THE ART, SCIENCE, AND MAGIC OF THE DATA CURATION NETWORK

A RETROSPECTIVE ON CROSS-INSTITUTIONAL COLLABORATION

DATA CURATION NETWORK

THE ART, SCIENCE, AND MAGIC OF THE DATA CURATION NETWORK

A RETROSPECTIVE ON CROSS-INSTITUTIONAL COLLABORATION

AUTHORS

Mara Blake
Jake Carlson
Joel Herndon
Heidi Imker
Lisa Johnston
Wendy Kozlowski
Cindy Xuying Xin
Sophia Lafferty-Hess
Hoa Luong
Wanda Marsolek
Jennifer Moore
Mikala Narlock
Dorris Scott
Cynthia Hudson Vitale
Briana Ezray Wham
Sarah Wright

EDITORS

Jake Carlson
Mikala Narlock

ACKNOWLEDGMENTS

This work is the result of many hands. The authors would like to express their gratitude to the Alfred P. Sloan Foundation, especially program director Josh Greenberg, and the Institute of Museum and Library Services for funding this work and providing support throughout. The authors would also like to acknowledge their institutions, who supported and provided space for this project, as well as colleagues at these institutions who encouraged us. We are also grateful for Shawna Taylor for her insightful comments and feedback on previous drafts of this report. The authors would like to especially thank every member of the Data Curation Network, for embarking on this experiment, for valuing trust and community, and for building a network we are all proud of.

Copyright © 2023 by Jake Carlson and Mikala Narlock
Some rights reserved

This work is licensed under the Creative Commons Attribution 4.0 International Public License. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/ or send a letter to Creative Commons, PO Box 1866, Mountain View, California, 94042, USA.

Published in the United States of America by
Michigan Publishing
Manufactured in the United States of America

DOI: http://doi.org/10.3998/mpub.12782791

ISBN 978-1-60785-813-3 (paper)
ISBN 978-1-60785-814-0 (open-access)

An imprint of Michigan Publishing, Maize Books serves the publishing needs of the University of Michigan community by making high-quality scholarship widely available in print and online. It represents a new model for authors seeking to share their work within and beyond the academy, offering streamlined selection, production, and distribution processes. Maize Books is intended as a complement to more formal modes of publication in a wide range of disciplinary areas.
http://www.maizebooks.org

DATA
CURATION
NETWORK

CONTENTS

Executive Summary v
About this Report ix

Data Curation Network Background 1

Radical Interdependence 6

Top 10 Successes 9

#10 Research and Experience to Inform Our Practice 9
#9 Data Curation Primers 10
#8 Actionable Curation Practices and Standardized CURATE(D) Workflow 11
#7 Broadening Our Expertise by Leveraging the Collective 13
#6 Radical Interdependence to Address Wicked Problems 14
#5 Engagement in Support of Advancing Data Curation Work 14
#4 Training and Education 15
#3 Documentation Improvements for Member Institutions and in the Network 16
#2 Models of Data Services 17
#1 Community to Learn From 18

Collaborative Structures that Enabled Successes 20

Administrative Structures 20
Trust-based Structures 27
Tool-based Structures 30

Challenges and Future Directions 33

Balancing Local and Shared Curation 33
Homogeneity of DCN Member Institutions 34
Addressing Lack of Researcher Response 36
Managing Community Capacity 37
Ensuring Sustainability of the Network 38
In Hindsight. . . 39

Potential Collaborative Endeavors 42
Supporting Data Communities Through Data Management 42
Research Data Management Plans 43
Dealing with Restricted Data 43
Data Discovery 44
Applying Metadata Standards 45

Conclusion 47
References 49
Appendix A: Data Curation Network Members and Institutions 53
Appendix B: Retrospective Agenda 55
Appendix C: Collaborative Structures 59

EXECUTIVE SUMMARY

In March 2022, sixteen members of the Data Curation Network (DCN) gathered together in Washington DC to participate in a retrospective discussion about the development, activities and impact of the DCN. The sixteen attendees included both administrators and curators, some of whom had been with the DCN since its inception, while others had joined more recently. All attendees were asked to consider our experiences in developing, operating and growing the DCN. Our discussion was wide ranging and included identifying and unpacking our successes, an examination of the structures we developed to operate the DCN, as well as the challenges we faced and what we might have done differently in hindsight. This report captures the results of our project retrospective meeting, in the hope that the insights described will be of use to other collaborative efforts.

The DCN is an example of a successful collaborative effort to address shared challenges of developing, implementing and scaling services to curate, share and preserve research data. The DCN was formed in 2016 by six librarians who had taken on the responsibilities of establishing and running their libraries' data repositories to disseminate and preserve the research data generated at their respective institutions. Data curation was not commonly offered as a service provided by libraries at the time and there were few models to follow in launching and delivering these services. Instead of navigating through the process of developing their respective services on their own, these six decided to come together to define the work, create best practices and develop shared resources that would benefit them all. Although initially conceived of and established through grant funding, the DCN transitioned to a sustainable, member-funded organization in July 2021. The DCN is now composed of almost 50 data curators from 17 institutions.

The DCN has evolved into a complex organization. Our members have tightly woven the DCN into their local services and have come to depend on the DCN to carry them out. By increasing member access to data curators with a variety of expertise from different institutions, each member institution is able to provide a greater depth of service to their local constituencies than they could have independently. However, the DCN's model also carries a fair amount of risk. If our members do not deliver on their commitments to each other, our local services would be negatively impacted. Thus, the DCN depends upon developing and sustaining a community of highly collaborative members that are exceedingly invested and motivated in the success of the network, are deeply trusting of each other and hold each other accountable. We define this as our "radical interdependence" model.

The labor of building and sustaining an active community is an artform that requires humility, intuition, and vulnerability. In reflecting on the work and success of the DCN, a key theme that emerged was the DCN's ability to center humanity: in our curators, in our workflows, and in our leadership, particularly acknowledging and leaning into our feelings of vulnerability. Being vulnerable with each other and admitting when we need support has allowed us to further advance a sense of shared ownership, responsibility and community. The continued viability of the DCN depends on each individual member feeling comfortable enough to bring their most authentic self to the network.

At the retrospective meeting, attendees were asked to reflect on their experiences in contributing to and engaging in the DCN community, as well as the factors that contributed to its success. We began by defining our successes collectively. In addition to our "radical interdependence" approach, our successes include the resources we have developed, such as our data curation primers that serve as reference documents for working with particular type of formats of data (Hudson Vitale et al. 2020), or the CURATE(D) workflow that we use in curating data (Data Curation Network 2018); our actions, such as our research projects or our advocacy in support of data curation (e.g., Wright et al. 2022); and our investments, such as our training programs.

We then moved on to a discussion of the support mechanisms and structures developed by the DCN that enabled our successes. We grouped these components into three broad and interrelated

categories: administrative structures, tool-based structures, and trust-based structures. The administrative structures we identified as being important to our success included the grants that we received, which gave us the resources and space to pilot, define and create the DCN model, and our shared governance model, collaboratively designed with the continuing evolution of the DCN in mind. The tool-based structures are the resources that DCN members have created to support the services provided by the DCN, such as our workflow management system used by members to submit a dataset to the DCN be curated, or the tools developed to assist curators with their work, such as the data curation primers. The DCN's trust-based structures are intended to support the ongoing work of building community through radical interdependence. These include our annual All Hands Meeting where members come together to engage, train and share with each other, and our peer-to-peer comparisons, biweekly semi-structured discussions for members to discuss successes and challenges they are facing.

Although the DCN has been an overall success, we have encountered numerous challenges along the way, many of which are ongoing. The next phase of our retrospective meeting was to reflect on the challenges we faced in launching the DCN, as well as what we might change if the network were to be created today. Our conversation included addressing our shared local challenges, such as what to do when researchers do not respond to our communications, or how to handle datasets that include sensitive information. Retrospective attendees also discussed the organizational challenges inherent in maintaining and growing the DCN, including how quickly we should grow and how to meaningfully include institutions beyond resource rich, research-intensive institutions. We concluded the retrospective by examining other potential opportunities to tackle additional grand challenges in data sharing and curation via cross-institution collaboration.

We present this report of our retrospective meeting with the intention and hope that our insights and experiences in designing, launching and growing the DCN will be of use to others who are engaged in or considering developing cross-institutional collaborations and partnerships of their own.

ABOUT THIS REPORT

The DCN is a collaboration of academic research institutions and non-profit organizations that share data curation expertise and staffing to support researchers in sharing their datasets.[1] We define data curation as a set of actions that "enable data discovery and retrieval, maintain data quality, add value, and provide for reuse over time through activities including authentication, archiving, metadata creation, digital preservation, and transformation" (Data Curation Network n.d., "Our Mission"). Datasets, from the perspective of a curator, then, are a collection of information that has a key role in the long-term scholarly record, and needs to be well-described to maximize the utility of the content. Through the DCN, members can focus on curating research datasets, and coalesce around this specific component of research data management to develop and share recommended practices.

Based at the University of Minnesota, the DCN is a member-funded consortium that facilitates a shared curation workflow, in which datasets from one institution are matched with an expert at a different member institution (e.g., Johnston et al. 2018b). The DCN is also a thriving community of practice, in which members can exchange information and learn from one another. DCN members help shape the future of data curation by developing community-oriented educational resources, offering professional development opportunities inside and outside of the DCN, and facilitating conversations around data curation topics through special interest groups, webinars, and other virtual gatherings.

The DCN's accomplishments are notable since its inception in 2016 (See Figure 1 for timeline of the DCN). In addition to collectively curating nearly 300 datasets through a shared curation workflow, the DCN has, to date, published numerous peer-reviewed articles and

1. The Data Curation Network, at time of writing (May 2022), consists of 15 member institutions. See Appendix A for full listing of DCN member institutions and individual contributors.

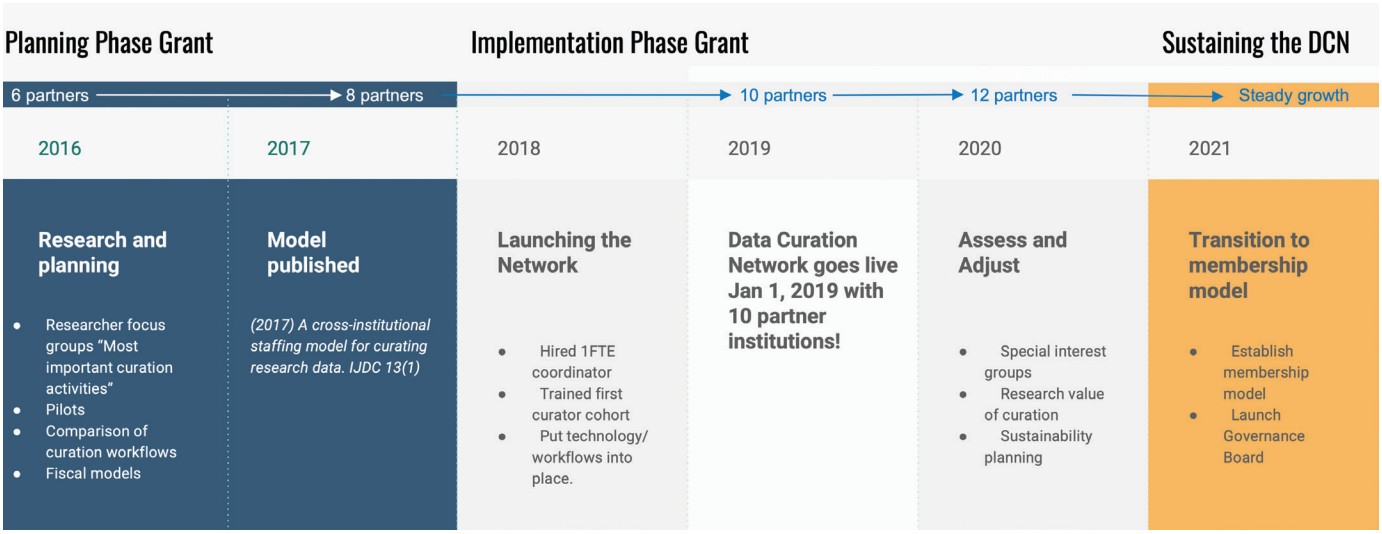

Figure 1: Development timeline of the Data Curation Network.

reports, presented more than 50 posters and presentations, created nearly 30 data curation primers, collaborated on several national-level education and research efforts, and trained nearly 200 curators through online and in-person workshops. Through its work and the shared vision of its members, the DCN has established an enduring community of practice.

As noted previously, in March 2022 members of the DCN held a project retrospective (see Appendix B for the retrospective agenda and attendee list) to reflect on the outcomes of its work. In particular, the project team was tasked with collaboratively unpacking the experience of establishing the network in order to identify the tools, structures, and support mechanisms that enabled the community's success. Feedback gathered during the retrospective was solicited through open discussion and interactive tools (e.g., Padlet, an online, interactive platform for capturing and discussing ideas in a "sticky-note" like format).

This report captures the discussion of the attendees during the retrospective, and is meant to provide in-depth information about our experiences in launching and operating the DCN. With this report, we seek to build on the existing community development literature within the field of library science, such as that provided by the Educopia Institute (n.d., "Research"). Our goal is to supplement this literature with an in-depth case study and analysis of our work. We hope that the approaches taken in this report are useful not only for understanding the art, science, and magic of the DCN, but in providing a model for building and implementing a cross-institutional community-based network.

Data Curation Network Background

The idea of forming a data curation network came from a casual discussion among peers at a conference. We were sharing stories about establishing our respective repositories and services at our libraries and quickly realized that we all were facing similar challenges. Many of our institutions were just getting started with managing datasets and curating research outputs, and our job responsibilities were shifting more toward focusing on preserving and providing access to research data. However, at the time, there were few models for us to refer to in conceptualizing and running a data repository and supporting services out of an academic library.

Over the course of our conversation we began to wonder: why couldn't we work together to address these challenges? And not just in sharing information, but in actually doing the work of curating the data that we would publish in our repositories? What started off as offhand comments led to an increasingly serious discussion. What if we could mitigate or even remove institutional and other barriers that prevented us from leveraging our collective knowledge and expertise for the benefit of all of us? Could we work so closely with one another that our services would come to depend on each other to function? Did we dare even try this, knowing the risks that it might not work? The original members of the DCN did not know how, or even if, such a network would work out, but we each saw the potential inherent in the idea and committed ourselves to giving it a try.

The first steps in this process were to identify how we wanted to work together, what we hoped to accomplish, and where we could secure resources to support our work. After sharing our vision for the DCN with the Alfred P. Sloan Foundation, they invited our team to submit a proposal. Our planning grant proposal centered on

developing a proof of concept for what a "data curation network" might look like. Specifically, we focused on five key areas:

- defining a staff and governance model that specified both the benefits and expectations of membership;
- building a shared submission workflow between the DCN and the data curators at each member institution;
- crafting an implementation plan to serve as a roadmap in growing curation offerings and membership in a deliberate and measured pace over time;
- developing a financial plan focused on the long-term sustainability of the DCN from its inception; and
- creating an assessment plan to demonstrate the benefits of a networked approach to curating research data in meaningful ways.

The Sloan Foundation approved our planning grant proposal in 2016, giving the six original member institutions—the University of Minnesota (lead), Cornell University, Penn State University, the University of Illinois, the University of Michigan and Washington University in St. Louis—the initial funding we needed to begin developing the DCN. More information about the planning grant and the results of our work can be found in our final report to the Sloan Foundation (Johnston et al. 2017).

In 2018, we received a second award from the Sloan Foundation to implement our plans. In this grant, we added two additional member institutions, Duke University and Johns Hopkins University. We also set our sights on transitioning to a sustainable economic model based on securing the resources needed to fund the DCN from our members, rather than relying solely on external sources such as grants. Information on the development and implementation of the DCN's sustainability model can be found in our sustainability plan we developed near the conclusion of the second grant (Johnston et al. 2020).

Six years after our initial planning grant was awarded, the DCN is thriving. We have grown from six curators from six institutions into an organization of nearly 50 curators from 17 institutions (see Appendix A for full listing of participants and supporting institutions). In looking back on our early days, we think the following

factors were critical in enabling the DCN to get off the ground and launch successfully:

- **Strong existing relationships between founding members**: The six original DCN members had established relationships as collaborators and colleagues. This reduced the barriers that come with establishing new relationships, and provided a solid foundation on which to build.
- **Building trust through establishing community norms**: In addition to a code of conduct (see trust-based structures), DCN members also dedicated working time early and throughout the grant to revisit and reestablish community norms. We held one another accountable for creating a safe environment and led by example.
- **A willingness to dedicate significant amounts of time and energy to this project**: Despite the uncertainty and potential risks of this work, each DCN member, with the support of their library administrations, was willing, able, and even eager to spend time on developing the network.
- **Grant funding to give us the resources (and bandwidth) to experiment:** In addition to funding an additional full-time project coordinator to help develop the DCN, the grants from the Sloan Foundation provided DCN members the space to create, iterate, and experiment.
- **Conducting applied research to create practical resources**: DCN members engaged in multiple research projects to understand the needs and practices of researchers and curators (e.g., Johnston et al. 2018a; Johnston 2018; Johnston 2020) and then applied what was learned into tools and structures to support its work.
- **A high tolerance for uncertainty and risk**: There was no guarantee this network would persist beyond the initial planning grant, however each of the original members of the DCN believed in the project and saw the potential for a great return on their investment.

All of these factors were critical for building the solid foundation of the network, for creating an environment conducive to collaboration, and for empowering the community to keep developing.

Additionally, our work has been informed by and grounded in international principles and best practices, which allowed our team to coalesce around aspects of data curation with a shared understanding. In particular, the DCN centered the *findable, accessible, interoperable, and reusable* (FAIR), and *collective benefit, honor the authority to control the data, share data responsibly, and operate ethically* (CARE) principles during its development. These principles have been adopted as best practice in the data sharing community, are increasingly being required by funders, publishers, and scientific communities, and are useful guideposts for the DCN in our work. We strive to make the datasets that we curate as FAIR as possible (Wilkinson et al. 2016). Applying the FAIR principles encourages the scholarly treatment of data by enabling the datasets to be reusable objects of scholarship that can be properly cited, maximizing its reusability and the impact of our researchers. Similarly, our curators also strive to follow the CARE principles of Indigenous data sovereignty (Carroll et al. 2020). The CARE principles are a lens to recognize the impact data has on human lives and the humans at the center of the data, even when the data is not necessarily about the people themselves, and they remind us that our job as curators is to ensure data is shared in ways that correct past harms and do not create new harm. In short, by aligning our work with both the FAIR and the CARE principles across the network, DCN members curate datasets that are shared ethically and with reuse in mind. This shared grounding in international principles was also critical for building the foundation of the community.

As the DCN has grown, its foundation rests on the trust of its members. It is worth highlighting that this trust was not formed by accident, but instead was the result of a conscious and deliberate effort to develop a community that empowers data curators. In both the planning and implementation of the DCN, team members supported open conversations and were dedicated to empowering one another. While the network was based on strong existing relationships between members, the collegiality continued to develop through discussions around a set of difficult challenges to address, clear benefits in collaborating with one another, a shared passion for advancing data curation, and a high tolerance for uncertainty and risk. Important in these efforts were regular working meetings, virtually and in person, to nurture existing relationships and establish

community norms. Particularly telling is the concluding sentence of the article "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services" (Johnston et al. 2017, 24) in which the initial members of the DCN write:

> by intentionally structuring our efforts to coordinate as a Network that can grow and incorporate new institutions over time, we hope to play a role in engaging and empowering the larger data curation community through sharing experiences and providing a platform for continued dialog and discussion in this area.

Additionally, the DCN has prioritized slow, thoughtful, deliberate movements in building our community. Instead of trying to move quickly and risk breaking trust, our leadership has been diligent in slowing down to ensure all community members can participate in the work of the network. In hindsight, some of our efforts reflect the *Slow Food* movement (Petrini 2003), of which its theoretical grounding has recently been applied to libraries (Glassman 2017; Farkas 2021) and archives (Christen and Anderson 2019). Our team has chosen to move strategically, even when that might feel slow. This time and space have allowed us to be thoughtful about our work, provided us time to reflect on success and challenges, and to continuously improve. This does not mean the DCN is not able to pivot or be agile, only that the team has embraced saying "No" to opportunities when the timing is not right in order to make space, in order to say "Yes" to the opportunities that are at the heart of our community. We move slowly so we can move together as a community.

The careful planning, thoughtful expansion, and intentional trust building have been critical efforts within the DCN, even if difficult to quantify or capture—and they have enabled the successes of our community.

Radical Interdependence

In the DCN, trust is the foundation of our work—specifically, what we term "radical interdependence." Radical interdependence is more than cooperation—it is relying on others outside of your organization to complete work for our individual institutions; it is providing and receiving learning opportunities; and it is leaning on one another for support when necessary. It is about collaboration, trust, vulnerability, and accountability, not from a punitive perspective, but from an intrinsic desire to show up for and serve one another based on clear expectations, commitments, and the shared goal of advancing the field. It is about pooling our resources, our knowledge, and our interests, to better accomplish our work and advocate for the role of data curators. It is about recognizing the interconnectedness of the work we are all doing. In our shared curation workflow—in which dataset curation is connected to our distributed network of experts to ensure data are curated robustly from disciplinary and format experts—we trust that the work we assign one another will be completed on-time and to a high-caliber level.

In our radical interdependence model, DCN members are invited to bring their most authentic selves to our work, to be vulnerable and trusting, and recognize that we are all continuing to learn and grow in this field. Members of the DCN are empowered to say "I don't know, can you help me?" and then trust that their fellow members will support them and follow-through on their commitment. This cooperation allows us to address complexities beyond our control, such as the limitations of our local expertise or capacity, to serve our researchers, elevate the roles of libraries, curators, and data stewards, and to continue advancing and advocating for open research. We have the benefit of relying on each other for support and education to learn together and improve our skills and expertise. As a community

of practitioners and experts, we recognize that each DCN curator brings a unique background and skill set that allows the DCN to collectively curate a broader range of datasets than any one of us could individually.

Through our model of radical interdependence, in which we educate and empower each other, the DCN also increases the reputation and trustworthiness of institutional members. By ensuring that datasets are curated by librarians with domain and format expertise, researchers at member institutions receive a thorough curation of their datasets and code, with a wider-array of expertise beyond what any institution could provide independently. This in turn increases the trust researchers have in DCN member repositories, as researchers can share their data with the assurance that their datasets will receive a high level of curation support. Based on internal DCN documentation, of the datasets submitted through the shared curation process (n=275), our curators intervened in 80% of datasets submitted, including remediating major curation issues (44.5%), minor curation issues (31.4%), and engaging essential curation activities (4.9%), as can be seen in Figure 2. Without this work, many of the datasets we steward, though openly available, would be less reusable.

The DCN is also seen as a leader in open science, and often called on to collaborate and partner in education and research efforts. This includes open educational efforts (Data Curation Network 2022a; NIH 2022), research on and in support of data communities (Ruediger and

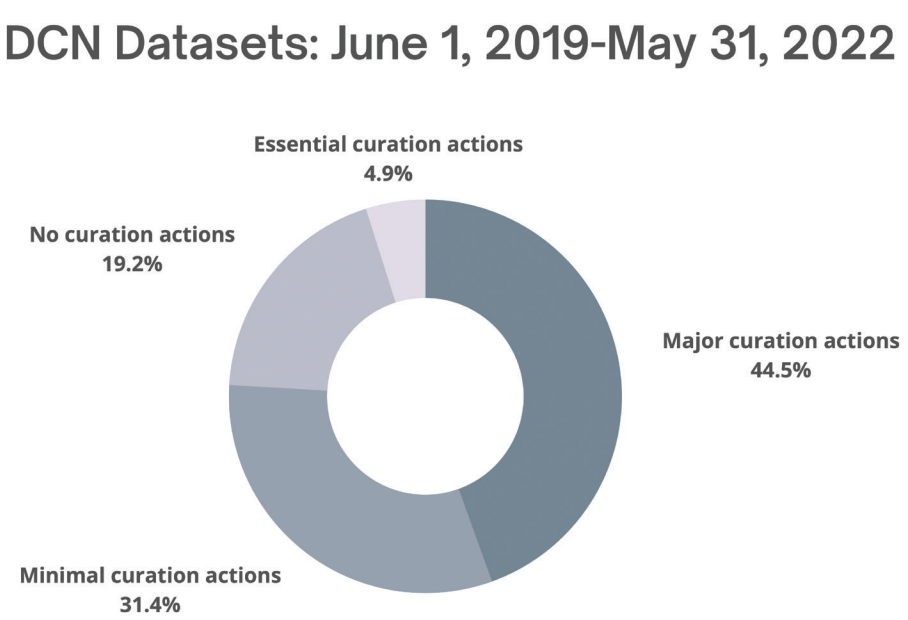

Figure 2: Level of curation support provided to datasets submitted to the DCN's shared workflow.

Cooper 2021), investigating the challenges and costs of academic data sharing (Association of Research Libraries 2022), and explorations of member practices (e.g., Data Curation Network 2022b). Because we are acting as a collective and leaning into our radical interdependence model, we can identify pain points beyond technological limitations and individual institutional nuances to address problems that could affect a broad array of research data management stakeholders, including researchers, data stewards, librarians, campus administrators, and future data (re)users. This work, and the other successes of the DCN, would not be possible without the radical interdependence model that is the cornerstone of our community.

Top 10 Successes

During our retrospective meeting discussions, we identified many successes of the DCN over the past six years. These ranged from concrete and quantifiable achievements such as datasets curated, to intangible yet equally significant accomplishments such as a sense of trust and radical interdependence. As a group, we identified the ten successes that were most impactful, for ourselves, our institutions, and our researchers. We present these successes below in a countdown format, but this ordering is not to suggest that one success is more important than another, because while each of the successes are listed as discrete items, many of them overlapped, intertwined, and in fact built off one another.

#10 Research and Experience to Inform Our Practice

The DCN's use of empirical research to assess needs, develop resources, and inform daily practices of data curators and researchers represents a core strength of the partnership. During the planning phase of this project (2016–2017) a literature review was conducted, complemented by surveys and interviews of research data librarians, data curation experts, academic library administrators, and researchers (Johnston et al. 2018a). The result of this early information gathering was a clearer understanding of researcher needs and curation best practices, and informed the curation workflow later developed and implemented by the network (Johnston et al. 2018b). This habit of informing practice with shared research

grounded in experience guided the early work of the network. Research remains a critical part of operations, particularly through special interest groups, sharing knowledge at the annual All Hands Meetings (AHM), and collaborating through a shared cloud-based workspace.

#9 Data Curation Primers

Data curation primers are detailed reference documents centered on a specific subject, disciplinary area, or curation task that can be used by curators when curating a dataset that falls outside of their immediate expertise. These step-by-step resources provide a shared knowledge base for a specific data format, method, or tool. The primers are based on a flexible template that was developed through a combination of faculty interviews and curator expertise (Sciolla and Borda 2018). The first primers were developed by teams composed of curators and participants at specialized data curation workshops offered by the DCN between 2018 and 2020 (funded by the Institute of Museum and Library Services [IMLS] Grant #RE-85-18-0040-18.) Over time, primer creation has expanded to welcome community contributions by domain data experts beyond this initial group, providing a valuable platform for publishing and sharing curation knowledge, as well as unifying community members around a shared task. After peer review, the primers are published in GitHub, an online version control platform, with archived PDF versions available in the DCN repository. While the archived versions provide a stable citation to credit the authors, sharing the primers through GitHub allows the outputs to be editable, living documents. Data curation is a rapidly evolving field, with changing best practices, new formats and technologies, and evolving standards. We recognize that the resources we create and provide today will need to be updated to keep pace with the profession; using GitHub allows us to refine and keep fresh advice for curators.

In addition to being a critical reference document that can be updated over time, the primers provide an opportunity to form a cohort around a topic of interest and importance to that data curator community (Hudson Vitale et al. 2020). The primers are designed

to support the work of all curators, both within and outside of the DCN. The primers are also being leveraged as instruction materials for future data curators and serve as a means to provide guidance for researchers themselves in preparing their data for sharing and archiving.

#8 Actionable Curation Practices and Standardized CURATE(D) Workflow

The definition of a curation workflow—a standardized set of actions each curator performs when reviewing a DCN-curated dataset—was a critical early success of the DCN. The DCN CURATE steps (later modified to CURATE(D)) is a genericized curation workflow that can be used when curating datasets, in managing and reviewing researcher files, and as an instruction tool. The creation of the CURATE(D) workflow came out of a recognition that we needed to develop a common understanding of the actions necessary to curate a dataset effectively and efficiently. This includes steps taken to curate all components of a dataset: the data files, code, metadata, and other possible content. The CURATE(D) steps are defined as follows:

C – Check files/code and read documentation
U – Understand the data (or try to)
R – Request missing information or changes
A – Augment metadata for findability
T – Transform file formats for reuse
E – Evaluate for FAIRness
D – Document all curation activities throughout the process

Since its inception, the CURATE(D) workflow has provided a grounding framework for DCN members. The initial workflow was drafted in the planning phase of the DCN and further enhanced by members of the DCN at the first annual AHM in July 2018 (Data Curation Network, 2018). This collaborative exercise of outlining a standardized workflow required a significant amount of time and energy from all DCN members. In order to create a workflow that is systematic enough to provide a minimum level of curation and is flexible enough to incorporate file-type-specific actions and

domain-specific knowledge, members worked together to articulate the key elements of curation practice.

The CURATE(D) steps are also a shared foundation for all curation that occurs through the network. In addition to the many DCN member institutions that leverage this model in their daily work with researchers and graduate students, the CURATE(D) steps ensure that all datasets submitted to the DCN through the shared curation model are curated consistently. In our workflow management system, where all DCN datasets are tracked, each curator is prompted to respond to every component of the CURATE(D) steps when providing feedback for another institution. This structured formatting is critical for maintaining a high-level quality of curation regardless of data format or type.

A great strength of the CURATE(D) steps is that they are designed to be actionable and implementable—but the individual curators, both within and outside the DCN, can continue to develop and improve the model. As noted, data curation is a rapidly evolving field, so it was important to design the workflow with room to iteratively adjust the steps as best practices change. Recently, the CURATE(D) steps were revised to include a preamble, defining the document and its goals, and key ethical considerations at each step. As with the initial creation of the steps, this required a significant investment of time and effort from DCN members (Data Curation Network 2022a).

Furthermore, with a flexibility that easily incorporates the addition of customized resources and tools, the CURATE(D) steps provide a framework for training that can be adapted for a wide range of audiences. For example, the CURATE(D) steps have been leveraged as a key mode of collaboration between the DCN and the National Institutes of Health (NIH). The data sharing requirements instituted by the NIH and scheduled to come into effect in January of 2023 are a shift from previous practice. The NIH will now require funded researchers to share data as openly as possible while respecting legal, tribal, and other regulations and ethical considerations. Recognizing the need to prepare researchers and other stakeholders for this change in policy, the NIH's Office of Data Science and Strategy has reached out to many community organizations, including the DCN, for support and assistance. DCN members recently partnered with the NIH to offer a series of training programs for medical librarians, program

officers and other NIH personnel (e.g., Carlson et al. 2022). This educational series leveraged the CURATE(D) steps and our experiences in curating data to demonstrate the processes that curators take when reviewing datasets.

#7 Broadening Our Expertise by Leveraging the Collective

The DCN has achieved its successes through a cooperative model that values interdependent collaboration and leverages the skills and expertise of members to benefit the collective. This was accomplished in several ways. First, the networked curation services enable individual institutions to practice data curation based on an established protocol. The DCN supports curators in developing additional expertise through the peer-to-peer network by curating datasets for other institutions, but also in sharing knowledge about how to curate specific types of datasets for future reference. Second, the collection of curation primers created by members with domain data expertise provides a foundational curation workflow for specific disciplines and formats. In creating this collection of resources, the DCN leverages the expertise of its members to create practical educational tools for data curation professionals. Finally, information exchanges between curators in a community are a critical example of harnessing the knowledge of members to have a greater impact within the broader data management and sharing community. This happens both asynchronously, such as via instant messaging and email discussions, and synchronously, at regular check-in meetings as well as at the annual AHM. In all of these examples, our team has relied on the experiences and knowledge of the community to establish a broader expertise than we could host alone.

In addition to leveraging the collective of the DCN, our community also relies on the knowledge and experience of related efforts. This includes partnering with entities such as the Digital Research Alliance of Canada, the GO-FAIR US Office, and the National Center for Data Services (NCDS) to keep one another informed and to increase our collective impact. Through this shared expertise, DCN members create invaluable resources and empower one another through learning opportunities.

#6 Radical Interdependence to Address Wicked Problems

The DCN has provided an opportunity for member institutions to collaboratively address shared and difficult challenges in data curation, to promote ethical data sharing, and to foster open science. Importantly, DCN members realize that what is beyond cooperation is interdependence. While the DCN is rooted in this radical interdependence, we also believe it is a key success of the network. The interdependence provides a venue for coping with challenging issues regarding data curation and data sharing in local institutions. This is because DCN's curation services model provides more than the shared curation workflow, in which members rely on one another to curate datasets. The DCN also allows members to share their local practices, such as administrative models that help guide both the shared curation workflow, while creating a safe and inclusive space for members to address shared challenges and issues collectively, to continue exploring the evolving field of data curation, and to learn from one another. As Heidi Imker, University of Illinois at Urbana Champaign, writes: "On one hand, sometimes it feels like we're not quite ready for the DCN, but on the other hand, the DCN has helped us become ready in ways that have made our curation services better right off the bat" (personal communication with Mikala Narlock, February 10, 2022). In other words, while most DCN members have established curation programs, the DCN provides a space to continue developing services or to receive a jump start when launching a new curation program, thanks in no small part to the radical interdependence that guides the network. While our interdependence is at the core of the DCN, it is also a critical success, one that, as was mentioned above, was intentionally cultivated and nurtured over time.

#5 Engagement in Support of Advancing Data Curation Work

Through the DCN, we have drawn attention to and demonstrated the potential for librarians and data curators to be essential partners in the data sharing process. We have conducted research to identify the value add that curation provides from the perspectives of

both researchers and repository managers (e.g., Wright et al. 2022; Johnston et al. 2021). Additionally, the DCN has presented on the work of its data curators and data curation more generally at conferences in the United States (e.g., the Research Data Access and Preservation Association, the Coalition for Networked Information), internationally focused conferences (e.g., the International Association for Social Science Information Services and Technology [IASSIST], the International Digital Curation Conference [iDCC]), and in special editions of journals (e.g., Hudson Vitale et al. 2021), all of which were intended to raise the profile of data curators and the importance of curation activities.

Outside of libraries and curation communities, the DCN has worked to raise awareness of the importance of curation to different stakeholders. The DCN's data curators are increasingly recognized as valuable partners in creating and offering training for researchers, program officers and other stakeholders on data curation topics. For example, as noted, the DCN recently collaborated with the NIH to offer a learning series for librarians, researchers, and program officers centering on data curation and review prior to research data publication. Through this and other outreach and engagement initiatives, the DCN has contributed to national and international efforts that help to raise the profile of data curators and stewards.

#4 Training and Education

The DCN has made education and training in data curation one of its primary focus areas. Data curation workshops were initially funded by grants, which were crucial for bringing a team of DCN members together to develop relevant and informative educational offerings. Over time, leveraging the CURATE(D) steps and first-hand experiences from curators, the DCN has continued to refine and implement training and education materials for the evolving data curation profession. The DCN education curriculum, typically presented in the form of workshops, is in high demand as a workforce development offering, and has trained nearly 200 students in-person and virtually. Library schools are increasingly offering courses to provide the education needed for students to assume positions in data services, however many librarians graduated before working with research data

was included in the curriculum. The DCN workshops are designed to fill this gap through presenting a case study and having students work through it with their instructors and peers. The hands-on experience presented in the workshops means that participants are able to apply the skills learned immediately after completing the training.

The success of the education efforts of the DCN is rooted in the individual contributors from the DCN. Each curator brings a unique perspective in developing the curriculum, which is invaluable when establishing the learning environment of the workshop. The efforts of the workshop developers and instructors have ensured that each workshop is centered around creating meaningful engagements between instructors and attendees. The culture and ground rules for using training and education materials within the network are presented in a peer-to-peer tone, all of which helps to establish a space conducive to learning, trust, and community. This approach is a key factor in the success of the education efforts, as attendees feel welcomed and empowered to ask questions, and are able to connect with colleagues to support one another long after the workshop's end.

#3 Documentation Improvements for Member Institutions and in the Network

A benefit of the DCN's radical interdependence model is the creation of shared and consistent documentation and templates for data curators. As mentioned previously, many member institutions have integrated the DCN's established CURATE(D) steps into their local workflow and training. This is particularly useful for nascent data services programs, as it provides a baseline for curation expectations. This method is also helpful for established programs, as it provides a digestible and implementable foundation for onboarding new members to the team.

In addition to generating documentation tools and templates collectively, the DCN has leveraged the work done by member institutions in support of the network. The README template that was created by Cornell University, for example, has been a useful resource for other peer institutions to build on and reference (Research Data Management Service Group 2022). README files are critical pieces of documentation that provide context into project datasets and other

outputs. Specifically, these files provide information on how data was collected, analyzed, and what contents are in a dataset, all of which are necessary for future reusers and are required for data to be understandable, trusted, and reused. The need for such a tool was a shared challenge across DCN institutions, and Cornell's template was readily adoptable by DCN members as a means of promoting consistency in README files. It also serves as a teaching tool for researchers, who can revisit the template when submitting new datasets.

A final example of the documentation created by or adopted by the network is a standardized email for requests for additional information or missing files from researchers, or to resolve any other issues uncovered during the curation review. This standardized email removes the challenge of how to structure communication pertinent to data curation, and instead provides a clear and consistent mechanism for engaging researchers. All of these tools are utilized at individual member institutions as well as within the network to promote consistent documentation practices and to remove the burden of each institution creating their own tools.

#2 Models of Data Services

Coming together as a data community allows member institutions to discuss different models of data services, engage in peer-to-peer comparisons about shared topics of concern (e.g., policy, workflow, preservation, etc.), and ultimately enhance local services by learning from one another. As a complement to the documentation templates discussed above, these informal discussions provide an overall sense of the research data landscape, including identifying mutual challenges and opportunities to collaborate in improving, refining, and expanding data services. The DCN as a space to openly share both successes and failures in a trusting environment allows for efficiencies of scale: instead of building our workflows, services, tools, and practices in silos, we leverage what is already being used effectively elsewhere within the network. Conversations like this primarily happen within DCN's special interest groups, which bring together members around a shared topic. The focus of these groups range from the highly technical, such as discussions on implementing Globus for large datasets, to those focused more on social and ethical

issues, such as the Racial Justice Working Group. While each group is left to define their own projects and goals, many have found value in discussions about recent projects, tools, and ideas that have been successful at their institution. For example, in the Institutional Outreach and Communications Interest Group, each institution voluntarily shared the marketing efforts that they have been using to raise more awareness on their campus about the local data services. By learning from others' successes and failures, member institutions are able to enhance their local practices and mature their data services to meet researcher needs. This success draws on the expertise of the collective to improve work at our individual institutions.

#1 Community to Learn From

Perhaps the primary success of the DCN is the active and engaged community of practice. Each institution brings a unique workflow with different staffing and technologies, all of which have allowed the DCN to foster a space for learning while developing a technology-agnostic shared curation service. Each curator brings different skills and interests to the network, which expands not only our shared curation expertise, but also contributes to research opportunities, new areas of engagement, and fresh perspectives within the network. The DCN creates a space for data professionals and stewards to exchange and broaden their knowledge across disciplines. Regular curator meetings provide an opportunity for members to share goals, discuss challenges, or ask questions that arose during the curation process. Additionally, instant messaging and email groups provide the opportunity to collaborate asynchronously and receive support. Being a part of this community of practice motivates innovations through information exchange among institutions.

The expertise of DCN curators, ranging from social sciences to life sciences, humanities to physical sciences and engineering, has helped member institutions fill gaps in their local expertise. This provides immediate curation support when necessary, and can be understood as a safety net for curators who may want to grow their expertise and upskill by working with a DCN curator to practice curating datasets that might fall outside of their disciplinary or format expertise. This support is two-fold: individuals receive the immediate support

of curation and benefit long-term by learning how to curate new and different types of research outputs from colleagues across the United States. This peer-to-peer networking allows individuals to receive a proverbial fish (the curated dataset) while learning how to fish (learning to curate new types of data). This has been a great professional development opportunity for many curators in the network: the benefits of having the specific type of datasets curated by fellow DCN curators increases member confidence in interacting with the researchers. Similarly, as one curator stated: "[the DCN] helped me overcome my fear of running the code and looking at this part of the data." This kind of support extends beyond professional development opportunities, which often end when the workshop, webinar, or other training session ends, and instead provides a space to continue learning, growing, and implementing the knowledge gained. The community is the foundation of the DCN, and will remain instrumental in supporting the network as it continues to develop and welcomes new members.

Collaborative Structures that Enabled Successes

In order to achieve these successes, the DCN has invested a great deal of thought into collaboratively designing administrative, trust-based, and tool-based structures to keep us organized, in communication, and fitted with resources, both as a larger community of practice and within our home institutions. At our core is a group of savvy collaborators who have put in place structures which can radiate out as we grow. Our lightweight administrative structure was designed to allow a high level of engagement, but as we have matured and expanded, we have added additional administrative structures and support to meet the demands of scale and sustainability. This includes developing shared, accommodating, flexible processes which provide a strong foundation for inter-institutional work, but allow plenty of room for local variation, replication, and divergence. When members need to accomplish something that is outside of our area of expertise, we reach out to one another for assistance. This could include developing a list of relevant tools to address an aspect of curating a specific type of data to recruiting a member to write up a grant in support of an initiative. The following sections provide a non-exhaustive list of structures that fostered community and advanced the work of the DCN. Although that have been split into three categories—administrative, trust-based, and tool-based—many of these structures overlap, intersect, and rely on one another, as has been visualized in Figure 3.

Administrative Structures

Implementing and running the DCN has required the development of multiple administrative structures, including collaborative efforts

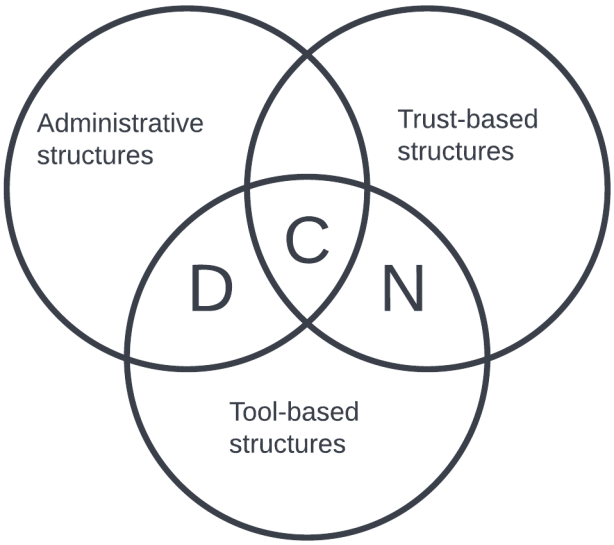

Figure 3: The overlapping structure types were essential for supporting the DCN.

when applying for **grants**, the curator **onboarding process**, leveraging **consultation services** to develop key administrative tools, and the **shared governance model**. Much of what has been successful with these administrative structures is that they foster many of the trust-based structures that we will draw attention to in the following section.

Grants and external funding opportunities have been a critical first step to establishing and building the DCN. They have allowed the DCN the space to plan, pilot, implement and adjust initiatives. Grants have also provided the resources for dedicated time to initiate and evaluate key areas of the DCN, such as the ability to hire a project coordinator to focus on the DCN full time. Both private and federal funding initiatives have funded planning, research, and implementation of various DCN initiatives into full-scale production.

The planning phase of the DCN (2016–2017) was generously funded by the Alfred P. Sloan Foundation and led by principal investigator (PI) Lisa Johnston then at the University of Minnesota. During this planning period, the six initial members of the DCN collaborated to develop a list of curation activities, conduct information exchanges with key partners, understand researcher needs for curation services, and more. This planning grant also provided the opportunity to pilot a shared staffing model for data curation among our six original institutions. In other words, this initial grant provided the opportunity to develop a conceptual model of the DCN and to see if it could work in practice. After the planning

phase of the grant, the DCN received a further three-year grant from the Alfred P. Sloan Foundation to implement the network, which was again led by PI Lisa Johnston. Over the course of the three years, the DCN implemented a workflow management system for the coordination of exchanging datasets and expertise among our members. Additionally, the project coordinator, Liza Coburn, led efforts to refine the networked curation workflow, streamlining dataset submissions as much as possible. DCN members also continued critical community development work, such as collaborative research and frequent meetings. This second grant allowed the team to implement the DCN shared curation model fully while providing capacity for the DCN team to plan for the transition from a grant-funded organization to a self-sustaining member-based organization. These grants were critical in jumpstarting these efforts, but the team was aware that a viable, sustainable community could not be dependent on grants. DCN members therefore planned for organizational sustainability by developing a shared governance model (outlined below) and hired an assistant director to support the network in its transition from a grant-funded to a member-funded organization.

Grant funding also played a central role in the DCN's education and outreach efforts. A 2017 event grant from IASSIST allowed members of the DCN to pilot a peer-to-peer, hands-on workshop based around the CURATE(D) workflow. In 2018, the education efforts of the DCN, led by PI Cynthia Hudson Vitale, then at Pennsylvania State University, received additional support in the form of an IMLS Laura Bush 21st Century Librarian grant titled "Building the Digital Curation Workforce: Advancing Specialized Data Curation." (#RE-85-18-0040-18) The grant allowed the DCN to further develop and expand the IASSIST-funded workshop to a larger, broader workshop model with the capacity to include more interested librarians and library staff from across the United States. Workshop attendance was not, and is not currently, limited to DCN member institution staff. From 2018 to 2022 this grant provided three in-person workshops to over 75 librarians across the country, resulted in the creation of nearly 30 data curation primers, and funded the development of online curation modules for broader library engagement and training. While the grant afforded the time and financial support to develop freely available educational

resources to support a growing and evolving workforce, the DCN leadership acknowledged the need for and planned a sustainable path forward that would enable it to continue to offer and develop its programs. In building out our sustainability framework, the DCN took into account the need for an annual workshop in our financial planning and provided a member tier to explicitly support additional workshops. These funds, as well as collaborations developed during and after the grant period, have ensured the educational efforts of the DCN are sustainable.

Another beneficial administrative structure that supports the DCN is a thorough and hands-on **curator onboarding process.** The onboarding workflow offers an entry point for new members and provides the information necessary to quickly and successfully engage with the DCN community. The curator onboarding process provides new curators with background knowledge of the DCN, outlines the shared curation practices and workflows, and ensures each member receives an overview of other engagement opportunities such as DCN special interest groups, resource development (e.g., primers, online modules), and collaborative research opportunities. More specifically, the curator onboarding process offers data curation practitioners a deep dive through the CURATE(D) training, in person or virtually. This training is then reinforced through hands-on experience curating a dataset and the opportunity to test out the shared data curation workflow through the workflow management system. During onboarding, new members are also asked to fill out the expertise assessment form which is then reviewed with DCN staff during a one-to-one meeting. The expertise assessment form offers new members the opportunity to highlight the strengths they bring while also providing a space to request skill development from others in the community. Finally, all new DCN members are introduced to the community via Slack, an instant messaging platform, and email, so existing community members can connect with new community members. Overall, the success of the onboarding process comes from the fact that it systematically welcomes new members to the community, provides hands-on curation experiences which boosts new member confidence in the process, and offers human interaction and feedback. This process provides all new members the same grounding in the network and helps to promote understanding and engagement with community norms.

Over the course of its development the DCN brought in several **consultation services** to provide an external perspective in addressing our internal challenges. External consultants provide expertise, facilitation, and ideas for taking action around a particular topic. Consultants can be short-term team members or more hands-off. For a cross-institutional collaboration, a consultant can help surface any unspoken expectations or hidden assumptions that, if unrealized, may slow or impede the collaboration. Additionally, consultants play a pivotal role in helping the team consider how they might approach developing administrative structures based on goals.

In 2016 the DCN hired an external facilitator for our kick-off meeting during the grant-funded planning phase. The facilitator was essential to help our new team of PIs from six institutions come to a shared understanding and articulation of our goals, the metrics for success in meeting those goals, the barriers, how we might overcome those barriers, and in the development of a strategic plan for how to start doing the work. In this case, the facilitator focused on discussion and consensus building to help launch the network.

In 2018 the DCN administrative team consulted with Lyrasis, a non-profit member organization that supports library, archive, and museum communities with a variety of services, to plan for the sustainability of the DCN. These consultants were tasked with providing different options for ensuring the organizational and fiscal success of the DCN after the end of the implementation grant. This resulted in the *Data Curation Network Sustainability Plan, Final Report* (Arp, Clareson, and Egan 2020), which provided the team a clearer understanding of the market and potential membership models for long-term sustainability. This work was instrumental for the team when building out the post-grant membership model over the following year in a way that would meet community needs and ensure the fiscal stability of the organization. Without the support of a consultant unifying the team members, developing these key administrative structures would have been significantly more challenging.

When collaborating with external consultants, the team found a formal request for proposal process useful, not only to receive written bids, thereby making the hiring process more transparent and ensuring the company or individual would fit into the project scope and budget, but also to allow the administrative team of the DCN to better articulate and define the role that a consultant might play. In

other words, the sheer process of writing a statement of work, including expectations and goals, was beneficial for aligning team members' understandings and objectives. Having a clear set of expected outcomes and meeting those expectations as a condition of the consultancy is key, otherwise the potential for the work to be open-ended, poorly scoped, or outside of the stated need is significant.

Last, the **DCN shared governance model** is a critical administrative structure for ensuring the daily operations and achieving the long-term goals of the network. With the transition away from grant funding toward a member-based model, DCN leadership designed, vetted, and implemented a model for collective governance (Data Curation Network 2021). This governance model is a living document that will evolve as the organization does, balancing providing enough supporting structures to meet the needs of the community while leaving enough flexibility for growth. There are several overlapping boards, committees, and groups that support one another. The main governing body of the DCN is the Governance Board, which is composed of one representative from each sustaining institution. This group votes annually on the budget, governance model revisions, and on accepting new community members based on recommendations from the DCN's Membership Committee. Recently, the DCN formally implemented an Executive Committee, which is a subset of the Governance Board composed of four DCN representatives and the DCN director. These two groups manage the day-to-day activities of the DCN and continuously articulate and refine long-term strategic goals. Supporting the Governance Board is an Advisory Board of 10 to 12 stakeholders at the Associate University Librarian level, a group that meets quarterly. Within this tiered-approach to governance, the DCN continues to promote collaboration through transparency. Our team uses fully open agendas, open leadership slack channels, broadly shared meeting notes and meeting summaries, and an open voting structure that promotes trust and accountability.

The DCN Governance Board is heavily supported by its committees and interest groups (see Figure 4). The Education Committee coordinates all workshops, primer creation, and training opportunities. The Membership Committee is tasked with reviewing and refining the membership model and ensuring community needs are met. The DCN has numerous interest groups, which are community-driven groups around a common topic, such as big data

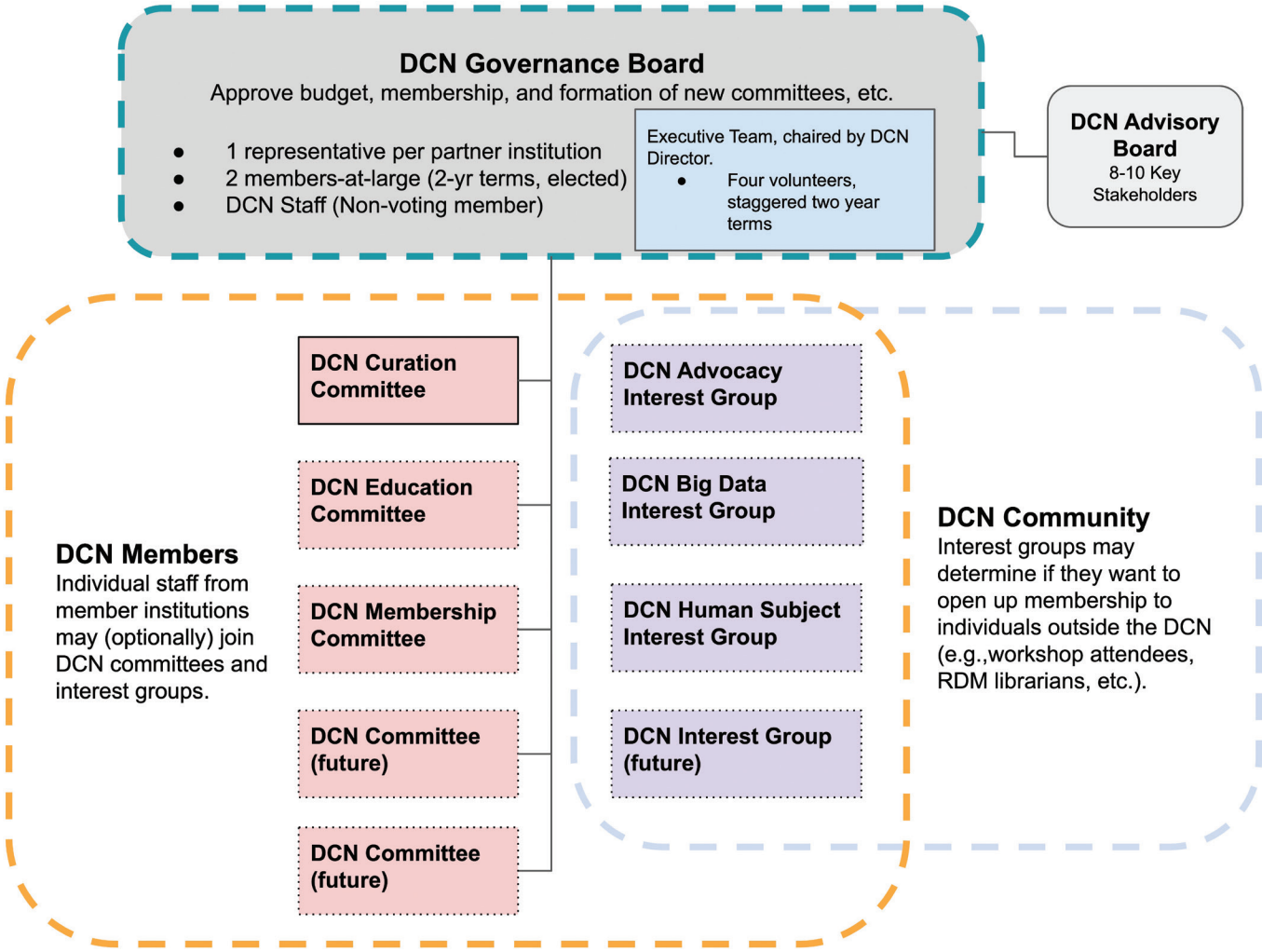

Figure 4: Relationship between DCN committees and groups through the shared governance model.

or institutional outreach. These groups have a significant degree of autonomy to decide the goals or desired outcomes of the group. Many of these groups have overlapping members, but they are all opt-in: we want each member of the community to be empowered to engage to the degree they wish depending on their interests and capacity.

This shared governance model was collaboratively developed with feedback from all members of the community. The governance structure was designed in large part to facilitate communication, keep members connected with each other and the wider DCN, and provide leadership with the space to make informed decisions and take action. The governance structure of the DCN is still new and in some ways a work in progress. Work continues

to ensure that the DCN governance structure best reflects and addresses the needs of its membership. In this governance model, the DCN prioritizes and celebrates a shared leadership approach to managing the network. This is essential for ensuring the long-term sustainability of the network. For example, after the departure of PI and DCN director, Lisa Johnston, in early 2022, the DCN adapted our governance model. The assistant director transitioned to the role of director, we phased out the assistant director role, and formalized an Executive Committee that meets weekly with the new director. These changes are a testament to the resilience of our organization and the shared leadership we embrace: even following the departure of key personnel, the network continues to serve one another and develop. Each of us have a vested interest in ensuring the success of the DCN—it is not just one person, it is all of us, because the organization belongs to all of us.

Trust-based Structures

Another key type of structure enabling the success of the DCN is trust-based structures, or those that support the network in establishing trust and building relationships between DCN members. The DCN shared staffing model and community of practice are supported by our trust-based structures as the operation of the DCN necessitates that both member organizations and individuals trust that each member will approach the work according to the values and norms established by the network. There are a number of communication, community building, and information sharing structures that help build trust within the network including the annual **AHM**, **special interest groups**, **peer-to-peer comparisons**, **collaboratively developed values**, our **code of conduct**, and **a shared leadership and ownership of the organization**.

The **AHM** is our flagship yearly event, held either in-person or virtually, that brings together member representatives and curators from each participating institution to share and reflect on the previous year's work, provide training opportunities, build community, celebrate the year's accomplishments, and plan for the upcoming year. The AHM is planned and developed by a volunteer committee

and has been an effective structure to form relationships, engage in deeper conversations, and brainstorm and develop new initiatives with a sense of shared ownership. While connecting with colleagues is different in a virtual environment, we have found that more members can attend the virtual AHM than can in person and that it increases equity among members. Moving forward, the DCN will balance remote and in-person community-building activities to ensure all who are interested in participating can do so to the degree to which they are interested and in a mode suitable to individual comfort levels.

The DCN has also used **special interest groups** as a mechanism for advancing shared goals through specific projects as well as flexible spaces for open sharing among peers, all of which rely on each group member collaborating and trusting one another. The process for developing special interest groups within the DCN has largely been organic and flexible with topics of interest arising at the AHM or through shared research goals of the network. Some special interest groups have produced discrete projects or deliverables and then disbanded, while others have been longer-standing with shifting goals and projects that are determined and scoped by group participants, and others still are open groups that prioritize discussion and learning over creating an output. Current special interest groups are detailed on the DCN website and updates are shared quarterly via our monthly newsletter.

Peer-to-peer comparisons is another structure grounded in trust that the DCN has used to share practices with the purpose of finding similarities and learning from one another. These often take the shape of semi-structured conversations around a single topic such as policies for curating data with human subjects, curation workflows, or metadata. Sharing our practices and being open with each other about how we "do curation" within our local contexts, helps to establish trust, realize efficiencies of scale, and avoids "reinventing the wheel" among members. These comparisons foster communication and entail a great deal of vulnerability—each member had to trust that the others would abide by Chatham House rules when sharing examples of failures or setbacks. These moments of candor and trust were critical for strengthening the sense of community.

The work of the DCN is built on **collaboratively developed values** that guide the work of the organization. The values of the network are:

- **Trusted:** We add value to data while maintaining data integrity and upholding the ethical responsibilities of data sharing.
- **Collaborative:** We build an innovative community of practice for data curation.
- **Open:** We share data curation techniques that are practical, transparent and available to all.
- **Inclusive:** We foster an accessible and welcoming environment to equitably support all participants.
- **Empowering:** We advocate for data professionals by providing a pipeline for training data curators, promoting data curation practices and enriching capacities for data curation writ large.

The mission, vision, and values of the DCN were initially developed during the planning phase of the DCN to help guide the work in the implementation phase. As the DCN neared the end of the implementation grant, DCN leadership designed and ran an activity at the fourth AHM in June 2021 to refine and rearticulate DCN values that would support the organization in the transition to a member-funded organization. In this virtual activity, members were invited to vote on the value statements that most resonated with them. Following this exercise, DCN members synthesized the results into the values listed previously. These values, as well as the mission and vision statements, were critical in developing a strategic framework to advance the organization (Data Curation Network 2022c).

The DCN's **code of conduct** is an essential structure for establishing and maintaining trust within a community. The code of conduct, based on similar codes from other community-driven organizations, lays out the behaviors that are expected to foster an inclusive and respectful community in all venues of collaboration (Data Curation Network n.d., "Code of Conduct"). The code addresses enforcement procedures in the event of violations. In addition to identifying the ideal behaviors to foster an inclusive and respectful community, the code of conduct also serves as a mechanism to reinforce the values of the DCN in practice. The process of drafting and adopting the code of conduct afforded the members of the DCN the opportunity to clearly identify goals for member interactions and articulate them internally

and externally, hold one another accountable, and ensure the community remains a welcoming space.

Finally, the trust within our community is based upon a sense of **shared leadership and ownership**. The DCN does not belong to any one of us: we all own it, we are all invested, we all want to see it succeed. This is due in part to our ethos of cooperation, but is also the result of our intentional design: in drafting our mission, vision, values, strategic framework, governance model, and other guiding documents, all members of the DCN were invited to actively participate and own the process. The founding members' experience and success working together laid a bedrock of trust and respect, which carried forward as the DCN grew. The founders were slow, thoughtful, and intentional in growing the network and in identifying new contributors to ensure the fabric of trust and respect expanded. This is a great example of what is meant by a healthy, functional community of practice, but this is an area of concern as our growth continues. We ask ourselves, what does it mean for our shared leadership and ownership as we expand? How will we include more institutions of varying size and focus and maintain our fabric of trust and respect? To date, we continue to grow slowly and intentionally so that we can thoughtfully address this issue of scale.

Tool-based Structures

Finally, the DCN developed a number of tool-based structures to enable data curators within the network to carry out their work. In addition to their utilitarian purpose, the tool-based structures also served as a means of enabling successful collaboration among geographically distant members as well as to establish or reinforce our administrative and trust-based structures. These tools included those that were collaboratively developed, such as the **DCN website** and **data curation primers,** and those that enable collaboration, including **collaborative office tools** and a **shared curation workflow**.

The **DCN website** enables the network to maintain a public presence. The website provides information about the organization including the mission, vision, values and structure of the network, as well as how to become a sustaining member of the DCN. The DCN values transparency around our efforts, and this includes frequent

sharing of education opportunities, research conducted, and other outputs created by members of the network via newsletters and blog posts. The website provides a way to share out the community aspects of the DCN, including the people involved, the activities of the special interest groups, and how individuals can participate in the DCN. The DCN website also provides a space to demonstrate the datasets that have been curated through the network as well as the work of DCN curators behind the curation through curator profiles. Curator profiles provide an image of the curator, a brief biography, details each member's curation expertise broken down by subject, discipline, and data type, and list all of the datasets they have curated through the DCN. It is not yet common practice for curators to receive recognition and credit for the labor they invest in preparing a data set to be shared, understood and used by others. The DCN curator profiles seek to make what is often invisible labor more visible. This is also indicative of the DCN's ability to recenter the human in curation: the datasets have not been curated by an unknown entity, but by an individual bringing with them their own perspectives, interpretations, skills, and expertise.

Data curation primers, as described previously, are brief reference documents that provide basic background information about specific data types or formats as well as practical approaches for curating the data type or format. The data curation primers are an exemplar of the DCN's "by the community, for the community" approach. The primers address common needs of data curators in that they provide guidance and structure for curators to use when curating unfamiliar content, both within and outside of the DCN, helping the entire research data management community upskill itself. Additionally, the primers serve as an extended learning opportunity for attendees of the DCN workshops to expand on what they learned through the creation of a new resource with the support of other attendees and the guidance of a peer mentor and peer reviewers. This helps to reinforce the knowledge gained at the workshops, foster community, produce a tangible result that benefits the wider community and provide professional recognition for information professionals who author a primer. Moreover, because the primers addressed a need in the curation community for reference guides describing actionable steps that one could take in curating data, they strengthened the reputation of the DCN in the larger data community.

Collaborative office tools, which allow DCN members to write in the same cloud-based documents, work on presentations and publications asynchronously, and share files with one another, are also key in supporting the work of the DCN. Having a single, centralized location that all members can access is critical for not only working together, but also for promoting and ensuring transparency. Documentation in the DCN, with the exception of HR information, is open to all. Additionally, instant messaging is essential for a network distributed across the United States in different time zones. While these tools are not replacements for community building, these structures help foster collaboration even while members are geographically distant.

The DCN also developed a **shared curation workflow** through **Jira**, a workflow management system, which allows DCN staff and members to track each dataset submitted for shared curation. In this system, members provide basic information about the dataset, namely format and discipline, and DCN staff connect the dataset with an appropriate curator. Leveraging a workflow management system also provides a dashboard for easy data visualization and analysis, such as which institutions are submitting or curating data, datasets by subject area, and datasets by format. This system helps ensure the DCN has a consistent way to communicate about datasets in the shared curation workflow. Moreover, since all of our curation work happens in this same space, we have a significant amount of data about the shared curation model that we can use to refine and improve our processes.

Collaborative Structures Conclusion

Discussions during the retrospective meeting provided insight into the importance of our structures as a core element of what makes the DCN successful. Each of these structures served a specific function in the launching and sustaining of the DCN. While administration and tool-based structures may seem like obvious needs in supporting the work of an organization, the critical work of creating trust-based structures should not be ignored. The DCN works because we built all of our structures together, for our unique community's needs and purposes, with an emphasis on fostering respect, mutual-benefit, and empowerment. These structures, and the resulting trust among members, were, and remain, critical for the success of our collaborative effort.

Challenges and Future Directions

While the DCN has achieved a great deal over the past six years, the process of building and sustaining the network came with a series of challenges. In particular, alongside a global pandemic, the community had to contend with **balancing local and shared curation**, a relatively **homogeneous set of member institutions**, **addressing lack of researcher response** to curator recommendations, **managing community capacity**, and **ensuring the sustainability of the network**. Below, we describe these challenges in more detail and discuss potential reasons for and solutions to the issues. We conclude by reflecting on how we might change our approach if we were to rebuild the DCN again today.

Balancing Local and Shared Curation

The shared workflow and data curation exchange services have been the core to the foundation of the DCN. Over time, however, we have found that they have not been as widely used by data curators as we had expected initially. Each institution must balance local curation work with leveraging the DCN for its shared curation services in a way which most appropriately meets their needs. Member curators often make use of the DCN's shared curation services during times of limited capacity or when the dataset falls outside of their curatorial staff expertise. However, there are multiple reasons why a curator might decide to curate a data set locally rather than use the DCN, even if they may not have expertise in working with the type, format or subject matter of the data set. First, researchers may request expedited curation turnarounds to meet particular deadlines they may have. In these situations, bringing in another curator from an outside institution to review the data set

through the DCN may make it more difficult to complete the curation review in a timely manner. Second, although the DCN has developed a well-supported set of workflows and structures to enable collaborative data curation, making use of the DCN still requires curators to break out of their local workflows and structures to use DCN services. The time and labor involved in submitting a data set to the DCN and then reincorporating it into local processes can dissuade curators from making use of the DCN's curation services. Third, DCN member institutions are also working to increase the expertise of its local workforce through providing opportunities for the library's subject specialists, and others in the library, to learn how to work with research data. Creating situations in which library staff can gain hands-on experience with data curation practices is an effective means of developing local talent. The imperative to develop local capacity to work directly with research data is likely to increase as the new NIH data sharing mandates come into effect in 2023 and researchers seek more assistance from their librarians in responding. Research data sharing is likely to continue to become the norm in most disciplines, necessitating more staff development and investment in local institutional infrastructures. The DCN serves as an effective supplement for local investments, but it cannot replace them. While the shared curation model is used less regularly than we initially anticipated (see Johnston and Narlock 2021), the quantitative metrics are not the complete story. While the network receives fewer datasets than anticipated, we recognize that the benefits of the DCN are more than its data curation services, as each institution leverages the DCN in ways that best suit their needs. Even as we continue to monitor dataset submissions and increase in the number of curators engaged in the community, we are looking beyond quantitative measurements and gauging if we are meeting member needs in other ways through curator engagement and discussion.

Homogeneity of DCN Member Institutions

Currently, the DCN is primarily composed of institutions with well-resourced libraries with institutional repositories and curation programs, with the notable exceptions of Dryad and the Michael J. Fox Foundation which are not library based. Our planning grant was centered on six R1 institutions, defined as academic institutions with very

high degree research activity by the Carnegie classification system. These six institutions were well-equipped with curatorial staff with a broad range of disciplinary expertise, and were therefore well-positioned to pilot an experimental program. This early work has resulted in a network of peers that have similar challenges and opportunities. However, we recognize that this reinforces, and in some ways exacerbates, structural inequities inherent in systems of higher education. Now that the DCN has demonstrated that a member-driven cross-institutional model for curating data is possible, we need to dedicate more attention toward understanding the data curation needs of non-R1 institutions and in making DCN membership feasible and attractive to a more diverse array of institutions.

As the DCN has transitioned from a grant-funded pilot project to a membership organization, the community has made a concerted effort to grow its membership in a sustainable fashion and encourage participation at different levels. We have intentionally grown the network by a limited number of institutions per year, to ensure we can support the increased demand on resources. While this deliberate growth has been beneficial in affording the DCN's financial stability, we are also aware that our current membership model is not accessible for many institutions. For example, currently membership in the DCN requires that an institution contributes a percentage of a data curator's time to curate data sets submitted by other member institutions. The requirement prevents institutions who do not have the capacity to dedicate curator time to the DCN, or who do not have a data curator on staff. Additionally, while we are eager to include other institution types, we want to be sure that membership in the DCN meets the needs of and is of value to non-R1s, Minority Serving Institutions (MSIs), and other types of research institutions not currently represented in the membership of the DCN, such as disciplinary repositories or government agencies. Instead of developing a membership option that we think might meet the needs of other institutions, the Membership Committee has been tasked with gathering information over the next year to understand what data curation support—either in services or education—non-R1, MSIs, and other curation repositories might need, and, equally important, whether it is in scope for the DCN to try and address these needs. Is the lack of institutional diversity due to the DCN not communicating the value of curation outside of R1 institutions, or that our current model fails to serve the needs of non-R1 institutions? If it is the

latter, it will be important to determine where and to what extent the DCN can meet those needs.

Addressing Lack of Researcher Response

Curator feedback with requests for additional information on submitted datasets do not always result in a response from data depositors, challenging curators' ability to implement changes to the dataset which would result in the increased findability and usability of the research output. Many DCN member institutions follow the self-submission model with auto-acceptance, which means researchers are allowed to submit datasets to their institutional repositories at their convenience without waiting for approval from the repository staff. These policies were made to, typically, accommodate the tight deadlines that researchers normally face, thereby creating flexibility in data publication timelines and easing the burden of sharing data. Curators send their feedback for dataset improvement to the data depositors, but on occasion do not receive a response. Without the data creator's permission, only very limited changes can be made to improve the data.

Additionally, when datasets are submitted to the shared curation workflow of the DCN, upon curation resolution, DCN staff add the dataset to our website, noting which curators helped improve the dataset, which actions were taken, and the level of change required (none, minimal, major, or essential). When researchers do not respond to curation requests, the DCN is unable to formally resolve the task and share it on a curator's profile, leaving these datasets listed as "incomplete," in our internal system when in reality, curation has been completed on the DCN side. Conveying this, and appropriately recognizing the work of our curators even when a researcher does not respond, has been a challenge.

Despite this frustration, it is clear that curation support is highly valued by the researchers we have worked with. In 2021, representatives from six institutions in the DCN collaborated on a study which surveyed researchers who deposited their data between January 2020 and May 2021 on their satisfaction with data curation services (Data Curation Network n.d., "Value of Curation"). The survey was sent out to researchers who had deposited content with

the participating institutions' data repositories between January 2020 and May 2021. The results of our researcher survey from the six participating DCN institutions were overwhelmingly positive with an overall response rate of 42% (n=239/568) and consistently laudatory feedback, including many free-text responses testifying to the value of curation services. Researchers in the survey valued and appreciated the curatorial service that was provided by their local institution staff. Almost 50% of the survey participants did not expect that their data would be curated by the repository staff but were enthusiastic about the service and the improved quality of their dataset (Wright et al. 2022).

Managing Community Capacity

The DCN is recognized by our peers as an innovative organization advancing the capability of information professionals to apply their knowledge and expertise in curating research data through our advocacy, education, and research. This is evidenced by frequent requests for collaboration with national organizations (e.g., the NIH) and disciplinary societies (e.g., the Federation of American Societies for Experimental Biology), especially around educational efforts. Our success is rooted in our members, and stems directly from the willingness of individuals in the DCN to devote their time and energy to DCN initiatives and programs. However, each member also has responsibilities to their home institution: member engagement in the DCN comes on top of their daily work. Additionally, many librarians cannot focus on data curation full time, due to the niche skills and disciplinary expertise required to curate data, the fact that salaries are frequently not commensurate with the knowledge and skills needed to do this kind of work, and that job descriptions may or may not explicitly include curation, especially for those serving as liaison and data librarians.

Balancing our community interests with available resources and time is increasingly difficult. As the DCN keeps growing there is a risk that the number of initiatives that it develops and supports will outweigh the amount of time and resources available. Even staying informed about the work of the DCN becomes more challenging as it continues to expand and take on new work. Managing the capacity

of community members to avoid burnout while remaining actively involved in the wider research data community has been, and will continue to remain, a delicate balance.

The DCN has been working to address this very challenge. As membership has grown, the need for structure and guidance in operating the DCN has become apparent. As a complement to the shared governance model designed by DCN leadership, the DCN Governance Board, in conversation with all DCN community members, developed a strategic framework for the years 2022 to 2025 (Data Curation Network 2022c). While considered a "living document," meaning it is open to feedback from members and subject to change, the framework provides guidelines for the Governance Board to determine which opportunities fit well within the network's mission, vision, and values, and which opportunities might not. This allows the entire community to be conscious of our capacity and empowers us to say "No" to opportunities that are not a good fit.

Ensuring Sustainability of the Network

The DCN was started as an experiment between six librarians to develop a model for sharing curation expertise across institutions. We had always intended for the DCN to be an organization that would grow and thrive beyond our grant funding, which included developing a governance model and financial path to sustainability. Actually developing these paths and models to ensure the sustainability and long-term success of the community has proven to be challenging for several reasons. First, there are few models to follow in developing a cross-institutional network built on members contributing a portion of their working time like the DCN, and even fewer that have been sustainable after grant funding has ended. Second, the value proposition of the DCN continues to evolve over time and with growth in its membership. The DCN began as a means to support the work of curating data sets, but it has become a strong community engaged in advocacy, research, and education, in addition to the distributed data curation work. This evolution is a testament to the DCN's flexibility in responding to the needs of its members, however its expanding list of initiatives and activities may increase the efforts needed to ensure its sustainability. Finally, the DCN is dependent upon the

administration of its members to recognize the value of the DCN and to support it through membership fees. Institutional buy-in to the DCN overall has been positive, but not uniformly so. Some members are faced with the challenge of having to articulate the value of the DCN to skeptical audiences in an era where money is tight for many institutions. As the DCN continues to mature, explore membership models, and articulate a value statement, ensuring the sustainability of the organization will be a community-driven effort.

In Hindsight. . .

The DCN has largely been successful in achieving much of what we set out to do. The network has evolved into a valued community of practice that many of its members rely on to help address our data curation needs, whether it is to discuss a specific question or tool with another curator, to meet with a special interest group, or to submit a dataset for curation by a network curator—or even just to connect to someone to commiserate!

However, we know that we have made several missteps or fallen short in some of our goals as an organization. We are a group that is constantly looking for ways to improve, to learn from our mistakes, and to try doing things differently. This includes:

- **Making space for and building partnerships** (with publishers, professional organizations such as ITHAKA and OCLC, and researchers): We are open to collaborations with groups, but have only recently made that clear on our website, with information about how other organizations can collaborate with the DCN. Additionally, as more collaborative opportunities emerge, our leadership and community are ensuring we understand more about what partnerships can—and cannot—deliver.
- **Developing paid fellowships and other workforce development programs**: Learning how to be a successful curator requires hands-on experience working with datasets and researchers. This invaluable experience would significantly benefit library science students while providing additional curatorial support to DCN members. The DCN recently collaborated with the NCDS, based at the Network of the National Library of

Medicine, to host three interns over a ten-week period during Summer 2022. While this partnership provided a solid foundation for the DCN community to better understand the needs of students, we wish we had supported students earlier in the DCN through a fellowship or residency, which would have provided more opportunities to mentor students of different disciplines considering a career as a data curator.

- **Planning for long-term success:** Although we have considered issues around sustainability since the beginning of the DCN, much of our early efforts were focused on forming and defining the DCN in the short term. Questions about how to sustain the DCN over the long term, including whether the DCN should become a non-profit organization, remain. It is unclear to what extent the DCN can scale for true long-term success if that success depends on continuing to increase the number of institutions, curators, and so on. Additionally, as a community that is rooted in trust, transparency, mutual respect, and radical interdependence, any future growth will need to prioritize and center these values. This may become more difficult to achieve as our membership increases.
- **Incorporating more peer learning opportunities**: At the first AHM in 2018, one of our favorite outcomes was the positive impact of getting to see and hear each other's descriptions and drawings of our respective data curation workflows. We learned much from the information sharing sessions focused on the details and specifications of curatorial actions taken and we continue to learn more from tours of each other's repositories, discussions of practices and procedures, and other more informal information exchange opportunities that happen at AHMs. The challenge is enabling these meaningful conversations virtually. Understanding how we can best facilitate meaningful interaction between members when we are challenged by distance will continue to be a hurdle as we balance providing in-person and virtual opportunities to accommodate all members of our community.
- **More mindfully planning for diversity, equity, inclusion, and accessibility from the outset of a project:** In 2020, we formed a Racial Justice Working Group, which aims to help end racial injustice and create a more diverse, equitable, accessible, and

inclusive environment, both through our work as data curators and as individuals in academia. In 2020 the DCN as a collective also refined our values to be: trusted, collaborative, open, inclusive, and empowering. This work led to collaborating with Dr. Fay Cobb Payton as a consultant, who ran a workshop for DCN members to learn and engage around racial justice, intersectionality, and data stewardship as part of creating our organizational roadmap. We are continuing to apply this work, and have just published a revised version of the CURATE(D) steps that includes key ethical considerations at each step (Data Curation Network 2022a). However, there is much yet to be done, and planning for this critical ethical work earlier in the DCN would have helped the community center diversity, equity, and accessibility in all education, training, and curation materials from the outset.

Potential Collaborative Endeavors

Through the work of the DCN, the team has identified some additional shared challenges and are considering how we might try and address them collaboratively. As the need to manage, share and preserve data become increasingly important as a normative part of conducting research, collaborative efforts to address shared problems will also be of increasing importance. To that end, we present the following list of potential data sharing challenges that might be addressed via cross-institutional collaboration as representative, not exhaustive, of potential future directions that the DCN, or other interested communities, could follow. Any activities performed in these areas should acknowledge the work of experts in these fields. Additionally, any collaborative effort must ensure that leaders are building with the broader perspectives of communities in mind, not just for themselves. This work also requires institutional members of any collaborative effort to be empowered to make decisions and implement local changes.

Supporting Data Communities Through Data Management

Data communities are "formal or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries" (Cooper and Springer 2019). In addition to spanning disciplines, data communities span institutions and geographies when collaboratively sharing data and information. While institutions can and do provide support for *individuals* involved in the data community, a collaborative effort to provide consultation and data management support that extends the cross-institutional work of the

DCN would help ensure consistency in data sharing practices. This sort of radical collaboration, providing support to data communities regardless of disciplinary or institutional affiliation, could improve the quality of data to be shared and incentivize researchers to partner with data stewards in providing long-term access to and preservation of their scholarly outputs.

Research Data Management Plans

Research data management plans (DMPs) can be invaluable tools for guiding the creation and sharing of research outputs. However, in our experience, researchers tend to see creating a DMP as more of a burden than as a useful document to guide their work in managing and sharing their data and other scholarly outputs. Most funders now require DMPs, so supporting researchers in this effort is critical for institutions, labs, and departments that rely on funding to conduct their research. While there has been a significant amount of work done to support and assess DMPs, more work in understanding the potential utility of a DMP from the researcher perspective is needed. Efforts to make DMPs digital and machine actionable (maDMPs) are particularly promising as they could be embedded in the institutional systems and infrastructures used to manage grants, making the information they contain more accessible (Sims and Jones 2017). The assessment rubric developed by the Data Management Plans as a Research Tool (DART) Project might also be of interest (Whitmire et al. 2017). This multi-institutional effort to evaluate DMPs and design an analytical rubric to assess and improve DMPs provides a potential foundation for future work in this area.

Dealing with Restricted Data

Restricted datasets require more attention to work with properly. Researchers working with personally identifiable information, protected health information, or commercial data often need extensive support to curate, preserve, and provide appropriate access to their content. This challenge will continue to grow with increased

funder requirements for data management and sharing plans, and the increased attention to data during public health crises (e.g., Nelson et al. 2022). Many researchers and institutions that have not been trained to evaluate data, de-identify data, or share appropriately may assume that all data must be either entirely open access or entirely closed. However, this all or nothing approach leaves valuable data inaccessible. Collectively creating and implementing a unified approach to teaching and consulting on sensitive data would enable institutions to collectively and consistently support researchers working with sensitive data.

Data Discovery

As data deposits to repositories continue to grow, discovering datasets for reuse becomes increasingly important and challenging: the current repository landscape is fractured and siloed, meaning researchers are not able to search across the thousands of data records from multiple repositories to find what they are looking for. Instead, they have to know the repository landscape well enough to identify which systems might have the relevant datasets that would address their information needs. In order to facilitate data discovery, and therefore enable data access and reuse, institutions could work more collaboratively in support of data discovery. As repositories continue to mature—especially disciplinary repositories that provide domain-specific support to researchers—and integrate with linked identifiers such as open researcher and contributor identifiers (ORCiDS) and research organization registry identifiers (ROR IDs), harvesting dataset records and providing robust searching and filtering capabilities will increase the findability and reusability of datasets. Efforts in this area are well underway, as demonstrated by the New York University Data Catalog and the Dataset Search application (Sheridan et al. 2021). Libraries can and should continue to address this problem collaboratively. By adopting a collaborative approach, each institution will not need to develop their own tool or metadata crosswalk, but instead can build on the work of others to improve their systems.

Applying Metadata Standards

Similarly, each institution has nuances and variations in how metadata standards are applied. While standards might provide some guidance on how metadata fields should be used, often each institution adopts their own approach to how the metadata fields are utilized in their institutional repositories. For example, fields like Date, Keyword or Subject, or even Creator can be used differently depending on how a repository defines and applies them. "Date" might refer to when an item was created, when it was curated, or when the data was collected. "Keywords" might be a structured or unstructured field. "Creator" might be a free-text field or used to link an item to an author's ORCiD. Moreover, institutions often have different approaches in connecting their metadata to the global infrastructure (Taylor et al. 2022). These incongruities make it much more difficult to materialize the vision of FAIR data and connected research infrastructure. By addressing these issues through a collaborative effort and aligning our metadata schemas, academic institutions can leverage the expertise of information professionals to not only uncover the challenges of creating a more unified approach to applying metadata standards, but also work toward solving the problem through a connected and interdependent community of practice.

Conclusion

The experience of launching the DCN allowed us to collectively explore the specific challenge of curating research data through a shared workflow that relies on radical interdependence. When we were conceptualizing the DCN, there were few models for us to consider how we might build a network that would not only enable us to harness our collective expertise in addressing our common challenges, but to knit our work so tightly together that we depend on each other to get our work done. The successes of the DCN are rooted in this radical interdependence between members and in the values of our community. From developing shared tools and resources, creating and offering educational opportunities, and developing a community based on mutual respect, trust, and a willingness to collaborate, each and every success of the DCN can be attributed to the individual participants. Early efforts by DCN leadership to establish key structures of collaboration—namely administrative, tool-based, and trust-based structures—facilitated the launching and sustaining of the network. While the community has challenges to contend with, there are numerous future directions, both for the DCN and other interested communities, to continue growing and supporting open access to research data.

The work of bringing together a community, launching a collaborative effort, and finding appropriate capacity levels to avoid burnout is challenging labor, which will need to be carefully balanced as the DCN grows. Community development, especially for an organization grounded in radical interdependence, is not particularly speedy—in fact, this work is often slow, thoughtful, and deliberate, as it requires careful planning to ensure the fabric of a community grows with new members. However, this work is critically important to the mission of the DCN and it will remain our primary focus in continuing to

develop our organization. The degree to which the DCN is successful remains our ability to live up to our model of each member sharing ownership of the community.

REFERENCES

Arp, Laurie, Tom Clareson, and Carissa Egan. 2020. "Data Curation Network Sustainability Plan". Final Report by Lyrasis. Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/211865.

Association of Research Libraries. "Realities of Academic Data Sharing (RADS) Initiative." Accessed August 9, 2022. https://www.arl.org/realities-of-academic-data-sharing-rads-initiative/.

Carlson, Jake, Jen Darragh, Dave Fearon, Mikala R. Narlock. 2022. *Towards Authenticity: Critical Appraisal of Data Management Plans.* Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/229000.

Carroll, S. R., Ibrahim Garba, Oscar L. Figueroa-Rodríguez, Jarita Holbrook, Raymond Lovett, Simeon Materechera, et al. 2020. "The CARE Principles for Indigenous Data Governance." *Data Science Journal* 19 (1): 43. http://doi.org/10.5334/dsj-2020-043

Christen, Kimberly and Jane Anderson. 2019. "Toward Slow Archives." Archival Science 19 (2): 87–116. https://doi.org/10.1007/s10502-019-09307-x.

Cooper, Danielle and Rebecca Springer. 2019. "Data Communities: A New Model for Supporting STEM Data Sharing." *Ithaka S+R* (blog). https://doi.org/10.18665/sr.311396. Published May 13, 2019.

Data Curation Network. n.d. "Code of Conduct." https://datacurationnetwork.org/about/code-of-conduct/.

Data Curation Network. n.d. "Our Mission." https://datacurationnetwork.org/about/our-mission/.

Data Curation Network. n.d. "Value of Curation." https://datacurationnetwork.org/value-of-curation/.

Data Curation Network. (2018). "Checklist of CURATED Steps Performed by the Data Curation Network, v.1." Data Curation Network. Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/226891.

Data Curation Network. (2021). "Data Curation Network Governance Model." Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/225742.

Data Curation Network (2022a). "CURATE(D) Steps and Checklist for Data Curation, version 2." https://hdl.handle.net/11299/241454.

Data Curation Network. (2022b). "Peer Compare Preservation Practices Benchmark Report." Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/227211.

Data Curation Network. (2022c). "Data Curation Network Strategic Framework 2022–2025 v1.0." Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/226182.

Educopia Institute. n.d. "Research." https://educopia.org/research/.

Farkas, Meredith. 2021. "Slow life, slow librarianship." *Information wants to be free* (blog). August 8, 2017. https://meredith.wolfwater.com/wordpress/2021/09/08/slow-life-slow-librarianship/.

Glassman, Julia. 2017. "The innovation fetish and slow librarianship: What librarians can learn from the Juicero." *In the Library with the Lead Pipe*. http://www.inthelibrarywiththeleadpipe.org/2017/the-innovation-fetish-and-slow-librarianship-what-librarians-can-learn-from-the-juicero/

Hudson Vitale, Cynthia, Hannah Hadley, Jennifer Moore, Lisa Johnston, Wendy Kozlowski, Jake Carlson, Mara Blake, and Joel Herndon. 2020. "Extending the Research Data Toolkit: Data Curation Primers." *International Journal of Digital Curation* 15 (1): 1–14. http://www.ijdc.net/article/view/713.

Hudson Vitale, Cynthia, Jake Carlson, Hannah Hadley, and Lisa Johnston. 2021 "Introduction to the Special JeSLIB Issue on Data Curation in Practice", *Journal of eScience Librarianship* 10(3): 13. doi: https://doi.org/10.7191/jeslib.2021.1222

Johnston, Lisa R. 2018 "Barriers to Collaboration: Lessons Learned from the Data Curation Network." Research Library Issues, no. 296: 37–43. https://doi.org/10.29242/rli.296.5.

Johnston, Lisa R. 2020. "How a network of data curators can unlock the tremendous reuse value of research data." *OCLC Next* (blog). https://blog.oclc.org/next/data-curators-network/

Johnston, L. R., J. R. Carlson, P. Hswe, C. Hudson Vitale, H. Imker, W. Kozlowski, R. K. Olendorf, C. Stewart. 2017. "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services." *Journal of eScience Librarianship* 6 (1): e1102. https://doi.org/10.7191/jeslib.2017.1102.

Johnston, L. R., J. Carlson, C. Hudson Vitale, H. Imker, W. Kozlowski, R. Olendorf, and C. Stewart. 2018a. "How Important is Data Curation? Gaps and Opportunities for Academic Libraries." *Journal of Librarianship and Scholarly Communication* 6 (1): e2198. https://doi.org/10.7710/2162-3309.2198.

Johnston, Lisa R., Jake Carlson, Cynthia Hudson Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart et al. 2018b. "Data curation network: A cross-institutional staffing model for curating research data." https://doi.org/10.2218/ijdc.v13i1.616.

Johnston, Lisa R., Mara Blake, Jake Carlson, Joel Herndon, Cynthia Hudson Vitale, Heidi Imker, Greg Janée, Wendy Kozlowski, Jennifer Moore, Katie Wissel, et al. 2020. "Data Curation Network Sustainability and Transition Plan." Retrieved from the University of Minnesota Digital Conservancy. https://hdl.handle.net/11299/225741.

Johnston, Lisa R., Renata Curty, Sophia Lafferty-Hess; Hannah Hadley, Jonathan Petters, Hoa Luong, Susan Braxton, Jake Carlson, and Wendy Kozlowski. 2021. "Value of Curation Survey, January 2021." Retrieved from the Data Repository for the University of Minnesota, https://doi.org/10.13020/04ee-q089.

Johnston, Lisa R and Mikala Narlock. 2021. "Annual Report 2021: End of Year Highlights of the Data Curation Network." Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/225819.

Nelson, Alondra, Carrie Wolinetz, Andrew Hebbeler, Amanda Corcos. 2022. "A Call for Public Access to Monkeypox-related Research and Data." The White House Office of Science and Technology Policy. Accessed November 20, 2022. https://www.whitehouse.gov/ostp/news-updates/2022/08/04/a-call-for-public-access-to-monkeypox-related-research-and-data/.

NIH. 2022. "(New) Data Curation Network – Event Series (ODSS, NLM)." Accessed August 7, 2022. https://datascience.nih.gov/data-curation-network-event-series.

Petrini, Carlo. 2003. *Slow food: The case for taste*. Columbia University Press.

Research Data Management Service Group. n.d. "Guide to writing 'readme' style metadata." Accessed August 11, 2022. https://data.research.cornell.edu/content/readme.

Ruediger, Dylan and Danielle Cooper. 2021. "Leveraging Data Communities to Advance Open Science." *Ithaka S+R*. (blog). Accessed August 2, 2022. https://sr.ithaka.org/blog/leveraging-data-communities-to-advance-open-science/

Sciolla, Sam and Susan Borda. 2018. "Data Curation Format Profile: netCDF." Deep Blue Documents. University of Michigan. https://hdl.handle.net/2027.42/145724.

Sheridan, Helenmary, Anthony J. Dellureficio, Melissa A. Ratajeski, Sara Mannheimer, and Terrie R. Wheeler. 2021. "Data Curation through Catalogs: A Repository-Independent Model for Data Discovery." *Journal of eScience Librarianship* 10 (3): 4. https://doi.org/10.7191/jeslib.2021.1203.

Sims, S. R., and S. Jones. 2017. "Next-Generation Data Management Plans: Global, Machine-Actionable, FAIR." *International Journal of Digital Curation* 12 (1): 36-45. https://DOI:10.2218/ijdc.v12i1.513.

Taylor, Shawna, Sarah Wright, Mikala R. Narlock, Ted Habermann. 2022. "Think Globally, Act Locally: The Importance of Elevating Data Repository Metadata to the Global Infrastructure." Boulder, CO: Open Repositories. Retrieved from the University of Minnesota Digital Conservancy, https://hdl.handle.net/11299/228001.

Whitmire, Amanda L., Jacob Carlson, Brian Westra, Patricia Hswe, and Susan W. Parham. 2017. "The DART Project: Using Data Management Plans as a Research Tool." Retrieved from OSF. DOI: 10.17605/OSF.IO/KH2Y6.

Wilkinson, Mark, Michel Dumontier, IJsbrand Aalbersberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship." 2016. *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wright, Sarah, Wanda Marsolek, Hoa Luong, Susan Braxton, Sophia Lafferty-Hess, Jake Carlson. 2022. "Researcher Approved: A Multi-institutional Survey of Depositors to Six Academic Data Repositories." Denver, CO: Open Repositories. https://hdl.handle.net/11299/227932.

APPENDIX A: DATA CURATION NETWORK MEMBERS AND INSTITUTIONS

Since its inception in 2016, the DCN has been supported by many individuals and institutions. Below, we list the past and present participants in the network, up to date as of November 2022.

INDIVIDUAL CONTRIBUTORS

Adi Ranganath
Alexis Logsdon
Alicia Hofelich Mohr
Andrea Ogier
Andrew Battista
Andrew Johnson
Ashley Hetrick
Briana Ezray Wham
Charles Watkinson
Chen Chiu
Chris Prom
Cindy Xuying Xin
Claire Stewart
Clay Ford
Cynthia Hudson Vitale
Dave Fearon
Debra Fagan
Dorris Scott
Elizabeth Hull
Elizabeth Lorang
Erica Johns
Erich Purpur
Erin Clary
Greg Janée
Hannah Hadley
Henrik Spoon
Hoa Luong
Jacob Goldstein-Greenwood
Jake Carlson
Janice Jaguszewski
Jen Darragh
Jen Huck
Jennie Burroughs
Jennifer Moore
Jennifer Gibson
Jenny Coffman
Jess Herzog
Joanna Thielen
Jody Gray
Joel Herndon
John Russell
Jon Petters
Jon Wheeler
Jordan Wrigley
Josh Gottesman
Katie Barrick
Katie Wissel
Leslie Delserone
Leslie Kirsch
Lisa Johnson
Liza Coburn
Mara Blake

Marley Kalt
Matt Chandler
Melinda Kernik
Moira Downey
Neggin Keshavarzian
Nick Wolf
Padma Carstens
Rachel Woodbrook
Reina Chano Murray
Renata Curty
Rich Yaxley
Ricky Patterson
Rob Olendorf
Sandi Caldrone
Sarah Wright
Sayeed Choudhury
Scout Calvert
Seth Erickson
Shaan Hamilton
Shadae Gatlin
Shanda Hunt
Shane Coleman
Shawna Taylor
Sherry Lake
Simeon Warner
Sophia Lafferty-Hess
Susan Borda
Susan Mae Braxton
Talya Cooper
Tim McGeary
Tracy Teal
Trisha Cruse
Vicky Rampin
Wanda Marsolek
Wendy Kozlowski
Wind Cowles
Yasir Karim

MEMBER INSTITUTIONS

Cornell University Libraries
Dryad Data
Duke University
Johns Hopkins University
Michael J. Fox Foundation
New York University
Pennsylvania State University
Princeton University
University of California – Santa Barbara
University of Colorado – Boulder
University of Illinois at Urbana Champaign
University of Nebraska
University of New Mexico
University of Michigan
University of Minnesota
University of Virginia
Virginia Tech University
Washington University in St. Louis

APPENDIX B: RETROSPECTIVE AGENDA

Project Retrospective Meeting Agenda
March 9–11, 2022

Location: Hotel Madera 1310 New Hampshire Avenue, N.W., Washington, DC 20036

Objective: Convene a meeting for project personnel of the Sloan funded "Launching the Data Curation Network" project to reflect, assess, and close-out the grant.

Attendees: 10–20 individuals with the shared experiences of working on the project from its early days. Since this is a hands-on working meeting, everyone attending will be asked to actively participate. There will be other opportunities for others to weigh in once we have a draft.

GOALS

1. Analyze our outputs of the DCN, with a focus on the work of the planning and implementation grants. What has worked well? Where were our biggest challenges/barriers?
2. Draft a report that presents lessons learned from our experiences to an external audience

PRE-READINGS

- Guide to Accelerating Public Access to Data Sharing (https://www.aplu.org/wp-content/uploads/guide-to-accelerate-access-to-public-data.pdf)
- Future Themes and Forecasts for Research Libraries and Emerging Technologies (https://doi.org/10.29242/report.emergingtech2020.forecasts)
- Emerging Technologies for Research and Learning: Interviews with Experts (https://doi.org/10.29242/report.emerging-tech2020.interviews)

- Mapping the Current Landscape of Research Library Engagement with Emerging Technologies in Research and Learning (https://doi.org/10.29242/report.emergingtech2020.landscape)

MEETING AGENDA

Wednesday March 9, 2022	
12:00–12:30	Lunch - Catered in room (sandwiches/salad//beverages)
12:30–1:15	Welcome Session (30 m) - Introductions - Goals for the meeting - Process/ground rules for this meeting - Code of conduct - Audiences for this report
1:15–2:30	Breakout #1 Taking the 10,000-foot view - What are academic researcher needs in the data space? - Who is actively trying to find solutions for these challenges?
2:05–2:20	Break
2:20–4:30	Breakout #2 DCN Success - What successes has the DCN had? What researcher challenges have we solved and what impact has that made? - How did we accomplish these successes? What structures enabled success?
4:30–5:00	Debrief Jam Board and Round Robin
5:30–6:00	Happy Hour Hotel Madera Hotel offers complimentary HH
Thursday March 10, 2022	
8:30–9:00	Breakfast - In room catered
9:00–9:15	Welcome Back - How were we successful yesterday in addressing our goals?
9:15–12:00	Breakout #3 - DCN Challenges and overcoming barriers - What challenges/barriers did the DCN encounter? - Break - What structures did we set up to enable individuals to overcome barriers? - What structures got in the way?
12:00–1:00	Lunch Break - On your own (leave the room)
1:00–2:30	Breakout #4: Bringing outside in - What other outside DCN projects do we admire? - Blue sky
2:30–3:30	Writing session - Prioritize structures - Draft a few paragraphs

3:30–4:00	Day 2 Wrap up - Next steps with this report
5:30–6:00	Happy Hour Hotel Madera Hotel offers complimentary HH
Friday March 11, 2022	
8:00–8:30	Breakfast - In room catered Round Robin: Where were we successful yesterday? Are there areas we need to come back to?
8:30–9:00	What do we want this report to say about the DCN? - Round Robin
9:00–9:45	Other models we admire
9:45–10:00	Break
10:00–11:00	Breakout topic # 5 - Circle back to researcher needs - How could the structures identified by our collaborative model be applied to other needs faced by researchers? - What might the DCN take on in support of RDM, open science, and beyond? - What's next for collaborative cross-institutional networks?
11:00–11:30	Audience Communications Plan - Look at the audiences we brainstormed on day 1 - How will we share what we've learned with them?
11:30–12:00	Wrap up - Asynchronous peer review - Writing blocks for one month
12:00	Lunch Break - On your own (leave room)

ATTENDEES

- Mara Blake, North Carolina State University (formerly Johns Hopkins University)
- Jake Carlson, University of Michigan
- Briana Ezray Wham, Pennsylvania State University
- Joel Herndon, Duke University
- Cynthia Hudson Vitale, Association of Research Libraries (formerly Pennsylvania State University)
- Lisa Johnston, University of Minnesota
- Wendy Kozlowski, Cornell University
- Cindy Xuying Xin, Pennsylvania State University
- Sophia Lafferty-Hess, Duke University
- Hoa Luong, University of Illinois at Urbana Champaign
- Wanda Marsolek, University of Minnesota

- Jennifer Moore, Washington University in St. Louis
- Mikala Narlock, Data Curation Network / University of Minnesota
- Dorris Scott, Washington University in St. Louis
- Sarah Wright, Cornell University

APPENDIX C: COLLABORATIVE STRUCTURES

The following information about the collaborative structures that contributed to the success of the DCN was written during the project retrospective. Each retrospective participant was given an hour to draft information about one of the collaborative structures, which had been identified in a previous exercise. Participants were invited to leave the room, find a space to work individually or in pairs on the exercise, and asked to write a small amount about what the structure or tool is, and then articulate how it is used by the DCN. Participants were given the following example as a guide:

Slack

- Slack is a group chat tool that enables direct, asynchronous communication.
- The DCN used Slack as a communications tool to supplement our other channels (zoom meetings, etc.). What is interesting is that most (~90%) of the slack traffic in the DCN is direct messages rather than to the entire group.

This content drafted during the retrospective was then used in creating this published report. This Appendix provides supplementary information that may be of interest to readers. The text has been slightly edited for clarity, but has been left largely as written by the authors.

A Clear Process

One of the most impactful structures for the DCN involved developing a clear process for data curation.

While data curation has emerged as an area of rapid growth in academic libraries over the last decade, many institutional data services have embarked on a range of experiments designing workflows and creating infrastructure to meet local needs. One of the most impactful structures of the DCN is the development of a

shared, formal context / vocabulary / framework for data curation (the CURATE(D) steps) that allows institutions with disparate data curation workflows to leverage the best elements of curation across multiple institutions. By creating a shared vocabulary and technical workflow (through Jira) for data curation, the DCN not only enhanced the collaborative efforts of the member institutions, it also created the framework for a curation instruction program that allowed the DCN to onboard new members while training the larger curation community.

Consultants

- Individuals, external to the group, hired to provide expertise, facilitation, and/or consultation around a particular topic.
- The DCN worked with several consultants who brought an external perspective to our internal challenge or question. Consultants can be short term team members or more hands-off. For a cross-institutional collaboration, a consultant can help surface any unspoken expectations or hidden assumptions that, if unrealized, may slow or impede the collaboration.
- Example 1: We hired an external facilitator for our kick-off meeting during the planning phase grant phase. Santiago was ESSENTIAL to help our new team of PIs from six institutions come to a SHARED UNDERSTANDING of our goals, the metrics for success in meeting those goals, the barriers, and how we might overcome those barriers, and a strategic plan for how to start the doing. This type of facilitation is different from project management, it was consensus building and the facilitator really became part of the group over those two days.
- Example 2: For the sustainability project and the Diversity, Equity, Inclusivity, and Accessibility (DEIA) action plan project, we found a formal request for proposal process to be useful not only to get written bids making the hiring process more transparent and to fit well with our project and budget, but also to allow the collaboration team to better articulate and define the role that a consultant might play (e.g., writing a statement of work). Having a clear set of expected outcomes and meeting

those expectations as a condition of the consultancy was key, otherwise the potential for the work to be open-ended and not well scoped was huge.

Shared, Accommodating, Flexible Process that Allows for Local Variation

One sentence summary:

A workflow that is technology agnostic and highly adaptable, but is built on a common knowledge, language, and with a shared goal.

How the DCN uses this

The DCN uses this as the basis for a collaborative curation workflow model—while we all have local nuances with our repositories, curation practice, and connections with researchers, our shared workflow is easily integrated because it does not rely on a single technology. We have a shared vocabulary to discuss our work, though—our process and protocols—which means that even when there are differences in local implementation, we can have a productive conversation.

- this is really an agile approach. It's lightweight so it's incredibly portable (i.e., can connect into extant processes) but it's rigid enough to provide structure—folks just can't go off the rail or do their entirely own thing!
- It ensures consistency without requiring too much—provides a sense of cohesion, too. I'm part of this network, and I curate this way alongside my colleagues, even though we are using different technologies (e.g., dspace, fedora, etc.).
- This also takes off the mental load—don't have to reinvent the wheel every time!
- This also means that the process will be more sustainable because it does not have to be refreshed frequently as technology changes.
- We provide the "human in the loop".
- Example 1: CURATE(D): This provides the foundation for the shared curation—ensures all of our datasets are curated in a consistent manner; does not require specific technology. For example, while format and discipline specific primers and

CURATE(D) workflows might recommend tools to facilitate curation, all of the steps can be completed however best works for the individual curator.

- Example 2: Jira: This is where our process is actually lived out, but again, there is room for nuance. The form is built to capture standardized information, but also allows for flexibility. For example, not every institution can allow access to files that are embargoed or under peer review. So, Jira doesn't require a link to the dataset as it exists in the repository—just wherever is most convenient and logical for the submitting institution (e.g., Box).

How this might be translated into other communities / networks

- For others that might want to offer a service, I think this can and DOES translate.
- e.g., Girl Scouts of America—they have common structures, but they allow their troops a huge degree of flexibility. There are processes for earning merit badges—you complete these five steps—but each troop and individual can meet those however they want.

Primers

Write 1 definition sentence of the structure

Data curation primers are brief reference documents (~15 pages) that provide basic background information about specific data types or formats as well as information or approaches that can be used to help information professionals curate the data type or format.

How were they used by the DCN?

Data curation primers were first developed at the University of Michigan by Susan Borda as a part of an internship program to give a graduate student experience in approaches to curating netCDF files. Primers were adopted by the DCN as a part of our Advanced Workshop on Data Curation (not sure of the exact title) series that were developed with support from the IMLS. The idea behind adopting the primers was that they would give students attending DCN workshops an opportunity to apply what they had learned about

curation to a format or data type of interest. Students developed ideas for possible primers over the course of the workshop and then signed up to be a part of a team to work on them over the next six months. Each team was assigned one of the workshop instructors as a mentor, who could provide guidance and advice on developing the structure and content of their primer. Once the team had created a draft of the primer, the draft was submitted for peer review by other data librarians. Feedback from the peer review was then incorporated into the final draft. Teams then presented their primers in a webinar and completed primers were posted on the DCN website and deposited into University of Minnesota's data repository.

What made the primers a successful structure for the DCN?

- They address a common need of data curators: Data curators, particularly those affiliated with institutional repositories, often have to work with a wide variety of data that they may not be familiar with. Primers provide a ready reference guide and some structure for curators to use when curating unfamiliar content.
- They served as a means for librarians to apply what they had learned: One of the challenges of training is that the learning often stops at the end of the workshop. Working on the primers for an extended period of time after the workshops gave librarians an opportunity to dedicate time to further exploration of a topic of interest.
- Authors received professional recognition for their work: By presenting their work in webinars and having it published on the DCN web site, librarians who attended our workshop could point to a "tangible" output and receive presentation and authorship credit.
- Strengthening the connection with the DCN: The peer mentor who worked with the primer teams served as an ambassador, in a way, to the DCN as an organization.
- Strengthened the reputation of the DCN in the larger data community: The primers also addressed a larger need in the curation community for reference guides describing actionable steps that one could take in curating data.
- Within the DCN, they help to serve as a common anchor point for developing and delivering the workshop: Having this

anchor point, help us shape what content we would teach, how the workshops would be structured and our approaches to teaching our material.

Grants: External Funding

Write 1 definition sentence of the structure

Grants and external funding opportunities have been a critical first step to establishing and building the Data Curation Network.

How were they used by the DCN?

Both private and federal funding initiatives have been utilized to plan for, research, and ultimately move into production different initiatives within the Data Curation Network. The work to date of the DCN can be best split into two separate initiatives—a shared staffing model and an education initiative.

Grants for shared staffing

The planning phase of the Data Curation Network (2016–2017) was generously funded by the Alfred P. Sloan Foundation. This planning period allowed members of the DCN to develop a list of curation activities, conduct information exchanges with key partners, understand researcher needs for curation services, and more. It also allowed us the opportunity to pilot out what a shared staffing model for data curation might look like among our six original institutions.

After the planning phase of the grant was completed, the DCN, through the University of Minnesota, received an additional 3-year grant to implement the network. Over the course of the three years, the DCN implemented a technical project management tool for the coordination of exchanging datasets and expertise among our members.

Grants for education

The education initiative within the DCN was launched by a grant received from IASSIST in 2017 and further supported by DCN

member institutions. This grant allowed members of the DCN to pilot a peer-to-peer, hands-on workshop for the CURATE(D) protocol.

In 2018 the education initiative received an IMLS Laura Bush 21st Century Librarian grant, titled "Building the Digital Curation Workforce: Advancing Specialized Data Curation" to further develop and expand the IASSIST workshop for more library staff. From 2018–2021 this grant provided three workshops to over 77 librarians and is currently being built into a set of online curation modules for broader library engagement and training.

What made grants a successful structure for the DCN?

- Allowed us to plan and pilot out initiatives before jumping into them
- Provided us resources and dedicated time to initiate these key areas of the DCN

All Hands Meeting

In person or virtual annual meeting to bring PIs and curators together from each participating institution for sharing out the previous year's work and outputs, training opportunities, community building, and celebration of year's accomplishments and community.

How did we use it?

- The AHM was planned by a volunteer committee each year to develop programming. Each committee member or pairs of committee members were responsible for activities during 1–2-hour time blocks to help distribute the work. This also created multiple leads, who were all responsible for developing an informational, educational, and fun program of events.

What successes did it lead to?

- Lightly structured fostering of community, networking, community building, education. Because it was developed by our own, there was an additional sense of community introduced by

those who attended, giving up much of their time for a specified amount of time to be supportive of those who had done the heavy lifting of coming up with the program. Later AHM invited external members of the data community who had attended workshops hosted by the DCN, again nurturing the community and not just a one and done thank you ma'am. External folks were able to participate in information gathering as well as provide lightning talks on the primers they had created over the last three years

How did it help us overcome challenges?

- A distributed network such as the DCN saw the benefits of bringing the group together in the same space for team building and working harder/faster for a short period of time (2–4 days) together, similar to a research sprint. Early AHM involved a lot of brainstorming for how members viewed the future work of the DCN.

What other challenges might this structure help address?

- More than ever communities are feeling the effects of isolation. Actively bringing the community together helped form relationships that were missing from email updates, stand-ups, and the geographical distance of members.

Google Drive

Currently, Google Drive is the main storage solution that the DCN uses to store all its information, including administration documents, projects, notes, data, etc.—in short everything. The advantage of this choice is we have one central place to store all needed information. All archival notes are kept, which is super helpful to go back whenever needed. Access to folders can be easily controlled and handed off by the administrators. Having a central place makes it easier for interest groups to know where to store their information as well.

As we continue using Google Drive, there are multiple questions that arise from this:

1. Is the DCN Google Drive under institutional license or a private license?
2. Is there a migration plan in place in case we need to move away from Google Drive?
3. How do we determine what cloud storage to use in case Google Drive is no longer available for us?
4. Google note is the most convenient thing and we use it really often. Do we have a plan to convert them to open format for preservation and ease for migration?
5. Will we ever run out of space?
6. What are our retention plans for our documents in Google Drive?
7. Do we have another back-up mechanism besides the cloud?
8. Currently, who controls the access (ownership) and what are the access levels? In other words, who has access to which folder since administrative information is stored there.
9. Considering that we currently store all our results from our studies in this Drive, is it a practical way to handle research data?
10. Do we have a naming convention so all the folders/files can be easily sorted and found?

Network of Expertise

A peer-to-peer network with members with specialized knowledge and skills that can be shared to serve common goals, such as data curation.

DCN has been able to recruit curators in social sciences, STEM and other research areas to form a service network to provide cross-institutional curation support for member institutions. This has greatly helped local institutions to fill gaps in expertise needed for curating certain datasets. DCN has also generated data curation primers based on that collective expertise and shared them broadly within and outside the curation community in a sustainable way—shared them on GitHub. Forming a network of expertise has helped those within and outside the curation community promote FAIRness in research data for advancing open science initiatives.

Special Interest Groups

The DCN has used special interest groups as a mechanism for advancing shared goals through specific projects as well as flexible spaces for open sharing among peers. The process for developing special interest groups within the DCN has largely been organic and flexible with topics of interest arising at AHM or through shared research goals of the network. Some special interest groups have had discrete projects/deliverables and then disbanded, while others have been longer-standing with shifting goals/projects that are determined and scoped by group participants.

Example 1: The Human Subjects Interest Group came together at an AHM and determined as a collective that the primary output they wanted to produce was a primer on curating human subjects' data, generating as an output an open educational resource available to the entire community.

Example 2: The Institutional Outreach and Communications Interest Group has largely focused both on creating a space for sharing institutional approaches to outreach and advocacy through informal discussions or more structured "steal this idea" sessions, while also focusing on specific projects such as developing a framework for data stories and generating communication materials for the network.

Example 3: Two research focused groups—the End User Survey Interest Group and the Value of Curation Interest Group—fielded surveys to two key audiences, the end users of DCN repositories and information professionals working in repositories, to further answer the question "What is the value of curation?" Both groups are currently authoring papers to share the outputs with the broader community and plan to disband when those outputs are completed.

The special interest groups have enabled another point of collaboration among our community of practice that expands beyond our shared curation model to build relationships and contribute to the field.

Curator Onboarding Process

The onboarding process offers an entry point for new members and provides the information necessary for new members to quickly engage with the DCN community. The curator onboarding process was developed to provide new curators with background knowledge of the DCN, the necessary curation practices and DCN workflows to take part in both shared curation efforts, and also an overview of other engagement opportunities such as DCN special interest groups and resource development (e.g., primers, online modules). More specifically, the curator onboarding process offers data curation practices education through the CURATE(D) training or training video as well as hands-on experience curating a dataset with expert feedback and shared data curation workflow training through a walk-through of the Jira issue tracking platform. Additionally, during the curator training, new members are asked to fill out the expertise assessment form which is then reviewed with the assistant director during the one-to-one meeting. The expertise assessment form offers new members the opportunity to highlight the strengths they bring to the community while also providing a space to request skill development from others in the community. Overall, the success of this process comes from the fact that it systematically welcomes new members to the community, provides hands-on curation experiences which boosts new member confidence in the process, and offers human interaction and feedback.

Website

One of the ways in which the Data Curation Network maintains a public presence is through their website. The website gives information about the organization, datasets curated, the community of data curators at the heart of the DCN, along with educational outreach, research, and news.

The website provides various pieces of information about the mission, vision, and structure of the organization. In addition, the website has information about the datasets that have been curated and various visualizations about the datasets and the expertise of the curators. Examples of data visualizations are the growth of datasets that have been submitted over time and DCN data type expertise

based on the number of curators who identified as being an expert on a data type.

Community plays an important role in the DCN. The "Community" section includes the list of DCN curators in which you can filter by subjects, institutions, disciplines, data types, and software languages. This section also gives information about the resources and tools from DCN members, special interest groups and committees, the annual AHM agenda along with a section on the various ways in which one can get involved in the DCN such as joining a special interest group or registering for a workshop.

One can also see the list of previous workshops held and request for a workshop to be held at their institution. The website also lists the research that the Data Curation Network has been involved in by subject along with the data primers, data glossary, and the CURATE(D) model.

The website has a blog which has posts on various topics such as current happenings with the DCN along with posts that highlight data curators from various institutions. This blog is not only a good way to keep up with what is happening in the DCN but also a way to get to know the curators better in terms of their background, how they came to their current position, along with learning something interesting about the curator that is not work-related.

The website is an effective tool for outreach and can be leveraged by those who want to do outreach efforts about how to curate data, along with becoming familiar with a community of curators. In the future, once the membership tiers are more fully defined, there are plans to include an application form in which one can apply for individual membership.

Code of Conduct

The Data Curation Network's code of conduct clarifies norms and expectations for behavior of members in all venues of the collaboration.

The code of conduct draws on similar documents that apply to conferences, as well as those that structure contributions in open-source communities. It establishes expectations for member interactions in project and public spaces related to the work of the DCN. The

document also lays out steps for enforcement in the event of violations of the stipulations of the code of conduct. In addition to laying out behaviors to foster an inclusive and respectful community, the code of conduct also serves as a mechanism to articulate the values of the DCN. The process of drafting and adopting the code of conduct afforded the members of the DCN the opportunity to clearly identify goals for member interactions and articulate them internally and externally.

Shared Leadership Without Authority

We discussed in-depth what makes our shared leadership work. Did it come from the design of the project? Are like-minded folks attracted? Is the return on investment so large that it makes it worth the investment of time? All of these factors contribute to why the DCN shared leadership model works. We all own it, we are all invested, we all want to see it succeed. It's due in part to our ethos of cooperation that may be built into who we each are individually, but another part can be attributed to our intentional design. The founding members had experience and success working together and this laid a bedrock of trust and respect. The slow and intentional growth occurred as the founders sought out more contributors and the fabric of trust and respect expanded. This is a great example of what is meant by a healthy, functional, community of practice, but this an area of concern as our growth continues. We ask ourselves, what does it mean for our shared leadership and ownership as we expand? How will we include more institutions of varying size and focus and maintain our fabric of trust and respect? To date, we continue to grow slowly so that we can thoughtfully address this issue of scale. We look to peers for good examples and we talk about all of it openly, employing active listening.

"Necessity is the mother of inventions" and "motivation is born from success" are two essential factors, which cannot be overlooked. We needed each other, because none of us can solve our curation programs alone. That shared vulnerability made us dive into these uncertain waters and work together via our "radical independence." We learned quickly that we could progress together in a meaningful way, and this infused our continued work.

Peer-to-peer Comparisons

Definition: Documentation of a specific part of a workflow, practice or service with the purpose of information sharing and finding similarities, and learning from one another.

How we used them: The DCN is based on shared expertise, not shared infrastructure. Many of the DCN member institutions started curation at their local institutions at the same time the network was being established. This sharing of information of how we were each running our workflows and using various tools throughout the process allowed us each to not have to "reinvent the wheel." The summaries often prompted discussion on why something was being done a certain way and what challenges curators had overcome or didn't by doing things a certain way.

Sharing of our practices in this way served several purposes. By being open with each other about how we do curation, we built trust. It fostered communication between curators, strengthened the sense of community. Open sharing of what worked and didn't allowed for efficiencies of scale—we didn't each have to necessarily build things from scratch when we knew something was effectively being used elsewhere within the network.

Example 1: Jira—Jira was used at several institutions, which made it easier (though not easy) to put it in place for tracking of the DCN work.

Example 2: Curator response templates—initially shared as part of the CURATE(D) curriculum materials, additional response templates were shared and built on at an early AHM as part of the curator workflow comparison.

Coalescing Around a Shared Problem/ Challenge/Need

The Data Curation Network coalesced around a common problem: the need for data curation staffing and expertise to provide data curation services for data being submitted to institutional data repositories.

Network members chose to be part of the network and were there to solve a common problem, therefore we all have a sense of responsibility and ownership. Success for the network leads to success at our respective institutions and enables us all to benefit from the broader pool of expertise. Coming together around a shared problem also kept us focused on tangible results—the work was not theoretical or research-based, the network had work to do to be successful. The fact that we had work to do with tangible, measurable results also helped us to try out solutions and iterate to find out what worked and what didn't. The network approach may be useful for other challenges shared by a group of like-minded institutions, with a limited number of staff that can benefit from a community of practice to distribute and share the work.