



Digital Scholarly Record Working Group
Report to the Provost and Vice-Principal (Academic)
and the Vice-Principal (Research)

April 19, 2018

Don Aldridge, Centre for Advanced Computing
Brenda Brouwer, School of Graduate Studies
Amir Fam, Faculty of Engineering and Applied Science
Adam Grotzky, Society for Graduate and Professional Students
Cynthia Fekken, Office of the Vice-Principal (Research)
Karina McInnis, University Research Services
Martha Whitehead, Vice-Provost (Digital Planning) and University Librarian (Chair)
Resources, Queen's University Library: Rosarie Coughlan, Courtney Matthews, Heather
McMullen, Michael Vandenburg

1. Executive Summary

Digital access to information is profoundly changing research and enabling new innovations. Success in this environment is dependent on a strong digital research ecosystem, including strategic management of the digital scholarly record.¹ To enable that success, Queen's Digital Scholarly Record Working Group proposes a set of goals and these VOICE principles: Value, Openness, Inclusivity, Collaborative platforms, and Engaged researchers.

While the collaborative efforts of many individuals and units across Queen's University have created a strong foundation for giving VOICE to research, building a sturdy structure on that foundation relies on several integrated considerations:

- i. Access to appropriate levels of computing and storage resources
- ii. Human resources for managing, disseminating and preserving research outputs
- iii. Alternatives to prevailing unsustainable publishing models
- iv. Future development of digital research infrastructure

Requirements of research funding agencies are a key driver in this landscape. The *draft Tri-Agency Data Management Policy* has proposed requirements for the creation of an institutional research data management strategy, the creation of data management plans for research projects, and the deposit of data in an appropriate public repository or other platform that will ensure safe storage, preservation, curation and (if applicable) access to the data. The *Tri-Agency Open Access Policy on Publications*, which came into effect May 1, 2015, requires that journal articles be freely accessible within 12 months of publication.

The goals proposed in this report address these important issues and are deliberately aligned with the stated and anticipated policies of the Tri-Agency requirements. Queen's has an opportunity to continue its leadership in this area, as the future policy landscape emerges.

This report as a whole is intended to inform the university's digital planning, now under way. Digital planning pertains to the university's core mission of learning and research, as seen through a digital lens. The student learning experience and high impact research depend deeply upon the management, dissemination and preservation of the digital scholarly record.

¹ The term "digital scholarly record" refers to the elements needed to understand and replicate research results in the current digital environment. These are:

- Data used as evidence in the research process
- Computer programs and documentation used to generate and manipulate the data set
- Outputs of research and scholarship in digital form (e.g publications, creative works, digital environments, etc.)

2. Purpose

The purpose of the Digital Scholarly Record Working Group has been to further plan and raise awareness of services across Queen's University that support the management, dissemination and preservation of the digital scholarly record. This work was undertaken at a preliminary stage of planning for the development of the university's digital strategy, and was conducted with a view to related national and international initiatives.

The Digital Scholarly Record Working Group steered several initiatives to advance planning and awareness of services supporting research data management and open access publishing. The group also engaged in developing a training series addressing various data-related activities in the research lifecycle and advised on Queen's participation in the ORCID consortium in Canada, which is an international standard designed to improve data sharing.

Based on a review of the current state of data management services at Queen's, recent Queen's surveys regarding research data management needs, and relevant external policies, principles and initiatives, the group prepared a discussion paper² to engage the Queen's community.

Presentations were made to the Provost-Deans-Management Group, Associate Deans Research, Graduate Students Executive Council, Senate Advisory Research Committee, Senate Library Committee, Senate and several faculty boards. The presentations were well received; feedback helped to fashion the principles and goals outlined at the end of this report.

3. Current state

a) A strong foundation

Queen's has been engaged in the collaborative development of infrastructure and services in the digital scholarly record arena for several years and is seen as a leader in many ways. A Library working group, including representation from the Vice-Principal (Research) portfolio, laid the groundwork for research data management services at Queen's with a three-year action plan (2013-2016) and gained experience working with numerous researchers seeking to deposit data sets for future access and re-use. In support of the broad dissemination of research publications, the Library has offered scholarly communications services for many years, also in collaboration with the Vice-Principal (Research) portfolio, ITServices and the School of Graduate Studies.

To further advance research data management and scholarly communications, in 2017 the Library established its Open Scholarship Services division, which works closely with the Discovery and Technology Services division and with librarians responsible for liaising with particular faculties and departments. Within Discovery and Technology Services, a new position of Research Data Management Systems Librarian was established in collaboration with OCUL's Scholars Portal, whereby the librarian works on initiatives both at Queen's and regionally. This collaborative approach to the development and delivery of services and infrastructure is a key principle of the

² Queen's Digital Scholarly Record Working Group, <https://wiki.queensu.ca/display/QDSRWG/Queen%27s+Digital+Scholarly+Record+Working+Group+Home>

Library's approach, as seen also in the leadership roles that several Queen's librarians have played in establishing and advancing Portage, a national research data management network.

Closely related to research data management services is the role of the Centre for Advanced Computing (CAC) in providing researchers with secure, large scale platforms for manipulating and storing their data. The CAC has recently refreshed its infrastructure and expanded its team. The Library and CAC are partnering to develop services intended to enable ongoing access to research data curated for the long-term. This work builds upon a collaboration between Portage and Compute Canada, in which a scalable federated platform for research data management and discovery is now in limited production and is proposed as one aspect of digital research infrastructure strategy at the national level.

The group also provided advice on Queen's participation in ORCID-CA, the ORCID consortium in Canada.³ The Open Researcher and Contributor ID (or "ORCID") standard was established to solve name ambiguity and researcher identification problems by giving individuals a unique numeric identifier that persists over time. Major publishers, funders and research institutions have been adopting this international standard to support improved data exchange between scholarly search platforms and information systems. At Queen's, we are now encouraging researchers to register for an ORCID ID, and we can use the ORCID functionality in Faculty 180 to enable easy linking of researchers' publications.

b) A key driver

Open access to publications and the effective management of research data are not simply good ideas: they are becoming requirements of research funding agencies.

The *Tri-Agency Open Access Policy on Publications*⁴ came into effect May 1, 2015, requiring that grant recipients ensure that any peer-reviewed journal publications arising from Agency-supported research are freely accessible within 12 months of publication. The Library has worked with University Research Services to facilitate researchers' deposit of publications in the institutional repository, QSpace, but further service development is needed to promote and support these processes.

Regarding research data, the *draft Tri-Agency Data Management Policy* has three proposed requirements:

- each institution administering Tri-Agency funds could be required to create an institutional research data management strategy;
- grant recipients could be required to create data management plans (DMPs) for research projects supported wholly or in part by Tri-Agency funds;
- for all research data and code that support journal publications, pre-prints and other research outputs that arise from agency-supported research, grant recipients could be required to deposit these data and code in an appropriate public repository or other

³ ORCID-CA Consortium. <https://orcid-ca.org/node/1>

⁴ Government of Canada, "Tri-Agency Open Access Policy on Publications," 21 December 2016. http://www.science.gc.ca/eic/site/063.nsf/eng/h_F6765465.html?OpenDocument

platform that will ensure safe storage, preservation, curation and (if applicable) access to the data.

In response to this anticipated policy, the Portage Institutional RDM Strategy Working Group⁵ has released a draft template and supporting guidance document that are designed to assist Canadian research institutions in developing an overarching strategy for research data management. Many elements of this strategy development have been addressed at Queen's, but several integrated considerations remain to be addressed in our overarching digital strategy.

c) Integrated considerations

The collaborative efforts of many individuals and units across the university have created a strong foundation to support the management, dissemination and preservation of scholarly records, however building a structure on that foundation requires attention to the following important issues.

i. Access to appropriate levels of computing and storage resources

Data management relies on access to appropriate levels of computing and storage resources. Currently, ITServices offers network attached storage services with departmental allocations based on a 25GB/FTE calculation. The 600TB capacity is effectively 300TB, given that all data are replicated between two sites. This current capacity has reached critical levels of over 80% utilization and the equipment is at end of life. An option analysis has been conducted and a plan to transition to a scalable solution has been developed, however it will require capital investments to execute. ITServices offers compute services, however the capacity of this service (at CPU 25%, memory 60%, disk 80% capacity) is mainly to support the university's operations. The Centre for Advanced Computing has initiated new service offerings to meet some researchers' needs, and with coordination and sustainable funding more could be done. The Library offers the ability to deposit and store some digital assets in collaboration with ITServices, through local repositories, the Ontario Library Research Cloud and Scholars Portal services hosted remotely, however limited resources have made it difficult to advance these opportunities quickly.

Research data management surveys⁶ conducted at Queen's between 2015 and 2017, with 1004 respondents, provide some insights into the state of Queen's researchers' data storage needs, indicating that:

- 59.7% of respondents work with data requiring fewer than 50 GB of storage and 9.6% reported working with data requiring greater than 500GB of storage;

⁵ Portage, "Institutional RDM Strategy Working Group," <https://portagenetwork.ca/working-with-portage/institutional-rdm-strategy/>

⁶ Berish, Francine; Cooper, Alexandra; Druery, Jackie; Heil, Jeremy; Moon, Jeff; Maranda, Suzanne; Murphy, Sharon; Saleh, Nasser; Zaraiskaya, Tatiana. "Research Data Management Survey at Queen's University," 2018. <http://dx.doi.org/10.5683/SP/E6LSVQ>

- 69.5% of respondents expressed an interest in institutional repositories for long-term access and preservation of research data.

There has not been an equivalent survey with regard to researchers' needs for computing resources, but we do know through digital planning discussion groups and ongoing inquiries that there are Queen's researchers who feel compelled to rely on resources at their collaborators' institutions or commercial options such as Amazon Web Services. As well, those researchers observe that the projects undertaken by their students – undergraduates as well as graduate students – require resources beyond those available at Queen's.

ii. Human resources for managing, disseminating and preserving research outputs

Technologies are only part of the equation. Effectively implementing those technologies and providing training and support for their application requires skilled personnel.

The data-related activities of researchers are well illustrated in the figure provided by the Leadership Council on Digital Research Infrastructure, below. The activities in the centre of the circle (secure, discover, document and curate, and store) occur repeatedly throughout the research process. The activities in the outer ring (plan, create, process, analyze, disseminate, preserve, and reuse) are activities associated with a particular part of the research process and are often iterative. (For descriptions of each of the specific activities, see Appendix A).

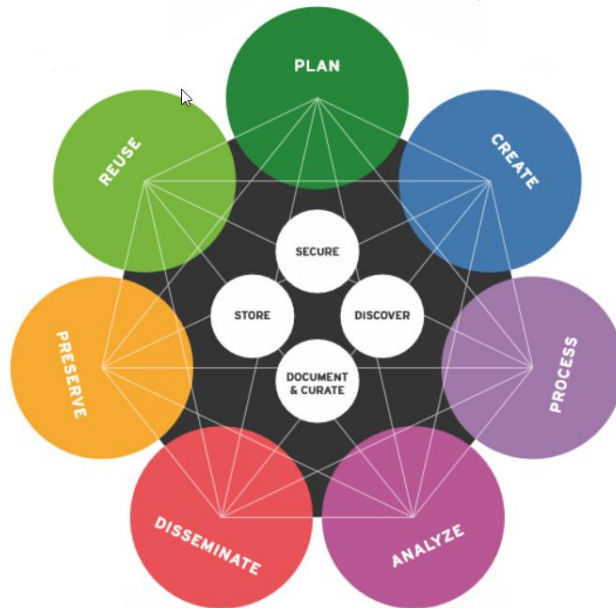


Figure 1: Data-related activities during the research process

In the past, research teams may have expected to have the capacity themselves to manage this cycle of activities, but as those activities have become more data-intensive they often seek the advantages of networked services. In the research data management surveys

mentioned above, strong interest in support for data management processes was expressed across all disciplines:

- 52.0% specified they would need guidance in drafting a data management plan for successful application for funding
- 60.7% need assistance in documenting and describing research data
- 63.9% are interested in personalized consultation on data management practices
- 65.7% are interested in workshops on best practices in data management for graduate students and 56.2% in workshops for faculty

iii. Alternatives to prevailing unsustainable publishing models

Technological changes have impacted radically how scholarly research is conducted and disseminated. Yet scholarly publishing continues to follow models that were conceived in the print era, and the means to effectively manage and preserve research data are just beginning to evolve.

Much of the scholarly publishing marketplace is profit driven, employing a business model whereby the publications resulting from publicly funded academic research are given to publishers and sold back to libraries, often at high profit margins. As well, libraries are unable to selectively purchase titles for their communities, because they are sold as bundles or “big deals.” Libraries are dealing with an oligopoly in which five publishers control over 50% of the market and above 70% in some disciplines.⁷

In response to these issues, and in the interests of open access to information, Queen’s, along with provincial, national and international partners, is working to leverage digital opportunities to advance innovative, cost-effective scholarly communication models in a global knowledge commons. This includes services such as hosting Open Journal Systems, Open Monographic Press, and next generation repositories that enable open discovery of publications across global networks.

There are two key challenges in advancing these new models. The first is the cost of providing bold new models, while still devoting considerable funds to acquiring publications through traditional models. These funding needs relate to the two matters raised above: technical infrastructure and services. The second challenge is cultural: the value systems that drive university rankings and promotion and tenure are heavily dependent on metrics derived from traditional publishing models. Raising awareness amongst researchers of the issues and opportunities for change is crucial.

iv. Future development of digital research infrastructure and services

Researchers are often not aware of the services available to support the management, dissemination and preservation of their digital scholarly records. Further development of those services would benefit from a coordinated approach that is user-centred rather than

⁷ Larivière, Vincent, Stefanie Haustein, Philippe Mongeon, “The Oligopoly of Academic Publishers in the Digital Era,” *PLoS One* 10 (6), June 2015. <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502>

unit-based, in order to ensure the most effective use of resources and a seamless user experience.

Another important issue, as noted in (i) above, is that the creation and dissemination of digital scholarly records is inextricably tied to the other elements of the digital research infrastructure (DRI) ecosystem. The effective functioning of this ecosystem depends on integration of its component parts (network, advanced research computing, data management, storage, advanced research software). In this image below from the Leadership Council on Digital Research Infrastructure, data is a common element across each aspect of the ecosystem. It places at the centre the highly qualified personnel required to manage the ecosystem, serve researchers and train the next generation. Furthermore, it argues the importance of strategic planning and coordination around the interconnected components.

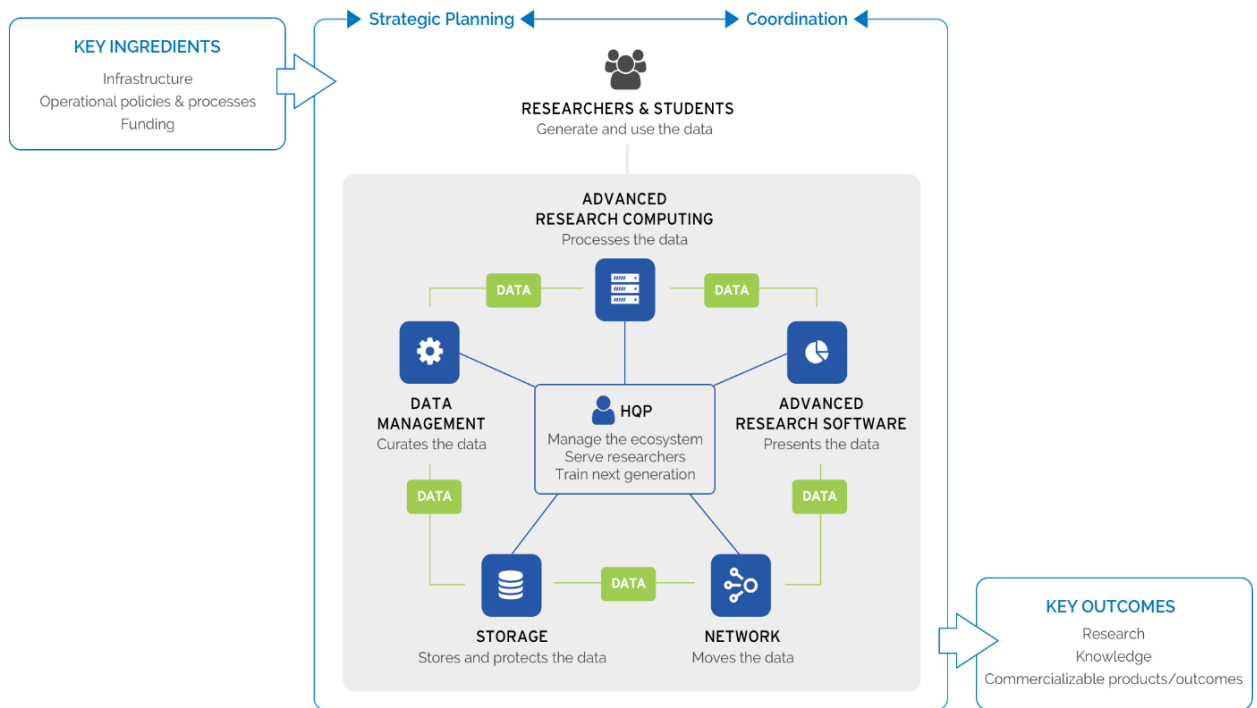


Figure 2: Components of the digital research infrastructure (DRI) ecosystem

4. Recommendations

In consultation with its stakeholders, the Digital Scholarly Record Working Group strongly recommends that the university’s digital strategy incorporate the following principles and goals for the dissemination of research results.

Principles

Queen's researchers and scholars aim to generate globally impactful research and scholarship, thus Queen's gives VOICE to research results according to the following principles:

Value: The university's evaluation of research impact includes a wide range of measures, including not only bibliometric analytics but also assessments appropriately based in qualitative analyses.

Openness: The results of research and scholarship should be disseminated as widely as possible for the advancement of research and the benefit of society.

Inclusivity: All disciplines, areas of study and human differences are supported, and it is recognized that they have unique and particular needs.

Collaborative platforms: The university supports the development and use of distributed, reputable platforms for research dissemination and preservation that reduce cost barriers and are guided by the [FAIR data principles](#) – that data must be Findable, Accessible, Interoperable and Reusable.

Engaged Researchers: The university supports researchers in their engagement in the research enterprise, including regaining control of the scholarly communication ecosystem (e.g. by intentionally licensing their publications and other research results so that they retain their ownership while sharing them openly).

Engaging with the services guided by these principles helps Queen's researchers and scholars ensure that their digital scholarly records are accessible and preserved in a global knowledge commons.

Goals

The following goals deliberately reflect priorities based on the proposed requirements of the *draft Tri-Agency Data Management Policy* (institutional strategy; data management plans; data deposit) and the requirements of the *Tri-Agency Open Access Policy on Publications* (journal articles must be freely accessible within 12 months of publication). Suggestions for those responsible for taking the lead in addressing these goals appear in square brackets.

1. Institutional Strategy

Coordinate and integrate strategic planning for all components of the digital research infrastructure ecosystem at Queen's, including institutional strategies for research data management and open access publishing.

- a. Engage key stakeholders in this strategy development, including researchers in a variety of disciplines (for example, health sciences, engineering and applied science, social sciences, humanities) as well as service providers (Centre for Advanced Computing, ITServices, Library, University Research Services). [Queen's Digital Planning Project Group]
- b. Consider the opportunities of regional, national or domain partnerships in establishing local infrastructure and services. [Queen's Digital Planning Project Group]
- c. Continue to raise awareness within the research community about the benefits of and best practices for good data management and sustainable publishing models,

for example through a coordinated web presence, training programs, etc. [Library with University Research Services]

2. Data management plans

Support researchers in creating data management plans [Library]

- a. Integrate the Library's DMPAssistant and training opportunities into researcher workflows, such as research funding program workshops delivered by University Research Services, the Expanding Horizons series for graduate students, etc.

3. Data deposit

Identify the required technical and human capacity associated with research computing, data storage and data services [Queen's Digital Planning Project Group]

- a. Analyze information being gathered currently for the Capital Asset Management Plan to assess current computing and storage capacity and options for leveraging it more effectively.
- b. Develop coordinated budget proposals for the 2019-20 to 2021-22 budget planning cycle, involving ITServices, the Centre for Advanced Computing and the Library, to address immediate gaps in platforms and the human resources required to deliver appropriate services.

4. Sustainable open access to publications

Further develop services supporting alternatives to prevalent unsustainable publishing models and enable open access to publications [Library]

- a. Engage in collaborative opportunities to develop networks of next generation open repositories and services.
- b. Refresh the current institutional repository offering as part of the Library's digital asset management planning.
- c. Further identify services to raise awareness of open access opportunities, support the deposit of publications in open repositories, and assist researchers in producing open publications.
- d. Develop a budget proposal for the 2019-20 to 2021-22 budget planning cycle to address immediate gaps in platforms and the human resources required to deliver appropriate services.

Appendix A: Data-related activities during the research process



The Data Management Working Group of the Leadership Council on Research Infrastructure developed the following descriptions to clarify what specific research actions take place during each of the data-related activities.

Store: writing (accessing/placing/putting) and reading (retrieving) digital content on a variety of physical media. Content is used for different purposes across the lifecycle, including copies for active use, archival use, and dissemination. Active storage enables the use of digital content within research currently being conducted. Archival storage enables long-term access and preservation by protecting the digital integrity of content for current and future uses.

Secure: protecting, controlling, and complying with legal and ethical conditions on the use of the data.

Document and Curate: describing, explaining, and communicating the context in which the data exists, the workflow of the data, and the technical details of the data (the metadata). Employing metadata standards provides structure and form in documenting data and enables the interoperable use of the documentation and the data it describes. There are many levels of metadata that can be articulated (e.g., study level description, sample description, variable description, instrument description, etc.), describing, identifying, and explaining the data for discovery purposes.

Discover: searching, finding, mobilizing, locating, interpreting, assessing, and visualizing metadata, and retrieving research data.

Plan: identifying resources, expertise, and services required to develop, manage, and share high-quality data; planning for resource, time, and cost management; preparing/arranging/identifying/planning for the deposit of data so that it is discoverable in a data

repository and reusable; developing and implementing policies at universities, and reviewing policies for compliance at project level; determining data management requirements from the research design; identifying data management practices and workflow; planning consent for sharing.

Create: identifying, acquiring and creating (i.e., experimenting, observing, measuring) data for the research; generating data (e.g., simulating) and corresponding metadata; consulting policies that structure or define procedures for collecting data, such as ethics procedures and approval.

Process: preparing research data for analysis; checking, validating, cleaning, describing, transforming, aggregating, manipulating, reducing, anonymizing; describing workflow and processing; consulting expertise in data wrangling and statistics, as well as policy and procedures in domain-specific instances.

Analyze: describing, comparing, interpreting, and modelling patterns within the data; deriving variables/data; consulting expertise for statistical research methodologies and domains; as well as policies and procedures in domain practices.

Disseminate: sharing the data; transferring data from project to repository; consulting policy for data deposit agreements, licenses, access (conditions for reuse), and preservation; promoting and discovering; consulting expertise in data curation, submission information package management (i.e., formats, identification of files, manifest that identifies types of files); ensuring metadata is developed robustly.

Preserve: preparing, enhancing, and storing data, metadata, documentation, and code for long term access and re-use; migrating data to the best formats and through suitable media; storing data and back-up; applying best practices for digital preservation processing (i.e., characterization and normalization); consulting policies to apply to a trusted digital repository (i.e., format, selection, etc.), and expertise for digitization, digital preservation, and metadata.

Reuse: combining existing datasets to create new data for research and analysis; ensuring data discovery, access, and re-use of existing data (secondary data analysis); using resources to perform reuse; consulting policies concerning attribution, provenance, and licensing, as well as expertise on data wrangling, search skills, and secondary analysis skills.