Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data

Lisa R. Johnston
University of Minnesota

Jake Carlson
University of Michigan

Wendy Kozlowski
Cornell University

Joel Herndon
Duke University

Cynthia Hudson-Vitale
Penn State University

Robert Olendorf
Penn State University

Elizabeth Hull
Dryad Data Repository

Heidi Imker
University of Illinois
at Urbana Champaign

Mara Blake
Johns Hopkins University

Timothy M. McGeary
Duke University

Claire Stewart
University of Minnesota

Elizabeth Coburn
University of Minnesota

Abstract

Funders increasingly require that data sets arising from sponsored research must be preserved and shared, and many publishers either require or encourage that data sets accompanying articles are made available through a publicly accessible repository. Additionally, many researchers wish to make their data available regardless of funder requirements both to enhance their impact and also to propel the concept of open science. However, the data curation activities that support these preservation and sharing activities are costly, requiring advanced curation practices, training, specific technical competencies, and relevant subject expertise. Few colleges or universities will

Received 27 October 2017 ~ *Accepted* 20 February 2018

Correspondence should be addressed to Lisa Johnston, University of Minnesota, 108 Walter Library, Minneapolis, MN, 55455. Email: ljohnsto@umn.edu

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see https://creativecommons.org/licenses/by/4.0/

http://dx.doi.org/10.2218/ijdc.v13i1.616
DOI: 10.2218/ijdc.v13i1.616

Abstract (continued)

be able to hire and sustain all of the data curation expertise locally that its researchers will require, and even those with the means to do more will benefit from a collective approach that will allow them to supplement at peak times, access specialized capacity when infrequently-curated types arise, and stabilize service levels to account for local staff transition, such as during turn-over periods. The Data Curation Network (DCN) provides a solution for partners of all sizes to develop or to supplement local curation expertise with the expertise of a resilient, distributed network, and creates a funding stream to both sustain central services and support expansion of distributed expertise over time. This paper presents our next steps for piloting the DCN, scheduled to launch in the spring of 2018 across nine partner institutions. Our implementation plan is based on planning phase research performed from 2016-2017 that monitored the types, disciplines, frequency, and curation needs of data sets passing through the curation services at the six planning phase institutions. Our DCN implementation plan includes a well-coordinated and tiered staffing model, a technology-agnostic submission workflow, standardized curation procedures, and a sustainability approach that will allow the DCN to prevail beyond the grant-supported implementation phase as a curation-as-service model.

Introduction

Well-curated data are valued by the scholarly communities that produce them. Professionally curated data are easier for fellow scholars and future collaborators to understand, are more likely to be trusted, and the research they represent more likely to be reproducible (Roche, Kruuk, Lanfear and Binning, 2015; McNutt et al., 2016; Smith and Roberts, 2016; Beagrie and Houghton, 2014). As researchers worldwide face emerging mandates and altruistic pressures to share their research data, curation activities can help make data findable, accessible, interoperable, and reusable, or FAIR (Wilkinson et al., 2016). For example, funders increasingly require that data sets arising from sponsored research must be preserved and shared, and many publishers either require or encourage that data sets accompanying articles are made available through a publicly accessible repository. Often, reproducibility is a driving factor for these policies (Stodden, Guo and Ma, 2012). Additionally, many researchers wish to make their data available regardless of funder requirements both to enhance their impact and in general support of the concept of open science. Some disciplines have embraced the open data movement as a positive development that will foster expanded practices in validation and replication (Munafò et al., 2017), and may even safeguard against scientific fraud or the dissemination of erroneous results (Fecher, Friesike, Hebing and Linek, 2017).

Curation staff are the ‘human layer’ in the repository technology stack that bring the disciplinary knowledge and software expertise necessary for reviewing incoming submissions to ensure that the data are FAIR. The skills and expertise required to curate data (to prepare, arrange, describe, and optimize data for reuse) cannot be fully automated nor reasonably provided by a few experts siloed at a single institution. Multiple data curation experts are needed to effectively curate the diverse data types a repository typically receives (Bloom et al., 2016; Johnston, 2014) and to keep up with changing trends and emerging tools that support research data best practice.

The **Data Curation Network (DCN)** addresses the challenge of scaling domain-specific data curation services by collaboratively sharing expert data curation staff across a network of partner institutions and data repositories beyond what any single institution might offer alone. The DCN will ensure that institutional repositories (IRs) and non-profit data repositories can draw from a pool of expert data curators for a wide variety of data types (e.g., GIS, tabular spreadsheets, statistical survey, video and audio, software code, etc.) and discipline-specific data sets (e.g., genomic sequence, chemical spectra, qualitative survey, etc.) while also providing normalized curation practices and professional development training.

The DCN planning phase began in 2016, with support from the Alfred P. Sloan Foundation, and brought the perspectives of researchers, librarians, administrators, and data curation subject experts from six U.S. academic institutions: the University of Minnesota, Cornell University, Penn State University, the University of Illinois, the University of Michigan, and Washington University in St. Louis. The planning phase team ran a baseline assessment of local services, held focus groups with faculty researchers, ran controlled data curation pilots, and surveyed the library curator community to understand existing support and future plans for services in these areas.

Our community-vetted planning phase report is grounded in the measurable metrics and observed demand for data curation services across six planning phase institutions.[1]

This paper will present our next steps for implementing the DCN, scheduled to launch in the spring of 2018 with new DCN partner institutions including Duke University, Johns Hopkins University, and the Dryad Data Repository. Our DCN implementation plan includes a well-coordinated and tiered staffing model that incorporates data curator expertise across a wide variety of domains, a technology-agnostic submission workflow that accommodates the various repository technologies in use (e.g., Samvera/Hydra, DataVerse, DSpace, BePress, etc.), standardized minimum levels of curation that enable DCN Curators to prioritize their work, a sustainable financial plan to support the DCN beyond the grant-supported implementation phase, an assessment plan to evaluate how a networked approach to curating research data is more efficient and scalable, and a professional development program that enables the Data Curation Network partners to train and recruit new data curators and keep up-to-date with data best practices across domains, communities, nations, and beyond.

Literature Review

Data curation is a subset of the broader suite of research data management services (Kouper, Fear, Ishida, and Williams, 2017). A number of studies and surveys have explored the extent of research data services provided by academic libraries and found that support for research data management, including data curation, has increased steadily over time (Soehner, Steeves, and Ward, 2010; Tenopir et al., 2011; Tenopir et al., 2015). More recent explorations by Lee and Stvilia (2017) found that support for data curation in libraries is mainly built upon existing and local IRs. IRs only account for a small percentage of the data repositories available to researchers, while discipline-specific data repositories (e.g., ICPSR, GenBank) and general-purpose repositories for data (e.g., FigShare, Zenodo) are enjoying growing use (Kindling et al., 2017).

The Data Curation Network builds on a rich history of well-established collaborative service models in libraries. Not unlike our vast interlibrary loan networks that deliver books, articles, and other library collections across networked libraries, or the collective contributions of catalogers adding unique and specialized MARC records to national and international cataloging databases (Weber, 1976), or the more recent response to on-demand web-based user needs with the successful implementation of 24/7 library reference chat services, the DCN builds from our common need to provide scaled services and expertise in a shared way. The appeal for a network of expertise model for delivering unique library services has been expressed through recent research on centers of excellence. Kirchner et al. (2015) recommend "...a pilot project in which experts at multiple institutions consciously create a shared approach to address specialized information needs or to solve a common problem" (p17). Additionally, Erway (2012) calls for a collaborative expert network for handling the variety of born-digital media managed in the nation's libraries.

Collaborative networks that specially address data and metadata curation issues provide a great foundation for the DCN to build from. The Research Data Alliance, launched as a community-driven international organization in 2013, provides a venue

[1] An expanded version of our DCN model is available with detailed curation workflows, staffing roles, draft MOUs, and tool tracking functional requirements as "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data" originally published July 27, 2017 online at https://sites.google.com/site/datacurationnetwork/results

for developing and establishing standards for data curation with special interest groups like the Publishing Data Workflows group (Bloom et al., 2015) and the newly formed Assessment of Data Fitness for Use working group.[2] The Curating for Reproducibility Consortium project combines staff and best practices for social sciences data.[3] Recent projects related to research data repositories and preservation (though not specifically focused on data curation services) are also underway. The Stewardship Gap project reported looked at how sponsored research data gets preserved for future generations (York, Gutmann, and Berman, 2016). The Portage Network[4], Canada's emerging shared data archive service, and the UK-based Jisc Research Data Shared Service Project[5], seek to build shared software and repository infrastructure for higher education institutions in Canada and the UK, respectfully. Finally, educational preparation for data curation services, like the DigCCuRR Professional Institute[6] and the CLIR data curation post-doctoral fellowship program[7], as well as information sharing networks such as the Digital Liberal Arts Exchange[8] and the DataQ Project[9], are leading the way in training data curators on relevant best practices in the field as well as providing valuable forums for community building and networking.

Methodology

The planning phase to develop a Data Curation Network model ran from 2016–2017 with support from the Alfred P. Sloan Foundation and brought together research data librarians, data curation experts, and academic library administrators from six academic institutions that each, separately, provided repository and curation services to their campuses. The initial six institutions were: the University of Minnesota, Cornell University, Penn State University, the University of Illinois, the University of Michigan, and Washington University in St. Louis. Core research activities performed in the planning phase that directly informed the DCN model development included:

- a baseline assessment of the six institutions to understand the existing levels of support for data curation and compare local policies and technologies already in place (Johnston et al., 2017);
- focus groups incorporating a total of 91 researcher perspectives across six institutions on the importance of data curation activities, their current habits, and needs (Johnston et al., 2017a);
- controlled data curation pilots with 17 curators to identify variations in local practice and potential implementation issues, including normalization of curation processes (Johnston et al., 2017b);

[2] Research Data Alliance: https://www.rd-alliance.org

[3] Curating for Reproducibility Consortium: http://cure.web.unc.edu.

[4] Portage Network: https://portagenetwork.ca

[5] Jisc Research Data Shared Service Project: https://www.jisc.ac.uk/rd/projects/research-data-shared-service.

[6] DigCCuRR Professional Institute: https://ils.unc.edu/digccurr/institute.html

[7] CLIR Postdoctoral Fellowship Program: https://www.clir.org/fellowships/postdoc

[8] Digital Liberal Arts Exchange: https://dlaexchange.wordpress.com

[9] DataQ Project: http://researchdataq.org

- community engagement with 124 US and Canadian-based academic research libraries to better understand levels of current support for data curation services (Hudson-Vitale et al., 2017);
- a cost model review to compare approaches to supporting sustainable data curation and repository services, supplemented with practical information exchanges with the leaders of major collaboration projects in order to learn from their past experiences (Johnston et al., 2017c);
- metrics tracking of the types, disciplines, frequency, and curation needs of data sets curated across our six institutions to understand the demand for data curation services over a one-year period (Johnston et al., 2017c).

Additionally, our team sought opportunities to broadly present our work and discuss our ideas with colleagues at relevant conferences. As a result of these conversations it became clear that although our planning phase work was focused on the needs of US academic research institutions similar to the six represented by the project team, this model would scale to a wider range of organizational make-ups and affiliations such as federal government agencies, international academic institutions, and small- and mid-sized liberal arts colleges. We very much welcome the opportunity to explore these and other avenues for broader interpretation of the DCN model.

A Cross-Institutional Staffing Model for Curating Research Data

The Data Curation Network harnesses the expertise of well-aligned institutions that collectively provide data curation services to researchers in a multitude of disciplines, ensuring that valuable scholarly datasets are findable, accessible, interoperable and reusable, or FAIR. Offered through a unique collaboration between academic libraries and general data repositories, DCN curators at distributed sites are matched with data sets according to their technical and disciplinary expertise, and conduct a rigorous review of the data using an established set of protocols that seamlessly fits within any local curation workflow.

The DCN will function through a well-coordinated and tiered staffing model that includes levels of participation allowing some institutions to join the Network by contributing in-kind data curation staff and others to utilize the Network's curation services as end-users. *Partner institutions* (e.g., academic libraries or general data repositories, etc.) contribute staffing and funds to sustain and offer central services to potential *users* (e.g., academic libraries with limited or no curation resource, general or domain repositories in need of a curation service layer, publishers with data sharing requirements, etc.). Stakeholders will gain access to data curation expertise in more disciplines/formats than locally available and contribute to a larger ecosystem of data curation practice (see Table 1). DCN users will be able to more efficiently work with investigators to capture as much context and description of the data as possible, expertly review data quality and validate code, assess risks and verify file integrity, and validate and transform files. DCN curators also provide guidance around secure storage, citation and persistent identification strategies, and data curated by the DCN may be deposited into the repository of the researcher's choice for ongoing stewardship.

Table 1. Benefits of Participating in the Data Curation Network.

Stakeholder	Benefits
Academic libraries with existing data curation services	• gain access to data curation expertise in more disciplines/formats than locally available; • contribute to a larger ecosystem of data curation practice; • participate in the development of shared standards; • build a pipeline for training data curators and establishing professional data curation practices; • inform and advance development of local curation services; • smooth and stabilize services during times of staff transition and shortage.
Academic libraries with limited to no resources for data curation services:	• are able to provide critical new data curation services when local resources are limited (without needing to hire); • have the opportunity for a local data curation specialist to join a larger, robust network; • benefit from a clear roadmap, presented by DCN partners, toward data curation services maturity and scale; • normalizing the practice of data ingest/deposits/archiving in library-hosted repositories.
Disciplinary- and general-subject data repositories:	• receive better, more valuable data submissions from DCN partner institutions and customers; • have potential to partner with the DCN to expand the scope of curation support for the disciplinary repository to new and/or less frequently encountered data types; • gain access to curation staff that are housed at external institutions thereby minimizing staffing overhead costs; • get more researchers directed to the disciplinary repository thanks to the broad network of participating institutions; • obtain potential new revenue stream as consumption scales, should the disciplinary repository seek to join as a partner.

DCN Staffing Model

The DCN will implement a well-coordinated and tiered staffing model. An important consideration uncovered in the DCN planning phase research was the need to maintain and strengthen local relationships between researchers and repository staff. Therefore, to reduce missed opportunity costs, our model incorporates several roles to better establish a chain of communication from the researcher to the DCN staff:

- *Local Researcher*: The individual responsible for the dataset. Often the author/creator of a dataset but may also be a representative acting on the author's behalf (e.g., a graduate assistant). The local researcher communicates with the…

- *Local Curator*: The staff member who submits a dataset from their home institution to the Network. The Local Curator continues to serve as the primary contact for all communications with the Local Researcher throughout the curation process. The local curator is also a …
- *DCN Curator*: Networked staff that provide expert curatorial services for datasets submitted to the DCN who each bring skills for specific file formats (e.g., databases, statistical survey data, video/audio files, computer code) and/or types of disciplinary data (e.g., 3D images, genomics, chemical spectra, ecological, etc.). DCN Curators take on the role of Local Curator when submitting data from their institution. DCN Curators benefit from annual training events and peer networking and work closely with the…
- *DCN Coordinator*: Centrally funded through the DCN, this role oversees the daily operations of the Network, tracks and monitors all datasets that flow through the Network, and reviews and assigns incoming data sets to the appropriate DCN Curator. The DCN Coordinator reports to the…
- *DCN Representatives:* Each partner institution will select one DCN Representative to participate in the Network as the institutional lead. DCN Representatives make up the governance body of the DCN and establish and enforce policy.

DCN curators, DCN Representatives and the DCN Coordinator will communicate on a regular, ongoing basis (e.g., bi-weekly conference calls) in order to share out on curation assignments and make adjustments and changes to the workflow as new situations arise.

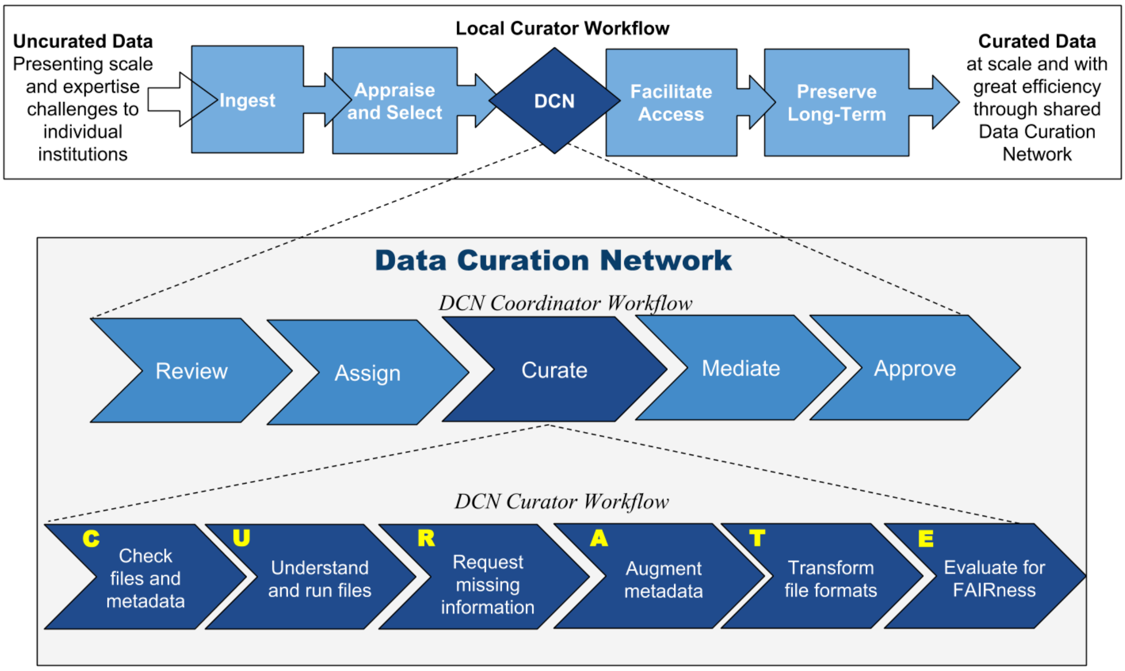

Figure 1. The Data Curation Network seamlessly interacts between local curation workflows and networked curator staff across the DCN partner institutions.

DCN Submission Workflow

The DCN incorporates a technology-agnostic submission workflow that fosters strong local connections between researchers and local curators and gives the home institution complete control to decide how to engage Network resources. For the implementation phase, the DCN submission workflow assumes that all repository functionality (ingest, storage, access, dissemination, and preservation) is the responsibility of the local institution. Therefore, local researchers may submit data to their local curation service like normal, then the Local Curator must determine if the dataset should be submitted to the DCN for expert curation and review. Datasets received by the Network will be handled via a submission-tracking tool to monitor a dataset's progress through the DCN workflow (see Figure 1).

Figure 2 shows the role-specific actions involved in the DCN submission workflow. All DCN submissions will receive a preliminary check (e.g., sensitivity risks, corrupted files, etc.) by the DCN Coordinator before assigned to an appropriate DCN Curator based on expertise match and availability. Once assigned a dataset, the DCN Curator is responsible for reporting any questions, changes, augmentations, and corrections for the data back to the Local Curator. We recognize that researchers may choose not to take recommend actions, therefore the last step in the DCN workflow is for the DCN Curator to assess the final result in order to determine if it meets standards for FAIRness (Dunning, de Smaele and Böhmer, 2017). Any issues (e.g., problems with a particular dataset) can be discussed at the regular curator virtual meetings where all DCN curators may participate. Here peers may recommend additional actions be taken or collaborate on resolutions for copyright issues, documentation, etc.

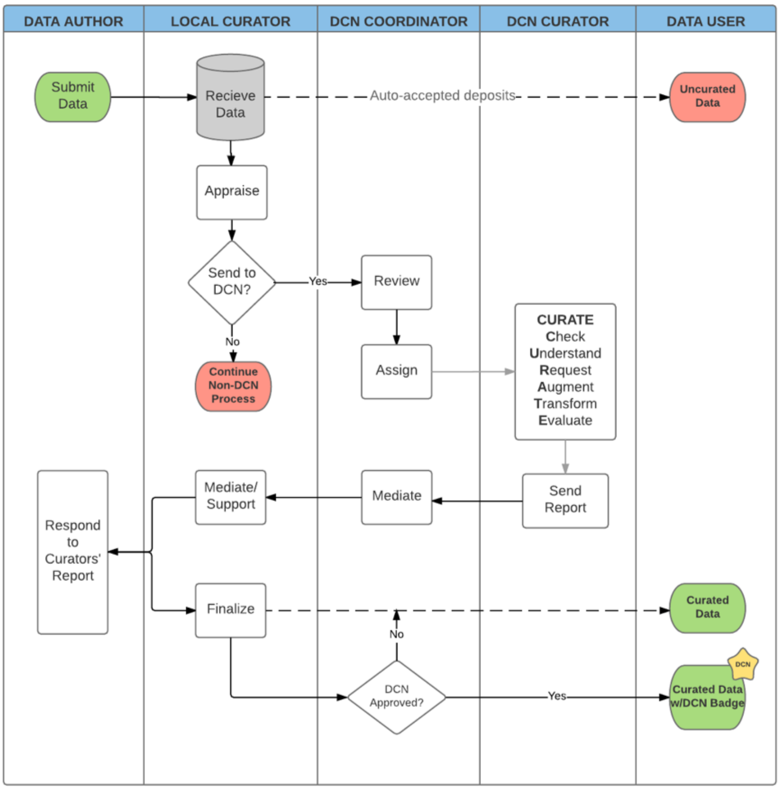

Figure 2. Role-specific actions in the DCN submission workflow.

The implementation phase of the DCN will track trends in the types of domains or file types that come to the Network and work to recruit new institutions that might fill any gaps in expertise support. Capacity for curating data in the Network will grow as new partners join. For example, we found from our 2016-2017 metric tracking that curators spend an average of two hours to curate a dataset (ranging from less than one hour to more than eight hours). In year three, if each institution contributes 10% of a DCN curator time (assuming 10% FTE = 16 hours/month) then with ten institutions the DCN will have roughly 160 curation hours or the capacity to curate an average of 80 data sets each month. Finally, the DCN will establish a public facing directory of datasets that were successfully curated by the Network. This web resource will be directional and link to the distributed and locally housed datasets providing a trusted source of well-curated data that are openly accessible.

Curation Procedures and Professional Development

DCN Curators will take standardized and file-type specific actions when reviewing the data for fitness for reuse using their expert skills and domain specific knowledge. Specifically, curators will be trained in minimum levels of curation for each data set that are summarized as DCN **C-U-R-A-T-E** steps (shown in Figure 1 and detailed in the Appendix), which stand for:

- **C** – **Check** data files and read documentation;
- **U** – **Understand** the data (try to), if not…;
- **R** – **Request** missing information or changes;
- **A** – **Augment** the submission with metadata for findability;
- **T** – **Transform** file formats for reuse and long-term preservation;
- **E** – **Evaluate** and rate the overall submission for FAIRness.

A hands-on training workshop will bring DCN Curators together annually to learn practical treatments for a variety of data formats and build peer relationships to ensure strong communications channels across the Network. Curators will be expected to contribute to a knowledge base of curation procedures and standards, as well as document their work (e.g., changes made to the data set in a provenance log) and generally complete data curation assignments in a timely fashion.

Financial Sustainability Plan

Our proposed model will allow the DCN to grow and sustain with controlled expansion into new service areas in the years to come. Following the implementation phase, the DCN will transition to a self-sustaining service where institutional and disciplinary partners contribute data curation staff and share the central operations costs.

The core partner institutions will share any central costs to allow the DCN to prevail beyond the grant-supported implementation phase as a fee-for-service model. Any financial support contributed by partner institutions (along with in-kind curator staff) will sustain a number of potential centralized services, including the hire of one full-time DCN Coordinator, hosting annual DCN Curator training events, and supporting administrative and technology services. Costs may be offset by potential revenue streams as fee-for-service users increase, and/or if the DCN becomes affiliated with a parent association to act as fiscal agent and cover some of the overhead burden.

The DCN planning phase team reviewed several governance documents of peer organizations, including the 2CUL project, arXiv, DataOne, HathiTrust, Portage, and the Texas Digital Library, in order to draft a Memorandum of Understanding for partner institutions. Our DCN draft MOU anticipates the need for a governance body that advises on any major issues encountered by the Network staff. However, details for the makeup and responsibilities of this governing board will be determined in the implementation phase of the DCN. An updated MOU will reflect any changes to the Network based on lessons learned from the implementation phase and will be used to normalize and sustain operations of the DCN moving forward.

Assessment Plan

Several key metrics will be used to track the impact and success of the Data Curation Network over time. From the start of the implementation phase our two-pronged assessment plan will measure:

1. *Scale*: The number of datasets curated by the Network, the frequency of submission (high-volume time periods, etc.), and the variety and types of data will be tracked in order to better understand the unique file formats and the range of disciplines that utilize DCN services. Plus, an important factor in our scale-based assessment will be to understand how a networked approach to curating research data is more efficient by tracking the time and costs involved at each stage of our curation workflow.

2. *Value-add*: The number of downloads, citations, alternative-metrics, and other use metrics for DCN-curated data sets will be gathered in order to assess whether curated data are more valuable. Our research and assessment of these trust markers for reuse will aid in understanding researcher attitudes toward the value of data curation generally.

Conclusion

Implementing the Data Curation Network will launch a valuable new service that will benefit researchers, their disciplines, and the end users of research data world-wide. The next phase starting in the spring of 2018 will bring together partners from US academic institutions (Minnesota, Cornell, Duke, Johns Hopkins, Illinois, Michigan, Penn State, and WashU) and the general-purpose Dryad Data Repository to pilot the DCN. Following the successful demonstration that a collaboratively-staffed network is more efficient and scalable and that data curated by the DCN are more valuable, our proposed curation-as-service model will allow the DCN to grow and sustain with controlled partner-driven expansion into new service areas in the years to come. Along the way the DCN will develop and openly share standards-driven data curation techniques, quantify the costs and measure the impact of data curation services, and provide essential training to a cohort of data curators. We release this model in the hopes that our vision may contribute to the discussion and implementation of collaborative networks even beyond the data curation topic for which it was designed.

Acknowledgements

The 'Planning the Data Curation Network' project was funded from 2016-2017 with the exclusive support of the Alfred P. Sloan Foundation. Project materials are openly available at the planning phase website.[10]

References

Beagrie, N., & Houghton J.W. (2014) *The value and impact of data sharing and curation: A synthesis of three recent studies of UK research data centres*. Retrieved from Jisc: http://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

Bloom, T., Dallmeier-Tiessen, S., Murphy, F., Austin, C.C., Whyte, A., Tedds, J., Nurnberger, A., Raymond, L., Stockhause, M., Vardigan, M., & Clarke, T. (2015). *Workflows for research data publishing: Models and key components*. International Journal on Digital Libraries-Research Data Publishing Special, 27. https://www.rd-alliance.org/system/files/Workflows_for_Research_Data_Publishing-_Models_and_Key_Components_submitted.pdf

Cronin, C., Laskowski, M.S., Mueller, E.K., & Snyder, B.E. (2017). Strength in numbers: Building a consortial cooperative cataloging partnership. *Library Resources and Technical Services*, *61*(2), 102-116. Retrieved from Knowledge@UChicago: http://hdl.handle.net/11417/305

Dunning, A., de Smaele, M., & Böhmer, J. (2017). Are the FAIR data principles fair?. Zenodo. doi:10.5281/zenodo.321423

Erway, R. (2012). Swatting the long tail of digital media: A call for collaboration. OCLC Research. Retrieved from http://www.oclc.org/research/publications/library/2012/2012-08.pdf

Fecher, B., Friesike, S., Hebing, M., & Linek, S. (2017). A reputation economy: How individual reward considerations trump systemic arguments for open access to data. *Palgrave Communications. 3*(17051). doi:10.1057/palcomms.2017.51

Hudson-Vitale, C., Imker, H., Johnston, L.R., Carlson, J., Kozlowski, W., Olendorf, R. K., & Stewart, C. (2017). SPEC Kit #354: Data curation. Association of Research Libraries (ARL). Retrieved from http://hdl.handle.net/11299/188643

Kindling, M., Pampel, H., van de Sandt, S., Rücknagel, J., Vierkant, P., Kloska, G., … Scholze, F. (2017). The landscape of research data repositories in 2015: A re3data Analysis. *D-Lib Magazine, 23*, 3-4. doi:10.1045/march2017-kindling

[10] Data Curation Network website: https://sites.google.com/site/datacurationnetwork

Kirchner, J., Diaz, J., Henry, G., Fliss, S., Culshaw, J., Gendron, H., & Cawthorne, J.E. (2015). The center of excellence model for information services. Retrieved from the Council on Library and Information Resources: http://www.clir.org/pubs/reports/pub163

Kouper, I., Fear, K., Ishida, M., Kollen, C., & Williams, S. (2017). Research data services maturity in academic libraries. In L.R. Johnston (Ed.), *Curating Research Data Volume One: Practical Strategies for Your Digital Repository*. Chicago: American Library Association, Association of College and Research Libraries. http://hdl.handle.net/10150/622168

Johnston, L.R. (2014). A workflow model for curating research data in the University of Minnesota libraries: Report from the 2013 Data Curation Pilot. University of Minnesota Digital Conservancy: http://hdl.handle.net/11299/162338

Johnston, L.R., Carlson, J., Hswe, P., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., & Stewart, C. (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. doi:10.7191/jeslib.2017.1102

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., & Stewart, C. (submitted). How important are data curation activities to researchers. *Journal of Librarianship and Scholarly Communication.*

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., & Stewart, C. (2017a). Results of the Fall 2016 Researcher Engagement Sessions. Retrieved from http://hdl.handle.net/11299/188641

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., & Stewart, C. (2017b). Results of the Fall 2016 Data Curation Pilot. Retrieved from http://hdl.handle.net/11299/188640

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R.K., & Stewart, C. (2017c). Data curation network: A cross-institutional staffing model for curating research data. Retrieved from https://sites.google.com/site/datacurationnetwork/results

Lee, D.J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PloS one, 12*(3), e0173987. doi:10.1371/journal.pone.0173987

McNutt, M., Lehnert, K., Hanson, B., Nosek, B.A., Ellison, A.M., & King, J.L. (2016). Liberating field science samples and data. *Science, 351*(6277), 1024-1026. doi:10.1126/science.aad7048

Munafò, M.R., Nosek, B.A., Bishop, D.V., Button, K.S., Chambers, C.D., du Sert, N.P., ... Ioannidis, J.P. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. doi:10.1038/s41562-016-0021

Roche, D.G., Kruuk, L.E., Lanfear, R., & Binning, S.A. (2015). Public data archiving in ecology and evolution: How well are we doing?. *PLoS Biol, 13*(11), e1002295. doi:10.1371/journal.pbio.1002295

Soehner, C., Steeves, C., & Ward, J. (2010). E-science and data support services: A study of ARL member institutions. Association of Research Libraries. Retrieved from http://www.arl.org/storage/documents/publications/escience-report-2010.pdf

Smith, R., & Roberts, I. (2016). Time for sharing data to become routine: The seven excuses for not doing so are all invalid. *F1000Research, 5*. doi:10.12688/f1000research.8422.1

Stodden, V., Guo, P., & Ma, Z. (2012, September). How journals are adopting open data and code policies. In *The First Global Thematic IASC Conference on the Knowledge Commons: Governing Pooled Knowledge Resources.* Retrieved from http://hdl.handle.net/10535/9584

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PloS one, 6*(6), e21101. doi:10.1371/journal.pone.0021101

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., ... Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One, 10*(8), e0134826. doi:10.1371/journal.pone.0134826

Weber, D. (1976). A century of cooperative programs among academic libraries. *College & Research Libraries, 37*(3), 205-221. doi:10.5860/crl_37_03_205

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., ... Bouwman, J. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*. doi:10.1038/sdata.2016.18

York, J., Gutmann, M., & Berman, F. (2016). What do we know about the stewardship gap? University of Michigan Deep Blue. Retrieved from http://hdl.handle.net/2027.42/122726

Appendix

Table 2. Draft checklist of DCN CURATE steps and FAIRness scorecard.

CURATE Actions	Curation Checklist
Check data files and read documentation • Review the content of the data files (e.g., open and run the files or code). • Verify all metadata provided by the author and review the available documentation.	□ Files open as expected □ Issues ___________ □ Code runs as expected □ Produces minor errors □ Does not run and/or produces many errors □ Metadata quality is rich, accurate, and complete □ Metadata has issues __________ □ Documentation Type (*circle*) Readme / Codebook / Data Dictionary / Other: ________________________ □ Missing/None □ Needs work
Understand the data (or try to) • Check for quality assurance and usability issues such as missing data, ambiguous headings, code execution failures, and data presentation concerns. • Try to detect and extract any "hidden documentation" inherent to the data files that may facilitate reuse. • Determine if the documentation of the data is sufficient for a user with similar qualifications to the author's to understand and reuse the data. If not, recommend or create additional documentation (e.g., a readme.txt template).	*Varies based on file formats and subject domain. For example….* Tabular Data Questions (e.g., Microsoft Excel) □ Organization of data well-structured □ Not rectangular □ Split tables into separate tabs □ Headers/codes clearly defined □ Define headers □ Clarify codes used ________ □ Clarify use of "blanks" □ Clarify units of measurement □ Quality control clearly defined □ Unclear quality control □ Update/add Methodology
Request missing information or changes • Generate a list of questions for the data author to fix any errors or issues.	*Narrative describing the concerns, issues, and needed improvements to the data submission*
Augment the submission • Enhance metadata to best facilitate	□ Discoverability sufficient □ Recommend (circle one) full-text index / file compression / file

discoverability. • Create and apply metadata for the data record, including descriptive keywords. • When appropriate, structure and present metadata in domain-specific schemas to facilitate interoperability with other systems.	reorder / file descriptions / zip Other ______________ □ Keywords Sufficient □ Suggestions _______________ □ Linkages Sufficient □ Link to Report/Paper □ Link to related data sets □ Link to source data □ Link to other ____________

Transform file formats
• Identify specialized file formats and their restrictions (e.g., Is the software freely available? Link to it or archive it alongside the data).
• Transform files into open, non-proprietary file formats[11] that broaden the potential audience for reuse and ensure that preservation actions might be taken by the repository in later steps. Retain original files if data transfer is not perfect.
	□ Preferred file formats in use
	  □ Recommend conversion from ___________ to ____________
	  □ Retain original formats
	□ Software needed readily available
	  □ Unclear version of software
	  □ Unclear software used
	□ Visualization of data easily accessible
	  □ Recommend graphical representation ____________
	  □ Recommend web-accessible surrogate ________________

Evaluate and rate the overall data record for FAIRness.[12] • Score the dataset and recommend ways to increase the FAIRness of the data and become "DCN approved."	Findable - □ Metadata exceeds author/ title/ date, □ Unique PID (DOI, Handle, PURL, etc.). □ Discoverable via web search engines like Google. Accessible - □ Retrievable via a standard protocol (e.g., HTTP). □ Free, open (e.g., download link). Interoperable - □ Metadata formatted in a standard schema (e.g., Dublin Core). □ Metadata provided in machine-readable format (OAI feed). Reusable - □ Data include sufficient metadata about the data characteristics to reuse without the direct assistance of the author. □ Clear indicators of who created, owns, and stewards the data. □ Data are released with clear data usage terms (e.g., a CC License).

[11] Format Recommendations, http://guides.library.cornell.edu/ecommons/formats

[12] Rubric evaluating the FAIR principles are based on the scoring matrix by Dunning, de Smaele and Böhmer (2017).