affirm and build upon existing cooperative efforts to share outreach and education resources among our member institutions.

Moving forward, the TDR Assessment Working Group recommends investigating additional assessment tools, other than Google Analytics and Miniverse, and how the TDR Dataverse services could be expanded to meet the needs of institutional members.

Appendix A: Survey

TDR Survey of Institutional Dataverse Liaisons

This survey was created by the Assessment Working Group of the TDR Steering Committee to aid in the prioritization of assessment needs for the TDR Dataverse. It will help the TDL/TDR Assessment WG to understand the different levels of institutional capacity to provide TDR services and manage TDR institutional Dataverses. The TDR Assessment WG hopes that the answers will help to anticipate potential usage in the future and guide further assessment, training, and growth of TDR.

*Required

Email address*

Name*

Institution*

Research Data Management at your Institution

These are general questions about research data management at your institution.

What is your job title?*

Which library department or unit are you a part of?*

To what position do you directly report?*

Do you have any other staff or student workers dedicated to research data management operations? (check all that apply)*

- ❏ Staff
- ❏ Student workers
- ❏ None
- ❏ Other:

If you checked ‘staff’, how many?

If you checked ‘student workers’, how many?

If you selected ‘other’, how many?

What is the Full-Time Equivalent (FTE) of all people working on research data management at your institution? (ex. a librarian working ½ time and two staff working ½ time is 1.5 FTE)*

- ❏ < 0.25 FTE
- ❏ 0.26-0.5 FTE
- ❏ 0.51-1.0 FTE
- ❏ 1.1-2.0 FTE
- ❏ >2.0 FTE

What type of training, experience and/or education have you had in research data management topics? (check all that apply)*

- ❏ Classes (including MOOCs)
- ❏ Workshops
- ❏ Institutes
- ❏ On-the-job training/experience
- ❏ Self-guided online course or modules
- ❏ Webinar(s)
- ❏ Conference(s) or professional organization(s)
- ❏ Committee(s) or Working Group(s)
- ❏ Other:

Wherever possible, please provide identifying information or a description for the checked above (ie MOOC = ODUM Institute)

What percentage of your time is devoted to providing research data management services?

- ❏ 0-25%
- ❏ 26-50%
- ❏ 51-75%
- ❏ 76-100%

Besides the TDR, which research data management tools and software do your libraries currently manage or provide access to? (check all that apply)*

- ❏ Open Science Framework for institutions (https://osf.io/search/?q=institutions&filter=institution&page=1ion&page=1)
- ❏ Data Management Planning Tool partnership (https://dmptool.org/partners_list?e=z&s=a)
- ❏ ICPSR membership (https://www.icpsr.umich.edu/icpsrweb/membership/administration/institutions)
- ❏ ORCID membership (https://orcid.org/members)
- ❏ DataCite or EZID DOI minting service
- ❏ Other:

Which research data management services do your libraries currently offer? (check all that apply)*

- ❏ Consultation
- ❏ Training
- ❏ Education (including classes, workshops, lectures, etc.)
- ❏ Data curation profiles
- ❏ Other:

Please describe each service checked above. Include general topics covered (DMPs, metadata, repositories, etc.); estimated frequency (daily, monthly, yearly); and scale (individuals, labs, small classes, etc.)

Which computational services are offered at your institution, by the Libraries or other entities? (check all that apply)*

- ❏ Data storage facilities
- ❏ Big data storage capacity
- ❏ Computing support
- ❏ High performance computing
- ❏ None
- ❏ Other:

Which entities at your institution provide computational services? (Computational services include any services directed at helping researchers with data computation activities.) (check all that apply)*

- ❏ Libraries
- ❏ Individual departments or colleges
- ❏ Office of Research
- ❏ Institutional IT
- ❏ None
- ❏ Other:

Which groups do you anticipate focusing your research data management services and outreach efforts on? (check all that apply)*

- ❏ Faculty
- ❏ Research staff
- ❏ Post-Docs
- ❏ Graduate students
- ❏ Undergraduate students
- ❏ Research labs/institutes
- ❏ College/Schools
- ❏ Departments
- ❏ Institutional offices
- ❏ Subject/Liaison librarians
- ❏ Library administration
- ❏ Other:

Please estimate how many faculty your institution has?*

Please estimate how many (non-professional) graduate students your institution has?*

Please estimate how much external funding your institution has for research?*

Which of these provides the majority of funding: (check one)*

- ❏ Federal grants

- ❏ Philanthropic organizations
- ❏ Private companies
- ❏ Not applicable
- ❏ Other:

TDR Dataverse-Specific Questions

These questions relate specifically to the TDR Dataverse.

Is there a designated backup who could take over Texas Data Repository Dataverse operations in your place in case you cannot perform your duties for any reason?*

- ❏ Yes
- ❏ No
- ❏ Maybe

If yes, who is the backup? (name and/or position)

What type of deposit [will/does] your institutional Dataverse have? (check all that apply)*

- ❏ Self deposit
- ❏ Mediated deposit (by a trained user)
- ❏ Curated deposit (manipulated after the fact by TDR liaison)
- ❏ Unsure

Which individuals or groups on your campus do you anticipate submitting data to the Texas Data Repository Dataverse? (check all that apply)*

- ❏ Faculty
- ❏ Research staff
- ❏ Post-Docs
- ❏ Graduate students
- ❏ Undergraduate students
- ❏ Research labs/institutes
- ❏ Colleges/Schools

- ❏ Departments
- ❏ Institutional offices
- ❏ Other:

As the TDR liaison for your institution, which descriptive information about your institutional TDR Dataverse currently interests you? Please be aware, these aren't mutually exclusive. (select top five)*

- ❏ A hierarchy of dataverses, nested dataverses and datasets
- ❏ Listing links to all dataverses and datasets
- ❏ Listing dataset DOIs
- ❏ Listing Datasets that have been updated after publication (versioned)
- ❏ Size of dataverses
- ❏ Size of datasets
- ❏ Combined size of files within datasets
- ❏ Individual sizes of files within datasets
- ❏ Number of files within datasets
- ❏ Number of file upload errors
- ❏ Other:

As the TDR liaison for your institution, which usage information about your institutional TDR Dataverse currently interests you? Please be aware, these aren't mutually exclusive. (select top two)*

- ❏ Identifying institutional researchers creating dataverses and datasets
- ❏ Identifying external researchers creating dataverses and datasets (i.e. via Google login)
- ❏ Number of unique page visitors to dataverses and datasets
- ❏ Dataset download counts
- ❏ Individual file download counts
- ❏ Other:

We are currently investigating Google Analytics and Miniverse. Are you aware of other assessment tools or software that the Texas Data Repository Steering Committee should be investigating?

How should TDR services be expanded to meet current needs at your institution? (check all that apply or rank order)*

- ❏ Large datasets-large number of files
- ❏ Large datasets-large file sizes
- ❏ Bulk upload
- ❏ Allowing outside collaborators to edit/upload
- ❏ Training for librarians
- ❏ Training for users
- ❏ TDR usage assessment
- ❏ Individual user assessment
- ❏ Other:

Appendix B: Demographics

Participating Institutions:

- Baylor University
- Texas A&M International University
- Texas A&M University
- Texas A&M University-Corpus Christi
- Texas A&M University-Galveston
- Texas State University
- Texas Tech University
- Texas Woman's University
- University of Houston
- University of Texas at Austin
- University of Texas at Arlington

Job Titles:

- Catalog Librarian
- Data Management Coordinator
- Data Management & Metadata Assistant Librarian
- Data Management Librarian
- Digital Collections Librarian
- Digital Initiatives Librarian
- Digital Projects Librarian
- Director of Research Data Services
- Manager of Digital Services
- Reference and Scholarly Communication Librarian
- Science and Engineering Librarian

Library Department/Unit Affiliations:

- Cataloging Department
- Collection Services
- Digital and Web Services
- Digital Projects
- Digital Resources
- Digital Scholarship
- Digital Services Unit
- Liaison Services
- Office of Scholarly Communications
- Research and Data Services
- Research and Engagement
- Research and Instruction

Direct Report in Respondents' Positions:

- *4* - Department Head
- *1* - Assistant Dean of Academic and Engagement Service
- *1* - AUL, Scholarly Communications
- *1* - Digital Resources Unit Coordinator
- *1* - Director of Library and Learning Commons
- *1* - Director of the Office of Scholarly Communications
- *1* - Library Director
- *1* - Manager of Digital Services
- *1* - Scholarly Communication Librarian

Dedicated Staff or Student Workers to Research Data Management:

About half of the respondents indicated that they have no dedicated staff responsible for Research Data Management. 30% have a librarian, or hope to hire a librarian soon that will be fully responsible for RDM. 20% indicated having a student worker or interns dedicated to RDM.

Percentage of Time Dedicated to Research Data Management:

About half of the respondents indicated that they dedicate 0-25% of their time to RDM. 20% dedicate about 51-75% of their time. 30% spend 76-100% of their time.

Dedicated Backup Person for TDR Dataverse Responsibilities:

The majority (58%) of respondents do not have a designated backup person at their institution to perform TDR Dataverse duties in case the primary liaison is unable to for any reason. Four respondents (33%) do have a designated backup and 1 respondent stated maybe (8%). Of those who responded yes, that they have a backup, the position(s) of the backup person are:

- Acquisitions Librarian
- Associate Professor & Director, Digital Library Services and Systems
- Research, Instruction, and Outreach Librarian
- Scholarly Communications Librarian

Estimated Number of Faculty at each Institution:

Respondents	Number of Faculty
1	150
1	400
1	600
2	1000
3	1200

1	1600
1	2500
1	3000
1	3500

Estimated Number of Graduate Students at each Institution:

Respondents	Number of Graduate Students
1	100
1	150
1	2000
1	2250
2	4000
2	5000
1	6500
1	8000
2	Not sure

Estimated Amount of External Funding at each Institution for Research:

Respondents	External Annual Research Funding
1	2.4 million
1	5 million
1	5.8 million
1	8 million
1	15 million
1	23.5 million

1	35 million
1	85 million
1	600 million
3	Unsure/Not applicable

Provider of the Majority of Research Funding at each Institution:

Respondents	Funding Provider
8	Federal grants
1	Philanthropic organizations
0	Private companies
1	Not applicable
2	Other: (Not sure)

Appendix C: Individual TDR Dataverses

Type of Deposit in the TDR at each Institution (all that apply):

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Self Deposit		✔	✔	✔	✔	✔	✔		✔			
Mediated Deposit (by a trained user)	✔	✔							✔		✔	
Curated Deposit *(manipulated after the fact by TDR liaison)*	✔	✔	✔		✔			✔	✔			✔
Unsure										✔		

Descriptive Information about Institutional TDR Dataverse of Interest (top five):

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
A hierarchy of dataverses, nested dataverses, & datasets			✔	✔	✔	✔		✔	✔	✔	✔	✔
Listing links to all dataverses & datasets		✔			✔	✔	✔		✔		✔	✔
Listing dataset DOIs	✔	✔	✔	✔		✔				✔	✔	✔
Listing datasets that have been updated after publication (versioned)	✔			✔	✔							✔
Size of dataverses		✔	✔	✔		✔			✔	✔		✔
Size of datasets		✔	✔	✔		✔		✔	✔	✔	✔	✔
Combined size of files within datasets		✔									✔	✔
Individual sizes of files within datasets		✔								✔		✔
Number of files within datasets	✔	✔				✔		✔	✔			✔
Number of file upload errors					✔	✔						✔

Types of file upload errors			✔		✔	✔			✔			✔

Usage Information about Institutional TDR Dataverses of Interest (top two):

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Identifying institutional researchers creating dataverses and datasets		✔	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔
Identifying external researchers creating dataverses and datasets (e.g., via Google Login)	✔								✔		✔	
Number of unique page visitors to dataverses and datasets			✔		✔	✔		✔	✔			
Dataset download counts	✔	✔		✔		✔	✔	✔	✔	✔	✔	✔
Individual file download counts						✔			✔		✔	

Suggested Expanded TDR Services to Meet Institutional Needs (all that apply):

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12
Large datasets-large number of files	✔	✔	✔	✔	✔		✔	✔	✔	✔		
Large datasets-large file sizes	✔	✔	✔	✔	✔	✔		✔	✔	✔	✔	✔
Bulk upload		✔	✔	✔	✔			✔	✔	✔		✔
Allowing outside collaborators to edit/upload	✔		✔		✔				✔			
Training for librarians	✔	✔			✔	✔	✔		✔	✔		✔
Training for users	✔	✔	✔		✔	✔			✔	✔	✔	✔
TDR usage assessment				✔	✔	✔			✔		✔	
Individual user assessment				✔	✔				✔			
Other		✔					✔					

Other Responses:

- HIPAA and FERPA security support

Appendix D: Open-ended Survey Responses

Training, Experience, and/or Education on Research Data Management:

- *3* - TDL Fall Data Symposium (2016)
- *1* - Research Data Alliance 10th Pneary in Montreal, CA (2017)
- *1* - ALA Data Management Course
- *1* - Harvard-Purdue Data Management Symposium
- *1* - New England Collaborative Data Management Curriculum (NECDMC)
- *1* - DataOne modules
- *1* - RDM Rose Learning Materials
- *1* - Greater Western Library Alliance Research Data Management Task Force
- *1* - Handling Restricted and Sensitive Data Webinar (SLA Data Caucus)
    - https://bit.ly/2sTf3jQ
- *4* - Research Data Management and Sharing MOOC, University of North Carolina at Chapel Hill and MANTRA team at the University of Edinburgh (Coursera)
- *1* - Certification in Digital Curation and Data Management, University of North Texas
- *1* - Data Curation Profile Workshop

Describe each research data management service offered at your library:

- We offer consultation and individualized training to faculty, staff, and students. We've consulted on basic steps to deposit research data, organizing the datasets, including appropriate metadata, and explaining this on a data management plan. It is on an as-needed basis and over the last four months we've only delivered that twice. Both were via the telephone and email.
- Consultations vary between years, but on average we are responsible for 4-6 a year. This consultation usually is specifically focused on helping faculty realize what kind of data they are generating during the life cycle of their award and how it's going to be made available and preserved. During a consultation process, file formats, amount of data, metadata, repositories, publishing, and short/long-term preservation are discussed.
- Education tends to be small workshops with faculty groups but sometimes one-on-one sessions after a consultation. These have been rare in the past, but may increase in frequency.
- Metadata always seems to be a difficult concept to grasp for faculty who aren't familiar with it. At our institution, we've created an online client that will allow the creation of metadata in the set schemas of Dublin Core and Darwin Core. We've just rolled out this option on our website, and the application layout and training materials for it are still in development.

- Individual consultations on DMPs, consultations with labs to create workflows for using TDR. Training researchers to use TDR. Teaching an introductory RDM workshop series for graduate students, developing on-demand RDM workshops for labs and research groups, and developing a recorded RDM series of presentations for online masters students.
- 10-20 general or open workshops per semester (usually 20-50 attendees each); 5-10 consults per week on everything: DMPs, file formats, database design, GIS support, tools, metadata, TDR, etc.; 5-8 smaller scale tailored sessions with research groups, classes, or labs.
- So far, we have only signed up to give one hour introductory session on the data repository.
- Consult with users as needed regarding TDR.
- One-on-one consultations with faculty interested in submitting to the data repository.
- Data management consultations that include plan review, editing, and general consulting. Training is provided for library staff.
- General data management workshops (2-6 per year), and review of DMPs (10-20 per year).
- We don’t have any services yet, but would like to offer some. I anticipate offering consultation to individuals as needed, as well as education and/or training in small classes on at least a semesterly basis. I would also like to support ORCID registration for undergraduate researchers and anyone else needing it. I believe our institution’s main campus has our faculty and graduate students covered, but would be glad to help them too.
- Still on planning stages to offer education and consultation.
- DMP reviews, data management workshops, and data repository offered in classes, through consultation requests, and during workshop series.