OPEN ACCESS

CrossMark

A survey of the state of research data services in 35 U.S. academic libraries, or "Wow, what a sweeping question"

Matthew Murray[‡], Megan O'Donnell[§], Mark J. Laufersweiler[|], John Novak[¶], Betty Rozum[#],
Santi Thompson[¤]

‡ University of Guelph, Guelph, Ontario, Canada
§ Iowa State University, Ames, Iowa, United States of America
| University of Oklahoma, Norman, United States of America
¶ University of Maryland, College Park, United States of America
# Utah State University, Logan, United States of America
¤ University Libraries, University of Houston, Houston, United States of America

Abstract

This report shares the results of a Spring 2018 survey of 35 academic libraries in the
United States in regard to the research data services (RDS) they offer. An executive
summary presents key findings while the results section provides detailed information on
the answers to specific survey questions related to data repositories, metadata,
workshops, and polices.

Keywords

research data, data management, research data services, academic libraries, survey,

Background

The Greater Western Library Alliance (GWLA) is a consortium of 39*1 academic research libraries located across the United States. Its members includes 31 public and 8 private institutions from 20 states, ranging from Delaware to Hawaii. The schools that these member libraries serve vary widely in size as seen in Table 1.

Table 1.

A breakdown of the GWLA member institutions at the time of the survey based on full-time student enrollment, land grant status, and research ranking. Enrollment numbers were provided from GWLA; land grant status and research rankings were obtained from 2017 IPEDS 2017 data U.S. Department of Education (2017). R1 is assumed to be equivalent to the "Doctoral Universities: Highest Research Activity" classification given in IPEDS.

*As The Claremont Colleges are not coded as a single entity in IPEDS they were excluded from the research ranking analysis.

Full-time enrollment	# of institutions	# land grant	R1 Ranking
7,000 - 9,999	3	0	2*
10,000 -19,999	7	4	7
20,000 - 29,999	15	8	15
30,000 - 39,999	8	1	8
40,000 - 99,999	4	2	4
100,000 and above	1	0	1

GWLA activities include programs for scholarly communication, interlibrary loan, shared electronic resources, cooperative collection development, digital libraries, staff development and continuing education.*2 In 2016 the Library Deans/Directors of these institutions created a Data Management Task Force to investigate issues related to research data management and to identify potential collaborative projects. A librarian from each GWLA institution served on the Task Force.

If the GWLA membership were to work collectively in the future to, for example, share expertise, develop shared repositories, or advocate for minimal core competencies in data management and curation, they would need information about the data services and resources offered by each institution in the consortium. A subgroup of the Task Force was established in 2017 to discover and document the data practices and policies of each GWLA institution.

The subgroup agreed to develop and administer a survey to collect this information from member libraries. The survey was developed in Fall 2017 and sent to the 38 GWLA Library Deans/Directors in February 2018. The final report, included here, was submitted to the GWLA Deans/Directors in May 2019.

Executive Summary

Key findings from the survey are presented here. These include library instruction, data repositories, digital preservation, metadata, polices and plans, campus research data services beyond the library, and library organization.

Library instruction

- All of the GWLA institutions who answered the survey (35 of 38), offer Research Data instruction services in the form of course-related instruction, workshops, and/or consultations.
- The top three topics for data workshops are general RDM (research data management) , specific tools, and programming languages.

Data repositories

- Over 85% stated that their institution had a repository that accepted research data, however 70% ofinstitutions have a single repository that is used for both data and other scholarly material. 30% have dedicated data repositories.
- 86.7% of repository policies address who is eligible to deposit, 73.3% address the type of files, 60% address the inclusion of sensitive date, 53.3% address the size of files, and 56.7% have a terms of deposit document.
- Institutions were split over mediated vs. unmediated deposit models, with 65.2% offering mediated approaches.
- Creative Commons licenses (CC-0 and CC-BY) are the most popular forms of licenses available in respondents repositories.
- 79% of respondents have repositories that produce persistent identifiers.
- Institutions were evenly split over using Software as a Service (SaaS) vs. self-hosted. DSpace and Digital Commons were the most popular software platforms used.
- 70% of libraries do not share staff across both their data and institutional repositories.
- Only 30% of repositories accept data that is still being updated.
- 80% of repositories allow for embargo periods. 75% of those have no limit on the length of embargo.

Digital preservation

- Digital preservation practices were nearly evenly split: 30% of respondents only backup content, 40% of respondents engage in strategies that exceeded backing up content, and 26% have no digital preservation strategy.
- Software used to implement preservation strategies, include Rosetta, DuraCloud, LOCKSS, Arkivum, Digital Preservation Network (DPN), and Amazon S3 services.

Metadata

- 79.3% of institutions use only one record level metadata schema in their data repositories. Dublin Core is the most prominent schema with 75.9% of repositories using it.
- 66% help researchers understand and identify metadata and metadata standards related to research data however only 49% help researchers apply metadata standards to their research data.
- 90% with a data repository have staff that create or assist researchers with the creation of "record metadata."
- 62% assist researchers with the creation of data documentation (i.e. README files or codebooks)
- 14% do not provide research data metadata services.

Policies and plans

- 80% of GWLA libraries do not have policies that address research data. However, 57.1% have campus research data policies and 61.7% said that their library or institution has a strategic plan or mission that addresses research data.

Campus Research Data Services (RDS) beyond the library

This section only asked about groups unaffiliated with the library. There is likely overlap between services offered by libraries and other groups on campus. However, the survey did not ask for this information.

The most common RDS (research data services) offered by groups on campus unaffiliated with the library are:

- 65.7% - Statistical software support
- 65.7% - Data analysis support
- 62.9% - Active research data storage and backup solutions
- 60.0% - GIS and geospatial analysis
- 57.1% - Data visualization support
- 57.1% - Assistance locating data storage and backup solutions

The least common RDS offered by groups on campus unaffiliated with the library:

- 5.7% - Metadata assistance
- 20.0% - File organization and naming conventions
- 22.9% - Topic or How-To Guides
- 28.6% - Locating and using existing data; including identifying and suggesting repositories

The most common groups offering RDS on campus beyond the library are:

- Research and/or service centers/institutes (number of institutions=23)
- Information technology units (number of institutions=20)
- Academic departments or colleges (number of institutions=14)
- High Performance Computing units (number of institutions=13)

Library organization

- The number of library staff dedicated to RDS ranges from 0 to 3.5 FTE (AVG 1.3 FTE). Meanwhile the number of staff who provide RDS, but are not dedicated to it, shows a much wider range (0-15). Together these two numbers indicate that how libraries staff RDS varies widely. The survey did not ask to differentiate between tenure track, permanent, or contract employees.
- Library staff who provided RDS fall within many different departments and job descriptions. Some institutions have dedicated RDS staff while many appear to expect library staff (such as liaison librarians) to preform RDS in addition to their other duties.
- Half of the libraries surveyed said they had a committee or group dedicated to RDS.

Survey data analysis and availability

The survey and its results are organized into five blocs:

1. **Demographics bloc.** This section asks respondents to provide their contact information and demographic information about the size of their institution. Contact information was gathered only for quality control purposes and will not be made public.
2. **Research Data Services bloc.** This section asks respondents to provide information on their institution's RDM teaching activities and services offered, as well as information on other campus groups that offer RDM services.
3. **Data Repository: General bloc.** This section asks respondents questions related to the scope and governance of their institution's data repository. Respondents who indicated that their institution did not have a data repository skipped this and the following section.
4. **Data Repository: Details bloc.** This section asked respondents to provide more technical details about their data repository's operations, costs, and metadata capabilities.
5. **Library Organization bloc** This section addressed library staffing for Research Data Services.

The full list of survey questions can be found in our OSF repository.

Distribution and responses

The survey was distributed in March and April of 2018. Invitations to participate were sent to the deans/library heads of the 38 GWLA institutions. Thirty-six of the thirty-eight institutions responded to the survey. However, two of the institutions only provided partial responses, one of which had to be discarded as only 12% of the survey was completed. The other partial response was complete enough to include in the majority of the analysis.

Data analysis

Analysis of the survey answers was only performed at the question level. This was done for two primary reasons. First, the survey was not designed for research or to preserve institutional or individual anonymity. The task was to find out what 38 institutions were doing and how they were doing it. Second, fact checking of major outliers (e.g. a reporting of 5000 library staff) indicated that self-reported statistics were found to be inaccurate and had to be discarded which prevented cross-analysis by library and parent institution demographics. The survey also did not yield sufficient data for any meaningful analysis for annual software and storage costs. Even with these restrictions, the authors believe that the data presented in this report are useful to other academic libraries who are exploring or building up research data services.

Data availability

As the survey was not anonymous the authors have decided that access to raw, and most coded data, should be restricted to GWLA institution members. Only answers to questions which contain public information have been shared. Links to these data sets can be found within the text.

Results

This section provides a detailed summary of the survey answers and offers no interpretation or conclusions upon the results. Only positive and negative answers were counted, blank responses were discarded. All percentages are rounded to one decimal place. Where indicated 'n' is equal to the number of institutions who answered the question and/or the number in the subset the question applied to. The results presented here can be used by GWLA members and other academic libraries as a baseline snapshot of the RDS offered by U.S. academic libraries at the time of the survey deployment.

Demographics bloc

This section asked for contact information and library and parent institution employment numbers. Results from the demographics section show that GWLA members vary widely in size and staffing levels. However, the data gathered from survey respondents in this section contained multiple major outliers (e.g. 5000 FTE library employees or 38 FTE university faculty) which prevented analysis by institution or library size. An overview of the

GWLA member institutions is available in Table 1. This data was not obtained through the survey and should only be used to gain a "ballpark" prespective of the membership.

Research Data Services bloc

This section asked respondents to provide information on their institution's RDM teaching activities and services offered, as well as information on campus groups external to the library that offer RDM services.

Which Research Data Instruction services are offered at your institution? Table 2

Table 2.

The number and percentage of libraries that provide various types of research data instruction. (n=34).

	# institutions	% institutions
Course-related instruction	34	100.0%
Workshops	34	100.0%
Consultations	33	97.0%

Please supply titles or topics for the research data workshops your institution provides.

Answers from thirty-two institutions were analyzed (n=32). Workshop titles/topics were assigned up to 2 topic codes based on the information provided (Fig. 1). The most workshops/topics provided by an institution = 15 (one institution), the least workshops/ topics provided by an institution = 1 (five institutions), and the average number of workshop/topics = 5. Workshops/topics devoted to specific tools or programming languages were coded and tracked separately from topic codes (Fig. 2). The coded data, codebook, and analysis for this question is available through Open Science Framework, see files labeled "Q08" in Murray et al. (2019)

Which of the following research data metadata services does your library provide? (see Table 3)

Table 3.

The number and percentage of libraries that provide different types of RDS metadata services (n=35).

Metadata services	# institutions	% institutions
Help researchers understand metadata and standards to describe their research data	24	68.6%
Help researchers identify appropriate metadata standards	23	65.7%
Help researchers apply metadata standards	17	48.6%

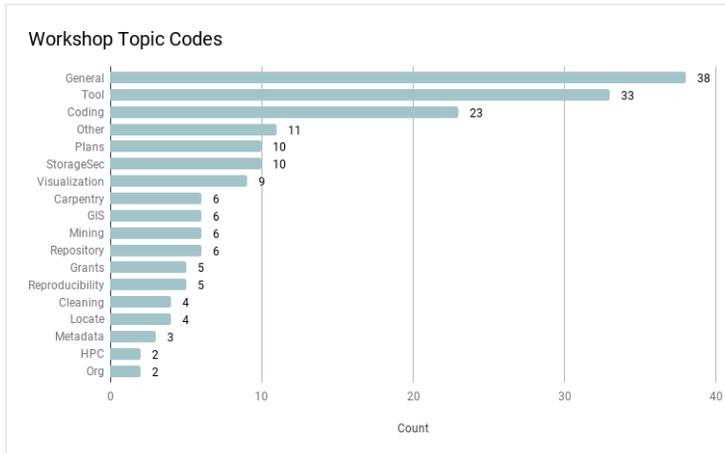| Metadata services	# institutions	% institutions
Other	9	25.7%
None	5	14.3%

Figure 1. doi

Workshop topic code frequencies. Up to two topic codes were applied to each workshop (n=160). Topic codes are defined as follows: *Carpentry*: a data or software Carpentry workshop; *Cleaning*: data cleaning and related techniques; *Coding*: how to work with data via command line or in a specific language; *General*: the basics of data management; *GIS*: geographic information system or spatial data/tools; *Grants*: the word "grants" or the name of a funding agency was explicitly mentioned in the workshop's title or description; *HPC*: high performance computing; *Locate*: focused on how to search and locate datasets; *Metadata*: metadata and data documentation; *Mining*: focused on text and data mining; *Org*: data organization; *Other*: misc. topics or unclassifiable; *Plans*: data management plans; *Repository*: addresses a specific repository, how to use a repository, or data repositories in general; *Reproducibility*: focused on research reproducibility; *StorageSec*: data storage and/or security tools and topics; *Tool*: focused on how to use tools related to data and data management (see Fig. 2); *Visualization*: data visualization.

Please identify which, if any, research data services are currently provided by other groups on your campus. Table 4

Table 4.

The number and percentage of institutions that have RDS services offered by groups external to the library (n=35).

Service	# institutions	% institutions
Statistical software support	23	65.7%
Data analysis support	23	65.7%
Active research data storage and backup solutions	22	62.8%
GIS and geospatial analysis	21	60.0%

Service	# institutions	% institutions
Data visualization support	20	57.1%
Assistance locating data storage and backup solutions	20	57.1%
Dataset purchase, acquisitions, subscriptions	14	40.0%
Database design and management	13	37.1%
Data mining	12	34.3%
Templates or boilerplate for Data Management Plans (DMPs), such as DMPTool...	12	34.3%
Assistance with completing Data Management Plans	11	31.4%
Locating and using existing data (includes identifying and suggesting repositories)	10	28.6%
Topic or How-To Guides	8	22.9%
File organization and naming conventions	7	20.0%
Other (with free-text entry)	3	8.6%
Metadata assistance	2	5.7%
None (i.e. none of these services are offered)	2	5.7%

Figure 2. doi

Breakdown of the workshops or topics with a tool or programming language code applied (n=47). Only tool codes that have a frequency >1 are shown. Tool code names are self-explanatory (i.e. the name of tool).

If Research Data Services are currently provided by other groups on your campus institution please identify the groups offering the services.

Groups were assigned a type code based off the names and descriptions provided (Fig. 3). Groups that could be identified as belonging to a specific discipline were coded and tracked separately from the type codes (Fig. 4). The coded data, codebook, and analysis for this question is available through Open Science Framework, see files labeled "Q11" in Murray et al. (2019).
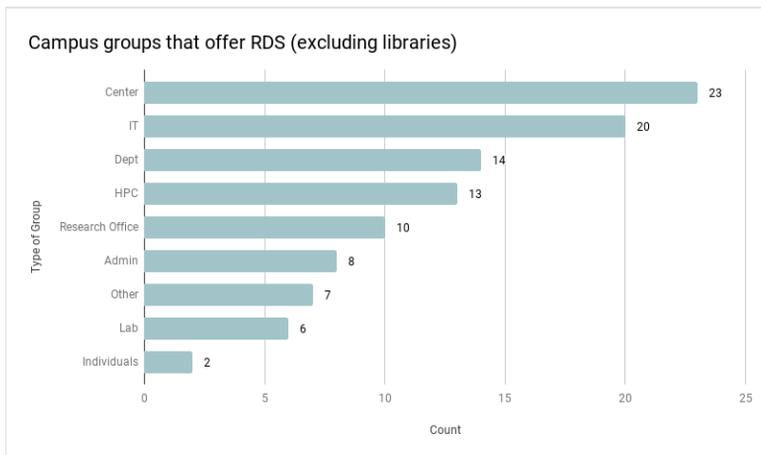
Figure 3. doi

Types of campus groups that provide RDS (n=103). Type codes are defined as follows: *Admin*: a campus administrative unit that does not fall into any other category; *Center*: research centers or institutes excluding HPC groups; *Dept* = Departments or colleges; *HPC*: High Performance Computing and research computing units including HPC run by IT units; *Individuals*: Individual staff, faculty, students, etc.; *IT*: Information Technology associated with the entire campus, colleges, or departments excluding HPC groups; *Lab*: Various labs on campus that do not fall into any other category; *Research Office*: Groups that oversee university research; *Other*: Groups that cannot be categorized under any other code.
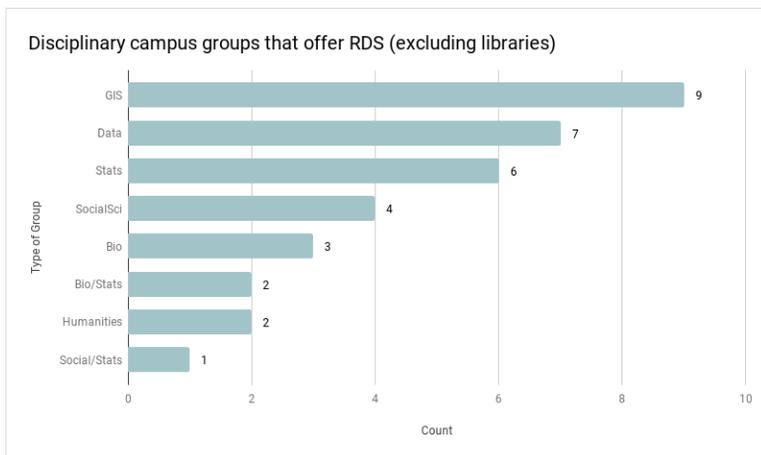
Figure 4. doi

Disciplinary categorization of campus groups that provide RDS (n=34). Discipline codes are defined as follows: *Bio*: Groups that specialize in biology, including health and medicine; *Bio/Stats*: Groups that specialize in biology and statistics; *Data*: no specific discipline but has the word 'data' in the name; *GIS*: Groups that specialize in spatial and GIS (Geographic Information Systems) data; *Humanities*: Groups specializing in humanities; *Social/Stats*: Groups that specialize in statistics and social science; *SocialSci*: Groups specializing in social science; *Stats*: Groups specializing in statistics.

Library and institutional research data policies Table 5

	# institutions	% institutions
Table 5. The number and percentage of libraries and institutions (university/college) that currently have policy(ies) that address research data (n=35).		
Libraries that currently have policy(ies) that address research data	7	20.0%
Institutions that currently have policy(ies) that address research data	20	57.1%

Follow up questions asked for links to library and institutional policies if they were public. These are presented in Suppl. material 1.

Does your library's or institution's strategic plan or mission address research data services? (n=34)

- Yes, 61.8% (21)
- No, 38.2% (13)

Data Repository: General bloc

This section asks respondents questions related to the scope and governance of their institution's data repository.Thirty institutions responded to all the questions in this bloc, while 5 institutions (14.3%) indicated that they did not have a repository that accepted data and skipped to the last question bloc (Library Organization) .

Do you have a research data repository or a repository that accepts research data? (n=35)

- Yes, accepts research data = 30 institutions / 85.7%
- No repository or repository does not accept data = 5 institutions / 14.3%

Does your institution have a dedicated repository for research data or is the same platform used for both data and other scholarly materials? (n=30)

- Dedicated data repository = 30.0%
- No, same platform = 70.0%

Do your institutional repository and data repository system share staff? (n=30*)

- Yes, share staff = 30.0%
- No, do not share staff = 70.0%

* This question was accidentally given to all survey respondents, regardless of if they had a separate data repository or not.

Which of the following are addressed in the policies and/or information pages for the data repository? Table 6

Table 6.

The number and percentages of institutions that cover various use case topics in their data repository policies or information pages (n=30).

Policy/information topic	# institutions	% institutions
Who can deposit	26	86.7%
File types / file formats	22	73.3%
Sensitive data	18	60.0%
File size limits	16	53.3%
Volume or number of file limits	9	30.0%
Retention periods	9	30.0%
Required files or documentation	8	26.7%
Other criteria	7	23.3%
None (i.e. no policies)	2	6.7%

Which of the following are addressed in your data repository's policies and/or information pages? Table 7

Table 7.

The number and percent of institutions which include legal documents in their data repository policies or information pages (n=30).

Legal topic	# institutions	% institutions
Terms of Deposit	17	56.7%
None	7	23.3%
Other	6	20.0%
End User Terms of Agreement	1	6.0%

Which "stages" of data does your data repository accept? Table 8

Table 8.

The number and percent of institutions that accept data in different lifecycle stages (n=30).

Data stage	# institutions	% institutions
"Live" / "active" / "raw" data	9	30.0%
"Final data"	27	90.0%
"Published data"	27	90.0%

Are embargo periods available for deposited data? (n=30)

- Yes = 80.0%
- No = 20.0%

Does your institution limit how long data may be embargoed? Table 9

Table 9.

The number and percent of institutions that allow various embargo lengths. Only institutions which have a repository that accepts data and embargo periods answered this question (n=24).

Embargo periods	# institutions	% institutions
0-6 months	0	0.0%
7-12 months	1	4.2%
13-24 months	3	12.5%
More than 24 months	2	8.3%
No limit	18	75.0%

What is your library's preservation strategy for the data in the repository? Table 10

Table 10.

The number and percent of institutions with preservation strategies. (n=30). The number and percent of institutions with each level of preservation practice.

None: We placed Libraries in this category if they indicated that they had no strategy or their strategy was under development

Low: We placed Libraries in this category if they indicated that they backed up data in some way, but were taking no other active preservation measures to ensure the ongoing viability of the data. Example responses include "respository content is backed up and check sums are run nightly"

High: We placed Libraries in this category if they indicated that they placed their data into a preservation system such as the now defunct DPN or if they provided a description of processes to verify file and format integrity.

Level of Preservation	# institutions	% institutions
High - Strategies that exceed "backing up" content	12	40.0%
Low - Strategy of "backing up" content	9	30.0%
None	8	26.7%
N/A	1	3.3%

Data Repository: Details bloc

This section asked respondents to provide more technical details about their data repository's operations, costs, and metadata capabilities. Only respondents that indicated

that their institution has a repository that accepts data answered this bloc. For this section n=29 as one of the respondents did not complete this section of the survey.

What deposit model is used for research data? Table 11

Table 11.

The number and percentage of institutions that use different deposit models for their repositories. Some institutions selected more than one answer for this question (n=29).

Deposit model	# institutions	% institutions
Mediated	19	65.5%
Mediated only	14	48.3%
Unmediated	13	44.8%
Unmediated only	8	27.6%
Both	5	17.2%
Other	3	10.3%

For the purpose of this question, *mediated* was defined as "subject to review/changes and approval" and *unmediated* was defined as "no review or approval needed." The "other" choice was a free-text box. Two of the free-text answers indicated that there were plans to move to a mediated deposit model and one provided details on a hybrid model.

Which licenses are available? Choose all that apply. Table 12

Table 12.

The number and percentage of institutions that reported using various licenses for data in their repositories (n=29).

Licenses	# institutions	% institutions
CC-0	23	79.3%
CC-BY	17	58.6%
Other	17	58.6%
(c) All Rights Reserved	8	27.6%
GNU General Public License (GPL)	4	13.8%
MIT	3	10.3%
Apache	2	6.90%
Mozilla Public License 2.0	2	6.90%

Does your repository assign persistent identifiers to datasets? (n=29)

- Yes = 23 (79.3%)
- No = 6 (20.7%)

What type of software does your data repository use? (n=29)

• Software as a Service (SaaS), aka cloud-based software: 14 (48.3%)

• Self-hosted, aka local installation: 15 (51.7%)

What is the name of the platform (software) that the data repository runs on? Table 13

Table 13.

The number and percentage of institutions that use specific software platforms for their data repositories. Twenty-nine institutions responded to this question however, two institutions reported two different software platforms so n=31 for the percentages calculated in this table.

Name of Software	# institutions	% institutions
DSpace	10	32.3%
Digital Commons	9	29.0%
Dataverse	5	16.1%
Samvera	3	9.7%
Islandora	2	6.5%
Figshare	1	3.2%
Home grown	1	3.2%

Two institutions reported using two platforms for their repositories. So, while the number of institutions responding to this question is 29, the number of repositories is 31.

Please provide the names of metadata schema(s) used to describe data the repository. Table 14

Table 14.

The number of institutions reporting the use of metadata schemas or vocabularies in their data repositories (n=29). Six of the 29 institutions that answered this question reported using more than one metadata schema. For this reason only frequencies are reported inTable 14

Metadata Schema	# institutions
Dublin Core	22
DataCite	8
DDI	6
ISA-Tab Specifications	6
ISO 639-1	6
ISO 3116-1	6
NCBI Taxonomy	6
OBI Ontology	6

Metadata Schema	# institutions
Virtual Observatory	6
EML	4
Qualified Dublin Core	4
Custom	2
N/A	2
Domain	1
FGDC	1
ISO 19115	1
METS	1
PREMIS	1

Table 14

Do library staff create or assist researchers with the creation of record metadata? (n=29)

For the purpose of this question: *Record metadata* is metadata that is searchable and harvest-able.

- Yes = 26 (89.7%)
- No = 3 (10.4%)

Do library staff create or assist researchers with the creation of documentation metadata? (n=29)

For the purpose of this question: *Documentation metadata* is metadata that exists to help others comprehend and reuse the data, such as a readme file.

- Yes = 18 (62.1%)
- No = 11 (37.9%)

Library Organization bloc

This section addresses library staffing for Research Data Services and the job titles of library staff who provide RDS.

How many staff at your library provide research data services? Table 15

Table 15.

The number of institutions reporting the number of library employees that provide RDS (n=34).

# of Employees	0	0.5	1	2	3	4	4.5	5	6	7	8	15
# of Institutions	1	1	4	7	8	3	1	3	3	1	1	1

Answers from thirty-four institutions were analyzed (n=34). The most frequent number of staff providing RDS was 3 (reported eight times) while the average was 3.58. The highest number of staff was reported at 15 and the lowest was 0.5 for those institutions with staff that provide RDS.

Please provide an estimate of the combined Full Time Employee (FTE) dedicated to research data services in your library, accounting for the time of all staff involved.
Table 16

Table 16.

The number of institutions reporting on dedicated full time employees (FTE) dedicated to research data services (n=34).

# of FTE	0.00	0.10	0.25	0.50	0.60	0.70	0.96	1.15	1.20	1.50	1.80	2.00	2.50	3.00	3.50
# of Institutions	2	2	4	2	1	1	1	1	1	7	1	6	2	2	1

Answers from thirty-four institutions were analyzed (n=34). The most frequent amount of combined FTE dedicated to RDS was 1.50 (reported by 7 times) while the average was 1.34. The highest amount of combined FTE was reported at 3.5 while the low was 0.10 for the institutions providing RDS.

Titles, positions, and departments of RDS staff.

This data was analyzed two different ways. Table 17 shows a word count analysis of the free text answers provided by the respondents while Table 18 analyzed the same data by assigning each position up to three codes.

Table 17.

The frequency of words found in job titles from 33 libraries for staff who provide RDS (n=104). The words "and", "of", and "library" were excluded from the analysis. Only words that appeared five or more times are included in this table.

Title word	Frequency
librarian	56
data	30
services	24
digital	17
research	15
science	11
specialist	10
coordinator	10
scholarly	8
head	8

Title word	Frequency
sciences	7
engineering	7
subject	6
management	6
librarians	6
metadata	5
liaison	5
gis	5
director	5

Table 18.

Coded analysis of the job titles of library staff who provide RDS (n=104). Topic codes are defined as follows: *Data* = included "data" in title; *Digital Collections* = included "digital collections" in title; *Digital Other* = "digital" in title but did not include "collections", "research", or "scholarship"; *Digital Research* = included "digital research" in title; *Digital Scholarship* "digital scholarship" in title; *Engineering* = included "engineering" in title; *GIS* = included "GIS" or "geospatial" in title; *Government Documents* = included "government" in title; *IR* = included "IR" or name of repository in title; *IT* = included "IT" in title; *Medical/Health Sciences* = included "medical" or "health" in title; *Metadata* = included "metadata" in title; *Other* = miscellaneous titles that did not fit into other categories; *Research* = included "research" in title; *Scholarly Communications* = included "scholarly communication" or "scholarly publishing" in title; *Science* = included "science" but excluded "health" or "social" in title; *Subject/Liaison Librarian* = included "subject", "liaison", or a discipline (e.g. "social science") in title; *Visualization* = included "visualization" in title.

Code	Frequency
Data	29
Subject/Liaison Librarian	26
Research	12
Science	12
Other	9
Scholarly Communications	9
Digital Other	8
Engineering	7
GIS	7
Metadata	5
Digital Scholarship	4
Government Documents	4
Digital Collections	3
IR	3
Medical/Health Sciences	3

Code	Frequency
Digital Research	2
IT	2
Visualization	2

Is there a Library Committee or Group dedicated to research data services? (n=34)

- Yes: 17 (50.0%)
- No: 17 (50.0%)

Conclusion and Future Directions

The authors believe that this report provides a baseline that other institutions can use to compare and measure the research data services they provide at their institutions. Since the survey was limited to the GWLA Libraries, future work could employ similar surveys to capture research data services in other academic libraries in order to gain a fuller understanding of the landscape.

This report, as a snapshot in time, could also be used as a marker for the development of research data services in academic libraries in North America. The GWLA member libraries could be surveyed again in a few years to determine changes in practices regarding research data management in this population of academic libraries.

This report and survey tool have limitations that should be corrected in future work. Some questions were unintentionally ambiguously worded and resulted in data that was difficult or impossible to analyze. Improved testing of the survey by a wider audience and soliciting the services of survey designer would improve question response and corresponding data quality. That being said, we believe it is important to understand the current state of research data services in order to monitor activity and measure future progress.

Glossary

Active storage: Fast access to storage space, usually used during the active portion of the data life cycle. Example devices/services: external drive, NAS, Cloud DropBox, OneDrive

Amazon S3: Amazon Simple Storage Solution. One of the services provided by Amazon Web Services (AWS). (https://aws.amazon.com/s3)

Arkivum: Cloud service and software that offers long term data management and digital preservation. (https://arkivum.com)

AWS: Amazon Web Services Cloud computing, storage, and other cyber-infrastructure solutions offered by Amazon.

Digital Commons: Cloud hosted solution for publishing, management and showcasing of researchers scholarly output. (https://www.bepress.com/products/digital-commons)

Digital Preservation Network (DPN): A now defunct service for preserving research outcomes.

DSpace: An open-source software system for creating and hosting an institutional digital repository. (https://duraspace.org/dspace)

Dublin Core: A metadata standard used by libraries consisting of a small set of vocabulary terms that can be used to describe digital and physical resources

DuraCloud: An open source, hosted service that makes it easy to control where and how your organization preserves content in the cloud. (https://duraspace.org/duracloud/)

FTE: Full Time Equivalent. A unit that indicates the workload of an employee and used in this survey to measure a worker's involvement in RDS activities. An FTE of 1.0 is equivalent to a full-time worker (i.e. 40 hr/week). Paraphrased from Wikipedia.

GWLA: Greater Western Library Association. (https://www.gwla.org)

Institution: GWLA member. (https://www.gwla.org/about-gwla/members)

LOCKSS: "Lots of Copies Keep Stuff Safe" Program hosted by the Stanford Libraries promoting best practices for digital preservation. (https://www.lockss.org)

Passive storage: Slow access to space which may require long waits for reading and writing. Used mainly for end of project storage of digital content or for recovery of catastrophic data loss. Example devices/services: tape media, AWS Glacier.

RDM: Research Data Management.

RDS: Research Data Services.

Rosetta: End-to-end digital asset management and preservation solution for libraries, archives, museums and other institutions from ExLibris (https://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation).

Acknowledgements

Ethics and security

Survey respondents were informed that results of the survey would be published and made publicly available. This project and survey did not meet the definition of human subject research. As such, it was not subject to Institutional Review Board and oversight.

References

- Murray M, Novak J, Thompson S, Rozum B, O'Donnell M, Laufersweiler M (2019) GWLA DMTF Project 2: Skill/RDS Program Levels: Public Data and Codebooks. osf.io/bpurj. Accessed on: 2019-8-27.
- U.S. Department of Education (2017) Integrated Postsecondary Education Data System (IPEDS). Land Grant Institution; Carnegie Classification 2015: Basic. National Center for Education Statistics. Release date: 2019-8-29. URL: https://nces.ed.gov/ipeds/datacenter/InstitutionList.aspx

Supplementary material

Suppl. material 1: Appendix A: Policy links `doi`

Authors: Megan O'Donnell
Data type: appendix
Brief description: Links to library and university/college research data management policies.
Download file (68.62 kb)

Endnotes

[*1] At the time of the survey there were 38 member institutions.
[*2] Greater Western Library Alliance, https://www.gwla.org/ and https://www.gwla.org/about-gwla