

Institutional Research Data Management Services Capacity Survey INSIGHTS Report #2

Current Capacity within Institutions: Highly Qualified Personnel, and Infrastructure and Services

Prepared by the Portage Network, Research Intelligence Expert Group (RIEG) on behalf
of the Canadian Association of Research Libraries (CARL)

FEBRURY 2021

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage

SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF
RESEARCH LIBRARIES ASSOCIATION DES BIBLIOTHÈQUES
DE RECHERCHE DU CANADA

Table of Contents

- Introduction..... 2
 - Respondent Demographics3
- Highly Qualified Personnel..... 4
 - HQP in the Plan Stage of Research Data Lifecycle 6
 - HQP in the Create, Process & Analyze Stages of Research Data Lifecycle7
 - HQP in the Disseminate & Reuse Stages of Research Data Lifecycle7
 - HQP in the Preserve Stage of Research Data Lifecycle8
- RDM Infrastructure..... 8
 - Data Storage Infrastructure8
 - Active Storage8
 - Data Repositories..... 9
 - Long Term Preservation9
 - Specialized Data Infrastructure..... 11
 - Data Transfer 11
 - Sensitive Data12
 - High Performance Computing12
 - Research Software12
- RDM Services16
 - Breakdown of Service Providers 17
 - Organizational Structure and Service Delivery19
 - HQP and Service Delivery20
 - Budgets and Service Delivery 22

Introduction

The Research Intelligence Expert Group (RIEG) is a pan-Canadian working group of the Portage Network that gathers evidence to guide the development of best practices in Research Data Management (RDM) in Canada, and communicates with stakeholder communities about existing and arising issues in related policies and practices.¹

Recently RIEG administered a bilingual questionnaire that surveyed the current state of Canadian research institutions in developing infrastructure and allocating human, organizational, and fiscal resources for RDM on their campuses. The survey also solicited suggestions for additional support that the Portage Network and other stakeholders could provide to assist these efforts. The goal of the survey was to assess institutional capacity to support RDM in response to the draft Tri-Agency RDM policy. The online survey was administered from September 3 to October 18, 2019 using SimpleSurvey software. It was distributed through several RDM-focused Canadian listservs and a contact list of identified institutional stakeholders. Each institution was asked to gather information from across campus and provide a single coordinated response, although multiple responses from a single institution were also accepted. The survey instrument consists of 27 questions ranging from general demographic information to detailed questions about current infrastructure and services in place across institutional stakeholder groups.² We received 77 responses from various institutions (mainly universities and colleges) across Canada.

An Executive Summary of the survey results was released in January 2020³ followed by the first Insights report in June 2020⁴ with subsequent Insights reports taking a deeper dive into key topics. In particular, this document explores the following topics:

- Highly qualified personnel (HQP)
- Infrastructure
- Services

Figures shown summarize findings across respondents from all institutional categories. Since a significant split in respondents were from universities and colleges (see Table 1), specific differences between these two categories are also highlighted in the summary text.

¹ <https://portagenetwork.ca/network-of-experts/research-intelligence-expert-group/>

² The survey questions can be found in the Data Dictionary in the Appendix of the Institutional research data management services capacity survey: Executive summary by Cooper, A., Perry, C., Szwajcer, A., Wang, M., & Khair, S. (2020), Portage Network, <https://doi.org/10.14288/1.0388722>.

³ <http://hdl.handle.net/2429/73607>.

⁴ Abel, Jennifer, Cooper, Alexandra, Dearborn, Dylanne, Perry, Carol, Szwajcer, Andrea, & Wang, Minglu. (2020, June 24). Institutional Research Data Management Services Capacity Survey INSIGHTS Report #1. RDM Support within Organizations: Budget, Structure, and Strategies. Zenodo. <http://doi.org/10.5281/zenodo.3906443>

Respondent Demographics

The survey received 85 responses from 77 institutions across Canada (Table 1) including universities, colleges/CÉGEPs⁵, research centres, and governments organizations. Eight institutions submitted two separate responses, which were combined into one response per institution⁶. The responses from the 77 institutions are summarized in this report.

Based on the institutions' names, we classified the responses into the institutional types of universities, colleges/CÉGEPs, research centres, and government organizations by geographical region:

- West (British Columbia, Alberta, Saskatchewan, Manitoba)
- Ontario
- Quebec
- Atlantic (New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland & Labrador)⁷

| | Atlantic | Quebec | Ontario | West | Total |
|----------------------|----------|--------|---------|------|-------|
| Types of institution | | | | | |
| University | 9 | 14 | 14 | 15 | 52 |
| College/CÉGEP | 0 | 9 | 6 | 6 | 21 |
| Research Centre | 0 | 0 | 0 | 1 | 1 |
| Government | 0 | 0 | 1 | 2 | 3 |
| Total | 9 | 23 | 21 | 24 | 77 |

Table 1. Number of institutional responses by region and institution types. Region and institution type are derived from Q1. "Name of institution."

⁵ The principles for the consolidation are 1) If the original values are consistent from the two respondents, the value for the combined case is the same as the original; 2) Text answers from the original two responses are combined into the value for the combined case; 3) If one of the original values is either "not chosen" or "don't know", the value for the combined case will be the same as the other original value which was either "chosen" or other affirmative answers, for example, "yes" or "no".

⁶ CEGEPs, or Colleges d'enseignement general et professionnel, are publicly funded, post-secondary, pre-university, collegiate technical colleges exclusive to the Quebec provincial educational system.

⁷ Institutional responses from provinces other than Ontario and Quebec were combined geographically into 'West' and 'Atlantic' to anonymize the relatively low number of responses.

Highly Qualified Personnel

HQP within institutions have specialized skills and expertise essential to supporting researchers in the area of RDM. Librarians and library-based functional specialists support a range of RDM expertise. This report focuses on RDM expertise and skill sets including:

- Knowledge of national policies
- Data security and risk management
- Legal (licensing, intellectual property, etc.)
- Advisory
- Technical aspects in the area of e-infrastructures
- Research software development
- Research data management
- Technical aspects of managing sensitive data
- Data analysis and visualization
- Data curation
- Metadata creation
- Data preservation

Figure 1 summarizes skill levels in these categories across respondent institutions. Most institutions reported not having a full complement of HQP in all areas of consideration. Specific shortfalls appear to exist in the skill areas of managing sensitive data, data curation, and research software development, data preservation, researchers' data management skills and technical aspects in e-infrastructure. Across all skill categories, institutions overwhelmingly indicated a need for more support. In this section, skill capacity is reviewed according to categories based on a data lifecycle model defined in the Portage RDM Primer document. It is important to acknowledge here many of these skills support data throughout its lifecycle. In this report we grouped some stages described in the data lifecycle model⁸ as follows:

- Plan
- Create, process, analyze
- Disseminate, reuse
- Preserve

⁸ Data lifecycle model as outlined in the Portage RDM Primer, <https://doi.org/10.5281/zenodo.4000999>

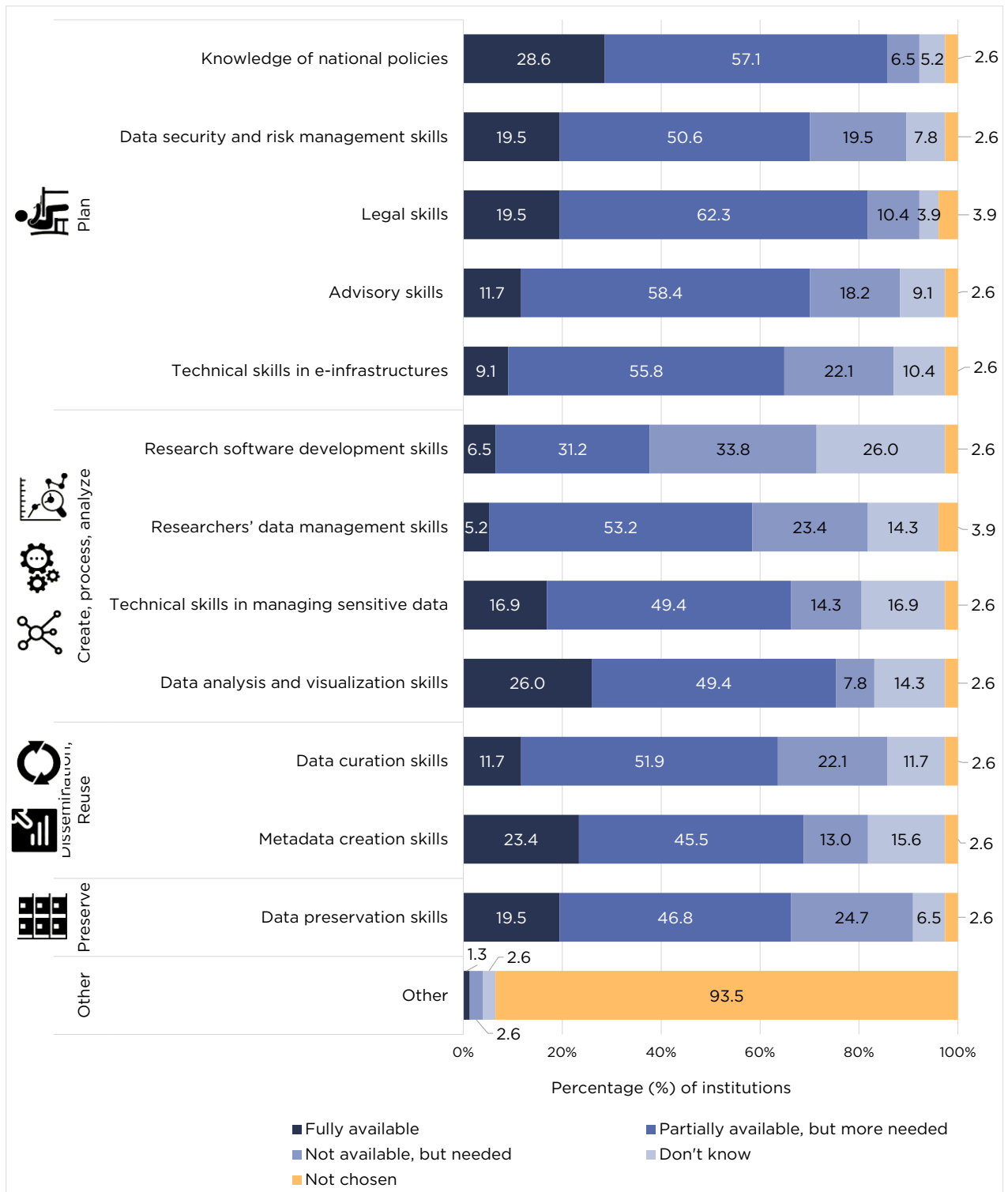


Figure 1. Level of skills at different stages of the research lifecycle. Summary of responses to across all institutions Q12A. "Different skills are needed to further support research data management. At what level are the following skills available at your institution? Check all that apply". (n=77)

HQP in the Plan Stage of Research Data Lifecycle

During the planning stage of the research data lifecycle, support skills could include: knowledge of national policies; legal skills related to intellectual property, licensing etc.; and advisory skills on technical, organizational and operations matters. Data security and risk management skills, as well as technical skills in the area of e-infrastructures, are also valuable when developing research data management plans early on during the research process.

Knowledge of national policies appears to be well developed in the 52 university respondents with 34.6% indicating that fully developed skills were available,⁹ 59.6% reporting knowledge is partially available but requiring further development, and 1.9% indicating that this skill was not available but needed. Only 9.5% of the 21 colleges reported fully available skill development related to national policies with another 52.4% partially developed and another 19% reporting not having these skills available but needing them.

Similarly, universities had a higher developed capacity for legal skills related to intellectual property and licencing with 21.2% reporting fully available skills, 67.3% reporting partially available skills with more development required, and 5.8% not available (but needed). Colleges reported a lower level of legal skill availability with 19% reporting fully available, 47.6% partially available and 23.8% reporting not available but needed legal skills.

Advisory skills on technical, organizational and operational matters is an area where few institutions reported full availability of skills (universities 9.6% and colleges 9.5%). Partially available skills were identified in universities (65.4%) and colleges (47.6%) with a significant number reporting no skill availability but needed (universities 15.4% and colleges 23.8%).

Fully available technical skills in e-infrastructure were limited at universities (17.3%) but were even less developed at the college level (9.5%). Both colleges (52.4%) and universities (51.9%) indicated that although partially available technical skills were available more work needed to be done with 11.5% of universities and 14.3% of colleges reporting that the skill was not available but needed.

Similar figures were reported by universities for data security and risk management skills (19.2% fully available, 55.8% partially available, 19.2% not available but needed). Colleges also reported low numbers for data security and risk management skills (14.3% fully available, 38.1% partially available, 23.8% not available but needed).

⁹ Fully available skills may indicate little or no need for further development. These were self-reported responses with no definition provided in the questionnaire.

HQP in the Create, Process & Analyze Stages of Research Data Lifecycle

During the active phase of the research data lifecycle (create, process, analyze) specific skills are helpful in order to support researchers. These include data analysis and visualization, technical skills in managing sensitive data, and research software development skills as well as the data management skills of researchers.

Most institutions reported fully available skills in most areas were limited with the exception of data analysis and visualization skills. Universities reported 21.2% fully available, 55.8% partially available skills and 7.7% not available but needed in this area. Colleges reported their strongest technical skills availability in the areas of data analysis and visualization (28.6% fully available, 38.1% partially available but more needed, and 9.5% not available but needed).

Only 7.7% of universities reported having fully available technical skills in managing sensitive data. Another 59.6% indicated that these skills were partially available, but more were needed and 23.1% indicating the support skills for managing sensitive data were not available at all (but needed). This skill was also low at colleges (9.5% fully available; 42.9% partially available, 23.8% no skills available but needed).

Fully available research software development skills were very limited (3.8% for universities; 9.5% for colleges), as were researchers' data management skills (1.9% for universities; 4.8% for colleges). Partially available research software skills were more widely available at universities (44.2%) but were still limited to only 4.8% of colleges, while partially available researchers' data management skills were more readily available at both institution types (59.6% for universities; 42.9% for colleges).

HQP in the Disseminate & Reuse Stages of Research Data Lifecycle

Skills in metadata creation and data curation support research data throughout its lifecycle, and are critically important to ensure that data is findable and accessible over the long term. Colleges and universities were limited in the full availability of metadata creation skills (25% for universities and 19% for colleges) and data curation skills (11.5% for universities and 4.8% for colleges); however, universities reported significantly higher partially available metadata creation (57.7%) and data curation skills (67.3%) while colleges were limited in their capacity (14.3% metadata creation and 14.3% data curation skills). Colleges were more likely not to have either of these skills available compared to universities. In colleges, 23.8% indicated metadata curation and 47.6% for data curation skills were not available (but needed) and universities reported 7.7% for metadata curation and 13.5% for data curation skills.

HQP in the Preserve Stage of Research Data Lifecycle

Preservation refers to the long-term storage of data post-project following the active phase of the research cycle. There are a small number of universities (17.3%) and colleges (14.3%) reporting that they have services fully available for data preservation. Universities appear to have more preservation skills partially available than colleges, with 55.8% of universities respondents indicating partially available and 22.2% of universities indicating no preservation skill availability. Colleges reported only 28.6% partially available skills in this area and 38.1% noted that they do not have preservation skills, but that these are needed.

RDM Infrastructure

A range of technical infrastructure supports the data management needs of researchers throughout the lifecycle of a project. Storage infrastructure is a crucial preoccupation, serving three distinct purposes: active research, repository, and long-term preservation. Researchers also may rely on a range of specialized infrastructure to manage their data. Examples include infrastructure for high performance computing, large data transfer, to infrastructure for safely working with sensitive data with high-security risks. Research software supporting RDM is also included as infrastructure, supporting researchers with collaboration, data collection, and analysis.

Data Storage Infrastructure

Infrastructure capacity includes the ability of the institution to provide active storage, data repositories storage and long-term storage to support data preservation. More specifically, active storage corresponds to storage supporting the create, process, and analyze phases of the data lifecycle; repository storage supports the dissemination and reuse phases of the data lifecycle; and long-term storage supports data retention and preservation post-project completion.¹⁰

Active Storage

Active data storage (see Figure 2) can be provided to researchers directly by institutions (via individual departments and faculties, or through centralized institutional services), through infrastructure supported at a national or regional level via consortia, or through access to commercial infrastructure. Researchers within an institution may also be responsible for supporting their own infrastructure for their labs or projects.

¹⁰ Long-term storage was not defined within the survey and is open to interpretation by respondents. It could include 'dark archives' or other storage options with limited access.

Universities provide a greater range of support for active data infrastructure than colleges. At the institutional level, 46.2% of universities provide support within departments or faculties and 50% provide centralized support compared to colleges at 23.8% for departmental or faculties and 28.6% centralized support. At the regional/consortial level, 44.2% of universities reported infrastructure support and only 14.3% of colleges reported this sort of support. Over half of the universities (57.1%) reported that researchers take on infrastructure support themselves while at colleges, 4.8% of researchers provide their own infrastructure support.

Data Repositories

Data repositories are online collections of deposited data (and related materials) that can facilitate discovery and appropriate access. Data repositories can be institutional, disciplinary, multidisciplinary, regional or consortial, national, or offered through a journal or publisher.

Figure 3 shows respondent access to data repository infrastructure. Looking further into respondent differences, universities are providing more infrastructure support for data repositories than colleges. Almost half of the colleges (47.6%) reported not knowing about infrastructure support for data repositories, while only 15.4% of universities reported not knowing. Universities identified high use of disciplinary/multidisciplinary data repositories (71.2%). Universities also indicated substantial use of regional or consortial repositories (51.9%), journal repositories (51.9%) and institutional repositories (44.2%). National data repositories were reported being used at 25% and other repositories at 5.8%. At colleges, institutional repositories use was reported at 19%, with disciplinary/multidisciplinary (9.5%), regional or consortial (9.5%), and other repositories (4.8%) reported being used.

Long Term Preservation

Long term preservation is the process through which some data may be stored for the longer term post-project. Figure 4 details respondent access to long term preservation infrastructure. This type of infrastructure is less supported in both universities and colleges. These institutions reported at a higher level of not knowing about long-term preservation (36.5% of universities and 47.6% of colleges) compared to other options in the survey. Infrastructure support for long term preservation in departments or faculties (universities 11.5%; colleges 14.3%) had a lower response than institutional long term preservation support (universities 32.7%; colleges 28.6%). Colleges reported that very few researchers were using regional or consortial infrastructure (4.8%) compared to 26.9% of universities. Similar reporting is seen in researchers' use of national infrastructure for long-term preservation, where universities reported 21.2% were using this while colleges reported no known usage.

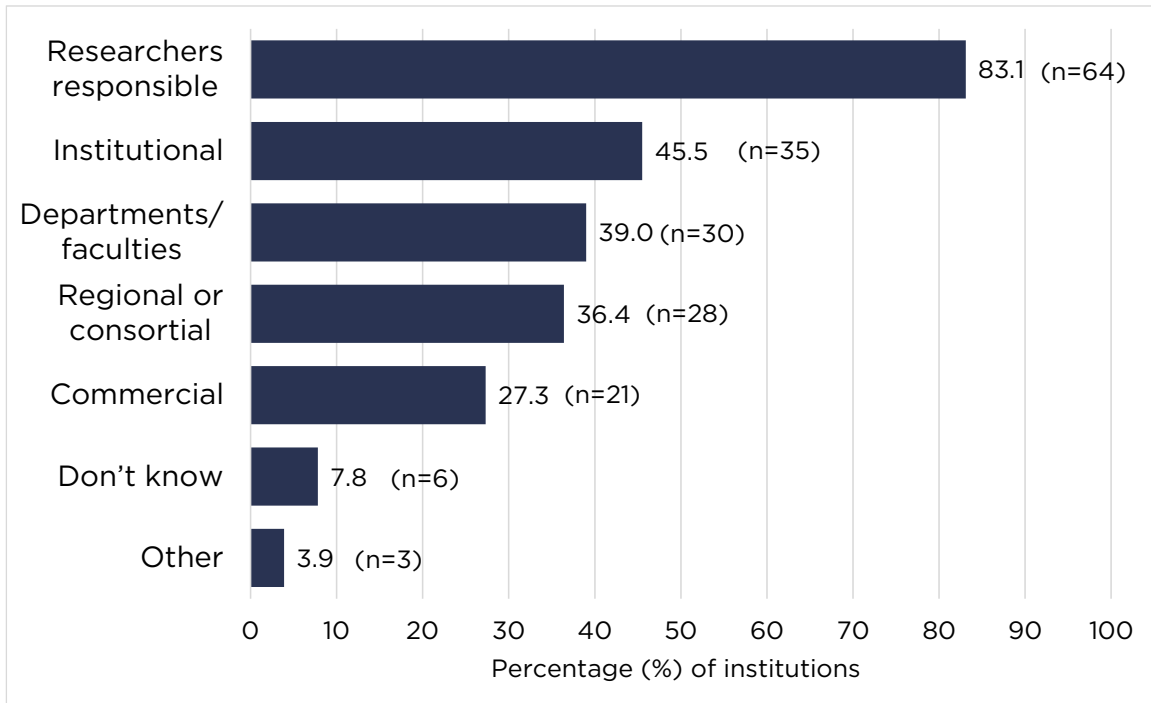


Figure 2. Active data storage services (all institutions). Q10A. “What research data management infrastructure does your institution have access to? Check all that apply. Active data storage services (storage used during the actual research process).” (n=77)

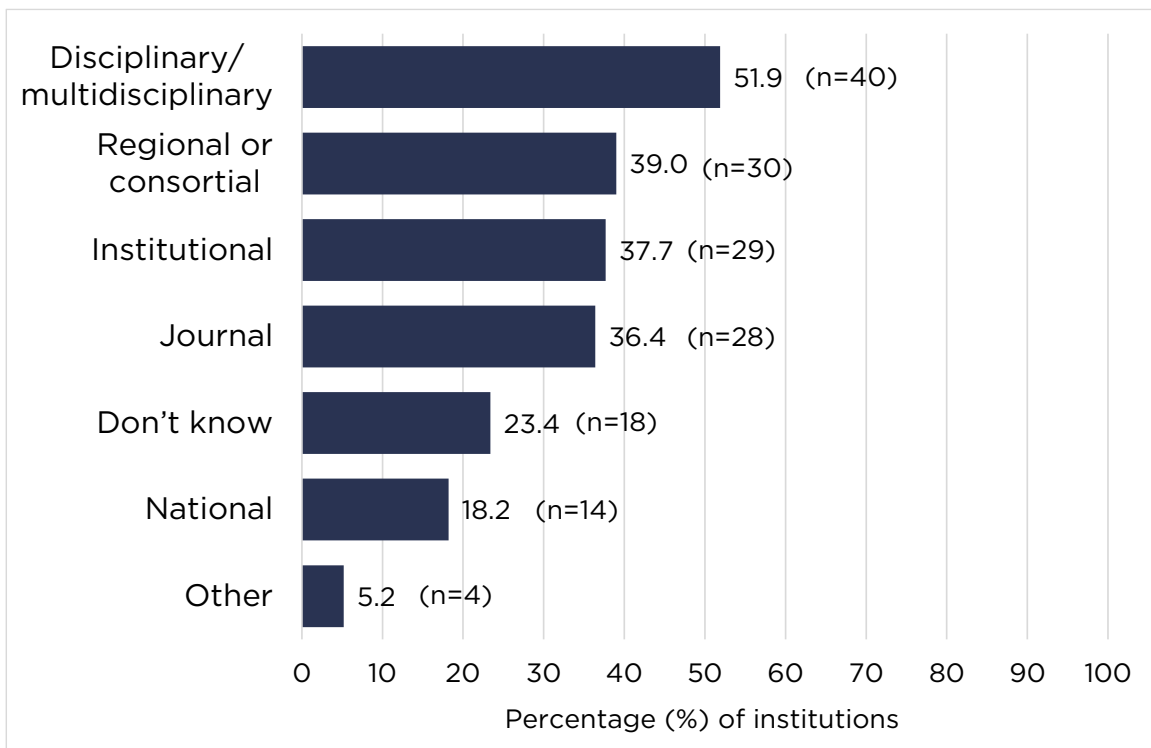


Figure 3. Data repository (all institutions). Q10B. “What research data management infrastructure does your institution have access to? Check all that apply. Data repository (where data are deposited/published for discovery and/or appropriate access).” (n=77).

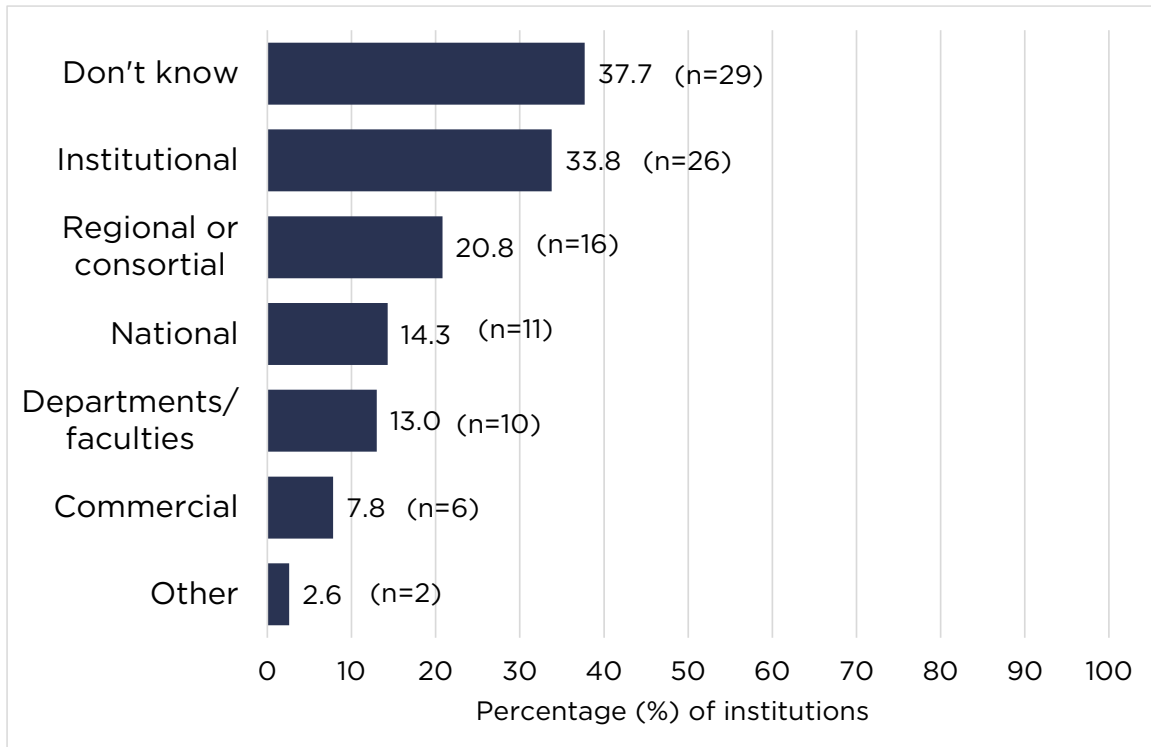


Figure 4 Long term data preservation (all institutions). Q10C. “What research data management infrastructure does your institution have access to? Check all that apply. Long term data preservation (the process through which some data may be stored for the longer term post project).” (n=77).

Specialized Data Infrastructure

Data Transfer

Data transfer refers to moving or replacing moving or replacing data from one location to another via a communication channel or network. Support for transfer of either small (<5GB) or large data sets (>5GB) offered by the institution was almost nonexistent in college respondents (19% for both small and large datasets). In universities, 28.8% provide institutional infrastructure for small data (<5GB) transfer and 21.2% provide institutional infrastructure for large data transfer. Figure 5 shows data transfer infrastructure across all respondents. All institutions reported low use of external infrastructure supports, with 1.3% indicating external support usage for small data transfer and 2.6% for large data transfer. It may be that more infrastructure does exist, as response rates indicated that respondents were unsure of what may be available for small (45.5%) and large data transfer (44.2%).

Sensitive Data

Infrastructure support for sensitive data both during the active phase of a project and post-project can be overwhelming. There are numerous requirements such as ethical, legal, and policy that researchers and institutions must navigate. Institutional infrastructure support for sensitive data was identified by 40.4% of universities, while 33.3% of colleges identified having institutional infrastructure support in place. Colleges identified no external infrastructure support for sensitive data, and only 5.8% of universities indicated external infrastructure support. Approximately a third of all respondents (32.5%) indicated their institutions had no known infrastructure support in place for sensitive data (see Figure 6). These levels of support should be viewed in light of the results discussed earlier regarding the low levels of fully available skills in managing sensitive data at both universities and colleges.

High Performance Computing

High-performance computing (HPC) refers to a class of infrastructure that performs computationally intensive operations across multiple computing resources. Of the 77 Institutions, 40.3% reported access to HPC infrastructure (see Figure 7). There was a notable difference in the breakdown of infrastructure available to universities and colleges, with 50% of university respondents indicating access to HPC infrastructure compared to only 14.3% of colleges. It should be noted that HPC infrastructure is offered nationally through Compute Canada¹¹ or other regional groups (e.g. Centre for Academic Computing). Access to these resources may not have been considered “institutional support” and therefore may not have been captured in the responses.

Research Software

Respondents were also asked about access to software tools through their institution, supporting a range of research needs (see Figure 7), including:

- Commercial quantitative research software (e.g. SPSS, Stata, etc.)
- Commercial qualitative research software (e.g. NVivo, Dedoose, etc.)
- Electronic data collection tools (for example, Qualtrics, RedCap, etc.)
- Research collaboration tools (e.g. institutional OSF, GitHub, etc.)
- Electronic lab notebooks (e.g. LabArchives, LabGuru, etc.)

Across all institutions, access to commercial quantitative software (64.9%) was most common and access to electronic lab notebooks (6.5%) was least expected, however there appeared to be notable differences in access between university and college respondents.

¹¹<https://www.computecanada.com>

University access to institutional support for qualitative (55.8%) and quantitative (75%) commercial software was fairly high but access was not as prevalent at colleges (qualitative commercial software at 28.6% and quantitative commercial software at 38.1%).

Institutional access to research collaboration tools was limited at universities (26.9%) and very limited at colleges (9.5%). There was some level of institutional access to data collection tools at universities (36.5%) and colleges (33.3%). University respondents also indicated some level of institutional access to research collaboration tools (26.9%), though college respondents reported lower institutional levels of access (9.5%). Electronic lab notebooks did not appear to have institutionally coordinated access, with no colleges and only 9.6% of universities reported that they support researcher access.

It should be noted that among all respondents, 16.9% indicated that they do not know whether there is institutional access to these supports. Also, some access to these types of infrastructure supports may be coordinated at the department or research group level rather than institutionally, which may not have been captured by this survey. Researchers may also be using these supports individually, rather than through institutional or departmental access coordination.

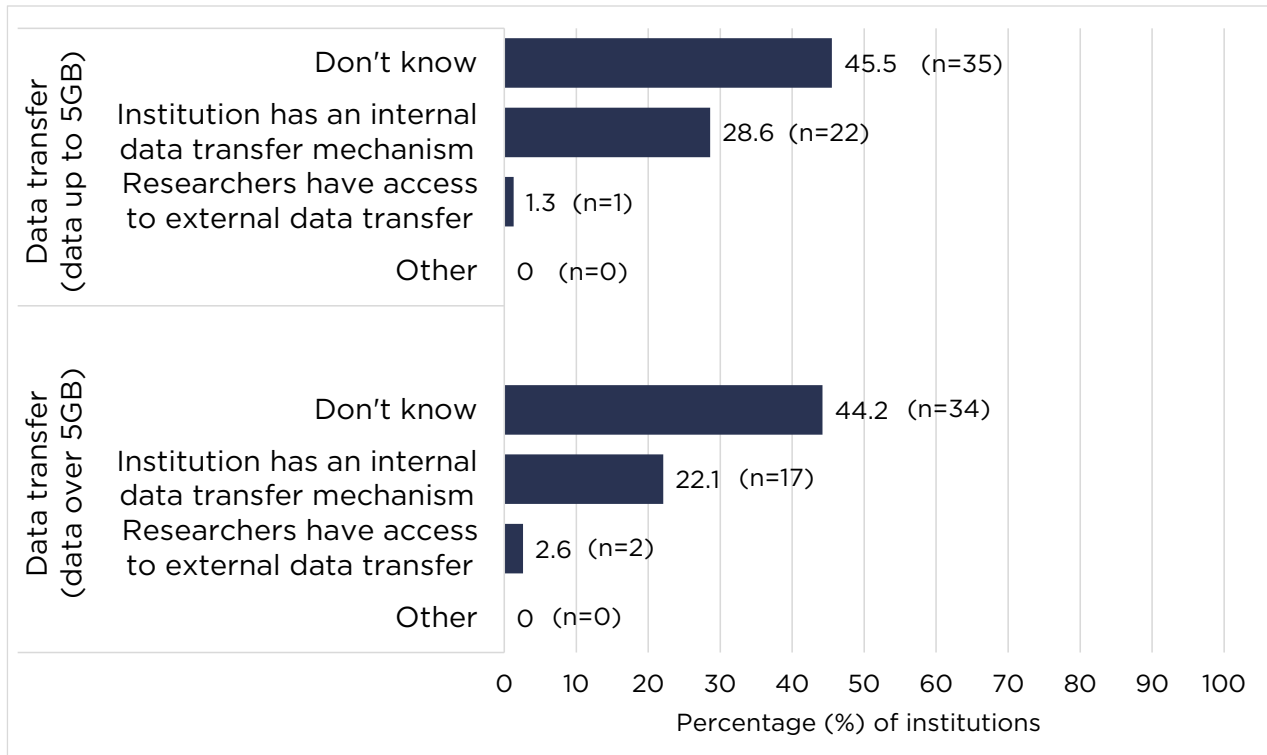


Figure 5. RDM infrastructure that institutions have access to for data transfer up to 5GB (Q10D) and over 5GB (Q10E) (all institutions). Q10D. “What research data management infrastructure does your institution have access to? Check all that apply. Data transfer (for data size up to 5 GB).” (n=77); and Q10E. “What research data management infrastructure does your institution have access to? Check all that apply. Data transfer (for data size over 5 GB).” (n=77).

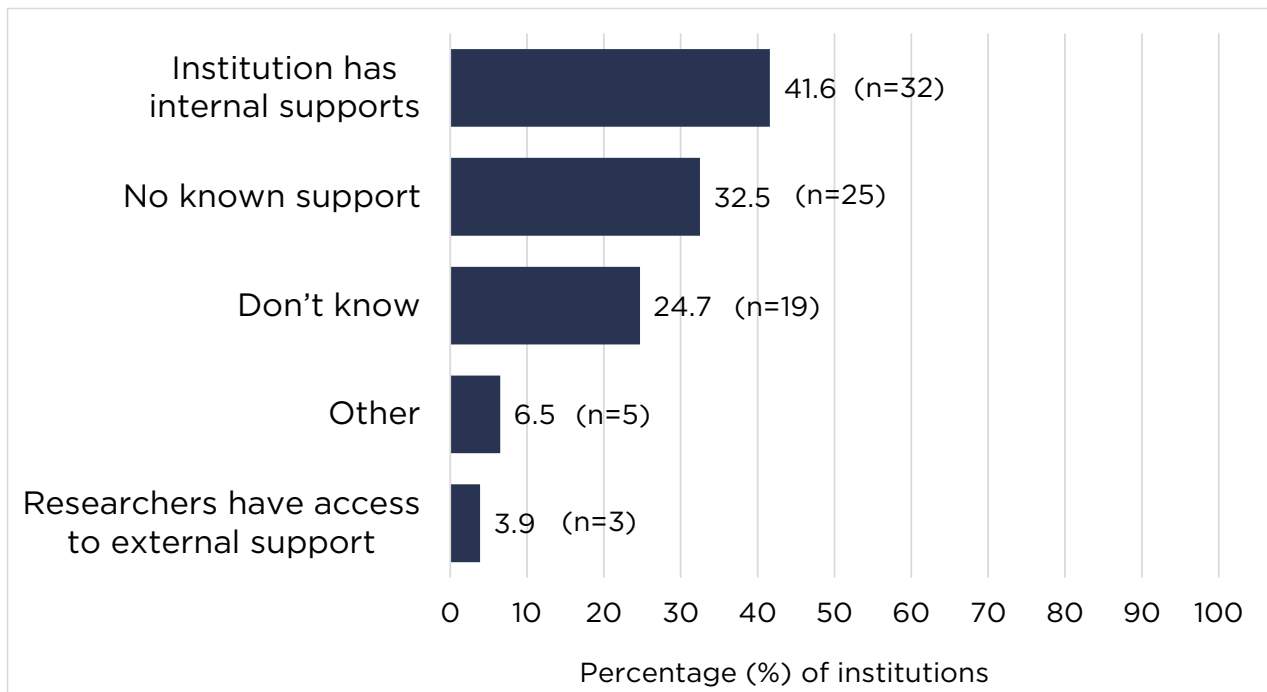


Figure 6. Infrastructure support for sensitive data (all institutions). Q10F. “What research data management infrastructure does your institution have access to? Check all that apply. Infrastructure for sensitive data.” (n=77).

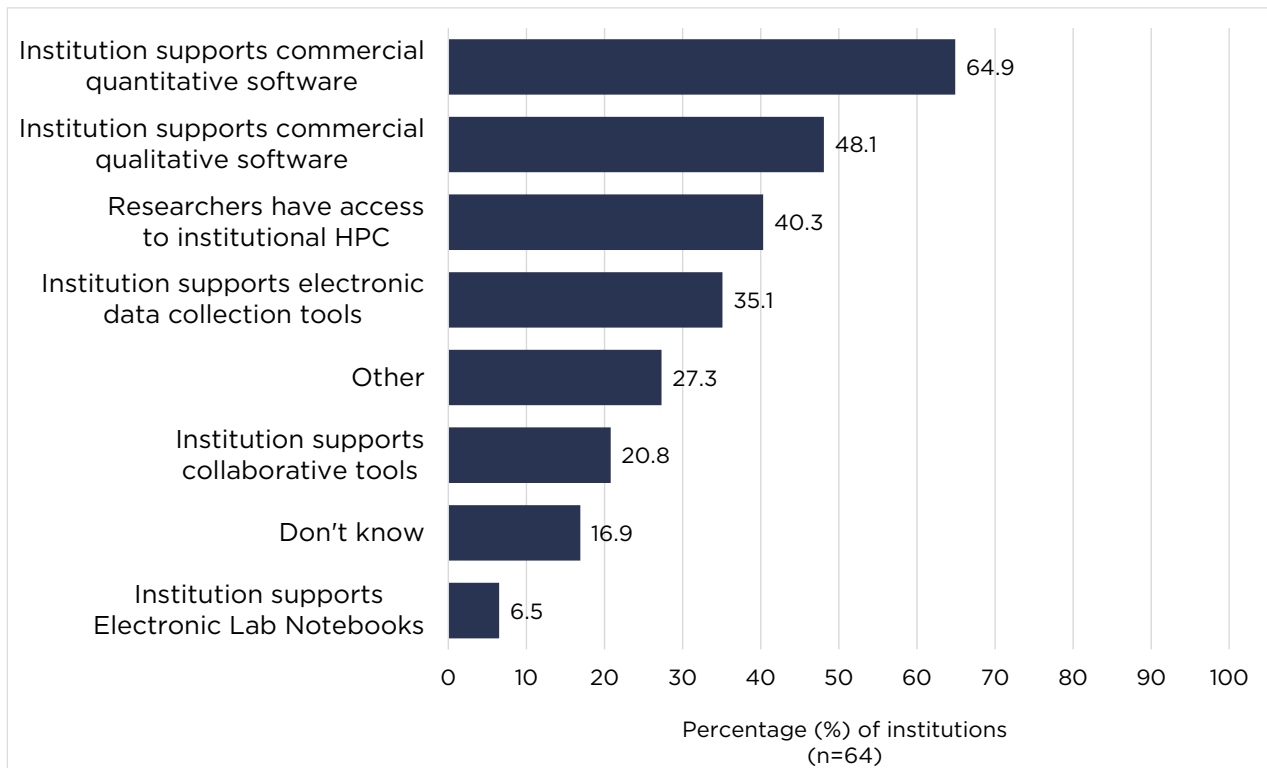


Figure 7. Other RDM infrastructure that institutions have access to (all institutions). Q10G. “What research data management infrastructure does your institution have access to? Check all that apply. Other research data management infrastructures.” (n=77).

RDM Services

Respondents were asked a series of questions related to their current capacity to support or provide informational websites, advisory services, financial support, data computing services, RDM training, curation support, and technical support for things such as data encryption and anonymization. Results show over half of the institutions which responded to the survey provide all of the above support to their researchers in some capacity (Table 2). While a range of RDM-related services are available in most institutions, many have not been fully realized. Existing services are primarily identified within university respondents, and typically offered through their libraries and/or in conjunction with available national resources. In contrast, colleges mainly appear to be in the development and planning stage across all these sectors of capacity.

| RDM Service | University (%) | College/CÉGEP (%) | Other (%) | Total responses (%) |
|---|----------------|-------------------|---------------|---------------------|
| Informational website(s) on research data management | 88.5 (n=46) | 47.6 (n=10) | 100 (n=4) | 77.9 (n=60) |
| Advisory services | 82.7 (n=43) | 33.3 (n=7) | 75.0 (n=3) | 68.8 (n=53) |
| General RDM best practices or DMP training | 76.9 (n=40) | 38.1 (n=8) | 75.0 (n=3) | 66.2 (n=51) |
| Curation support | 73.1 (n=38) | 28.6 (n=6) | 50.0 (n=2) | 59.7 (n=46) |
| Targeted hands-on RDM workshops | 69.2 (n=36) | 19.0 (n=4) | 25.0 (n=1) | 53.2 (n=41) |
| Technical support with data encryption, anonymization, etc. | 55.8 (n=29) | 38.1 (n=8) | 75.0 (n=3) | 51.9 (n=40) |
| Data computing services | 55.8 (n=29) | 42.9 (n=9) | 50.0 (n=2) | 51.9 (n=40) |
| Specific financial support to researchers | 17.3 (n=9) | 23.8 (n=5) | 50.0 (n=2) | 20.8 (n=16) |

Table 2. RDM Services by institution type. Q13A. "Research data management support can be a composite of services provided by different units at an institution, as well as regional and national efforts. Infrastructure may also consist of open platform/not-for-profit and commercial/publisher packages. Which units or products within and outside your institution support each of the following services? Check all that apply." (Universities, n=52; Colleges, n=21; Other, n=4; Total responses, n=77)

Breakdown of Service Providers

We classified the sources of RDM support services into broad categories: institutional (library, IT, or through joint efforts), regional/national, commercial, and disciplinary providers to try to determine how different types of institutions are engaging with these various resources. Results show that respondents are primarily relying more on institutional resources to provide RDM support services. Libraries, IT departments, or joint services provided by combinations of internal units are assuming these responsibilities. The breakdown of RDM support services by provider is shown in Figure 8 across all institutional types however, only the differences between colleges and universities are detailed below.

Within the 52 universities who responded, over 60% mentioned that informational websites, training, workshops, advisory services, and data curation support are provided in some capacity internally. Within the 21 colleges some internal services are provided but not consistently across specific services.

In addition to providing RDM services through internal units, institutions are also drawing on external resources. The respondents from universities utilize regional and national collaborative resources to assist with service delivery of informational websites (51.9%), training (32.7%), workshops (28.8%), and advisory (25%) services.

Respondents from colleges did not report high use of external resources, though 23.8% did indicate using regional or national resources to provide informational websites to researchers.

For universities, use of disciplinary resources to support RDM services at the institutional level was reported as follows: 21.1% indicated use of disciplinary resources for data computing, and 15.4% for RDM technical support. College responses also indicated low levels of engagement with disciplinary resources to support institutional RDM services. The use of disciplinary resources may not have been fully captured in this survey. Engagement with these services can happen at the faculty, department or researcher level rather than at the institutional level and therefore may not be known by survey respondents.

Universities also indicated low use of commercial resources for institutional RDM support; only 11.5% mentioned they use commercial resources for data computing support and 7.7% mentioned they use commercial resources for informational RDM support. None of the colleges surveyed reported using commercial RDM resources.

As part of the survey, institutions were asked to report on how they are planning to provide RDM support. Colleges had higher percentages of planning to provide RDM supports than universities across most categories except for data computing services and technical supports where response rates are approximately similar between both.

Around 33% of colleges are planning to provide informational RDM websites compared to about 10% of universities. Plans to provide specific financial support to researchers and general RDM best practices or DMP training is higher with colleges at around 48% for both compared to universities at around 29% and 27% respectively. Again, 43% of colleges responded that they were planning to provide advisory services which is higher than universities at 21%.

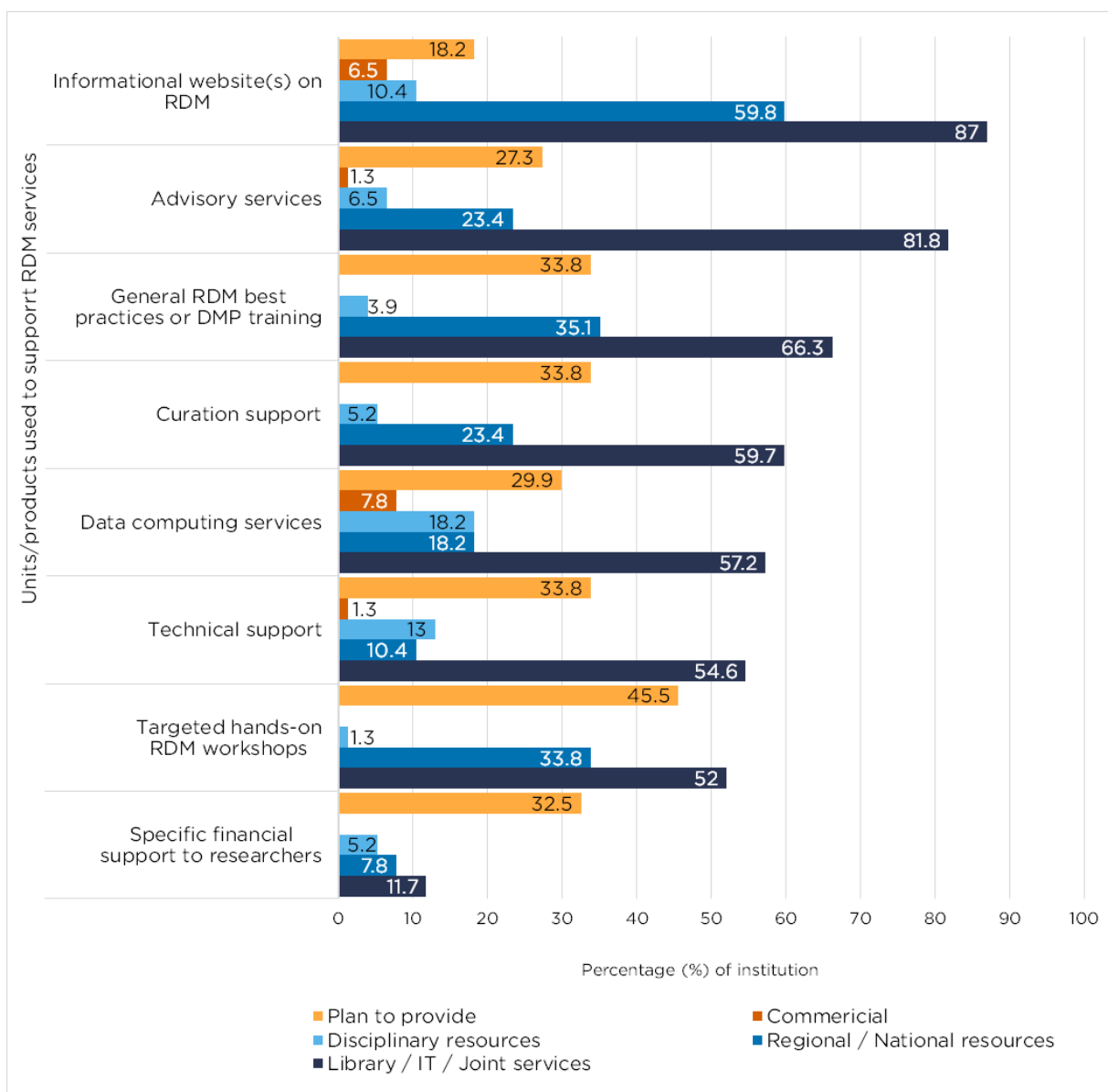


Figure 8. Unit or products used to support RDM services by institutional resources, disciplinary resources, regional/national resources, and commercial products (all institutions). Q13A. "Research data management support can be a composite of services provided by different units at an institution, as well as regional and national efforts. Infrastructure may also consist of open platform/not-for-profit and commercial/publisher packages. Which units or products within and outside your institution support each of the following services? Check all that apply." (n=77)

Organizational Structure and Service Delivery

The availability of RDM services is positively correlated with formality of institutional structures for RDM. Among the institutions analyzed, those with formal RDM campus stakeholder committees provided a higher level of services internally (through the library, IT and/or joint efforts) than those with informal campus stakeholder groups, while institutions with an informal group tended to provide a higher level of internal RDM services than those without any organized form of campus stakeholder collaboration. However, the internal structure of RDM stakeholder groups did not have much impact on utilizing external collaborations (through regional, national, disciplinary and/or commercial sources) for support. Figure 9 illustrates the breakdown of types of RDM support provided by the RDM groups created by institutions across all institutions.

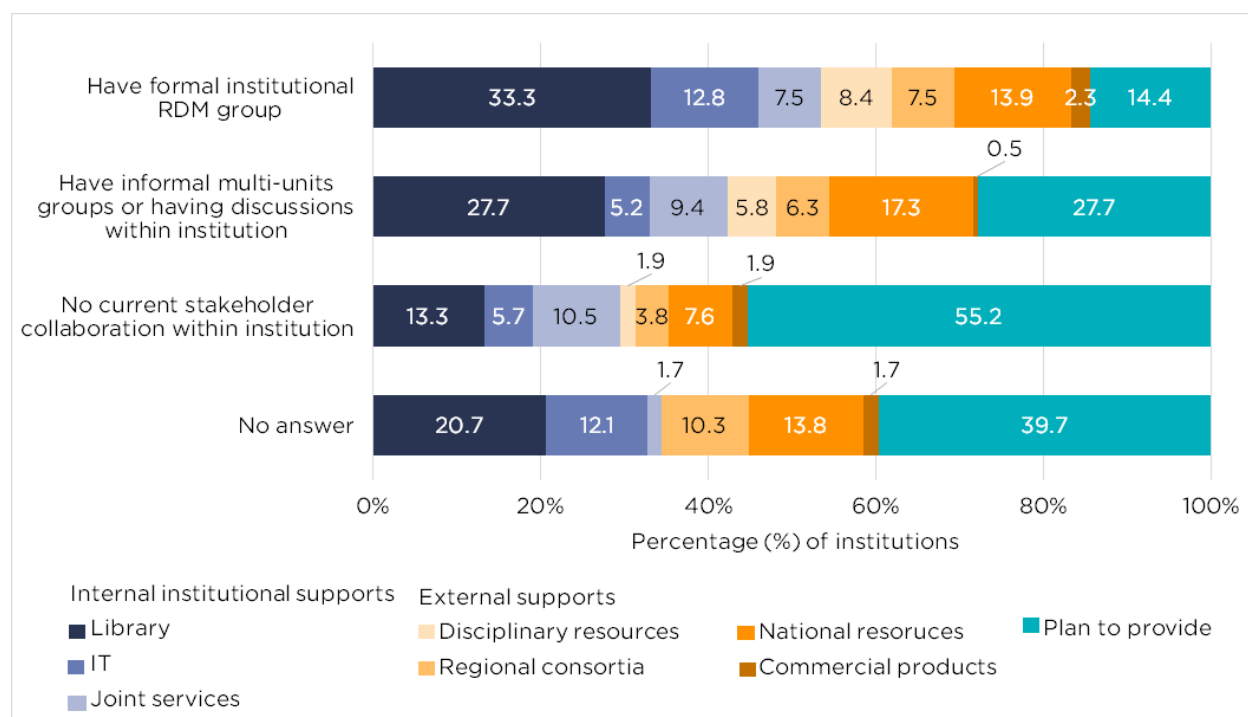


Figure 9. RDM service support providers and the types of RDM groups institutions have created (all institutions). Cross tabulation of Q13A. "Research data management support can be a composite of services provided by different units at an institution, as well as regional and national efforts. Infrastructure may also consist of open platform/not-for-profit and commercial/publisher packages. Which units or products within and outside your institution support each of the following services? Check all that apply." and Q9. "How are different stakeholders within your institution working collaboratively to tackle research data management challenges?" (n=77)

HQP and Service Delivery

Among the 77 institutions who answered our survey, there is a relatively equal distribution among those who have created dedicated RDM positions (n=26), those who have reassigned RDM responsibilities to be full or part of existing positions (n=22), and those who haven't created any RDM positions (n=25). Four institutions did not answer this question. Figure 10 illustrates the distribution of RDM services providers within institutions across their different RDM role creation or reassigning strategies across all institutional types.

For training, workshops, and advisory services, institutions with dedicated RDM positions demonstrate a high level of internal institutional support (73.1% - 88.5%), but a lower level of external support (<27%) in these areas. Whereas, institutions that reassigned RDM responsibilities to existing positions show a lower level of internal support in the areas (50% - 63.6%), but a higher reported use of external support (<41%). Institutions without any RDM positions provide less training, workshops, and advisory services internally (20% - 40%), and even fewer of them use external resources to provide these supports (<20%).

Information websites showed high levels of availability both internally and externally between institutions that had dedicated RDM positions and those with reassigned positions. Internal institutional support for informational websites was reported at 80.8% for dedicated RDM positions, 90.9% for reassigned roles, and 50% for institutions with no RDM roles. Externally, institutions used regional and national informational websites across all RDM positions reporting 50% in both dedicated RDM positions and those reassigned, and 44% with no RDM roles.

The presence of RDM-focused positions is correlated with presence of internal technical support (for encryption, anonymization, etc) or computing services for RDM (53.8% for both in institutions with dedicated RDM positions, 45.5% for both in institutions reassigned RDM positions, and 32% and 24% respectively in institutions without RDM positions). Lower levels of external resource use were reported for technical support and computing services. Of the external service providers, availability of both data computing and technical services from disciplinary resources was greatest in institutions with dedicated RDM positions (30.8% for both services). Reliance on national/regional technical support was also greatest for institutions without any dedicated RDM position (25.0%).

A high percentage of institutions with RDM-focused positions provide data curation support internally (76.9% with dedicated positions, and 68.2% with reassigned positions). Comparatively, only 32% of institutions without RDM positions provide this support internally. External collaborations in data curation are low across all institutions regardless of RDM-focused positions (<20%). No commercial resources are used for data curation support among any of these three groups.

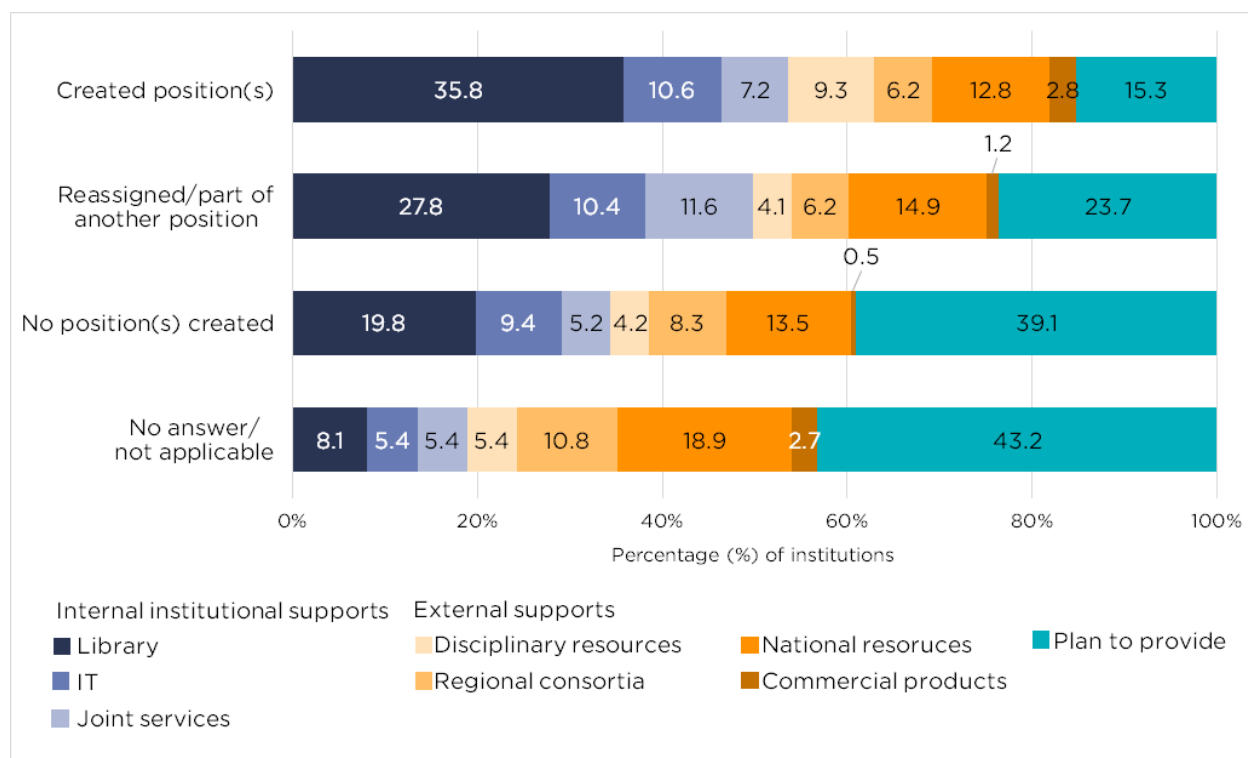


Figure 10. RDM service support providers and RDM positions created or reassigned (all institutions). Cross tabulation of Q13A. “Research data management support can be a composite of services provided by different units at an institution, as well as regional and national efforts. Infrastructure may also consist of open platform/not-for-profit and commercial/publisher packages. Which units or products within and outside your institution support each of the following services? Check all that apply.” and Q8. “What data management positions (if any) have your institution created or reassigned (please provide associated job titles)?” (n=77)

Budgets and Service Delivery

The survey asked a free text question about how RDM services are funded at institutions and coded answers based on the responses given into:

- dedicated institutional budget
- unit(s) budget within institution
- cross-institutional collaboration/consortium
- funding from outside of the institution (e.g. grants)
- no budget
- unknown/not answered.

The responses indicate that there is not typically a dedicated, institution RDM budget, but rather RDM is typically being funded by unit/departmental(s) efforts, if funded at all. There are regional (e.g. Scholars Portal Dataverse) and national (e.g. Portage) RDM efforts that are supported by several Canadian institutions; however, there were low responses recognizing participation in cross-institutional or consortial efforts. This may be because survey respondents do not perceive these contributions as the institutional strategy to support RDM.

Institutions with internal budgets that support RDM (both the one with a dedicated institutional budget and the 21 institutions that had support through unit/department(s) budgets) indicated higher levels of internal support for informational websites, training, workshops, advisory services, data curation, and technical support for sensitive data (71.4%-95.2%) than those that did not have some form of institutional budget for RDM (20.7%-58.6%).