

Building Research Data Management Capacity: Case Studies in Strategic Library Collaboration

Rebecca Bryant, Brian Lavoie,
and Amanda K. Rinehart

Building Research Data Management Capacity: Case Studies in Strategic Library Collaboration

Rebecca Bryant

OCLC Research

Brian Lavoie

OCLC Research

Amanda K. Rinehart

The Ohio State University



© 2023 OCLC.

This work is licensed under a Creative Commons Attribution 4.0 International License.

<http://creativecommons.org/licenses/by/4.0/>



December 2023

OCLC Research
Dublin, Ohio 43017 USA
www.oclc.org

ISBN: 978-1-55653-282-5

DOI: 10.25333/knxc-xy74

OCLC Control Number: 1402307712

ORCID iDs

Brian Lavoie, OCLC Research:  <https://orcid.org/0000-0002-7173-8753>

Rebecca Bryant, OCLC Research:  <https://orcid.org/0000-0002-2753-3881>

Amanda K. Rinehart, The Ohio State University:  <https://orcid.org/0000-0002-9987-0131>

Please direct correspondence to:

OCLC Research
oclcresearch@oclc.org

Suggested citation:

Bryant, Rebecca, Brian Lavoie, and Amanda K. Rinehart. 2023. *Building Research Data Management Capacity: Case Studies In Strategic Library Collaboration*. Dublin, OH: OCLC Research.

<https://doi.org/10.25333/knxc-xy74>.

CONTENTS

Preface	vi
Introduction	1
Library collaboration as a strategic choice.....	2
Strategic frames: Four economic concepts	2
Methodology	4
Recommendations for Libraries	6
Preparing to collaborate	6
Recognize that collaboration requires investment	6
Balance the trade-off between start-up and consortial models	7
Start collaborations early	7
Value the intangible benefits of collaboration	7
Establishing a successful collaboration	8
Leverage preexisting trust relationships.....	8
Find your people	8
Cultivate commitment.....	9
Stewarding for sustainability	9
Plan for success	9
Be strategically ambitious and tactically restrained	10
Align with institutional goals.....	10
Recognize that contributions and benefits may be asymmetrical	11
Case Studies	11
Case study organization	11
Texas Data Repository	12
Introduction.....	12
Origin story.....	12

Choosing collaboration	14
Strategic frames.....	15
Looking ahead.....	19
Lessons from the Texas Data Repository experience	20
The Portage Network	20
Introduction.....	20
Origin story.....	22
Choosing collaboration	24
Strategic frames.....	25
Looking ahead.....	30
Lessons from the Portage Network experience.....	30
The Data Curation Network	31
Introduction.....	31
Origin story.....	31
Choosing collaboration	34
Strategic frames.....	35
Looking ahead.....	40
Lessons from the Data Curation Network experience.....	41
Conclusion	43
Acknowledgments.....	44
Appendix: Semi-Structured Interview Protocol	46
Notes.....	49

FIGURES

FIGURE 1	Four economic concepts, four perspectives: Strategic frames for collaboration.....	2
FIGURE 2	Recommendations for libraries: Three stages of collaboration.....	6
FIGURE 3	Texas Data Repository timeline	14
FIGURE 4	Portage Network timeline	24
FIGURE 5	Data Curation Nnetwork timeline.	34

PREFACE

Library collaboration is an important option for sourcing capacity, and, when managed effectively, it can bring significant and sustainable benefits. A growing body of work from OCLC Research explores the contours of library collaboration from different perspectives and contexts. For example, in 2023, OCLC published *Sustaining Art Research Collections*, a series of two reports including *Using Data to Explore Collaboration*¹ and *Case Studies in Collaboration*,² that documents the importance of, opportunities for, and experiences pursuing collaboration in the art library community. The 2021 report *New Model Library: Pandemic Effects and Library Directions*³ highlights the central role of collaboration in helping libraries navigate the COVID-19 pandemic, as well as its aftermath and beyond. Most of these publications examine collaboration across institutions. However, the 2020 report *Social Interoperability in Research Support: Cross-campus Partnerships and the University Research Enterprise*⁴ examines how libraries partner with other units on campus to provide research support services.

Collaboration is one option among several sourcing strategies, the perceived costs and benefits of which must be carefully evaluated against both institutional priorities and the availability of alternatives. This topic was explored in the 2022 report *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*,⁵ which offers an economic framework that libraries can use to examine collaborative opportunities. These strategic frames help illuminate key features, potential challenges, and pathways to success and sustainability.

In the present report, we utilize those strategic frames as an analytical framework to explore the role of multi-institutional library collaboration in acquiring research data management (RDM) capacity. The three case studies that follow are instructive examples of how the concepts detailed in the strategic frames play out in real-world circumstances, and most importantly, how libraries worked together to address them. Our interview-based approach elicited a wealth of invaluable perspectives, insights, and advice on library collaboration that we synthesized into a set of recommendations for libraries contemplating future collaborations in RDM or other spaces.

Effective library collaboration is art as much as science. While concepts, frameworks, and theory are important for deepening our understanding of what makes collaborations successful and sustainable, we believe that sharing practical experiences of successful collaboration is also essential. We are grateful to the many individuals who spoke to us as part of our study, and candidly shared their experiences in library collaboration.

Rebecca Bryant
OCLC Research

Brian Lavoie
OCLC Research

Amanda K. Rinehart
The Ohio State University

INTRODUCTION

Research data management (RDM) is an emerging and increasingly vital category of services and expertise for universities and a significant area of strategic interest for academic libraries. Acquiring the capacity to support institutional RDM needs can be a daunting investment, with many strategic options available, ranging from building capacity internally to purchasing it from an external provider.* For many libraries, collaboration—working collectively with partners to provide a shared capacity—is an attractive strategy. However, successful and sustainable collaboration is not certain; many obstacles can hinder the way forward. Yet a few successful RDM collaborations are operating in the landscape today. Exploring the stories of these existing RDM collaborations and identifying key strategies that helped them succeed can provide libraries with insights to apply to their local context.

This report explores the role of multi-institutional library collaboration and collaborative decision-making in providing research data management (RDM) services. It documents the practical experiences of real-world collaborations and describes the challenges, opportunities, and trade-offs of choosing to collaborate to build RDM service capacity. To that end, it provides in-depth case studies of the following collaborations:

- **Texas Data Repository⁶ (TDR)**—An example of building RDM capacity by leveraging an existing collaborative consortium, the Texas Digital Library (TDL).
- **Portage Network** (now subsumed within the Digital Research Alliance of Canada⁷)—An example of establishing and coordinating RDM infrastructure at a national scale.
- **Data Curation Network⁸ (DCN)**—A rare example of a start-up RDM collaboration to fill a shared curation expertise gap.

We begin with the premise that collaboration should be a strategic choice. Collaboration requires a significant investment of time, effort, and resources to be successful and sustainable; therefore, the decision to make that investment should be weighed carefully. Moreover, once a collaboration is underway, it should be organized and managed strategically to maximize the prospects for achieving its goals.

* The OCLC Research report series, *The Realities of Research Data Management*, published in 2017 and 2018, provides examples of different strategies for acquiring RDM capacity, [oclc.org/research-reports/realities-RDM](https://oclc.org/research/research-reports/realities-RDM).

LIBRARY COLLABORATION AS A STRATEGIC CHOICE

This report is both informed by and a companion to the OCLC Research report *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*,⁹ which provides concepts, frameworks, and tools for supporting libraries as they evaluate and manage multi-institutional partnerships. *Library Collaboration as a Strategic Choice* draws heavily on an economics-based view, based on the idea that the decision to allocate resources to collaboration requires careful evaluation of how a partnership opportunity aligns with institutional priorities, the benefits and costs of participating, and the trade-offs that arise when chosen over other alternatives. This is a decidedly economic perspective on library collaboration, but an important one in that it captures the principal elements of what we mean when we say collaboration needs to be approached strategically. *Library Collaboration as a Strategic Choice* provides an investigative framework for exploring multi-institutional partnerships, enumerating the strategic aspects of choosing and managing partnership. The case studies in this report illustrate how these ideas manifest in practice through RDM collaborations, offering practical guidance on how managing the economics of collaboration impacts its future success and sustainability.

Four Economic Concepts, Four Perspectives: Strategic Frames for Collaboration

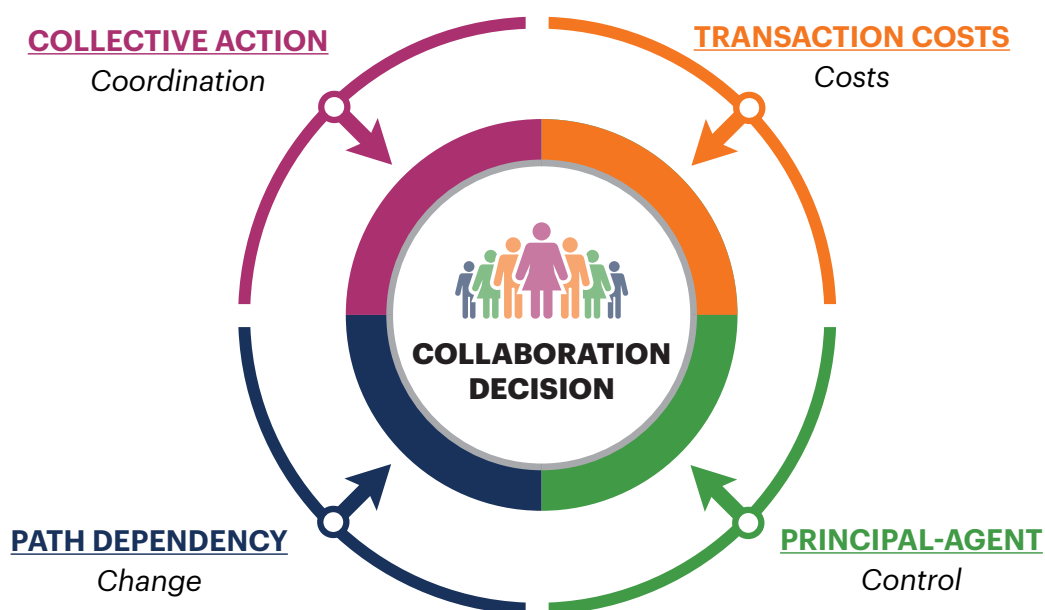


FIGURE 1. Four economic concepts, four perspectives: Strategic frames for collaboration

Strategic frames: Four economic concepts

Collaboration has received significant attention in the economics literature as a topic in a wide range of contexts and activities. Consequently, there is a rich resource of scholarly insight on collaboration that can be readily applied across many contexts—including libraries. *Library Collaboration as a Strategic Choice* applies four important economic concepts, or *strategic frames*, to deepen our understanding of factors impacting the prospects for successful, sustainable collaboration for libraries.

Each strategic frame represents a key aspect of organizing and managing collaborative effort: how to coordinate collaborative activity, how to minimize the costs of collaborating, how to change directions toward a collaborative approach, and how to manage decision-making control within a collaboration. The case studies in the present report are closely inspected according to these frames to better understand how each collaboration responded to a specific economic problem in deciding whether to collaborate and in organizing and managing their collective effort. A brief description of each strategic frame as it is applied to each case study in this report follows.

Coordination—This strategic frame examines the challenges of successfully coordinating group-scale efforts. It applies the economic concept of the *collective action problem*, which holds that groups with shared interests, working together toward a common goal, nevertheless can fail. So, steps must be taken to mitigate the risks that can threaten success. One example is *free riding*, in which one or more group members receive benefits without a fair contribution to the collective effort. The collective action problem is highly relevant to library collaboration in issues like assembling the right mix of institutional partners, establishing roles, engendering trust, obtaining commitments, and incentivizing partners in ways that draw out sufficient effort to achieve mutual goals.

Costs—This analysis focuses on the economic concept of *transaction costs*, which address the costs of organizing and conducting economic exchanges. In this case, the economic exchange of interest is collaboration, and the transaction costs to be considered are the costs of collaborating.

It is important to emphasize that our focus is on the costs of collaboration—i.e., organizing and managing the collaboration—not the operational costs of the shared capacity, including any technology or product choices.

In the context of library collaboration, transaction costs arise in activities like the effort to identify and convene a group of institutions to collaborate with, the labor to expand the membership of an existing group, the process of achieving consensus on methods and goals, and the investment in staff resources needed to participate in the governance and operation of the collaboration over time. It is important to emphasize that our focus is on the costs of collaboration—i.e., organizing and managing the collaboration—not the operational costs of the shared capacity, including any technology or product choices.

Change—This strategic frame applies the economic topic of *path dependency*, which examines how organizations cope with change, especially when they are invested in established relationships, workflows, and systems for getting things done. A key concept here is *switching costs*—the economic (and even non-economic costs) involved in moving from one choice to another. In this sense, path dependency is fundamentally about managing change. This can be important in a library collaboration context, touching on issues like the costs and obstacles involved in moving from a vendor-supplied to a collaboratively sourced solution or, conversely, the costs and disadvantages of maintaining the status quo instead of pursuing a collaborative approach.

Control—Finally, the *principal-agent problem* explores issues concerning control. When ownership of an organization is separated from management, the interests of these stakeholder groups may not align. For a library collaboration, this may involve issues like choosing governance arrangements—for example, whether decision-making authority will reside with a committee of members or a centralized executive. If the latter, oversight mechanisms may be needed to ensure executive decisions align with the membership’s interests.

While the framing of the case studies is predominantly economics-focused, the insight gleaned from them reveals that making this economic decision, or strategic choice, also requires full attention to a host of non-economic considerations that range from trust and community to commitment and the need to communicate intangible value. These insights are captured in recommendations that can help libraries deepen their perspective when considering whether to move forward with collaboration to acquire RDM capacity or, indeed, any capacity.

METHODOLOGY

The project team selected a case study approach as a way to examine multi-institutional library collaborations in the real world, offering a practical counterpoint to the more theoretical approach taken in the *Library Collaboration as a Strategic Choice* report.¹⁰ This results in case study narratives documenting the establishment and evolution of important new library initiatives that we believe are valuable exemplars for the broader library community.

We chose to explore case studies in research data management because it is a fairly new and rapidly growing area for academic libraries. Unlike longstanding library efforts in cataloging or resource sharing, RDM collaborations are an emerging practice. Therefore, these examples may be of high value to readers facing similar circumstances to the ones described in the case studies.

In selecting the case study subjects, our objective was to highlight different collaborative models operating in the RDM space, so we included examples that represented both collaborations through existing organizations (the Canadian Association of Research Libraries and the Texas Digital Library) as well as new efforts initiated by individuals unaffiliated with a guiding parent organization (Data Curation Network). From those institutions, we identified interview participants who served in leadership roles in establishing these collaborations. We limited our study to examples in the United States and Canada. While the case studies described here can be collaboration success stories, readers should not take away that these examples represent the only pathways to successful collaboration, nor are these case studies about technology or product choices. We are focused on the costs of organizing and managing a collective effort, not the operational costs of a particular RDM service.

Each strategic frame—coordination, costs, change, and control—represents an economic concept that is highly relevant to building successful, sustainable collaborations.

We based our interview protocol on the four strategic frames described in the previous report. Each strategic frame—coordination, costs, change, and control—represents an economic concept that is highly relevant to building successful, sustainable collaborations. We also developed questions that explored factors impacting the decision to choose collaboration, future outlook, and lessons learned.

We conducted one-hour, semi-structured interviews with individuals affiliated with each effort. Before each interview, participants received general information about the project and the interview protocol (see full interview protocol in the [appendix](#)). We conducted all interviews via videoconferencing technology with at least two researchers present—including a primary interviewer and a notetaker. Interviews were recorded and professionally transcribed for exclusive use by research team members. In total, we interviewed 20 individuals (seven for Texas Data Repository, seven for Portage, and six for Data Curation Network) in 18 separate interviews. The research team reviewed the transcripts using the four strategic frames for collaboration as a common template for analysis to identify both unique details of each case study as well as to identify common themes to incorporate into recommendations for libraries.

Because of the nature of this study, we are unable to ensure anonymity to interview participants. Instead, while participants are publicly identified in the Acknowledgments section of this report, we do not attribute any specific comments to any individual. We also utilize the non-gendered pronoun “they” when referring to or quoting our interview participants.

Information collected in these interviews was combined with a synthesis of the existing scholarly literature and considerable gray literature available in the form of policy documentation, blog posts, and web pages. Interview participants were invited to review a draft of the relevant case study to ensure accuracy, and their comments were incorporated into the final draft.

This study provides detailed examples of multi-institutional efforts to build RDM capacity through collaboration. These case studies provide a snapshot of these collaborations when we conducted the interviews in 2022; the services and scope of the three case studies have continued to evolve and change. Note that while this report does not address all types of collaborative RDM activities, we hope that it illuminates key practical aspects of library collaboration and will inspire future sharing of perspectives and experiences. Learning from one another is itself a form of collaboration, and one that can strengthen future efforts.

This study provides detailed examples of
multi-institutional efforts to build RDM
capacity through collaboration.

Recommendations for Libraries

The three case studies documented in this report serve as rich exemplars for collaborative efforts to acquire shared capacity to support RDM. Taken together, they provide practical guidance to others exploring the collaboration option. The following high-level recommendations, distilled by the authors from our interviews and research, provide concrete considerations for decision makers as they contemplate whether or not to form a collaboration as well as how to ensure the success of the collaboration, if undertaken. They are organized within three roughly chronological stages of collaboration: preparing to collaborate, establishing a successful collaboration, and stewarding for sustainability.

Recommendations for Libraries: Three Stages of Collaboration



FIGURE 2. Recommendations for libraries: Three stages of collaboration

PREPARING TO COLLABORATE

Recognize that collaboration requires investment

Collaboration is not free. The case studies in this report illustrate that significant investments in staff time and other resources are needed to facilitate the smooth operation of partnerships. These are collaboration's transaction costs or, more simply, the costs of collaborating. As our case studies illustrate, the scope of these costs can be broad: participation in governance structures, working groups, annual meetings, planning sessions, and so on. All collaborations must recognize and clarify transaction costs within their planning efforts, clearly articulating expectations for prospective participants.

In a professional environment where collaboration is esteemed as a core library value, investing effort in collaborations should be part of the recognized labor of librarians. To this end, it should be included in library job descriptions as an acknowledged aspect of advancing the library mission. But more than this, collaboration should be included in incentive and reward structures. Too often, librarians find they must wedge their contributions to collaborations between their regular duties as best they can, perhaps contingent on whether or not they have supportive supervisors. This can lead to neglect of the partnership or, worse, burnout. While contributing to collaborations should be included as a component of many library roles, it should not be simply added to existing responsibilities. Staff time is finite; investing time in collaboration may mean doing less locally.

Balance the trade-off between start-up and consortial models

In this report, we examine collaborations that are initiated in two different ways:

- Within an existing consortium or organization, with this body assuming some or all of the administrative functions and associated transaction costs
- Start-up efforts where no existing structures or support are available, and individuals must expend significant effort to create a collaborative apparatus, define roles, secure funding (likely through external grants), and dedicate resources to the effort

This suggests an important trade-off when choosing approaches for initiating a collaboration. Leveraging existing partnerships—an existing group of partners, collaborative infrastructure, and possibly a funding source—can be helpful, as evidenced by our Texas Data Repository and Portage Network case studies. However, membership may be limited to the existing partnership and other constraints imposed by the prevailing consortial setting. On the other hand, a start-up approach, like the one followed in the Data Curation Network example, can offer greater flexibility, nimbleness, and an opportunity to optimize the mix of partners, but with the added burden of organizing all the necessary operational structures and funding streams. One factor that may weigh heavily in this choice is the fit of existing partners to the collaboration's objective. A start-up approach may be the better option if the current grouping is not well-matched to the objective, despite the additional organizational work and overall uncertainty.

Start collaborations early

In dynamic and emerging areas of practice like research data management, there can be significant advantages to forming collaborations before deep investments are made in other approaches for acquiring needed capacities. For example, both the Portage Network and the Texas Digital Library (TDL) supported the development of shared data repository infrastructure, largely preceding widespread investment in local data repository solutions. As a result, partner institutions considering collaboration did not have to worry about switching costs—the costs of moving from one sourcing strategy to another.

The presence of switching costs raises the threshold of expected benefits needed to induce a prospective partner to join the collaboration because, in these circumstances, perceived benefits now need to exceed the perceived costs of switching to a collaborative solution. This enhances the incentive to form collaborations before deep investments are made in other approaches. It can be much easier to engage a partner to acquire a missing capacity rather than persuade them to abandon existing services in favor of collaboratively sourced alternatives. Collaboration may be an attractive option to early movers in emerging spaces, but perhaps less so for more mature areas with established offerings through other channels.

Value the intangible benefits of collaboration

Collaborations often release a wide range of benefits—some anticipated, others more serendipitous; some straightforward and quantifiable, others resistant to precise valuation. The most tangible benefits—for example, those that result in demonstrable cost savings or that fill a well-defined service gap—are, unsurprisingly, the easiest to measure, evaluate, and communicate, especially as part of justifying institutional participation to senior decision-makers. They are also often prioritized as the key benchmark of the value of the collaboration. While quantifiable benefits are undoubtedly important in evaluating the value of a partnership, it is important not to overlook other, more intangible benefits that flow from collaborative efforts. For example, the significant value collaborations can bring to strengthening professional networks and communities of practice may be overlooked or dismissed because this value is difficult to measure quantitatively. But an immeasurable benefit can still hold high value and contribute substantially to the collaboration's success.

Another example of an intangible benefit is the accumulation of trust through the shared experience of collaboration. Trust, in turn, is important to the success and the prospect of future partnerships. This intangible benefit features in the Texas Data Repository case study: one reason the TDR came to fruition within the Texas Digital Library consortium was the track record of successful collaboration that the latter had accumulated through past efforts. In this sense, collaborating is itself a benefit of collaboration, building up a shared foundation that can create an “option to collaborate” for the future.

ESTABLISHING A SUCCESSFUL COLLABORATION

Leverage preexisting trust relationships

Trust is a crucial ingredient for successful collaboration. Therefore, *preexisting* trust relationships—developed through previous partnerships, personal relationships, or other forms of interaction—can provide a solid foundation for new endeavors. For example, in Canada, the Canadian Association of Research Libraries (CARL) developed trust with library leaders by stewarding a national site licensing program that later became the Canadian Research Knowledge Network (CRKN). This successful precedent offered CARL a strong position to begin its efforts in the new RDM domain. Similarly, the Texas Digital Library had successfully executed many previous projects for the Texas library community, including providing infrastructure for institutional repositories, electronic theses and dissertations, and digital preservation. This built confidence and trust in TDL’s ability to follow through on RDM support.

Preexisting trust relationships—
developed through previous partnerships,
personal relationships, or other forms
of interaction—can provide a solid
foundation for new endeavors.

Research data management often requires librarians to work with non-library stakeholders in areas like information technology and research administration, but this can lead to indifference or skepticism if these stakeholders do not recognize the library’s value proposition in RDM. Therefore, building trust among collaborative stakeholders beyond other libraries is essential. In Canada, CARL leaders recognized that building new trust relationships with these stakeholder groups was necessary and were tireless in engaging with them through national groups and CARL-convened committees. In time, librarian perspectives gained the respect of the other stakeholders. This creation and maintenance of working relationships between individuals and organizational units is essential for the development of multi-stakeholder research support services and something OCLC Research has previously defined as social interoperability.¹¹ Through this effort, the Canadian library community has cultivating trust relationships that, as mentioned earlier, may serve as an option to exercise on future collaborative efforts.

Find your people

Interinstitutional partnerships offer the opportunity to connect with others with similar roles, responsibilities, and challenges. The DCN case study illustrates this vividly: in this case, a group of individuals from different institutions but with similar responsibilities found it efficacious

to band together and pool their expertise to overcome gaps in curation skills. However, this original goal was soon extended by the formation of a cross-institutional community of practice, which, to many DCN participants, represents the most valuable aspect of the partnership.

While connecting to colleagues at other institutions is valuable for any librarian, it can be especially so for those involved in emerging areas—like RDM—where there is little or no local peer network to draw upon, creating feelings of isolation and even alienation from others locally.¹² A cross-institutional peer network can ease the difficulties of taking on a role in a new area characterized by ambiguity, rapidly changing conditions, and indifference, ignorance, or even hostility from local stakeholders. It is an opportunity to pool uncertainty, provide mutual support, and, ideally, find shared solutions. In approaching new partnerships, one DCN participant advises to “feel the fear and do it anyway.” Deep personal networks allow librarians to face their fears together. When no internal colleagues are doing similar work, the peer network must be built beyond the boundaries of the local institution.

Cultivate commitment

It is difficult to pinpoint a single cause-and-effect relationship between commitment and success: Does preexisting commitment and enthusiasm for collaboration make it successful, or do these qualities emerge from early successes? Most likely, both avenues are operative, but as our case studies illustrate, there is at least a correlation between the personal dedication of participants in the collaboration and success.

An important element of building this dedication is cultivating the sense of being part of something that transcends the individual interests of the partners. Collaboration is one of the likeliest library sourcing strategies¹³ to elicit a deep commitment and personal dedication, leading to a sense of ownership, stake, and identity. While this can help drive a collaboration forward to success, there is also a risk that too much personal commitment can cloud judgment on the prospects of the collaboration, leading to a reluctance to exit or dissolve the partnership even when doing so is the optimal strategy.

STEWARDSHIP FOR SUSTAINABILITY

Plan for success

Collaborations are strengthened when they are well-planned from the outset. In fact, the importance of careful planning was exemplified in all three case studies. The DCN established a firm foundation through the time and effort invested in the activities of a planning grant, including a facilitated planning meeting among the founding members. Participants in the TDR noted that the data repository’s speedy implementation resulted from clear objectives. Portage built momentum through a series of quick wins in a planned gradualist strategy.

Collaborations are strengthened when they are well-planned from the outset.

But there are nuances to this insight. Planning is a bespoke activity based on the unique circumstances of each partnership; simply adopting another project’s plans does not account for local circumstances and partners. Planning should also not be excessively prescriptive but allow flexibility in response to community input. Sometimes, the best strategy might be to “lead and follow”: providing overall direction for the collaboration while being flexible and responsive to community interests. For example, even as the Portage Network developed

the Federated Research Data Repository (FRDR), it flexibly responded to community interest in developing a shared Dataverse model. This effort resulted in complementary and robust repository functionality for Canadian research while cultivating greater trust and goodwill for the Portage project.

Be strategically ambitious and tactically restrained

Collaboration benefits from an overarching strategic vision and even ambitious aspirations. But practical success ultimately depends on accomplishment and progress. Attempting to take on too much too early can lead to strategic goals that become untethered from the realities of resource constraints and partner commitments. To make collaborations successful, it is often better to think big but operate small—to have a macro vision but achieve it through micro steps.

A helpful safeguard against overreach is to launch the collaboration based on a well-defined, tractable problem that collective action can solve rather than simultaneously attempting to implement a grand strategy. As demonstrated in the Portage case study, gradualism can effectively build momentum and solidify the partnership. Similarly, DCN's initial efforts focused on the practical goal of establishing a shared staffing model for curatorial expertise; the collaboration later expanded to incorporate a broader strategic aim of building a peer community for data curators.

Accumulating quick wins on circumscribed goals helps build momentum and a sense of progress for the partnership. It also creates a track record of successful collaboration that can serve as a foundation for expanding the scope of the partnership's activities toward a more ambitious strategic vision. Moreover, early successes of this kind can enhance the sense of pride and personal stake in the collaboration's accomplishments.

Accumulating quick wins on circumscribed goals helps build momentum and a sense of progress for the partnership.

Align with institutional goals

Library collaborations benefit when they are understood to provide practical solutions for broad, campus-wide challenges in alignment with institutional goals—not just library values or interests.¹⁴ All three collaborations described in this report benefit from the growing awareness among campus administrators of the importance of research data management spurred by external mandates. In this sense, the RDM-focused collaborations described here serve to advance institutional interests in one form or another.

To support institutional alignment and communicate the value of the library to non-library stakeholders, libraries must increasingly engage individuals and units across the university. OCLC Research has described these efforts to work across internal silos as social interoperability, and a previous report explores the social and structural norms that shape interactions across campus units and offers strategies and tactics for establishing and maintaining successful relationships.¹⁵ Transparency about the collaboration and how it may benefit local stakeholders can help build an institutional interest in its success and sustainability. It can also enhance the library's institutional reputation and highlight its value proposition to campus stakeholders.

Recognize that contributions and benefits may be asymmetrical

Not everyone or every institution has the same capacities, resources, needs, or priorities; therefore, asymmetries in contribution will exist in collaborations. This can seem contradictory to assumptions that equal partnership is necessary to avoid unfair usage of collaborative resources—in economic parlance, free riding. Each of our case studies demonstrated a threshold of minimum expectations regarding participant contributions. Still, the collaborations typically offered significant opportunities beyond this threshold to contribute additional effort through working groups, governance, or leadership capacities. Members will spread themselves across this spectrum of contribution as their capacities and needs permit.

Asymmetries in benefits realized from collaboration may also exist. Some individuals may accrue significant personal benefits from their deep engagement in the form of greater knowledge, personal networks, and career mobility. Likewise, some institutions may benefit by gaining greater access to expert knowledge. Some of these individuals and institutions may make relatively small contributions to the collaboration in return. But this, in a way, is a benefit in itself. In fact, interview participants from each of the case study organizations emphasized the goal to broadening RDM services and participation beyond the most well-resourced research-intensive universities, recognizing that smaller institutions may have fewer resources to contribute and may even disproportionately benefit. As one Canadian interviewee said, “It’s part of your responsibility to build these services that . . . the entire community can derive benefits from” to “raise the level of performance for themselves and everyone else.”

Case Studies

The following case studies document how three RDM collaborations began and were sustained. While there are similarities, each case study is also distinctive in its collaborative approach. The Texas Data Repository (TDR) case study demonstrates how an existing consortium expanded its service offerings to include RDM support. In Canada, the Portage Network case study is a compelling example of how a library organization established RDM services that later became a nationally supported infrastructure. And in the case of the Data Curation Network, individual data curators came together to create a brand-new organization to support shared data curation.

CASE STUDY ORGANIZATION

The three case studies in this report follow the same format, drawing from the concepts presented in the *Library Collaboration as a Strategic Choice* report.

Introduction—A brief overview of the collaboration that will be the focus of the case study.

Origin story—A brief synopsis of how the collaboration originated, its major functions, and key milestones in its evolution.

Choosing collaboration—A description of the environmental factors shaping the collaboration decision, along with a discussion of priorities and trade-offs associated with that decision.

Strategic frames—An analysis of the case study using the four economic concepts of coordination, costs, change, and control. Each frame provides insight into how the collaboration responded to a specific economic issue in forming and managing the partnership and its activities.

Looking ahead—A speculative discussion of what lies ahead for the collaboration, including potential obstacles.

Lessons—General advice on achieving successful, sustainable library collaboration, based on the experiences of those involved in the collaboration.

Taken together, our case study framing offers a unique perspective on library collaboration in action:

- Why collaboration was chosen as the strategy for engaging in the RDM space
- What participation in the collaboration entails
- Investments in staff time and other resources needed to initiate, administer, and sustain the collaboration
- Pathways and obstacles encountered in switching to a collaborative approach
- Allocation of decision-making authority in governance structures
- Speculation on future opportunities and challenges
- Lessons learned from the collaborative experience

While the focus is distinctly economic and omits other important elements that combine to form the full collaboration story, we believe that highlighting the economic facets of collaboration will help libraries manage it strategically, and ultimately, maximize the likelihood of successful outcomes.

Texas Data Repository

INTRODUCTION

The Texas Data Repository (TDR) is a venue for faculty, staff, and students from institutions who are members of the Texas Digital Library (TDL) community to openly publish their datasets. TDR is a service of the TDL, which began in 2005 and today exists as a consortium of 30 institutions with a mission to build “capacity among its membership for ensuring equitable access to and preservation of digital content of value to research, instruction, cultural heritage, and institutional memory.”¹⁶

The TDR consists of a platform based on the Dataverse software, as well as a community of practice within a collaborative structure. Its aim is to curate and make freely available research datasets from any scholarly discipline. The TDR allows users to comply with funding requirements, ensure reliable long-term access to datasets, increase their scholarly impact, and collaborate with research teams before open publication. All this is accomplished with support from their campus librarians, who provide assistance, education, and guidance.

For member institutions, the TDR offers a centralized service for the significant tasks of providing data management and curation support. The governance structure of the TDL includes a member board that meets annually, a governing board that consists of library leadership from seven TDL institutions, and the executive committee. The executive committee is a subset of the governing board. In addition, user groups, working groups, committees, and affinity groups allow members to discuss and move forward on issues that are pertinent to them.¹⁷

ORIGIN STORY

Conversations about research data management within TDL long preceded any national public access mandates.¹⁸ Indeed, interest in supporting research datasets existed from the beginning of the TDL, with a mention in their 2006 business case publication.¹⁹ In 2013, the OSTP Public Access Memo²⁰ served as a substantial motivation to move forward with RDM, and TDL invited nine representatives from Texas institutions of varying sizes, public and private, and different research foci to form a Data Management Working Group. They

were empowered to self-educate regarding the storage and accessibility of data sets, recommend services, and nominate pilot projects for implementation. In August 2015, this group recommended the implementation of a Dataverse platform as a pilot project, based on its robustness, usability, scalability, and congruency with TDL's commitment to open source software. They further recommended creating the Dataverse Implementation Working Group (DIWG) to pilot this data repository.²¹

The DIWG was composed of 14 librarians and technical staff members from six universities who evaluated the costs, possible funding models, technical configuration, workflows, outreach, policy, governance, and metadata needs. Working within four subgroups—budget and business plan, policy and governance, technical configuration, and workflows and outreach—the DIWG drafted documentation and built the infrastructure.

In May 2016, the group beta-tested the new data repository by inviting librarians and other volunteers to complete tasks and provide feedback. The data repository is a single shared Dataverse instance that can be locally branded by each member, a desired configuration for efficiency and centralized maintenance. Following testing, they created information sheets for administrators, researchers, and librarians, expanded recommendations and guidelines, delineated the roles and responsibilities of users, revised navigation and branding of the infrastructure, created a website, and christened the repository the “Texas Data Repository” (TDR). The group also made recommendations for future efforts, including a TDL Data Symposium, in-depth training sessions, a statewide steering committee, and the need to address unmet digital preservation needs. The speed at which this effort was accomplished was extraordinary, partly due to a focus on the planning and the existing collaboration infrastructure of the TDL. One interviewee reflected:

Looking back, I cannot believe we did all of this in a year, from August [2015] to September [2016]. There was nothing and then there were a set of services, policies, and a repository. It was because we knew exactly what we wanted to do, and we had the people power to get it done. It also doesn't hurt that a lot of the barriers that often come up with collaborative processes were managed by TDL.

Importantly, the DIWG also explicitly articulated the TDL's service model as it applied to research data management: “We believe a consortial implementation best supports the development of new data management services and programs at individual institutions and provides curation services at a lower cost per institution.”²² The TDR was formally launched in December 2016, and in addition to the software platform, it is very much focused on the community of practice supporting it, evolving to meet the needs of the librarian community to develop mature RDM services. As one interview participant emphasized, the TDR “is not equal to [only] our Dataverse installation. The Texas Data Repository is the steering committee, the Dataverse implementation, and then the whole curation world that we create around that together, the shared resources, the documentation, the webinar series, all of that.” To date, this collaboration has resulted in TDR's 14 member institutions contributing over 1,819 datasets as of November 2023.²³

Texas Data Repository Timeline

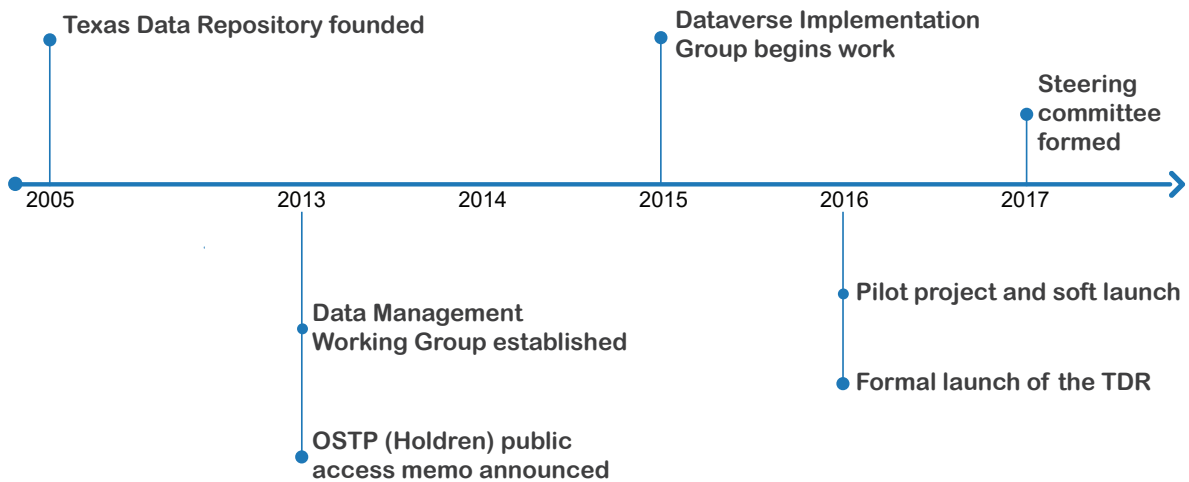


FIGURE 3. Texas Data Repository timeline

CHOOSING COLLABORATION

One of the most significant factors cited for choosing to collaborate was practical budgetary considerations. Interviewees generally acknowledged that budget and resource constraints created a strong desire for collaboration. Structuring the TDR as a single shared repository service resulted in some key efficiencies, including fewer staff to maintain it and a common environment for all institutions (even though they may be locally branded), resulting in common documentation and peer learning among the member institutions. Centralizing much of the work also allows institutions to put forth minimal initial investments or sunk costs, particularly if they lack the staff to run a repository. One curator referred to themselves as a “department of one for a very long time, and the only way I could be successful was because the Texas Digital Library existed.” For both the individual institutions and their local campus librarians, this collaboration is a cost-effective solution for developing RDM capacity.

The TDR collaboration was both a venue for professional development and a substantial source of community.

Participants also cited shared values and established trust as important collaboration incentives. The TDR was designed and operated by the TDL, a trusted entity with a track record of building community and shared capacities among its membership. In addition to this preexisting trust, many interview participants cited community as a primary motivator or reward of the TDR service: “A big reason that community is so valued around the research data repository, is because the skill level in Texas was still relatively low, and folks saw an opportunity to learn together, share resources, share strategies. And so, I think that was a draw [for people to join].”

The TDR collaboration was both a venue for professional development and a substantial source of community. While not as easily measured in the budget as other software or staffing support aspects, these components foster commitment and contribution to the effort.

While interviewees acknowledged both budget and values-based considerations as motivators to choose collaboration, some noted that the balance of these two influences may differ by institution. Some interview participants noted that the motivation to collaborate may differ based on the size of the institution because larger institutions often have greater resources than smaller ones: “The budget is the strongest motivator for collaboration when you’re looking at our member institutions that are smaller . . . [but] it’s more value-based when you look at some of our larger consortium members.” Another participant summed it up best as “especially some of the smaller institutions, it was either joining the Texas Data Repository or [not] get a data repository at all.” For these institutions, this lack of any other alternatives removed any costs associated with switching from an established data repository to the TDR.

However, there are trade-offs to structuring a data repository within a collaborative setting, as opposed to institution scale implementations. This includes the inability of individual institutions to freely incorporate preferred features, in addition to a potential slower speed of deployment and decision-making. One example of a trade-off was the need for members to define the parameters for datasets that are accepted for curation. Some institutions wanted to be able to accept larger datasets or data with more sensitive elements than is currently possible. Such capabilities are currently being explored.

To date, the TDR user base consists of 14 institutions,²⁴ most of them large and heavily focused on research. With 30 potential members in the TDL consortium, this is a lower uptake than some had anticipated and seems to indicate potential as-yet-unidentified trade-offs. As one interviewee reflected:

[It was] never my hope that less than half of the membership would be joining this. I thought they would be coming out in droves . . . and I don’t know why others aren’t, other than maybe they just don’t see it as a need.

While the lack of high levels of participation may be due to unidentified trade-offs that make the TDR less attractive to some potential members, it is also possible that sufficient RDM demand is only present in a portion of the TDL membership. Another interviewee speculated:

There are some institutions in TDL that are quite small, maybe they’re undergraduate-focused and don’t really have the large amount of research expenditures that are necessitating the incredible need for data archiving and sharing over time.

These trade-offs—the lack of customizability, slower speed of deployment and decision-making, technical limitations, or as-yet-unidentified ones—may lead institutions to find solutions other than the TDR to meet their data management needs. However, it is also possible that non-participating institutions may not be experiencing strong incentives to address RDM needs at all, which is itself a form of trade-off: RDM vis-à-vis other priorities.

STRATEGIC FRAMES

The TDR benefited from established institutional relationships and trust from the preexisting consortium TDL, clear roles and responsibilities, and intentional development of community. Part of this clear intentionality was the explicit expectation that beyond minimal obligations, there would be unequal participation. This prioritized diversity of membership over free riding concerns and lowered the barrier to participation. Today, decision-making control is centralized, and institutional budgetary costs are clearly articulated. This well-organized coordination minimizes resistance to change and increases buy-in from local librarians. However, buy-in from stakeholders outside the library has been slower, and less visible costs, such as staff

time, are still overlooked and often significant. The TDR case study illustrates the benefits of careful planning and building on existing collaborative relationships but faces challenges of low participation and communicating the true value of the collaboration to those outside the library.

Coordination

The TDR successfully coordinates its group-scale effort by defining institutional responsibilities, ensuring minimum institutional obligations, encouraging flexible participation, and demonstrating a strong commitment to supporting the community of practice. Their service level agreement clearly defines roles and responsibilities, from the liaison librarian role to the service and subscription fees.²⁵ At its most basic, to be a member in good standing, the minimum responsibility includes a subscription fee and the designation of at least one liaison to represent the institution.

Beyond the minimal requirements, there are flexible expectations regarding individual participation and benefits. The TDR reflects its parent organization, the TDL, in this respect. This is contrary to more traditional assumptions that equal participation and benefit accrual is the ideal, as it prevents members from free riding, or unfairly gaining more benefits than their contribution merits. Explicitly acknowledging that some institutions may gain more than they contribute is essential when different individuals have different resources. As one interviewee phrased it, “having the flexibility, realizing that . . . not everything has to be distributed evenly . . . realizing that some people just have less capacity because of where they’re coming from.” As one study interviewee stated, “A goal with all of our [TDL] services [is to] provide services that a variety of different institution types can take advantage of flexibly.” That means that some level of participation is available to institutions, whether they are large or small, private or public, well-resourced or not. “We think about ways that those different types [of institutions] can . . . flexibly interact with the services we’re developing. . . . We don’t have a single service [offering] that every one of our members is engaged with.” The TDL, and subsequently the TDR, has been very intentional in designing services that are inclusive of different institution types and accommodating participants with differing levels of support.

This flexibility, combined with clear baseline expectations, contributes to the strong community aspect of the overall TDR community. Examples of the social components of TDR include training, policy, governance, workflow, management, and needs assessment. Common themes among interviewees included learning from others in the collaboration, receiving both technical and intellectual support, and being passionate about this work. Building this strong sense of community was intentional: “If you’re first starting something, you need people who are enthusiastic, and people who realize that the sum is more than the parts and that we’ll get more out of it if we work together.” Even after the initial work was completed, TDR has maintained a welcoming and nurturing culture: “I feel like people were so open with me when I first joined and very, very helpful, [and that makes] you want to return that favor to the next person.”

“Library consortia . . . need to champion the social side as much as they champion the IT side.”

Again, this perspective mirrors the TDL's strong sense of commitment among its members:

Library consortia like this need to champion the social side as much as they champion the IT side . . . because it is on the social side that really makes TDL work so well. . . . My colleagues really enjoy collaborating with TDL. They get credit for collaborating with TDL, they see it as part of their main mission, they enjoy it thoroughly, and [have] moved things forward.

Therefore, even though all members of TDR are considered members in good standing if they meet minimum requirements, the community is a highly valued component of the TDR, and individuals are strongly motivated to work together. Indeed, the themes of prioritizing diversity and cultivating a strong, welcoming, and nurturing community appear to be common in the TDL approach.

Costs

The financial costs of utilizing the TDR service are fairly straightforward and delineated in the service level agreement.²⁶ However, the transaction costs or costs of collaborating are not as straightforward. There is clearly an investment of time, particularly at the beginning of the collaboration and when issues arise. One interview participant cautioned against underestimating the amount of work it takes to collaborate and ensuring sufficient capacity to manage the collaboration:

Collaboration doesn't just happen, it requires a lot of care and feeding . . . a lot of coordination, a lot of pushing, a lot of persuading, and a lot of wrangling, and so I think that we underestimate constantly how much work it is, and we undervalue the people who do that work.

Additionally, interviewees cited the less visible cost of the additional roles that member liaisons are expected to play outside of their primary duties, including sitting on the steering committee that meets monthly, one annual meeting, routing or answering help desk requests, and direct inquiries. There is also the option to sit on sub-committees, which may require additional time, and periods when more time is required to fix unforeseen issues. While harder to quantify than the membership fee, there are clearly transaction costs that must be assessed when evaluating the return on the investment of the collaboration.

Collaboration doesn't just happen, it requires
a lot of care and feeding.

Change

As anyone who has tried to implement change knows, trust is key to overcoming resistance. In our interviews, preexisting trust was cited repeatedly as the overriding factor for librarians in deciding to collaborate in the TDR, and thus buy-in to joining the collaboration was easy to cultivate. The TDL has been providing services such as institutional repositories, open access journals, electronic thesis and dissertation workflows, and digital preservation to the Texas library community since its inception and has already demonstrated a strong track record of successful collaboration.

Because of this trust in TDL's ability to deliver shared resources, the local environment was very favorable to making the change to collaboration. One interviewee even stated, "I don't think that the [Texas] data repository would have happened if TDL hadn't already existed. I think individual institutions would have either just not done anything or [the larger institutions] would have had their [own repositories]." Furthermore, the initiative was at an advantage because it began relatively early in the development of RDM infrastructure; few institutions had developed RDM repositories, meaning that there were fewer switching costs for institutions to consider when choosing to invest in the collaborative TDR option.

Because of this trust in TDL's ability to deliver shared resources, the local environment was very favorable to making the change to collaboration.

While trust within the library sphere was well established, selling the TDR to stakeholders outside the library is a different story, with some participants reporting indifference or lack of understanding. One study interviewee described their IT units as lacking in interest:

The prior [high-performance computing unit] was not focused on these things . . . they're very much about the active part of research, and once you're done with that active part where you are to the point where you can begin to archive and share your data; that's outside their scope . . . [while campus IT is] all about . . . making [sure] your data is secure and protecting intellectual property and all those sorts of things.

Similarly, some researchers were averse to sharing data or believed they had adequate solutions in place. As one interview participant related: "We found that a lot of departments and disciplines already had repositories going. It was like Math or Economics or something, and they said, 'Everybody in Math puts their stuff here. We don't really need you. We're taking care of ourselves.'" However, at times the "repository" in question was actually a locally maintained file management system that did not meet the accepted standards of a true data repository, and when IT units and researchers realized that the libraries were providing a more robust solution, they welcomed their assistance in changing their practices. Some stakeholders outside the library may need to see the TDR in action to completely understand how the libraries' decision to collaborate benefits them.

Interview participants reported a high level of trust from researchers in both their institution's libraries and the TDR, with one interviewee stating that they found that minimal promotion was necessary to ensure good usage. Although some external units struggled with understanding the role that TDR plays, no one actively opposed the library providing this multi-institutional solution to address the challenges around data management and sharing. Indeed, university administrators see data sharing as a compliance issue and are very happy to have the libraries contribute to the institution's prestige and reputation, a key value when considering whether to collaborate.²⁷ Overall, change was not an issue for members of the TDR as trust was so prevalent in large parts of the community.

Control

The TDR collaboration is structured in a way that clearly establishes who has decision-making control. Collaborations can inevitably result in conflicts around priorities and resources. To be successful, there must be an individual or group that serves as the locus of control and is empowered to resolve conflicts effectively, ensuring that the decisions of the collaboration are aligned with the goals and needs of its constituencies. If the locus of control is unclear or ineffective, the collaboration risks stagnation, or worse, collapse.

While interview participants generally acknowledged that the needs of the users drive TDR decisions, they also have a structure by which those ideas are elevated, as it is not possible to accommodate all ideas. Annual roadmap discussions result in sub-committees that require volunteers, so collaboration members must commit time to champion the idea. The governing board, which sets initial priorities, the TDR liaisons on the steering committee, and the TDL Service Manager all influence what moves forward. Group note-taking, extensive documentation, and explicit roles and responsibilities ensure that institutional memory is captured and that newer members can come up to speed quickly. However, the ultimate decision maker, or locus of control, is the director of the TDL. This single source of authority ensures that there are realistic prioritizations and actionable next steps to balance the participatory environment.²⁸

LOOKING AHEAD

The TDR hopes to add more member institutions in the future as they grow in capacity, including the ability to accept different sizes and types of datasets, improve services for smaller or less-resourced institutions, and increase their collaborative activities with the larger Dataverse community. This seems likely as more libraries see the value of greater productivity and efficiency from distributed expertise and labor.

“[T]he institutions that solve this problem first are going to gain significant prestige and reputation. Because their data will be accessible, discoverable, and useful. . . .”

In addition to the increasing pressure for researchers to openly share their data from federally funded research,²⁹ successfully sharing research data supports the larger vision of open science, and in the opinion of one interview participant, “the institutions that solve this problem first are going to gain significant prestige and reputation. Because their data will be accessible, discoverable, and useful. . . .” If this prediction is true, then this may be an important driver for other TDL members to engage in the TDR offering.

LESSONS FROM THE TEXAS DATA REPOSITORY EXPERIENCE

- **Be intentional.** Being intentional, as opposed to being reactive, results in a clear understanding of the vision and greater continuity of action. One interview participant commended the TDL Executive Director's intentionality and vision in terms of scoping discussions, gathering expertise, and assembling the right people around the table. This clear purpose results in focus and action, both necessary components for getting things done in a successful collaboration.
- **Create collaboration infrastructure.** Examples of collaboration infrastructure include user groups and steering committees, clearly articulated roles and responsibilities, and extensive documentation. This infrastructure benefits the collaboration by supporting the continual engagement of users and stakeholders, promoting consistency through personnel changes, providing accountability for outcomes, and may even be repurposed for other collaborative efforts.
- **Plan.** TDL's planning efforts were clear in terms of overall goals, yet flexible and phased in terms of actions taken. TDL progressively leveraged working groups to support planning at all stages of TDR development: "We knew clearly what we wanted to do. And we knew that planning, [while] not exciting . . . [is] so critical to this process."
- **Cultivate community.** TDL actively engaged its members through working groups in the planning, implementation, and maintenance stages of the project. While technology-focused efforts can be seen as separate or niche, they are actually pervasive and fundamentally entwined in librarianship endeavors.³⁰ This interdependent relationship between the social and technical aspects of library efforts means that cultivating community is as essential to the success of collaboration as getting the technology in place.
- **Support inclusive participation.** Contributions to the TDR come from a diverse set of institutions with differing resources. Explicitly acknowledging that TDR members do not benefit equally, nor are responsibilities distributed equally, allows different types of institutions and individuals to contribute and ultimately, accomplishes the overall goal of responsibly stewarding datasets.

The Portage Network

INTRODUCTION

The Portage Network was an initiative led by the Canadian Association of Research Libraries (CARL) from 2014 until its integration into the Digital Research Alliance of Canada (referred to as the Alliance elsewhere in this report) on 1 April 2021.³¹ Under the leadership of CARL, Canadian libraries pooled financial and staffing resources to develop shared infrastructure and a network of experts to support research data management in Canada. CARL's membership includes 29 university libraries, plus Library and Archives Canada and the National Science Library. As the most research-intensive libraries in the country, these libraries are all concerned with the growing need to support data management practices at their institutions.

The Portage Network narrative that follows is complex, documenting a wide array of stakeholder groups and organizations as well as the significant challenges of working in an emerging area with such an intricate stakeholder environment. It's a powerful example of how a consortial-initiated project was developed into the national infrastructure of the Alliance.

Glossary

This glossary lists and describes key organizations that played a direct role in the development of Canadian RDM infrastructure highlighted in this case study. Readers can use this glossary as a reference tool as they read the Portage Network's story.

CANARIE, formerly known as the Canadian Network for the Advancement of Research, Industry and Education, operates key infrastructure supporting Canada's national research and education network (NREN). It served an intermediate funding role to Portage as it transitioned into the Alliance.

Canadian Association of Research Libraries (CARL) established and led the Portage Network until it was incorporated into the Alliance in 2021.

Compute Canada previously coordinated advanced computing infrastructures for Canadian researchers. It was subsumed within the Alliance in 2021.

Digital Research Alliance of Canada, or the Alliance, is a non-profit organization funded by the Government of Canada that provides advanced research computing, research data management, and research software infrastructure. The Portage Network was folded into the RDM directorate of the Alliance.

Innovation, Science, and Economic Development Canada (ISED) is a department of the Government of Canada with responsibility and oversight for many federal functions, including the promotion of science and innovation. It convenes the LCDRI.

Leadership Council for Digital Research Infrastructure (LCDRI), comprised of senior executives from research universities and governmental organizations, was established and funded by ISED. During the 2010s, it provided thought leadership through position papers that shaped what would become the Alliance.

National Research Council (NRC) is an agency of the Government of Canada that seeks to support innovation and R&D capabilities through strategic partnerships. It convened the RDSWG in 2012.

New Digital Research Infrastructure Organization (NDRIO) was the temporary name of the organization that was rebranded in 2021 as the Digital Research Alliance of Canada.

Project ARC was the initial name of the Portage Network from 2014-2015.

Research Data Alliance (RDA) is an international research organization established in 2013 with the goal of facilitating open data sharing through international collaboration.

Research Data Canada (RDC) was established in 2012 to work at a policy level with Canadian stakeholders nationally, and internationally with the Research Data Alliance (RDA). It was subsumed by CANARIE in 2015.

Research Data Strategy Working Group (RDSWG) was a short-lived but strategically important group established by the NRC in 2008. Comprised of stakeholders from many groups, the group led discussions and events that recognized the need for greater RDM cooperation in Canada. It was reorganized into RDC in 2012.

Tri-agency is the umbrella term used to describe the three Canadian governmental research funding agencies: the Canadian Institutes of Health Research (CIHR), the National Sciences and Engineering Research Council of Canada (NSERC), and the Social Sciences and Humanities Research Council (SSHRC).

ORIGIN STORY

Before piloting the Portage Network in 2014, there were several other initiatives intended to advance research data management efforts and infrastructure in Canada, reflecting the growing interest of the Canadian government and federal granting Tri-Agencies in increasing the accessibility and transparency of Canadian scientific research.³² Of particular relevance was the establishment of the Research Data Strategy Working Group (RDSWG) by the National Research Council in 2008, composed of stakeholders from federal government institutions, universities, libraries, funders, and researchers. In 2011, the RDSWG convened a Research Data Summit that resulted in the report, “Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit,” which recognized the need for greater coordination of RDM activities in Canada.³³ RDSWG was reorganized into Research Data Canada (RDC) in 2012, with a mandate to work at a policy level with Canadian stakeholders nationally, including representing Canadian interests to the newly established international Research Data Alliance (RDA); in 2015 the Canadian Network for the Advancement of Research, Industry and Education (CANARIE)³⁴ assumed oversight for the RDC.³⁵

The genesis of the Portage Network began in March 2014 when CARL launched a pilot effort to establish a library-based RDM network in Canada, initially called Project ARC. It was steered by a working group of library leaders from across Canada and included representation from each of Canada’s four regional library consortia, as well as the Canadian Research Knowledge Network (CRKN) and the RDC.³⁶

Over the next several months, the Project ARC working group defined goals and developed an organizational framework, operational plan, and initial funding model. The Project ARC pilot identified two main goals with the formal launch of the Portage Network in March 2015:

- The development and stewardship of a library-based network of experts (NOE) in RDM
- The development of shared infrastructure to support data management, preservation, and discovery

In September 2015, CARL hired the inaugural Portage Network director. Over time, additional staff were added to help manage the Portage effort.³⁷ The Portage Network’s first project was developing the bilingual, online DMP (data management planning) Assistant. The Portage DMP experts group developed the Canadian DMP Assistant by building upon the existing DMP offering developed by the University of Alberta Library, which was based on the DMP Online tool created by the UK Data Curation Centre. The DMP Assistant was launched in October 2015 and continues to offer researchers information, examples, and a step-by-step guide for creating a data management plan; it provides institutions with the option to brand it as their own local tool.³⁸ Soon, a data preservation expert group was also established, and more groups followed, on topics such as curation, discovery and metadata, and data repositories.³⁹

Portage also turned its attention to developing a national preservation and discovery platform, with the goal of developing shared infrastructure that could enable participation from all interested Canadian universities. Software development of the platform that would later become the Federated Research Data Repository (FRDR) began in 2016 in partnership with Compute Canada—which provided initial funding in the amount of CAN\$2.2 million—followed by beta testing and a limited production launch in 2017.⁴⁰ Additional funding from national sources supported further development, and FRDR went into full production in 2021 and today is maintained by the Alliance.⁴¹

Following input from the Portage community, Portage also convened an expert group called Dataverse North in 2017 to explore a national, bilingual institutional repository service using the open source Dataverse platform.⁴² As with the DMP Assistant, to gain a “quick win,” Portage once again leveraged an existing local implementation by a CARL member institution—in this case, the existing Scholars Portal Dataverse managed by the Ontario Council of University Libraries (OCUL) and hosted by the University of Toronto. The Dataverse North expert group developed a guide to metadata best practices, training materials, and a proposal for a Canadian Dataverse Repository (managed by OCUL), along with information about business models and cost recovery that could support decision-making.⁴³ Universities, libraries, and researchers desired local branding of data repositories, but this potential point of friction could be addressed by the architecture of the Dataverse community itself. Inside the primary Dataverse instance supported by OCUL, additional Dataverses could be nested within each other, accommodating local branding requirements for a modest fee. This resulted in the national Borealis Canadian Dataverse Repository.⁴⁴

Today there are two general-purpose data repositories in Canada whose offerings are complementary. For example, FRDR, which is hosted on Canada’s national high-performance computing infrastructure, supports large data files, while institutional Borealis instances support files smaller than 5GB. Borealis supports versioning, local institutional branding, and some active data management functionality, while FRDR does not. Borealis eligibility also depends upon institutional participation; the use of FRDR does not.⁴⁵

Today there are two general-purpose data repositories in Canada whose offerings are complementary.

CARL provided the initial support for the Portage Network, allocating CAN\$200,000 per year from September 2015 through August 2017. Following the development of a comprehensive business plan in 2017, CARL members agreed to a special levy to increase the base Portage budget to about CAN\$300,000, which created new incentives for all CARL libraries to more deeply engage, including by dedicating their own library staff to participate in the growing network of experts. In 2019, Innovation, Science and Economic Development Canada (ISED) provided CAN\$2 million⁴⁶ to support Portage and the plan for its incorporation into the newly established national organization to operationalize national digital research infrastructures, particularly in support of advanced research computing, research software, and research data management. This organization was initially called the New Digital Research Infrastructure Organization (NDRIO), later rebranded as the Digital Research Alliance of Canada. This was a great influx of resources, removing the need for financial support from CARL library directors, and by 2021, Portage was fully integrated into the Alliance, which continues to support RDM nationally through tools, services, infrastructure, training, and a network of experts.

Portage Network Timeline

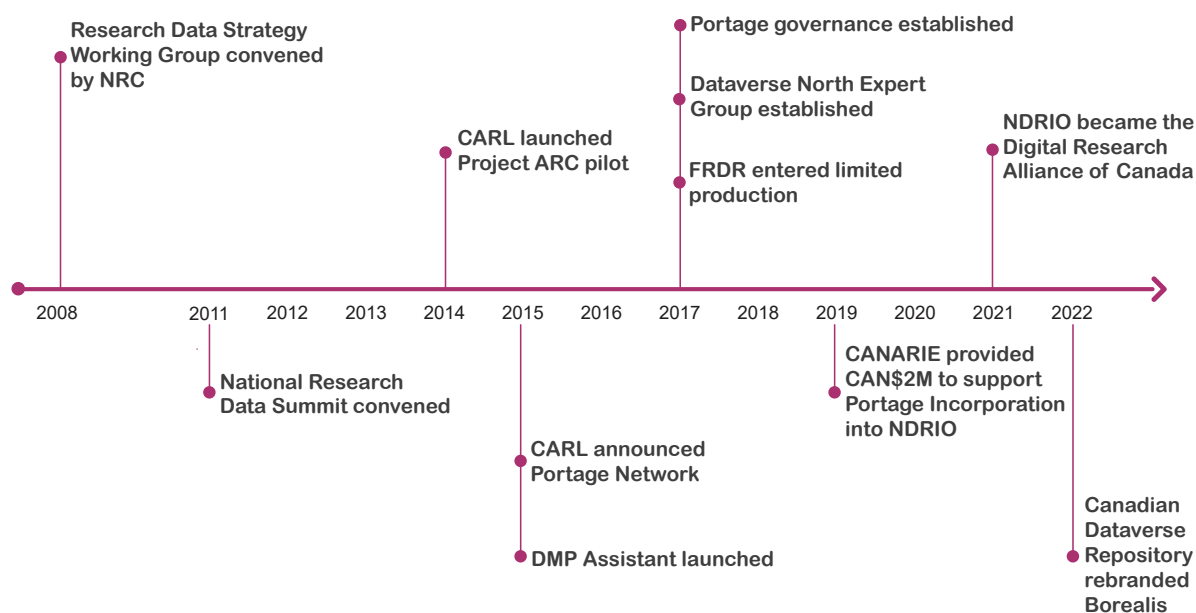


FIGURE 4. Portage Network timeline

CHOOSING COLLABORATION

There were several environmental factors that favored a collaborative approach toward the provision of research data management services in Canada. Perhaps the most powerful factor was a strong, preexisting belief in collective investment held by the Canadian library community. One interview participant commented that the Portage Network was “created originally from a collection of university libraries in Canada that were convinced that the only way forward is to be collaborative and create the network of experts and the partnership.” Another described how “there’s almost a sense in Canada that if you are larger and well-funded, it’s part of your responsibility to build these services that can support services the entire community can derive benefits from,” because networked investments are seen as a way to “raise the level of performance for themselves and everyone else.”

This ethos of collaboration seems systemic. An interview participant described how “most librarians or most academic librarians in Canada have a service component to their jobs where you’re expected to contribute to the academic community as part of your professional work.” In other words, participation in external networks and working groups like the Portage Network of experts is not an activity to be done in addition to local work responsibilities but instead is defined as a significant part of those local responsibilities. In some cases, the expectation to participate in collaborative activities is written into librarian job descriptions and also demonstrated through secondments to other organizations.

It’s also notable that discussions about a national approach to RDM services were initiated before many institutions had begun investing in locally provisioned services. With few legacy systems in place, this meant low or even non-existent switching costs—the costs incurred when changing from one sourcing strategy to an alternative—for most institutions. Instead, with no current solutions in place to a growing need, the partnership option was an attractive solution.

Finally, CARL was well positioned to lead because it was seen as a trusted leader and convener around the interests of Canadian research libraries. Several interview participants described CARL’s stewardship of the initial national site licensing program, which led to the establishment

of the Canadian Research Knowledge Network (CRKN), as a template for the success of Portage. In that effort, CARL took the lead in raising community awareness, developing relationships, and uniting 72 university libraries in Canada in a consortial purchasing arrangement that provides significant cost savings through cooperation.⁴⁷

CARL was well positioned to lead because it was seen as a trusted leader and convener around the interests of Canadian research libraries.

We asked our interview participants to identify the priorities and trade-offs they considered when deciding to collaborate on the Portage Network. Several of our informants described the highest priorities to be values-driven: considerations like community, trust, and reputation. Many described how the Canadian library community believes in collective investment. We also heard that collaboration was required to have enough influence to exert any kind of meaningful change at the national level. Collaboration offered a path to a national discussion on RDM service provision in a way that no individual library could achieve independently by offering “a collective voice that is much stronger than the voice of the individual institution,” and giving libraries a type of legitimacy with other stakeholder groups. This enabled the Portage Network—and by extension, CARL libraries—to position themselves as national stakeholders in the research data workflow, at the level of other national stakeholders, such as those in high-performance computing.

Working collaboratively could also support goals of quality, efficiency, and even speed. One interview participant remarked, “You can just do that a lot more effectively if you’re taking the best and brightest nationally, as opposed to just what you have in your own institution.” By providing leadership and convening a community of practice, the Portage Network facilitated rapid information sharing and learning among librarians, which they could, in turn, pass on to others across the broader community. And because the effort was led by CARL, which provided initial support and resources, it was still a fairly nimble effort, with centralized decision-making in the CARL executive director and board even as input and guidance were gathered from CARL members.

It’s notable that shared costs and cost savings were not primary drivers at the start of the effort; it began more as an effort to harness available but limited and distributed expertise to define a vision, guiding values, and initial quick-win actions. There was also little risk to CARL libraries at the beginning of the effort because CARL provided initial start-up support—there were no financial requests from CARL libraries until 2017. Portage was also able to reduce risks within specific projects by leveraging partnerships and building on preexisting local implementations. For example, Portage developed the DMP Assistant by leveraging the existing DMP Assistant developed at the University of Alberta; partnering with Compute Canada on FRDR is another example. By reducing risk and investment for member libraries early on, the collaboration option became a more compelling sourcing choice.

STRATEGIC FRAMES

The Portage Network began with the active leadership of just a handful of CARL libraries, increasing potential risks for collective action problems. Although there was an unequal distribution of resource contribution—larger, better-resourced institutions tended to be

more involved in the initiative—the project was able to gain momentum quickly by building upon existing efforts by member libraries, including the DMP Assistant and Dataverse efforts. Leveraging the existing trust it had built with the Canadian library community, CARL underwrote the initial costs for starting the project, with later investment from CARL libraries as confidence and buy-in increased. Because the collaboration was established when RDM uptake was relatively immature, there were fewer barriers to collaboration, instead offering significant incentives for choosing the collaboration option. CARL invested time in developing relationships and buy-in with a complex array of stakeholders by convening Portage committees for both librarians and non-library stakeholders, as well as by participating in national conversations convened by others. These efforts led not only to support for the project but also to a growing appreciation for the value proposition of libraries as well as the necessity of including RDM within a nationalized digital research infrastructure.

Because the collaboration was established when RDM uptake was relatively immature, there were fewer barriers to collaboration.

Coordination

With 31 libraries participating in CARL, the group was fairly large, increasing the risks to the collective effort because it can be more difficult to achieve consensus with more participants. This was offset, however, by the relative homogeneity of the participants as well as the strong Canadian ethos of collaboration. Furthermore, another mitigating factor was that the group of institutions were not strangers to each other and instead had strong preexisting trust relationships facilitated by CARL.

However, Portage was not immune to the collective action problem. Initially, there were only about 10 CARL institutions that were leading the effort through investments in time and money. One of our interview participants commented that these were primarily the larger, better-funded institutions that provided leadership on behalf of the whole and were able to contribute more due to greater staff and resources.

Project leaders sought to secure buy-in from others by establishing credibility and transparently articulating Portage's values in a clear set of organizational principles and operational guidelines as the project launched in early 2015.⁴⁸ One of our interview participants stated,

I think there is a tendency to respect action, and we were very careful . . . in terms of making sure that outputs were visible, credible, [and] respect-worthy . . . to try to make sure that what was getting talked about or presented was based on the wisdom of the collective.

This included developing an internal library directors' governance committee as well as an external stakeholder advisory committee and ensuring that skeptics and supporters alike had seats at the table.

The second part of their strategy was to secure well-articulated "small wins" to help build trust, commitment, and momentum in the collaborative effort.⁴⁹ The development of the DMP Assistant bilingual data management planning tool for the Canadian environment was their first small win.⁵⁰ This was a well-defined project that everyone agreed was needed. This encouraged

greater engagement, as demonstrated in this quote from one of our interview participants: “How do we get them to come in and join? You start small and succeed! And when you have a group of eight to twelve librarians working on a DMP Assistant, and they do their thing, then you say, ‘Okay, well, we need a group on preservation. Do you know anybody?’” Another said, “We started small, got the quick wins, and then just kept motoring along.”

By 2017, CARL library directors began financially supporting the effort. This wasn’t preordained and would have been unlikely to happen without the careful management of the Portage Network at its inception. Individual CARL libraries also invested in other ways, including by dedicating their own library staff to participating in the growing network of experts. This investment helped Portage continue to grow during a critical stage.

Costs

Organizing an effort like the Portage Network requires significant transaction costs. To begin with, the initial leaders of the effort gave “scores of presentations” to audiences across Canada to raise awareness, develop relationships, and encourage participation in the nascent initiative. Considerable thought was also given to developing a robust governance structure, requiring numerous meetings—both in person and virtually. There were also direct costs for tools to support collaborative work, such as Google Docs subscriptions, Zoom licenses, and translation costs.

Portage required significant investment to develop, convene, and steward a national community of practice around data management. CARL dedicated staff resources to coordinate the network of experts (NOE), supporting administrative tasks like scheduling, communications, notetaking and synthesis, and event planning. This included holding regular meetings of NOE and ad hoc working groups, coordinating bi-monthly meetings of the council of chairs, and quarterly meetings for the entire NOE. Considerable staff time was spent managing the NOE, but with significant dividends, measured by participants’ investment of time, expertise, guidance, input, and labor. Our interview participants described the NOE as the backbone of the Portage effort, exploring and sometimes defining best practices, helping to set and drive the Portage agenda, providing the labor to accomplish Portage goals, offering continuous feedback and insights from experts on the ground, and providing a valuable feedback loop of credibility, trust, and engagement from across the community. NOE participants are “the best and the brightest across the institutions, and it’s really free expertise from the point of view of [Portage].”

More than one of our interview participants emphasized their perspective that significant gains could be made by being personally involved, as well as by dedicating local institutional staff resources to this type of effort. While allocating institutional resources to a national effort like Portage might seem to offer direct benefits primarily to the project, there were also significant rewards for the home institution, as staff members returned with more expertise and contacts, as well as larger influence in national efforts. Furthermore, by positioning CARL libraries as stakeholders in the national RDM conversation, individual libraries could benefit locally, as campus stakeholders gained a greater recognition of library offerings and value propositions.

Change

While local and national-level discussions had already occurred for several years by the time CARL’s pilot Project ARC was initiated in 2014, the development of RDM infrastructure in Canada was still relatively immature. Some of the things that would have diminished a collaborative approach (such as existing local investments in data repository capacity) did not yet exist, and the lack of maturity in the RDM landscape meant that there were fewer switching costs that could impede the adoption of the collaboration option. Furthermore, status quo costs, or the costs and drawbacks of persisting with the current sourcing strategy, were also diminished because there was no previous sourcing strategy.

Instead, when local RDM investments did exist, they could—and would—be leveraged to build community infrastructure. This was true for both the development of the DMP Assistant, which built upon earlier efforts at the University of Alberta, and the national Borealis Canadian Dataverse Repository, supported by the Scholars Portal managed by the Ontario Council of University Libraries (OCUL).

Developing shared data repository infrastructure is particularly interesting in this case study. While Portage was working with Compute Canada to develop the FRDR repository environment, some institutions had also already established their own local data repositories using Dataverse, and there was interest from the CARL community in establishing a working group to examine the potential for a shared model for Dataverse. Portage leadership could have responded by directing library interest exclusively to FRDR development; however, it responded flexibly to the interests of the growing network of experts, a reaction that likely engendered greater trust and goodwill for the Portage project.

Building upon librarian expertise, enthusiasm, and existing capacity, the Dataverse North expert group was established in 2017 to examine how a national Dataverse instance would work as a complement to the FRDR infrastructure that was purpose-built to handle large datasets. By leveraging the preexisting Scholars Portal Dataverse into a national service, the bilingual Borealis Canadian Dataverse Repository has resulted, where each institution can have a locally branded instance.

While some institutions had already established their own Dataverse data repositories, many of these systems—such as the Dataverse at Dalhousie University—were sunsetted over time, and their data migrated to the new shared infrastructure. Here it seems that switching costs were less than the status quo costs, which included more staff resources to locally manage repository infrastructure.

Today Borealis is governed by the four regional academic library consortia and operated by OCUL.⁵¹ It also receives financial support from the Alliance, and it supports smaller datasets, some active data management, and offers some local control and branding. In contrast, FRDR, now managed by the Alliance, supports larger datasets and ensures long-term preservation.

Control

Since its beginnings as Project ARC, the Portage Network has benefited from centralized decision-making authority from the CARL Executive Director, who is informed by the leadership of the chair of the CARL Data Management Subcommittee, along with the CARL Research Associate. Soon, it became evident that a dedicated leader for the effort, “somebody with credibility, a credible voice, someone that everybody respected,” was needed, and the first Director of the Portage Network, Chuck Humphrey, was hired in September 2015.⁵²

With the appointment of a director, responsibility and decision-making authority passed from the CARL executive director to Humphrey. However, while this role provided leadership and, to some extent, supported project nimbleness, it was also highly consultative. One participant described it as, “We keep a light hand on the tiller . . . gently guiding things along with and through our coordinators.” This effort required close consultation with the council of chairs, the network of experts groups, and external stakeholders to ensure that goals, efforts, and expectations were aligned. Another interview participant similarly described this as “a balance between nudging the community in the right direction and being open to feedback.”

It’s the right balance between those two things—not imposing a vision on people but having enough of something that they can hold on to where they see . . . the value and they see themselves. And then allowing flexibility so that that can be shaped by the community that’s working together.

Through this arrangement, library and expert interests can percolate up to the executive and vice versa.

The project also received input from two governance groups established by CARL. One was an external stakeholder group, with representation from federal funding agencies, CANARIE, Compute Canada, and others. CARL library directors comprised the second governing group, as by 2017, they were contributing funds to support the nascent Portage Network. Several of our interview participants emphasize the importance of being inclusive of a wide variety of stakeholders, particularly those who “might be skeptical or might otherwise have their own agendas to work out so that they are hearing what’s going on and have a seat at the table.” Another said that having skeptics “in the tent as opposed to out of the tent is quite important.”

“Library voices gained the respect of the other stakeholders around the Council table, which worked well for RDM being seen as essential digital research infrastructure, and for libraries being seen as essential within RDM.”

CARL leaders were also invited into national conversations convened by others, particularly the Leadership Council on Digital Research Infrastructure (LCDRI). LCDRI was established and funded by Innovation, Science and Economic Development Canada (ISED) to provide community-based strategic leadership in shaping the Canadian digital research infrastructure (DRI) agenda. The leadership council was composed of senior executives from research universities and governmental organizations, and the council delivered three position papers on the topics of data management, advanced research computing, and the future coordination of the national layer of the DRI ecosystem.⁵³ These deliverables provided thought leadership for shaping what would become the Alliance. Several interview participants emphasized the pivotal nature of library participation in LCDRI, where “Library voices gained the respect of the other stakeholders around the Council table, which worked well for RDM being seen as essential digital research infrastructure, and for libraries being seen as essential within RDM.”

Engaging external stakeholder groups was also essential for minimizing misconceptions and securing buy-in. For example, the topic of “data storage” meant different things to different stakeholders—with CIOs concerned with active data storage while libraries sought to address repository storage (and discovery) needs. “Archive” was another term that meant different things to different people. By bringing diverse stakeholders together and clarifying points of confusion, the challenges began to dissolve, at the same time that the library’s role within RDM workflows became more visible. It also helped to dispel outdated stereotypes of the work of libraries, helping the broader community understand the essential role of libraries in data curation. Because research data management involves multiple stakeholders, building relationships across these groups and demonstrating the role of the library was essential for both the success of the Portage Network and for the eventual establishment of national RDM infrastructure. As one of our interview participants told us, “It was more successful because we tried to connect with those other groups than . . . because it was just CARL [libraries].”

LOOKING AHEAD

The Portage Network, which had grown to include 15 full-time positions, was formally transferred from CARL to the Alliance in 2021, bringing three core DRI activities together under a single operational umbrella: advanced research computing, research software, and research data management. The Alliance provides continuity, government commitment, and ongoing financial support for the national RDM efforts initiated by the Portage Network.

As the Portage Network director position transitioned to the Alliance as Director of Research Data Management, this role instantly became situated within a larger parent organization, with broader strategic goals, a more complex administrative hierarchy, an even more complex array of stakeholders, and the challenge of addressing long-standing issues in the national advanced research computing landscape. The position is now better and more stably resourced and directly connected with complementary research services but is less agile. Furthermore, governance structures have changed, as the internal library directors' governance committee and an external stakeholder advisory committee group have been retired and replaced by new governance structures at the level of the parent Alliance organization, including a board of directors and researcher council. This now positions CARL and its libraries further from the center of control and requires that the CARL Executive Director and the Chief Executive Officer of the Alliance maintain a strong relationship. Currently, the challenge is mitigated to some extent by the fact that four of the 11 university-affiliated directors serving on the Alliance Board of Directors are CARL library directors, and the current board chair is a library director.⁵⁴

Our interview participants expressed some anxieties that RDM needs—and the role of libraries—could be overshadowed by other Alliance activities, particularly advanced research computing. One interview participant said, “The bigger it gets, the further removed we get from having input or influence.” However, another remarked, “It’s a risk but it’s also an opportunity because RDM operating out of libraries on their own has much less potential for impact, and there’s so much greater potential for impact on research data if it is taken into consideration within the research software that researchers are using.”

Indeed, the work continues in earnest with the continuation of the network of experts, with participation from more than 70 institutions, “where the real work is happening.” Through this active community of practice, libraries and librarians remain connected with the Alliance and with each other as they continue to address RDM challenges.

LESSONS FROM THE PORTAGE NETWORK EXPERIENCE

- **Secure small wins.** Portage leaders strategically tackled discrete, well-articulated projects of a limited scope, which offered the opportunity for more rapid progress and success. This instilled momentum and trust in the effort. The DMP Assistant was the first of these “small wins.”
- **Leverage existing efforts.** Instead of starting from scratch, the Portage Network leveraged infrastructure already developed by the University of Alberta, which formed the backbone of the national DMP Assistant. The community similarly leveraged an existing Dataverse to work toward a national Borealis Canadian Dataverse Repository.
- **Start early.** CARL initiated the Portage Network fairly early in the development of RDM infrastructures in Canada—notably, before many institutions had developed their own local services. By beginning the conversation early, CARL made the collaboration option more attractive for cooperative members by removing the friction of potential “switching costs” from local to shared infrastructure.
- **Lead and follow.** CARL provided enough leadership, energy, and resources to launch the effort, and it could move nimbly because of its small, start-up project team. But it “kept a light hand on the tiller.” When librarians sought to establish a network of experts to

explore a Canadian Dataverse repository—when Portage was already pursuing FRDR—it listened and followed the lead of the library community. These actions built more trust and commitment from community members.

- **Strategically engage multiple stakeholders.** CARL leaders dedicated the time and planning to develop and steward relationships across the research community, engaging both CARL libraries and other non-library stakeholders. They established two Portage advisory committees—one composed of library directors and a second comprised of external stakeholders from government organizations like CANARIE, Compute Canada, and the Tri-Agencies. By getting everyone “inside the tent,” both supporters and skeptics alike could be informed in a more transparent environment that provided opportunities for building new trust relationships. Without this effort to engage outside the library community, it seems doubtful that the Portage Network would have grown from a library-led effort into a national infrastructure.
- **Articulate local expectations of national collaboration.** Librarian participation in the network of experts is strong, with 150 individuals participating from 70 institutions. This level of involvement happens when local participation in collaborative efforts isn’t just something done “off the side of your desk” over evenings and weekends, but is an institutional expectation, often explicitly written into library job descriptions. This requires the leadership and commitment of library directors.

The Data Curation Network

INTRODUCTION

The Data Curation Network (DCN) is a collaboration with 17 sustaining members, including 15 US universities, an independent data repository, and a philanthropic foundation, to pool expertise in research data curation. By sharing their collective expertise through a cross-institutional shared staffing model, DCN members can draw on a network of experts to cover a much broader range of research data types and formats than any single member would be able to support on their own. The DCN is composed of a “human layer”⁵⁵ of “professional data curators, data management experts, data repository administrators, disciplinary scientists and scholars” and strives to “build a trusted community-led network of curators advancing open research by making data more ethical, reusable, and understandable.”⁵⁶

Launched in 2016, DCN enables its members to send datasets to a partner institution with the requisite expertise to curate the data. Following curation, the datasets are returned to the original institution for deposit in a repository. In addition to the curation network, DCN also develops educational resources such as workshops and primers for data curators and other stakeholders in data management.⁵⁷ DCN offers significant community-building opportunities through working groups and other kinds of engagement with peers. While DCN was founded on the idea of a shared curation network, its mission now embraces the broader goal of cultivating best practices, training, and professional development within the data curation community. A full participant in DCN’s activities pays a \$10,000 annual fee and commits to contributing 200 hours of curation capacity to the network, as well as a modest time commitment by one staff member to serve as the institutional representative in DCN’s governance activities.

DCN is a rare example of an RDM-related collaboration “built from the ground up,” without the benefit of an existing collaborative structure. In this sense, it represents a scenario where a new grouping of partners forms to fill a shared gap in capacity.

ORIGIN STORY

The DCN⁵⁸ started with a conversation among a group of librarians at a conference. The conversation was inspired by an ambitious question: “Wouldn’t it be great if. . . .” At that time,

significant uncertainty surrounded data curation services in academic libraries—including what data curation meant, how curation services should be delivered, and what skills would be required. As one interviewee explained, there were “not a lot of road maps to follow.” At the same time, compliance with emerging data curation mandates was becoming a pressing issue.

One key concern was staffing models for data repositories in light of the diverse data types that could be deposited, each with different curation requirements. No single institution could realistically house the wide range of curatorial skills needed to address all forms of data sets. Additionally, staff departures could create unexpected skill gaps. As one interview participant described, the group pondered whether a network of curators could be created without a burdensome technological infrastructure that would allow curatorial expertise to be shared across institutional boundaries—just as research itself is conducted. In this sense, the idea of a shared staffing model for data curation “grew out of a very, very practical need.”

Following the conversation, the librarians wrote a proposal for a planning grant which was awarded by the Arthur P. Sloan Foundation in 2016 and took the effort from idea to project. Interviewees emphasized the importance of receiving the grant as a strong impetus for getting the idea off the ground. “It really changed the game. . . . And for some reason, it’s like a magic ‘you get to go do this’ card.” Receiving grant funding not only provides tangible resources with which to initiate activities but also represents a significant external validation of the ideas motivating the project—both important factors in translating a collaboration from concept to reality.

In addition to the principal benefit of providing funding, receiving a grant also helped crystallize formal commitments from participants, as well as a clear articulation of roles and responsibilities. Overall, it supplied a rudimentary infrastructure for the collaboration to organize efforts and secure accountability. Without the structure of a grant, collaborative partners would likely need to create their own accountability infrastructure through a memorandum of understanding or a contract.

The DCN effort launched its planning phase in 2016 with six partner institutions.⁵⁹ The University of Minnesota played a vital role as a leading institution for the effort, serving as the host for the project as well as contributing the leadership of Lisa Johnston, widely recognized as instrumental in getting DCN off the ground. A key catalyst to moving the effort forward was a kick-off meeting among participants with a professional facilitator. In this meeting, the founding members defined the kind of community they wanted to be, discussed the challenges to achieve that vision, and identified strategic priorities. Interviewees stressed the importance of these conversations as a means of building community early and achieving a collective view of where the partnership was headed. The time spent working through the planning grant was a critical investment in building the foundations that positioned DCN for long-term success.

DCN was constructed “from the ground up” as an entirely new partnership, instead of from within an existing library consortium. A key reason for this was that the group of partners already extended outside the boundaries of any existing consortia; moreover, moving into a consortium would likely limit the ability to be inclusive in expanding DCN’s membership. Preserving a sense of community was important, especially in a community that sprang up from long-standing personal networks.

Following the success of the planning grant, the Sloan Foundation awarded DCN a three-year implementation grant, and the network formally launched in January 2018, now with two additional partners.⁶⁰ At the conclusion of this grant, DCN transitioned to a membership-based organization, sustained by an annual membership fee. Currently, DCN provides two major offerings to its members:

- A coordinated workflow that members can use to share curation staff and expertise and manage data sets.
- A community of professional data curators who can share tools, provide training, and promote best practices.

As DCN continues to evolve, the scope of its activities has expanded well beyond the original vision of shared curation expertise. While this remains a core offering, it has been augmented by educational, training, and community activities. According to one interviewee, with the addition of these new activities, the shared curation aspect of DCN—that is, the exchange of data sets between institutions for specialized curation—“was . . . now ten percent of what the DCN is.”

DCN’s focus on education and training led to the award of an IMLS grant in 2018 to build an educational program for curators.⁶¹ The contours of this program were shaped by surveying curators to assess needs, part of a broader emphasis on encouraging engagement within DCN. For example, one educational resource that DCN produces is a series of primers that detail curatorial requirements for specific data formats, ranging from clinical trials data to Twitter data.⁶² A second IMLS grant was awarded in 2022 to develop educational resources focused on geospatial data, scientific images, code, and simulation data.⁶³ Beginning in 2022, the DCN was primarily sustained by member fees, rather than external grant support.

DCN is, fundamentally, a community for members to share knowledge and learn from one another. From its origins as a shared staffing model for curatorial expertise, DCN has evolved to provide what may be an even more important service: providing a learning and training community for its members where “shared experience has become as valuable as the shared expertise.”

Data Curation Network Timeline

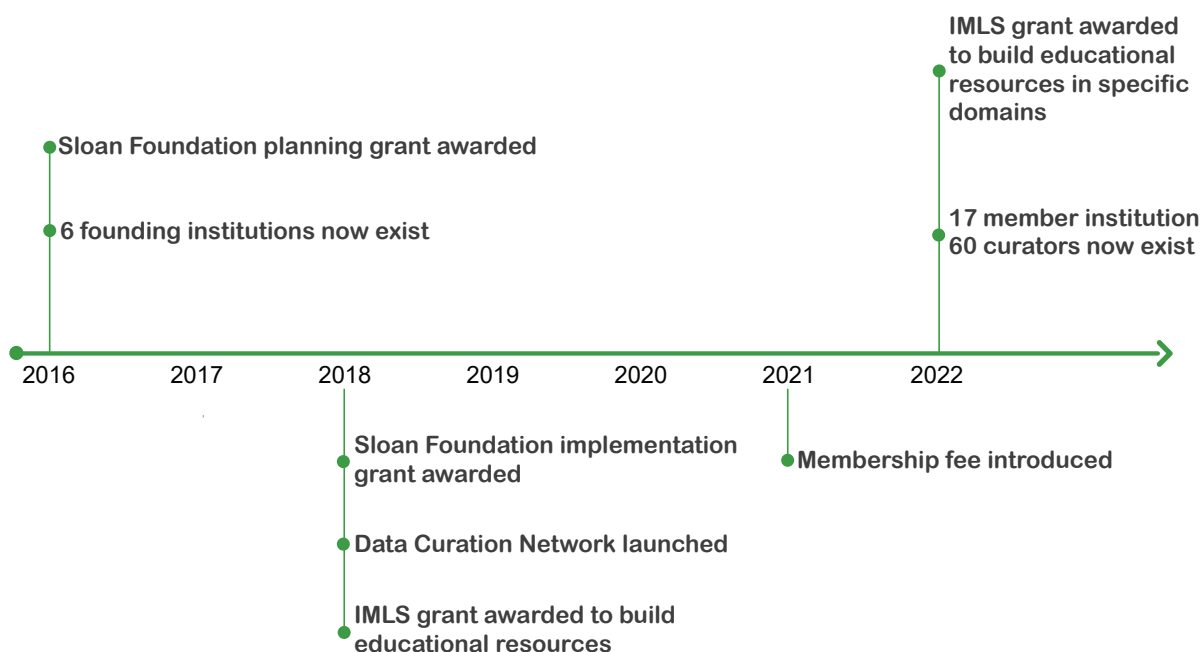


FIGURE 5. Data Curation Network timeline.

CHOOSING COLLABORATION

DCN represents a choice to adopt a collaborative approach to develop curatorial capacity and a community of practice. There were several environmental factors that favored this approach, such as the desire for a trusted community, innovation, and institutional prestige.

“I think community really has risen to be one of the defining characteristics of the DCN.”

Although the idea of a shared staffing model for curation was the original impetus for launching DCN, the trusted community of practice that developed alongside the curatorial network is perhaps DCN's greatest success. The DCN is seen as a group that gets things done through the community and produces value for the community. One interview participant remarked that community is “something that has really permeated the DCN since its inception, of coming together and experiencing things or crafting ways forward, not just individually, but as a group. I think community really has risen to be one of the defining characteristics of the DCN.”

The desire for innovation and quality were also factors in favor of DCN's collaborative approach. For example, DCN facilitates an exchange of information that allows best practices learned from partners to be used to improve local services, spark new ideas for services, and support quality curatorial work. Agility was also highlighted as a factor, in the sense of having the ability “to take on things that we wouldn't necessarily have otherwise been able to do.” Budgetary issues play a role as well: relieving economic pressures through collaboration can be attractive to administrators. “We're continuing to see this as a core part of how we describe the value of the DCN. . . . You don't have to have three people . . . to curate data [when] you can pay . . . \$10,000 [to be part of DCN] . . . and have that similar kind of expertise.”

Participation in DCN could also reflect positively on institutional reputation: one interviewee noted that their local faculty were impressed by the universities and organizations in the DCN network, while another mentioned that campus administrators were particularly interested in the partners involved and what had been accomplished, as well as the fact that the effort had secured two grants. Collaborating in the DCN network can lift the reputations of those institutions who feel they are lagging in curatorial capacity. As one interviewee observed, “We feel like we're playing catch up in this arena compared to our peers in the DCN, so there's just so much that we would stand to gain and learn, and there's a certain level of admiration we have for where they are in the maturity of building out repository services.”

An institutional mindset that embraces collaboration as a means of acquiring capacity was, unsurprisingly, an important driver influencing the decision to participate in DCN. One interviewee mentioned that their institution was very supportive of “radical collaborations”: “How can we break down the barriers of what traditional collaboration has been? And this fit that bill really well.”⁶⁴ It is worth noting that institutional views on collaboration can change dramatically when a partnership moves from being grant-funded to self-sustaining. “Universities aren't set up for . . . collaboration where we hand each other money,” observed one interviewee. “If I asked them to do a grant with me, I would've gotten a signature in ten minutes. If I asked for them to give me small sum of money, then we've got a bigger problem.” In light of this, choosing to make the transition from grant-supported to self-sustaining can be a critical decision point.

A number of interviewees cited uncertainty over data curation needs, standards, and practices, as well as how to move from one-off data curation support to standardized data services as key factors that enhance the value of collaborative options like DCN. In this sense, the timing

of DCN's arrival on the scene was opportune. Fueled by trends ranging from funder mandates to data scandals, there is a growing recognition across universities that RDM is a key aspect of responsible, impactful scholarship. Often, the library's ability to respond to emerging needs on campus can be enhanced through a collaborative partnership like DCN. As one person put it, "Our connections to our peers and the ideas that we get from being a part of this community, and the services that we're able to offer above and beyond what we could do locally, that's all a part of the story that we want to tell."

While participating in DCN's collaborative approach clearly had many benefits, there were also a few trade-offs. For example, although there are efficiency gains from distributing curation expertise over a network of institutions, there can be efficiency losses as well. It takes time and effort to collaborate across institutions, and this can reduce the value of the network when researcher needs are immediate: "At times we're not even able to take advantage of the shared expertise because sometimes when researchers need something, they need it yesterday." In addition, data curators want to retain and steward relationships with local researchers; DCN therefore has local curators serve as the go-between between researchers and the curation expertise on the network, instead of connecting researchers directly to data curators at other institutions. In this sense, DCN's desire not to get in the way of local relationships comes at the cost of some efficiency.

Risk was also mentioned as a possible trade-off of participating in the DCN network. For example, there was a risk that the partnership could fail, which could then harm the institutional reputations of those involved. Some institutions may even prefer to build up local capacity before entering into something like DCN—a cautious approach, one interviewee surmised, that may be partially motivated by the fact that the Digital Preservation Network, another collaborative effort, had recently ceased operations.⁶⁵ Another possible risk was that some researchers would be uncomfortable about their datasets being sent to other institutions. As a counterpoint, however, one person noted that risks of this kind can be mitigated because every DCN partner can control their use of the curatorial network.

STRATEGIC FRAMES

The Data Curation Network operates a multi-institutional curation network in which partners agree to curate datasets submitted by other partners. Although there are no requirements to curate or limits to submitting datasets, free riding on the network has not been a problem. A membership fee and a strong culture of engagement have helped diminish this risk. Active participation in DCN is bolstered by a strong culture of commitment among the members. While DCN was initiated to support shared curatorial capacity, today its members emphasize the peer network as a key value of the collaboration. This value is difficult to quantify, but being able to articulate its impact on members is an important aspect of the cost-benefit equation for DCN membership. A significant cost in establishing the DCN was the need to build a collaborative apparatus from the ground up, but this was mitigated by the benefit of extending membership across preexisting consortial boundaries. The fact that DCN filled a gap in capacity, rather than offering an alternative to existing capacities, eliminated the barrier of switching costs, which helped smooth the path toward adopting the DCN model. Today, DCN manages the tensions of moving beyond a small, tight-knit group of founders to a larger membership requiring greater structure. Despite the significant challenges inherent in starting up a new collaboration, DCN has demonstrated that this collaboration model can be both successful and sustainable.

Coordination

DCN collaboration is characterized by multi-faceted engagement: in governance, interest groups and working groups, and of course, actively participating in data curation by submitting datasets to the network for curation and accepting datasets in return. While in the strictest sense, a member in good standing is an institution that pays the annual membership fee

required of all partners, one interviewee suggested that an active member should participate in multiple aspects of the DCN network: in other words, moving beyond the core shared data curation staffing function to include engagement in DCN's community of practice. Greater participation in the collaboration creates a virtuous circle: "When folks are more engaged, they're more comfortable with the collaboration and the work being done."

There is a real risk of free riding in this type of arrangement where partners may utilize the curation network without reciprocating since there are no limitations on the number of datasets an institution can submit to the network, nor are there any requirements that an institution receiving a dataset from another partner must perform the requested curation. The threat of free riding is at least partly remedied by the annual membership fee that all DCN members must pay. In determining the fee, DCN adopted the approach of dividing the budget by the number of members. A key reason for this was the desire to create an equal partnership through the funding mechanism, rather than a more transactional model based on benefits received and contributions given. Additionally, a transactional approach, or "curation as a service," did not reflect the spirit in which DCN had been conceived, being perceived as commodifying their work in a way "that was highly unattractive to our curators who didn't want to just be a curator machine."

One important safeguard against free riding is to preserve DCN's "culture of trust, camaraderie, and esprit de corps." For example, one interviewee remarked that when the flow of curation requests to their institution was infrequent, they tried to increase their engagement in other areas, like governance and committees, to ensure their institution continued contributing toward making DCN a "healthy enterprise." There is a strong focus on making sure that participants in the curation network see DCN as something more than fulfilling external curation requests in addition to their normal work responsibilities. Noting that the DCN curators were all essentially volunteers, one interviewee observed, "We wanted to engage them in a new way and make sure that they were getting something out of this. So, we really started focusing on curator engagement, curator training, upskilling . . . [and] information exchanges with other peer groups."

Another safeguard for balancing curatorial requests to the DCN is the DCN director, who serves as a "clearinghouse" for distributing curation requests across the network. These assignments are made on the basis of curation expertise as well as availability and recent contributions to the network, with the goal of distributing the work evenly and avoiding overburdening individual curators with excessive requests. As one interviewee said, "There is somebody at the wheel who's keeping an eye on balance." While this clearinghouse mechanism is intended to manage the curatorial burden of curators participating in the network, it could conceivably serve as a means of counteracting any tendencies toward free riding that might arise. Interviewees also noted that within the network of DCN participants, some peer pressure or reputational risk could be incurred if a member is seen not to be contributing—an informal system for managing equality of contribution. While incentivizing contributions may need to be revisited in the future, for now, there is simply a prevailing expectation that members will participate.

DCN recognized that its members do not fit into a single mold, and different members will cite different value propositions for participating in the partnership. In response, DCN offers a wide array of participation possibilities. But the common theme across all modes of participation is contribution. For example, the DCN board has discussed creating a "paying member" status, where an institution can pay a fee to submit a certain number of datasets to the network for curation, without reciprocal curation obligations. But this would likely be contentious among existing members, in that purchasing access to curation on demand would contravene DCN's ethos of building community.

Individual engagement and contribution strengthen the sustainability of DCN's core function of shared data curation. As a result, the ability to draw forth contributions from

the DCN membership is, in large part, predicated on the strong commitment to DCN and its community. One interviewee noted that individuals have continued to contribute to and engage with DCN even if their institution ceases to be a member. As one person put it, the shared data curation model is important, but the broader community of practice DCN fosters is even more so: “If the DCN didn’t exist, we would have to find other people to talk to [to] help us figure out the answers.”

Costs

For DCN, building a new collaborative network from the ground up—or a start-up approach—involved significant transaction costs, often borne through the voluntary efforts of community members, in the form of time, effort, and resources. One reason for adopting this approach was that, at the time, existing library consortia and groupings were not focusing on data curation or offering programs or services related to research data management. There was also a concern that an existing collaborative grouping could change priorities, leaving a DCN-like initiative behind.

Another reason for choosing the start-up approach was the perception that new challenges, like data curation, sometimes require new partners. For example, one interviewee noted that many advances in data curation have not originated in libraries, and therefore much value can be gained by engaging with experts beyond the library domain: “I think we’re enriched by being able to talk with [the Michael J. Fox Foundation] and . . . seeing their experience and hearing their issues and their challenges.” And, of course, hearing about library experiences in data curation is valuable for those outside that domain.

The burden of building a new collaborative apparatus was eased when DCN hired a coordinator for the partnership, managing the network of collaborators with a big-picture view of the effort. The costs of collaborating have also lightened as DCN matured, with less need for organizational evolution as well as less decision-making about DCN’s collaborative structure in comparison to its earlier planning stage.

The minimum investment to participate in DCN is the annual fee and providing an institutional representative for the Board. Embracing the full range of opportunities available within the collaboration could require significantly more investments in time and effort from individual participants. On the other hand, it is important to remember that contributions to collaborative efforts like DCN create local benefits: in this sense, the costs of collaborating must be weighed against the value they generate for the participating institution, such as access to expert knowledge. The effort required to fully participate in DCN is like an investment with a return. In addition to benefiting member institutions, participation also benefits individuals. One interviewee stated:

For curators who . . . get very involved in working groups and build relationships and take on datasets that maybe they haven’t seen in the past, I would say that DCN is really a professional development enterprise for them as well. It’s definitely a CV line item, annual review line item.

Another cost of collaboration is the effort required to recruit new partners. While DCN does not actively market itself to new members, affiliated curators frequently speak about DCN at meetings and conferences, and also publish papers; raising awareness in this way can spark interest on the part of other institutions to reach out to learn more. There were some early discussions of recruiting members based on curation expertise, but a good balance naturally developed between curation needs and curation expertise within the network. It should be emphasized that membership growth is not an explicit goal for DCN; as one person expressed, “rather than saying, what we want to do is grow, I would say what we definitely don’t want to do is exclude.”

Broader environmental trends can impact the costs of collaboration. During the COVID-19 pandemic, DCN found that virtualized interactions often removed obstacles to engagement and equalized access. But these benefits can come at the cost of the dynamism, serendipity, and opportunities to meaningfully connect that in-person interactions often offer:

And that's why I value our in-person all-hands meetings . . . having that one time of the year to come together and have those conversations with individuals, what I find is that I really learn more about their background and I learn more about their interests and their skill sets and their expertise. And for the data curation network, that translates into trust.

Several interviewees discussed the transition to a membership model, with the introduction of the annual membership fee, as a trade-off. While promoting DCN's sustainability, the membership fee is a tangible cost of participation, which must be set alongside quantifiable benefits like the number of datasets curated by the network, as well as less quantifiable benefits such as the ability to engage with a community of practice and apply knowledge and experiences from that community to local services. In justifying DCN membership to administration, it can be difficult if not impossible to quantify these community benefits in ways that allow them to be set alongside the monetized cost of the membership fee. So, while it is often easier to make the case that the membership fee buys access to a network of curatorial expertise, as opposed to engagement with a community of practice, this is only a partial accounting of the benefits of DCN membership—a problem exacerbated by the low volume of curatorial work currently conducted on the network.

Change

Because the DCN was formed around RDM, a relatively new area of library activity, there were few switching costs that would impact the ability of member institutions to adopt DCN's shared staffing curation model, easing the transition to a collaborative approach. Indeed, issues involving conflicts or disruptions with local services, workflows, or sourcing strategies tended to be minimal. As one interviewee explained, "We are starting not quite but almost from scratch when we joined the DCN, so we didn't have legacy ideas or processes that we had to disrupt in order to . . . become a member."

However, interviewees did note other kinds of roadblocks that could impede a switch to a collaborative approach—especially with a new, untested partnership like the DCN network. Campus administrators often have reservations about collaborations as a means to acquire capacity, including how such collaborations might work smoothly and sustainably, and what benefits the collaboration will bring back to the local institution. The fact that data curation was a relatively new service area for libraries amplified the problem of communicating the goals and value of the collaboration. Uncertainty about the value of collaboration around data curation capacity extended not just to campus administrators, but sometimes even to librarians and library leaders.

In particular, a disconnect can arise between the type of measurable value administrators expect to see, and the less quantifiable value perceived by the data curators themselves. As one interviewee remarked, "I know our curators . . . feel like the community is the most significant part," but for campus administration, transactional value and efficiency may be a more compelling argument: "We're going to have to really show the impact and the value the DCN brings . . . less so for the community and more so for the shared staffing because I think that that's what library directors and administration are really keen on."

An important strategy for securing buy-in for shifting to a collaborative approach is recognizing and communicating different sources of value to different stakeholders. In the case of DCN, this often means communicating the return on investment regarding the shared data curation, even though much of the benefit—and labor—is in building community.

Transitioning into a collaboration can be difficult but transitioning *out* of a collaboration can also create problems without a clear exit strategy. For the DCN partners, the process of leaving is fairly easy: within the new membership model, memberships are renewed annually, providing an automatic and regular decision point to reevaluate ongoing participation in the collaboration. To minimize attrition, DCN devotes considerable effort to checking in regularly with its members and soliciting their feedback. If an institution does leave, individuals from that institution can continue to participate in working groups and other DCN activities but without voting rights. Indeed, DCN's interest groups are open to anyone who wants to participate, regardless of whether their institution is a member. While institutions may come and go in the formal DCN data curation network, the peer network that has emerged within DCN seems much more durable.

Control

Decision-making for the DCN partnership is vested in the board and includes representation from every member institution. In addition to its monthly meetings, the board convenes an executive group of four or five people that acts as a filter in regard to what issues require board attention, and to assist in preparing the board for subjects that require group discussion. DCN also has committees on special topics, such as membership or education, and regular all-hands meetings provide opportunities to engage the entire membership.⁶⁶

The DCN's governance model is very much like a legislative approach to decision-making—an assembly of stakeholder representatives tasked with making decisions and managing the collaboration—rather than an executive approach, in which decision-making authority is ceded to a centralized executive that acts on behalf of all. The decision-making process involves a great deal of discussion, collecting opinions, and then voting. This governance model fits with DCN's aspiration to treat its membership as equal partners, an approach clearly supported by the partnering institutions. "We have heard very strongly from our members that a sense of being a part of the DCN is the ability to have that voice, that we are all very much seen as equals," observed one person. The equal partner approach seems to be, in part, a byproduct of the strong personal network that runs throughout the DCN membership, or perhaps it is more accurate to say that the personal network helps make the "equal partner" approach work.

An important inflection point in the evolution of DCN's governance was the move from grant funding to a membership model, supported by the membership fee; this created new fiduciary responsibilities and necessitated a more structured approach to governance. The introduction of a membership fee provides a clear delineation of who can participate in governance. As a result, the community ethos has shifted from a more "experimental" approach under the grants phase to one that incorporates a greater concern with oversight and responsible stewardship of community goods.

The DCN began as a high-trust, tight-knit group that was intensely committed to forming the network, and as a small group, they could act quickly and nimbly. As a result of membership growth, decision-making now is more structured (with the Board and executive committee) and is based on a "majority rules" approach undertaken through voting. While the mindset to act in an egalitarian fashion is still evident in DCN governance, there is a recognition of the trade-offs this creates. "If we grow too big, too fast, we kind of lose something. And I feel . . . already . . . it's hard to make decisions as a 15-organization governance board." There is

a stark difference between running DCN as a six-person co-PI team that makes all decisions collectively, and a governance board nearly three times larger representing a diverse array of interests across the membership.

When considering possible changes to the DCN governance model, there are important trade-offs between the collegial, group-driven approach currently in place and one that, while perhaps more efficient, might diminish the level of engagement and personal connection to the collaboration. One interviewee emphasized the importance of accumulating trust in making a partnership like DCN successful; engagement is more forthcoming as the shared experience of working together deepens: “I think the interactions, the communications, the working side-by-side with one another is a really integral piece of the network that helps grow that trust, that helps the rest of the network continue to function and be sustainable.” If the DCN membership grows, and if, as a consequence, governance becomes more distant, centralized, or impersonal, a fundamental ingredient for DCN’s success may be lost.

LOOKING AHEAD

Moving forward, DCN has the opportunity to leverage its role as a unifying force and voice for data curators and libraries in the RDM domain. This includes advancing DCN’s core mission of maintaining a curatorial network and extends to its unique role as an educational and training resource for curators. This may include an expanded role for DCN as a leader in raising awareness of both the need for data curation and as a unifying voice for libraries in the broader data curation landscape: “To continue to have a role in advocacy on behalf of our researchers and in the service of fair data practices and open scholarship, . . . something that all of the members feel very strongly about.” DCN furthermore has the potential to be a national leader in supporting institutions in managing research data, similar to national organizations like the Digital Curation Centre in the UK and the Digital Research Alliance of Canada.

DCN might also serve as a means to bring together data curation stakeholders within an institution. Noting that membership in DCN is an *institutional* membership, one person suggested that this meant there was opportunity and value in encouraging more engagement from other campus units besides the library, such as campus IT or even researchers themselves. Finding a way to incorporate their experiences and expertise into the DCN community would be mutually beneficial.

DCN will continue to evolve and grow, particularly as new institutions—and new people—are added; to some current members, the prospect is both exciting and unsettling. The DCN culture has so far been able to accommodate the addition of new members, including several from outside the higher education and academic library context, but challenges may arise if the network continues to expand. In particular, further growth may require significant changes to governance processes. Would this reduce the highly valued personal touch of the network, and would the free rider problem start to become evident? Further expansion of the network would also necessitate additional funding, especially to support the hiring of additional centralized staff to augment the current DCN staff of one.

There are also questions about what services should be offered. For example, DCN could provide more centralized services, such as expert consultation and curation, especially for institutions that might not have a staff of data curators. But this raises an issue concerning contribution, because if an institution lacks any real data curation capacity, it can utilize DCN to fill that gap, but would not be able to reciprocate to other members. Increasing DCN’s complement of centralized staff might be a solution but would alter the nature of DCN’s current model of distributed contribution.

An important concern we heard from interviewees was the effect on DCN as the original founders eventually left, a concern made real by the recent move of one founder—

acknowledged by all as the central driver of DCN's creation and success—to a new position at a university that was not yet a DCN member. One interviewee suggested an analogy to an entrepreneurial startup: “[T]he kind of startups . . . that are successful tend to be around a core group of people, [but when] one of those people leaves . . . is this startup going to survive and become a real company, or is it going to fizzle out because we’ve lost one of our key founders? I think that’s where we are with the DCN.” While there is a sense that this challenge will be overcome, it nevertheless highlights a broader question of how portable DCN’s model is to other scenarios that lack a foundational core of deeply connected and committed individuals.

One clear signal of DCN’s success is the perceived value of the partnership as it currently stands. One interviewee put it this way: “It’s the inspiration, it’s the fun, it’s my favorite colleagues, and it’s also they’re helping me get my work done and helping me learn, so I would love for it to continue being that.”

LESSONS FROM THE DATA CURATION NETWORK EXPERIENCE

- **Begin with a well-defined problem.** The practical problem that catalyzed the DCN partnership was the need for curation expertise. Starting with a clear goal that can be addressed through collaboration helps generate interest, buy-in, and a foundation for expanding the partnership in new directions. Starting with the less precise ambition of building a peer community to promote interinstitutional engagement in data curation would likely have been more difficult for DCN’s founders to create and sustain. Instead, the peer community emerged as a byproduct of the initial, practical motivation for the collaboration.
- **Start small.** Starting small is a way to keep the collaboration tractable, flexible, and responsive to the needs of the collaborators. DCN began with a circumscribed goal, to meet the need for a broad range of curation expertise—and a proposed solution—a shared staffing model in the form of a curator network. Other aspects of the collaboration seen today, such as the educational resources program and the expansive community of practice surrounding DCN’s activities, emerged and developed as the partnership evolved.
- **Act nimbly.** DCN leaders began their efforts by acting flexibly and adapting rapidly to the changing needs and expectations of the growing partnership. They were unencumbered by the premature establishment of excessively formal procedures and governance structures, which might appear to supply order but would instead inhibit the ability to respond nimbly to changing goals and needs.
- **Act boldly.** Collaboration often requires working outside professional comfort zones, with unfamiliar colleagues, and adopting innovative but potentially risky approaches. The DCN effort necessitated working with external partners to develop a new, multi-institutional collaboration to meet local needs. By sharing knowledge, pooling uncertainty, and developing trust in the collective wisdom of participants, it is possible to act boldly and mitigate risk. As one interview participant described it, “Feel the fear, but do it anyway.”
- **Invest time in planning.** The activities that took place under DCN’s planning grant—especially the professionally facilitated discussions—were key ingredients for the partnership’s success. Taking the time to hear all the partners’ voices, identify a collective sense of priorities, and anticipate from the outset the likely challenges created a strong foundation for the growing organization.
- **Create collaboration infrastructure.** As the collaboration matures, it is important to establish the appropriate infrastructure and workflows. DCN created a centralized coordinator position responsible for creating and managing curation workflows. Dedicated, centralized staff are important to keep the collaboration moving forward—especially as the scale and complexity of the collaboration grows and relying entirely on volunteer effort from the partners becomes impractical.

- **Build relationships.** The DCN partnership benefited from the strong, preexisting personal network among its founding members. Having such a network in place at the outset may not be feasible for all new collaborations, but it nevertheless serves as an aspirational benchmark: a group of people who share a common purpose and trust in the collective effort to achieve it.
- **Check your assumptions.** The shape of DCN emerged from extensive discussions among partners that helped establish priorities and synchronize expectations. These discussions can help surface differing ideas of what the partnership should look like and begin to harmonize them into a mutually supported vision. Collective action is predicated on collective dialogue.

CONCLUSION

Our case studies have provided three rich narratives of collaboration in action, each exemplifying a group of institutions working collectively to acquire RDM capacity. While the circumstances of each case study are different—for example, one taking place within an existing institutional grouping, one built from the ground up by a group of practitioners, and one tapping into a broader national network of allied initiatives and agencies—all demonstrate the opportunities, challenges, and trade-offs of collaboration in the RDM space.

The economics-based focus of our case study analysis is a step toward uncovering some of the important factors that influence the choice to collaborate and, ultimately, make collaboration successful and sustainable.

The success of these collaborations in achieving their goals reinforces the idea that collaboration will be an attractive sourcing option for many institutions seeking to fill gaps and extend capacities in RDM and beyond. However, the decision to do so must be carefully considered, weighed against alternatives, and based on a thorough appraisal of the key benefits, costs, and trade-offs involved. If collaboration is chosen, it should be organized and managed in such a way as to maximize its chances for success. The economics-based focus of our case study analysis is a step toward uncovering some of the important factors that influence the choice to collaborate and, ultimately, make collaboration successful and sustainable.

While each example of collaboration will be unique in its origins, development, goals, organization, and management, our analysis of the case studies yielded general recommendations that can serve as a starting point for library leaders as they work through the decision of whether or not to collaborate, as well as providing insight on how to optimize the rewards of collaboration, whether in RDM or in other spaces. In addition, the lessons learned shared by our case study interviewees provide valuable on-the-ground perspectives gleaned from personal experience in collaborative efforts. We were inspired by the efforts of the Texas Data Repository, Portage Network, and the Data Curation Network, and we hope that readers will also be inspired by their achievements.

ACKNOWLEDGMENTS

The authors extend special thanks to the 20 interview participants who generously shared their experiences and expertise with us for this investigation. These individuals also helped us to ensure the accuracy and quality of the case studies by providing input prior to publication:

Data Curation Network

- Jake Carlson, University at Buffalo
- Cynthia Hudson Vitale, Association of Research Libraries
- Lisa Johnston, University of Wisconsin
- Wendy Kozlowski, Cornell University
- Katie Wissel, New York University
- Sarah Wright, Cornell University

Portage Network

- Donna Bourne-Tyson, Dalhousie University (retired)
- Susan Haigh, Canadian Association of Research Libraries
- Shahira Khair, University of Victoria (formerly with the Digital Research Alliance of Canada)
- Jeff Moon, Compute Ontario (formerly with the Digital Research Alliance of Canada)
- Kathleen Shearer, Confederation of Open Access Repositories (COAR) and Canadian Association of Research Libraries
- Martha Whitehead, Harvard University
- Lee Wilson, Digital Research Alliance of Canada

Texas Data Repository

- Reid Boehm, Purdue University (formerly with the University of Houston)
- Christina Chan-Park, Baylor University
- Bruce Herbert, Texas A&M University
- Courtney Mumma, Texas Digital Library
- Kristi Park, Texas Digital Library
- Laura Sare, Texas A&M University
- Santi Thompson, University of Houston

This report was also facilitated by a three-month Special Assignment research leave granted to Amanda K. Rinehart from The Ohio State University Libraries. Special thank you to the OSU Libraries Committee on Faculty Benefits, Responsibilities, and Research for awarding this leave and to the Science Librarian Cohort (Danny Dotson, Florian Diekmann, Belinda Hurley, and Jessica Page) for their support and assistance in fielding teaching and research consultation requests in Amanda's absence.

Several OCLC colleagues provided invaluable input as we prepared this report. Kendra Morgan, Diane Vizine-Goetz, and Cynthia Whitacre provided input on the publication proposal, and OCLC Librarian Kem Lang supported our literature review. J. Elizabeth Mills, Mercy Procaccini, and Brittany Brannon reviewed the manuscript in progress and generously offered detailed and thoughtful comments to strengthen the final report.

This project has also benefited from the expertise and guidance of an advisory committee comprised of:

- Donna Bourne-Tyson, Dalhousie University (retired)
- David Groenewegen, La Trobe University Library
- David Minor, University of California San Diego Library
- Judy Ruttenberg, Association of Research Libraries
- Michael Witt, Purdue University Libraries and School of Information Studies
- Maurice York, Big Ten Academic Alliance

This report could never have been published without the significant talents and professionalism of the OCLC Research communications team, including Amy Dalrymple, Erica Melko, Jeanette McNicol, and JD Shipengrover.

Finally, our work was made possible by the senior leadership of OCLC; we wish to particularly thank Lorcan Dempsey, former Vice President, Membership and Research at OCLC, for his inspiration and support of this effort. Additionally, we thank Rachel Frick, Executive Director, Research Partnerships and Engagement; Lynn Silipigni Connaway, Executive Director, Research; and Chip Nilges, Vice President, Business Development, Membership and Research, for their ongoing support of projects like these.

Author contributions:

- Lavoie: Conceptualization, Investigation, Project administration, Writing—original draft, Writing—review and editing
- Bryant: Investigation, Project administration, Writing—original draft, Writing—review and editing
- Rinehart: Investigation, Writing—original draft, Writing—review and editing

APPENDIX: SEMI-STRUCTURED INTERVIEW PROTOCOL

Library Collaboration in RDM project, OCLC Research

1. Please introduce yourself, your job title, and **briefly** share your interest/engagement with the <<organization/effort>>. (5 min)
2. [ORIGIN STORY]
In your own words, tell us about how this effort began. What were the initial goals? Plans? Participants? (10 min)

Question purpose: to understand the historical underpinnings of this collaboration and to understand the drivers, participants, scope, and key decisions that were made in bringing the effort about.

3. [ORIGIN STORY: Priorities]
In choosing to collaborate to develop this effort, some aspects were likely more important than others. Examples of possible aspects are listed below (but are not exclusive). Which aspects were most important to participants when choosing to collaborate? Which ones were they willing to sacrifice? (10 min)

Question purpose: to identify the trade-offs that were made in deciding to collaborate. We want to document which dimensions were prioritized in setting up the collaboration—what mattered more and what didn't.

Here is a starter list of “sourcing dimensions” that might help your reflections:

- 1) **Agility**
Ability to respond with flexibility and nimbleness as conditions evolve or change
- 2) **Speed**
Rapid deployment, immediately addressing user
- 3) **Risk**
Does the decision to collaborate mitigate risk & uncertainty? Or does it increase risk?
- 4) **Community**
Does the decision to collaborate incorporate the values, practices, and in some cases, the members of the library community?
- 5) **Innovation**
Does the sourcing strategy identify and take advantage of opportunities to innovate around existing systems/practices/technologies?
- 6) **Total cost**
What is the cost projection (i.e., “total cost of ownership”) over the lifetime of the effort?
- 7) **Sunk costs**
Are significant upfront investments required, that cannot be re-couped if the strategy flounders or is ended?
- 8) **Efficiency**
Does the investment reduce the per-unit costs?

9) **Autonomy**

Must the local institution cede significant decision-making authority to an external entity?

10) **Quality**

Does this effort offer “best in class” performance?

11) **Lock-in**

Does this effort lock the library into a technological monoculture or proprietary eco-system that reduces or constrains choice?

12) **Reputation**

Does this enhance the institution’s reputation among stakeholders or peers?

13) **Endgame/Exit plan**

Is there a clear plan or terms for exiting the strategy?

4. **[ORIGIN STORY: Environmental factors]**

The decision to collaborate is informed by the institutional and contextual environment in which decision-making takes place. Often this environment helps influence the relative importance of the factors we talked about in the previous question. **We provide you with a starter list of environmental factors (below)—were any of these particularly influential in establishing this effort? Were there other factors not listed here? (10 min)**

Question purpose: to understand, for this case study, what local/institutional factors exerted the most influence in shaping the effort.

Here is a starter list of “environmental factors” for consideration:

- 1) The need for customization to meet local requirements
- 2) Budgetary status
- 3) Fiduciary responsibilities
- 4) Institutional values or priorities
- 5) Issues of data ownership and portability
- 6) Inter-unit relationships on campus (e.g., does the central IT unit disapprove of “home-grown” systems?)

5. **[CURRENT: Coordination & governance]**

Let’s talk a bit about how the collaboration is organized as a group to achieve its goals. **How does this effort act collectively to provide the shared capacity? How has this effort been able to get all the partners to commit to make the collaboration work? (5 min)**

Question purpose: To identify challenges and solutions regarding how the group acts collectively to achieve shared goals.

6. **[CURRENT: Evaluating the costs of running the collaboration]**

What are the significant investments (time, effort, resources) that were/are needed to organize and participate in the collaborative arrangements? To clarify, we are asking about the costs for operating the collaboration, not the costs for operating the service provision/capacity. (10 min)

Question purpose: To identify the significant costs, effort, investments needed to build and maintain a collaborative arrangement (and not the costs of running the service/operation).

7. [CURRENT: Change and competing priorities]

There can be a lot of challenges in moving to a collaborative arrangement when the collaboration strategy conflicts with previous sourcing decisions, institutional sourcing philosophies, and existing stakeholder preferences and relationships. For example, an institution may have already made investments in legacy systems with proprietary technologies that would not integrate well with collaboratively-produced capacity. Another example might be that an institution could have existing relationships/providers that they privilege, meaning that they are reluctant to enter into new partnerships. **Did issues of this kind impact the organization and evolution of this project? (10 min)**

Question purpose: To understand how choosing the collaboration option interoperates socially and technically (or not!) with previous sourcing decisions or existing sourcing philosophies and stakeholders.

8. [CURRENT: Who makes the decisions]

Where does decision-making authority reside in this collaboration? What arrangements are in place to ensure that decisions align with the interests of the stakeholders? (5 min)

Question purpose: To identify who has the decision-making authority and how the collaboration tries to ensure that authorized decision makers are acting in the interests of the whole.

9. [FUTURE]

Looking ahead: (10 min)

- a. **What do you *want* the future of this collaboration to be?**
- b. **What do you *think* the future is most likely to be?**

Question purpose: to understand future plans and initiatives for the collaboration, as well as potential risks. This question helps us understand both opportunities and challenges that the initiative may encounter in the future.

10. [FUTURE]

If someone were to ask you for advice on library collaboration based upon this experience, what would you tell them? (5 min)

Question purpose: to collect and share lessons learned from collaboration experiences with the broader library community.

11. Do you have any questions for us? (5 min)

NOTES

1. Lavoie, Brian, Dennis Massie, and Chela Scott Weber. 2023. *Sustaining Art Research Collections: Using Data to Explore Collaboration*. With a foreword by Jon Evans. Dublin, OH: OCLC Research. <https://doi.org/10.25333/V636-PE98>.
2. Massie, Dennis, Chela Scott Weber, Mercy Procaccini, and Brian Lavoie. 2023. *Sustaining Art Research Collections: Case Studies in Collaboration*. With a foreword by Amelia Nelson. Dublin, OH: OCLC Research. <https://doi.org/10.25333/kc2z-an73>.
3. Connaway, Lynn Silipigni, Ixchel M. Faniel, Brittany Brannon, Joanne Cantrell, Christopher Cyr, Brooke Doyle, Peggy Gallagher, Kem Lang, Brian Lavoie, Janet Mason, and Titia van der Werf. 2021. *New Model Library: Pandemic Effects and Library Directions*. With a foreword by Janice Welburn. Dublin, OH: OCLC Research. <https://doi.org/10.25333/2d1r-f907>.
4. Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/wyrd-n586>.
5. Lavoie, Brian. 2022. *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/mt16-0c57>.
6. Texas Digital Library. n.d. "Texas Data Repository." Accessed 12 September 2023. <https://www.tdl.org/texas-data-repository/>.
7. Digital Research Alliance of Canada. n.d. "Research Data Management." Services. Accessed 12 September 2023. <https://alliancecan.ca/en/services/research-data-management>.
8. Data Curation Network. n.d. "Home." Accessed 12 September 2023. <https://datacurationnetwork.org/>.
9. Lavoie, Brian. 2022. *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/mt16-0c57>.
10. Ibid.
11. Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/wyrd-n586>.
12. Whitchurch, Celia. 2015. "The Rise of *Third Space* Professionals: Paradoxes and Dilemmas." In *Forming, Recruiting and Managing the Academic Profession Perspective. The Changing Academy – The Changing Academic Profession in International Comparative Perspective* (book series), edited by Ulrich Teichler and William K. Cummings, vol. 14: 77-99. Switzerland: Springer, Cham. https://doi.org/10.1007/978-3-319-16080-1_5.
13. Lavoie, Brian. 2022. *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*, 5-7. Dublin, OH: OCLC Research. <https://doi.org/10.25333/mt16-0c57>. [This publication explores four sourcing options: build, collaborate, outsource, and refer, along with the trade-offs of choosing each sourcing option.]

14. Cooper, Danielle M., Catharine B. Hill, and Roger C. Schonfeld. 2022. *Aligning the Research Library to Organizational Strategy*. Ithaca S+R. <https://doi.org/10.18665/sr.316656>.
15. See Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/wyrd-n586>.
16. Texas Digital Library (TDL). 2021. "About." Posted 20 August 2021. https://www.tdl.org/wp-content/uploads/2021/08/TDL_About_2021-08-20.pdf;

TDL. 2023. "About TDL." <https://www.tdl.org/about-tdl/>;

TDL also supports group-scale services for institutional repositories, electronic theses and dissertations (ETDs), open access journals, digital preservation, open educational resources support, and the Digital Public Library of America. See TDL. 2023. "Services." <https://www.tdl.org/services/>.
17. TDL. 2023. "Governance." <https://www.tdl.org/governance/>;

TDL. 2023. "Member & Community Groups." <https://www.tdl.org/members/groups/>.
18. Holdren, John P. 2013. "Increasing Access to the Results of Federally Funded Scientific Research." Washington DC: Executive Office of the President, Office of Science and Technology Policy (OSTP). Memo posted 22 February 2013. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf. [Today this is also known as the "Holdren memo," to differentiate it from the more recent August 2022 OSTP public access memo, frequently called the "Nelson memo."];

Kozlov, Max. 2016. "NIH Issues a Seismic Mandate: Share Data Publicly," *Nature* 16 February 2022, 602, no. 7898: 558–59. <https://doi.org/10.1038/d41586-022-00402-1>;

Nelson, Alondra. 2022. "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." Washington DC: Executive Office of the President, Office of Science and Technology Policy (OSTP). Memo posted 25 August 2022. <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf>. [Today this is also known as the "Nelson memo," to differentiate it from the earlier February 2013 OSTP public access memo known as the "Holdren memo."]
19. Leggett, John, Mark McFarland, and Drew Racine. (2005) 2006. "The Texas Digital Library: A Business Case." Revised 12 July 2006. <https://tdl-ir.tdl.org/server/api/core/bitstreams/f637aac9-8ade-4d57-8fe5-96ee10292d71/content>.
20. Holdren, John P. 2013. "Increasing Access to the Results of Federally Funded Scientific Research." Washington DC: Executive Office of the President, White House Office of Science and Technology Policy (OSTP). Memo posted 22 February 2013. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
21. Herbert, Bruce, Martha Buckbee, Jeremy Donald, Maria Esteva, Colleen Lyon, Christie Peters, Kristi Park, Ryan Steans, and Santi Thompson. 2015. *TDL Data Management Working Group Report*. Texas Digital Library. <https://tdl-ir.tdl.org/handle/2249.1/68438>.

22. Thompson, Santi, Bruce Herbert, Kristi Parks, Jeremy Donald, Denyse Rodgers, Sean Buckner, Wendi Kaspar, Cecilia Smith, Christopher Starcher, Todd Peters, Ray Uzwysyn, Ryan Steans, Nick Lauland, and Laura Waugh. 2016. *Texas Digital Library Dataverse Implementation Working Group Final Report*. Texas Digital Library. <http://hdl.handle.net/2249.1/76364>.
23. Texas Digital Library. 2023. "Texas Data Repository: A Statewide Collaboration of Texas Higher Education Institutions." Repository. TDL.org: Texas Digital Library. <https://dataverse.tdl.org/>.
24. Texas Data Repository members as of February 2023:

Baylor University; Texas A&M International University; Texas A&M University; Texas A&M University at Galveston; Texas State University; Texas Advanced Computing Center; Texas Tech University of Houston; University of Texas at Arlington; University of Texas at Austin; University of Texas Southwestern Medical.
25. "Texas Data Repository, Data Repository Service Level Agreement (SLA) between [Member Institution] and Texas Digital Library" (Microsoft Word template). 2016. TDR MOUs and SLAs. Texas Digital Library Wiki. Last updated 1 May 2018. <https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/629014529/TDR+MOUs+and+SLAs>. [To download the Word template select the hyperlinked "SLA" located on page];

"Texas Data Repository, Data Repository Memorandum of Understanding (MOU) between [Member Institution] and Texas Digital Library" (Microsoft Word template). 2017. TDR MOUs and SLAs. Texas Digital Library Wiki." Last updated 1 May 2018. <https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/pages/629014529/TDR+MOUs+and+SLAs>. [To download the Word template select the hyperlinked "MOU" located on page.]
26. Ibid.
27. Lavoie, Brian. 2022. *Library Collaboration as a Strategic Choice: Evaluating Options for Acquiring Capacity*. Dublin, OH: OCLC Research. <https://doi.org/10.25333/mt16-0c57>.
28. "Texas Data Repository User Documentation, Texas Data Repository." n.d. Texas Digital Library Wiki. Accessed 12 February 2023. <https://texasdigitallibrary.atlassian.net/wiki/spaces/TDRUD/overview>.
29. Nelson, Alondra. 2022. "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." Washington DC: Executive Office of the President, Office of Science and Technology Policy (OSTP). Memo posted 25 August 2022. <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf>.
30. Dempsey, Lorcan. 2023. "The Technology Career Ladder." *LorcanDempsey.net*. Posted 13 March 2023. <https://www.lorcandempsey.net/technology-is-not-on-the-outside/>.
31. Morin, Julie. 2021. "The Portage Network Joins NDRIO to Continue Advancing RDM in Canada." Canadian Association of Research Libraries (News), 01 April 2021. <https://www.carl-abrc.ca/news/the-portage-network-joins-ndrio-to-continue-advancing-rdm-in-canada/>.

32. The three largest research funding agencies in Canada—the Canadian Institutes for Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and the Social Science and Humanities Research Council (SSHRC)—frequently work to harmonize policy requirements, as a “Tri-Agency.” This includes policies such as:
- Government of Canada. 2016. “Tri-Agency Open Access Policy on Publications (2015).” Modified 20 December 2016. <https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/open-access/tri-agency-open-access-policy-publications>;
- Government of Canada. 2021 “Tri-agency Research Data Management Policy.” Modified 14 March 2021. <https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy>;
- Humphrey, Chuck, Kathleen Shearer, and Martha Whitehead. 2016. “Towards a Collaborative National Research Data Management Network.” *International Journal of Digital Curation* 11(1), 195–207. <https://doi.org/10.2218/ijdc.v11i1.411>;
- Former Portage Director Chuck Humphrey provided a rich summary of these efforts in his blog. See Humphrey, Chuck. 2012–2015. *Preserving Research Data in Canada: The Long Tale of Data* (Research data management blog series 2012–2015). <https://preservingresearchdataincanada.net/>.
33. Research Data Strategy Working Group. 2011. *Mapping the Data Landscape: Report of the 2011 Canadian Research Data Summit*. Carleton University Library data-centre. Government of Canada. https://library.carleton.ca/sites/default/files/help/data-centre/Mapping-the-Data-Landscape_Data_Summit_Report.pdf.
34. CANARIE (formerly the Canadian Network for the Advancement of Research, Industry and Education) is a non-profit organization that provides the infrastructure to support research at Canadian research and educational institutions, including Internet service, identity and access management, and much more. See <https://www.canarie.ca/about/>.
35. Humphrey, Chuck, Kathleen Shearer, and Martha Whitehead. 2016. “Towards a Collaborative National Research Data Management Network.” *International Journal of Digital Curation* 11(1), 195–207. <https://doi.org/10.2218/ijdc.v11i1.411>;
- Trevor, Lynn. 2011. “Canadian Research Data Summit convenes in Ottawa.” Government of Canada. News release, 14 September 2011. (Archived 15 September 2015). <https://www.canada.ca/en/news/archive/2011/09/canadian-research-data-summit-convenes-ottawa.html>.
36. “Project ARC.” 2013. *Canadian Community of Practice for Research Data Management in Libraries* (blog), 21 May 2013. <https://cancoprmd.wordpress.com/project-arc/>. [The four regional library consortia are l’Association des bibliothécaires du Québec / Quebec Library Association (ABQLA); Council of Atlantic Academic Libraries (CAUL); Council of Prairie and Pacific University Libraries (COPPUL); and the Ontario Council of University Libraries (OCUL).]
37. Humphrey, Chuck, Kathleen Shearer, and Martha Whitehead. 2016. “Towards a Collaborative National Research Data Management Network,” 198. *International Journal of Digital Curation* 11(1), 195–207. <https://doi.org/10.2218/ijdc.v11i1.411>.
38. DMP (data management planning) Assistant. n.d. “Welcome to DMP Assistant.” Digital Research Alliance of Canada (the Alliance) and University of Alberta. Accessed 17 September 2023. <https://assistant.portagenetwork.ca/>.

39. Current Expert and Working Groups are listed on the Alliance web page. See Digital Research Alliance of Canada. 2023. "Network of Experts." Services, Research Data Management. <https://alliancecan.ca/en/services/research-data-management/network-experts>.
40. Compute Canada was Canada's national high-performance computing system, providing resources to support analysis and active data management of large datasets. Like Portage, Compute Canada has been absorbed into the Digital Research Alliance of Canada. See:

Tourangeau, Robbin. 2022. "Compute Canada Operations Move to the Digital Research Alliance of Canada (the Alliance)." *Acenet and the Digital Research Alliance of Canada*. Posted 24 March 2022. [https://ace-net.ca/compute-canada-operations-move-to-the-digital-research-alliance-of-canada-\(the-alliance\).html](https://ace-net.ca/compute-canada-operations-move-to-the-digital-research-alliance-of-canada-(the-alliance).html).
41. CANARIE provided CAN\$2 million in funding to Portage in 2019, part of which helped steward FRDR toward full production. Product work continued with additional funding from the New Digital Research Infrastructure Organization (NDRIO), which was later renamed the Digital Research Alliance of Canada (the Alliance).
42. "Canada's CARL Portage Network Announces the Launch of Dataverse North." 2017. Dataverse Project Blog, 30 March 2017. <https://dataverse.org/blog/canada%E2%80%99s-carl-portage-network-announces-launch-dataverse-north>.
43. Ibid.
44. Borealis: The Canadian Dataverse Repository. 2022. "About." Updated 13 September 2022. <https://borealisdata.ca/about/>.
45. Borealis. n.d. "Partners." Accessed 18 September 2023. [https://borealisdata.ca/#partners](https://borealisdata.ca/#partners;);

Barsky, Eugene. 2022. "Where Should I Deposit my Data?" Data Deposit Decision Tree, version 2.2, February 2022. In *Public UBC RDM Documents*, by Doug Brigham and Eugene Barsky. University of British Columbia. Open Science Framework (repository). <https://osf.io/rc7de>;

Dalhousie University. 2023. "Research Data Management." Updated 18 September 2023. <https://dal.ca.libguides.com/rdm/chooserepository#s-lg-box-wrapper-19204059>.
46. The ISED support was administered by CANARIE (see also n. 26).
47. Canadian Research Knowledge Network (CRKN). n.d. "History of CRKN." Accessed 18 September 2023. <https://www.crkn-rcdr.ca/en/node/15581>
48. Humphrey, Chuck, Kathleen Shearer, and Martha Whitehead. 2016. "Towards a Collaborative National Research Data Management Network," 204. *International Journal of Digital Curation* 11(1), 195–207. <https://doi.org/10.2218/ijdc.v11i1.411>.
49. Ibid., 199.
50. Ibid., 198.
51. These are: Bureau de Coopération Interuniversitaire (BCI), Council of Atlantic Academic Libraries (CAAL), Council of Prairie and Pacific University Libraries (COPPUL), and Ontario Council of University Libraries (OCUL).

52. Haigh, Susan. 2015. "Chuck Humphrey Named Initial Director of Portage." Canadian Association of Research Libraries (News), 6 July 2015. <https://www.carl-abrc.ca/news/chuck-humphrey-named-initial-director-portage/>.
53. Tourangeau, Robbin. 2017. "LCDRI (Leadership Council for Digital Research Infrastructure) Presentation to CARL." PowerPoint presentation, high-level introduction to the research data management landscape in Canada at the CARL Fall General Meeting, Ottawa, Ontario, 8 Nov 2017. https://www.carl-abrc.ca/wp-content/uploads/2017/11/LCDRI-Presentation-Robbin-Tourangeau_EN.pptx.
54. Digital Research Alliance of Canada. 2023. "Board of Directors." <https://alliancecan.ca/en/about/board-directors>.
55. Johnston, Lisa R. 2018. "Barriers to Collaboration: Lessons Learned from the Data Curation Network," 38. *Research Library Issues*, no. 296: 37–43. <https://doi.org/10.29242/rli.296.5>.
56. Data Curation Network. <https://datacurationnetwork.org/>.
57. Data Curation Network. "Data Curation Primers." <https://datacurationnetwork.org/outputs/data-curation-primers/>.
58. For a detailed retrospective on DCN's founding and evolution, see:

Carlson, Jake, Mikala Narlock, Mara Blake, Joel Herndon, Heidi Imker, Lisa Johnston, Wendy Kozlowski, Cindy Xuying Xin, Sophia Lafferty-Hess, Hoa Luong, Wanda Marsolek, Jennifer Moore, Dorris Scott, Cynthia Hudson Vitale, Briana Ezray Wham and Sarah Wright. 2023. *The Art, Science, and Magic of the Data Curation Network: A Retrospective on Cross-Institutional Collaboration*. Michigan Publishing Services. <http://doi.org/10.3998/mpub.12782791>.
59. For more information on the planning phase, see University of Michigan Libraries Digital Conservancy. 2016–2017. "Data Curation Network Planning Phase (2016–2017)." Regents of the University of Minnesota. <https://conservancy.umn.edu/handle/11299/202808>.
60. For more information on the implementation phase, see University of Michigan Libraries Digital Conservancy. 2018–2021. "Data Curation Network Planning Phase (2018–2021)." Regents of the University of Minnesota. <https://conservancy.umn.edu/handle/11299/202809>.
61. Institute of Museum and Library Services. 2018. "Pennsylvania State University, Awarded Grant no. RE-85-18-0040-18." Laura Bush 21st Century Librarian Program. <https://www.ims.gov/grants/awarded/re-85-18-0040-18>.
62. Data Curation Network. "Data Curation Primers." <https://datacurationnetwork.org/outputs/data-curation-primers/>.
63. Data Curation Network. <https://datacurationnetwork.org/2022/08/01/duke-receives-impls-grant-to-build-on-dcn-training-2/>.
64. For more information on the concept of radical collaboration in RDM, readers are encouraged to consult the Association of Research Libraries' December 2018 issue of *Research Library Issues* no. 296: 1–68. <https://doi.org/10.29242/rli.296>. [This issue suggests that "not only is RDM an ideal scenario for exploring radical collaboration, but that this kind of collaboration has already resulted in demonstrable success in the RDM arena" (p. 3).]

65. Schonfeld, Roger C. "Why is the Digital Preservation Network Disbanding?" *The Scholarly Kitchen* (blog), 13 December 2018. <https://scholarlykitchen.sspnet.org/2018/12/13/digital-preservation-network-disband/>.
66. Data Curation Network. n.d. "Governance." Accessed 19 September 2023. <https://datacurationnetwork.org/about/governance/>.

For more information on our work related to
Library Collaboration in RDM, please visit:
oclc/collaboration-rdm-project.



6565 Kilgour Place
Dublin, Ohio 43017-3395

T: 1-800-848-5878

T: +1-614-764-6000

F: +1-614-764-6096

www.oclc.org/research

ISBN: 978-1-55653-282-5
DOI: 10.25333/knxc-xy74
RM-PR-216996-WWAE A4 2311