

IJDC | Conference Pre-print

Extending the Research Data Toolkit: Data Curation Primers

Cynthia Hudson Vitale
Pennsylvania State University

Hannah Hadley
Pennsylvania State University

Jennifer Moore
Washington University in St. Louis

Lisa Johnston
University of Minnesota

Wendy Kozlowski
Cornell University

Jake Carlson
University of Michigan

Mara Blake
Johns Hopkins University

Joel Herndon
Duke University

Abstract

Niche and proprietary data formats used in cutting-edge research and technology have specific curation considerations and challenges. The increased demand for subject liaisons, library archivists, and digital curators to curate this variety of data types created locally at an institution or organization poses difficulties. Subject liaisons possess discipline knowledge and expertise for a given domain or discipline and digital curation experts know how to properly steward data assets generally. Yet, a gap often exists between the expertise available within the organization and local curation needs. While many institutions and organizations have expertise in certain domains and areas, oftentimes the heterogeneous data types received for deposit extend beyond this expertise. Additionally, evolving research methods and new, cutting-edge technology used in research often result in unfamiliar and niche data formats received for deposit. Knowing how to 'get-started' in curating these file types and formats can be a particular challenge. To address this need, the data curation community have been developing a new set of tools - data curation primers. These primers are evolving documents that detail a specific subject, disciplinary area or curation task, and that can be used as a reference or jump-start to curating research data. This paper will provide background on the data curation primers and their content detail the process of their development, highlight the data curation primers published to date, emphasize how curators can incorporate these resources into workflows, and show curators how they can get involved and share their own expertise.

Submitted 22 December 2019 ~ Accepted 19 February 2020

Correspondence should be addressed to Cynthia Hudson Vitale, W312 Pattee Library, Pennsylvania State University, University Park, PA 16802. Email: cuv185@psu.edu

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020

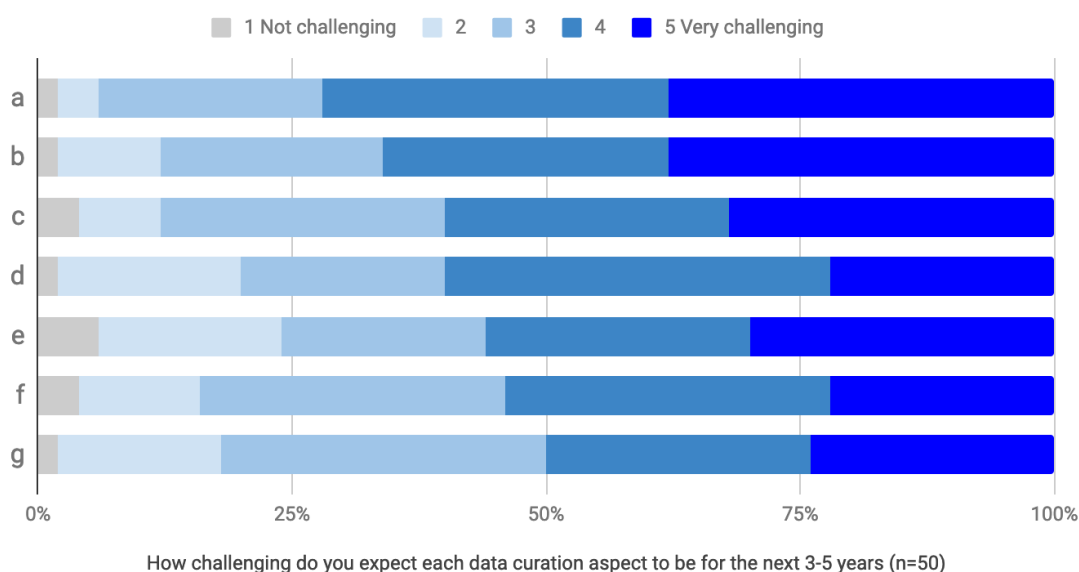
The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256.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

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



Introduction

Results from SPEC Kit #354 (Hudson-Vitale, et. al., 2017) indicate that expertise in domain data, scaling for curation demand, and library staffing for data curation are the top three most significant challenges facing academic libraries with regards to curating data (see Figure 1). Many institutions rely on the equivalent of less than one full time staff member to provide research data management and curation support. This indicates that while subject or format expertise may exist within an institution via subject liaisons, repository staffing or others, much of that expertise may not currently be leveraged for data curation purposes. Despite this lack of resources, data curation services have been repeatedly identified in reports and publications as a rapidly growing and critical service area for the transformation of academic libraries and necessary for research data reuse, transparency, and reproducibility (Hedstrom, 2015). Additionally, research indicates that faculty and researchers require a variety of curation treatments for their data, but often lack the support needed to satisfactorily complete the treatments (Johnston et. al., 2018).



- Expertise in domain data
- Scaling with increased demand
- Training & retooling existing staff
- Outreach/marketing
- Recruiting & retaining staff
- Keeping up with technology changes
- Keeping up with data sharing requirement changes

Figure 1. Results from ARL SPEC Kit #354 indicating greatest challenges to data curation.

A study by the Data Curation Network (DCN) ¹demonstrated that both subject and functional expertise are needed to richly curate data that are being deposited in institutional or data repositories (Johnston, 2017). During the DCN planning phase (May 2016-June 2017) six institutions observed 24 different file types from 176 datasets (see Table 1). These file types also spanned over 76 different disciplines, the most frequent being Anthropology (n=15), Crop Sciences (n=12), Civil, Environmental and Geo-Engineering (n=9), and Oceanography (n=7).

Table 1. DCN reported file types sorted by most to least frequent (n=176)

Primary File Type	Count	Primary File Type	Count	Primary File Type	Count
Tabular	51	Code Matlab	5	Database SQL	1
Sequence	25	Video	5	Database	1
Geospatial	19	Textual	4	Plots	1
Image 3D	16	Survey SAS	3	Autocad	1
Code R	9	Simulation	3	Image	1
Code Python	7	Database Access	3	Code C++	1
Code Java	6	Survey SPSS	2		
Audio	6	Spectra	2		
Image Biological	5	JSON	2		

While tabular data has widespread use, its curation is made more difficult depending upon the subject matter. For example, one of the tabular datasets received during this period was comprised of instrument readings of voltage-clamp fluorometry of heart valves. While the files are easy to open and view, without some type of discipline knowledge it is difficult to know if they are complete and well described, which complicates the curation. On the flip side, one of the datasets reflected in the numbers above is comprised of metadata about international roll call votes in a SQL database. While this type of metadata is relatively easy to understand, without knowledge of how to curate or understand a SQL file, they are inaccessible for curation. While extensive training and experience in SQL or voltage-clamp fluorometry is not required to richly curate research data, a baseline knowledge of how to open, check, and inspect the files is extremely helpful.

The data curation primers were developed as jump-start guides to assist curators in quickly getting up to speed on how to appropriately curate various data types and formats within the context of specific disciplines. These primers were initially developed through the Data Curation Network but have expanded and are seeing community contributions worldwide.

¹ <https://datacurationnetwork.org/>

Data Curation Network Background

The DCN project launched in 2016 in response to a common challenge for data repositories: we ingest dataset from a vast range of domains (from neuroscience to entomology) and in all conceivable file types/formats (such as Python files or 3D image formats). With such complexity, we cannot individually hire the expertise needed to properly ensure well curated data.

Therefore, with funding from the Alfred P. Sloan Foundation, the DCN team set out to create a cross-institutional network that would enable researchers to share their data in ethical and reusable ways, regardless of their final repository destination.

Our planning phase began with a deep comparison of the repository services across the (then) six institutions (Cite 2016 JESLIB). Each of our repository platforms are unique. We use a variety of software (including DSpace, Samvera, Fedora, Dataverse) and we have separate policies and ingest workflows. This is why we modeled the DCN as a seamless “human layer” in the local repository stack. Each partner institution receives their own data and archives it locally. They decide what to accept and if/when to send data to the DCN for curation. They also provide repository services of storage, DOI minting, and long term preservation.

The DCN transitioned from theory to operation in January 2019. In the implementation phase, we have grown to include 10 partners led by the University of Minnesota, with participation by Cornell University, Duke University, the Dryad Data Repository, Johns Hopkins University, New York University, Pennsylvania State University, University of Illinois, University of Michigan, and Washington University in St Louis. Since launching we collectively curated over 70 data sets, representing a small piece of the overall data archived across the partner data repository services (see Figure 2).

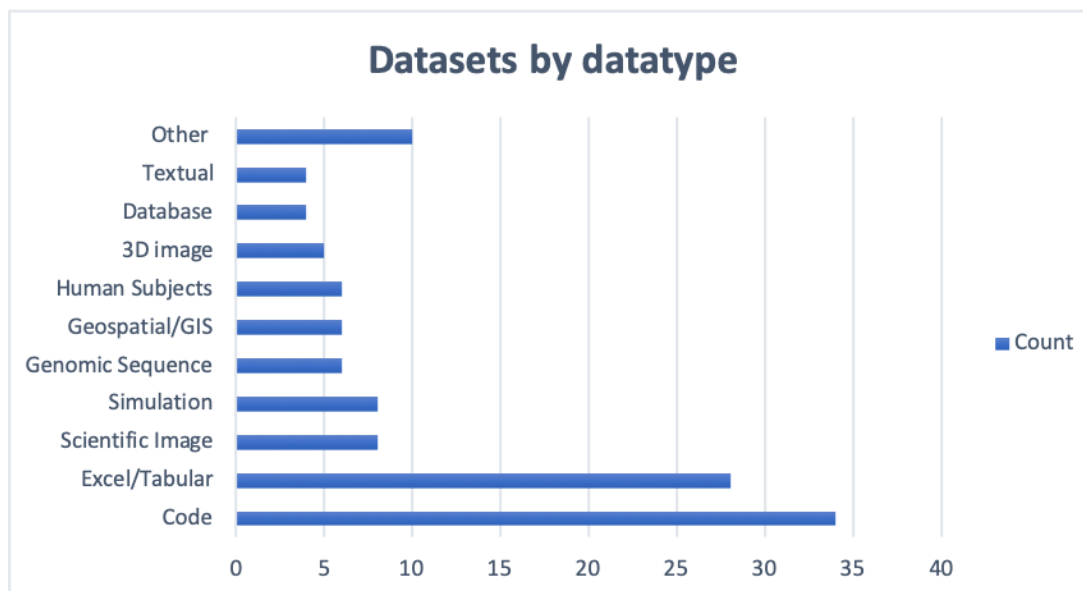


Figure 2. Number of datasets curated by the DCN since January 2019 by datatype.

Table 2: Researcher data curation activities

Data Curation Treatment	Definitions
Documentation	Information describing any necessary information to use and understand the data. Documentation may be structured (e.g., a code book) or unstructured (e.g., a plain text “Readme” file).
Chain of custody	Intentional recording of provenance metadata of the files (e.g., metadata about who created the file, when it was last edited, etc.) in order to preserve file authenticity when data are transferred to third-parties.
Quality Assurance	Ensure that all documentation and metadata are comprehensive and complete. Example actions might include: open and run the data files; inspect the contents in order to validate, clean, and/or enhance data for future use; look for missing documentation about codes used, the significance of “null” and “blank” values, or unclear acronyms.
Persistent Identifier	A URL (or Uniform Resource Locator) that is monitored by an authority to ensure a stable web location for consistent citation and long-term discoverability. Provides redirection when necessary. E.g., a Digital Object Identifier or DOI.
Discovery Services	Services that incorporate machine-based search and retrieval functionality that help users identify what data exist, where the data are located, and how can they be accessed (e.g., full-text indexing or web optimization).
Technology Monitoring and Refresh	Formal, periodic review and assessment to ensure responsiveness to technological developments and evolving requirements of the digital infrastructure and hardware storing the data.
File Audit	Periodic review of the digital integrity of the data files and taking action when needed to protect data from digital erosion (e.g., bitrot) and/or hardware failure.
Metadata	Information about a data set that is structured (often in machine-readable format) for purposes of search and retrieval. Metadata elements may include basic information (e.g. title, author, date created, etc.) and/or specific elements inherent to datasets (e.g., spatial coverage, time periods).
Versioning	Provide mechanisms to ingest new versions of the data overtime that includes metadata describing the version history and any changes made for each version.
Contextualize	Use metadata to link the data set to related publications, dissertations, and/or projects that provide added context to how the data were generated and why.
File Format Transformations	Transform files into open, non-proprietary file formats that broaden the potential for long-term reuse and ensure that additional preservation actions might be taken in the future. Note: Retention of the original file formats may be necessary if data transfer is not perfect.

CURATE Steps

When a dataset is submitted by a curator at a partner institution to the network, our curators make recommendations for the data author by applying a standard set of procedures, called CURATE(D) steps, to ensure that minimal level of curation is taken for every dataset (Data Curation Network, 2018). This conceptual workflow is the basis of our training (both for DCN curators and our outward-facing workshops for the community). Each step has its own checklist accompanying the step to represent actions that any dataset might need, such as auditing the contents of a dataset and verifying the accuracy of all metadata provided. The steps represent a baseline set of actions to take for any given dataset, but the specific approach may vary. The CURATE(D) steps were based upon the results of previous research (Johnston et al., 2018), in which faculty and researchers identified the most important curation activities for their work and the treatments for which they were least satisfied. This research indicated a number of high priority areas for data curation treatments (see Table 2).

The CURATE(D) steps represent a framework for how we standardize our work and provide a useful tool for our training (see Figure 3). But, again, due to the diversity of data we curate, they are not a blueprint or procedure. They are prompts, or a starting point for a professional curator with skills and expertise to scaffold their unique knowledge about a data set, its contextual domain and the file container, and aspire to a minimum level of curation.

C heck files
U nderstand or try to
R equest missing information
A ugment the submission
T ransform the format
E valuate for FAIRness
D ocument throughout

Figure 3. CURATE(D) steps used for curating data through the DCN and to teach curation practices.

DCN Education

The DCN supports workforce development and training for the community of professionals emerging around data curation. To this end, all member curators of the DCN receive an annual training developed around the CURATE(D) steps described above to provide them access to in-depth, hands-on exposure to data curation education. To extend access to this training beyond the members of the DCN, PI Hudson Vitale received an Institute of Museum and Library Services (IMLS) grant #RE-85-18-0040-18 to offer Specialized Data Curation workshops to other professionals in need of skills on the topic.

This grant allowed the provision of three two-day workshops across two years, reaching more than 150 participants. The participants consisted of librarians and data professionals from a variety of academic, government, and cultural heritage institutions. In advance of the IMLS grant, a pilot of the workshop was offered in partnership with the International Association for Social Science Services and Technology (IASSIST) in December 2018. Outside of the IMLS grant, two DCN instructors recently held an abbreviated workshop on the CURATE steps for the Portage network in Canada in November 2019.

The Specialized Data Curations Workshops have the following learning objectives:

1. Increase understanding of data curation practices and tools in various disciplines, data types, and formats.
2. Share expertise and enhance curation capacity for curation nationwide.
3. Meet like-minded colleagues who are interested in building and extending curation practices.

Featuring a peer-to-peer learning model, the structure of the workshop takes participants through alternating lecture by DCN instructors and hands activities of curation. Participants self-select into groups based on data types: tabular, code, geospatial, image and survey. The workshop facilitators provide test datasets from their own repositories for participants to work with. After receiving an overview of a particular CURATE(D) step from one of the DCN instructors, the participants work in their small groups to apply the step to the provided data set. The workshop also features lightning presentations on particular topics from DCN curators, as well as less formal time for networking and building connections between participants. In parallel, participants also divide into groups to create primers, further described below.

Data curation primers

As demonstrated in the ARL Data Curation Spec Kit 354 (Hudson-Vitale, 2017), many institutions struggled to establish diverse expertise which covers all functional and domain data types. A separate study which internationally assessed libraries' data services in 2014 and 2018 found data curation skills were among the most commonly cited in gap analyses (Cox 2019). This type of research is foundational to the emergence and development of the DCN and the educational initiative. While building the network was an important step in leveraging each others' expertise in the act of curation and education, there was also a need for more practical, curation-focused, network-informed, shared resources.

The data curation primers were conceived as practical capstone projects that expanded the in-person workshop. The goal was to increase the learning for attendees (and beyond) by facilitating each cohort in creating some kind of deliverable together. The first primer was designed by DCN curator Susan Borda and graduate student Sam Sciolla at the University of Michigan. The team interviewed faculty and came up with the major components that became our primer template we use today (Sciolla and Borda, 2018). Ultimately, primer capstones have transformed beyond the sharing of expertise of workshop attendees and have become a powerful community platform for publishing curation knowledge.

Primer Development Process

Primer creation begins at the specialized data curation workshops. Our curriculum provides an outline of data curation workflow based on the Data Curation Network's CURATE(D) steps. This method provides a common language and definitions for curation actions that may be termed or organized differently across institutions. Potential primer topics are brainstormed by the attendees on the first day of the workshop and groups are self-selected on the second day. We assign a mentor from the Data Curation Network to assist with each group. Each primer group has approximately six months to work on their project.

Each primer is created with oversight from our organization and the growing community of primer authorship. Mentors assigned from the Data Curation Network assist with organizing and supporting the growth of content for each primer. We also hold a peer review process in the fourth month that is refereed by the project manager. Peer reviewers may be from the Data Curation Network or attendees of any of the three specialized data curation workshops. Every primer must have a reviewer. Each review is carefully considered for the revision of the final primer draft.

Completed drafts are published to Github as “living” data curation primers. These primers are expected to grow with updates and new content added through pull requests by the primer authors or anyone interested in adding their expertise to the project. Pull requests are reviewed by the primer authors or members of the Data Curation Network. A contributors guide was created to help users navigate our process. The completed primer drafts are additionally published to the University of Minnesota’s institutional repository ²as static (version 1.0) copies that serve to archive this IMLS funded portion of the project.

Primer components

Primers are intended to be a resource for data curators about specific software or disciplinary considerations that may be encountered during the process of reviewing or curating a dataset. We acknowledge that different subjects will require different information and level of detail to be useful, but have laid out a suggested set of topics for authors to address in each document to guide the process. Authors are encouraged to follow the general outline, but adapt as appropriate for their topic. Suggested primer components are described in detail below. The prototype for the data curation primers, a netCDF profile, was developed at the University of Michigan (U-M) in the summer of 2018 and was published in U-M’s institutional repository, Deep Blue.

Overview

Each primer begins with a table that provides a basic overview of the topic in table structure. It provides the reader with a concise summary of relevant information, including highlights of key curatorial considerations as well as basic document provenance. Authors are encouraged to use only the table sections that are relevant to their topic, and add additional overview components, as appropriate. Suggested contents for the overview section include: file extension(s), MIME type(s), version(s), primary field(s) or area(s) of use, source and affiliation, metadata standards, key questions for curatorial review, tools for curatorial review, document creation date, document authors, and a record of changes made to the primer (see Table 3). Some primers include a table of contents following the overview table.

Description of format

Because software and even disciplinary datasets often contain multiple formats, this section details both the structure and logic behind possible formats curators are likely to encounter. If there are significant differences in versions of a format, or implications of running software on a specific operating system that would impact curatorial actions, that can be included in this section as well. For example, use of Open XML format options, available in Office 2017 and newer products, reduces the urgency of transformation to a non-proprietary format.

Examples and sample dataset citations

This section is an opportunity to provide citations for, and links to, example datasets that have content in the format described in the primer. Primers might also include in this section use cases where the format is likely to be employed. This gives curators a quick library of samples that illustrate the format in actual use.

Key questions

Like doing a reference interview, curating data is most effective when you have some understanding of the context around the dataset. This section (or sections, if broken down into separate curator / data provider sections) is meant to get the curator thinking about the critical elements of the format, and key factors that will impact curatorial actions. This can manifest as a

² <http://hdl.handle.net/11299/202810>

Table 3. Example data curation primer Overview section for Microsoft Access

Topic	Description
File Extension	mdb, accdb
MIME Type	None / application/octet-stream
Structure	Proprietary relational database
Versions	Access 1.0 - 2003, 2007 - 2016
Primary fields or areas of use	Storing relational data. Used in many disciplines. Surveys, data entry, geospatial, storing metadata for certain disciplines, etc.
Source and affiliation	Microsoft
Metadata standards	No widely adopted standards. Built-in documentation may be used
Key questions for curation review	Intended use: use as a db or just be able to view the data? Does the data exist in other formats? Are there “linked” tables? Is there documentation about each table and relationship?
Tools for curation review	A Windows PC with Access (Access does not run on Mac). Recent versions of Access (e.g., 2016) should be able to open down to Access 2000
Date Created	2019-01-18
Created by	Fernando Rios (University of Arizona Libraries, Office of Digital Innovation & Stewardship), Dave Fearon (Johns Hopkins University Libraries Data Services)
Date updated and summary of changes made	2019-02-01 - 1.0: initial version 2019-04-18 - 2.0: DCN peer review

checklist of attributes to evaluate and act on immediately, or it may provide prompts for the curator to explore together with the data provider. Some questions will be broad, for example, asking the curator to consider the audience and impact thereof; others may be very specific, yet critical to long term preservation of the content (e.g. evaluating file naming integrity and organizational structure of related files) (see Table 4)

Applicable metadata standard, core elements and readme requirements

This section is where dataset documentation specifics can be laid out. In some cases, formalized standards may be appropriate; elements of those standards that are key to

understanding or reuse of the data can be identified. In cases where no formal standards are available or typically employed, key elements to include in alternate documentation (eg. a readme file, a data dictionary or a codebook) should be described.

Table 4. Example Microsoft Access key questions for curators and researchers

Key questions for curators	Key clarifications to get from researcher
<ul style="list-style-type: none"> • Is documentation about the columns included? I.e., a codebook. The codebook is especially critical to understand the relationships between the tables • Are the primary/foreign key relationships described? Exporting individual tables to a spreadsheet format does not keep these relationships so it's critical to document them (at least include an exported version of the Relationships diagram alongside the exported tables.) 	<ul style="list-style-type: none"> • For data entry databases with forms, a description of the data entry workflow would help with understanding the underlying database structure <ul style="list-style-type: none"> ○ What is the main entry form? Where does the entry workflow start? Etc. • Documentation of the meaning of the relationships in the database, not just which relationships exist (the latter can simply be obtained via the Relationship Diagram which is auto-generated by Access) • Is the data available natively in a non-Access format elsewhere?

Software for reviewing, viewing or analyzing data

Many data file formats can be opened in multiple software packages. Some may allow limited functionality (viewing only, for example), and others may allow full interaction, including re-processing, analyzing or creation of key visualizations, models, or other outputs. A well-documented primer identifies the most popular software used to access and evaluate the files submitted for curation. Ideally, the primer will guide the curator to software that is readily available and allows the curator to examine files without modifying the data submitted for curation.

Preservation actions

Data preservation remains one of the overarching goals of any curation process. Ideally, the primer will try to provide guidance on preservation concerns associated with certain file formats, recommendations on data transformations that will extend access to the underlying data, and advice on these transformations. Some primers may provide additional guidance on strategies such as adding documentation, file treatments, and/or conversions that can help ensure the long-term preservation of the data.

What to look for to make sure this file meets FAIR principles

Overall, the primer should help guide a curator in evaluating and processing data deposit against the FAIR Data Principles (Wilkinson, 2016) and suggesting possible treatments that help make the data more findable, accessible, interoperable, and reusable. The FAIR guidelines are particularly useful in shaping primer recommendations about the appropriate amount of documentation to improve discovery and accessibility. The FAIR guidelines also offer a helpful

perspective for primer authors to consider file treatments and organization that can improve interoperability and reusability of the formats in question.

Ways in which fields may use this format

While the content of the data curation primers will vary by format or research method, all primers should include a section that describes the research context for material covered. This section can include information about the disciplines, fields of research, and communities where the format/method is used. Some primers may use this section to provide links to disciplinary resources and examples that would help a curator better understand the context for a given format or research method.

Documentation of curation process: What to capture from curation process

When curating a data collection submitted for publication, curators have a responsibility to document their curation workflow noting both their observations and the treatments on the submitted files. If the format or method in primer requires special consideration for this documentation process, the primer should endeavor to note particular steps or treatments that should be documented in the curation process. By noting these areas, the primer ensures that the curator can adequately document their curation process so that secondary data users can understand any processing or transformations in the dataset (see Table 5)

Table 5. Example Microsoft Access documentation section

Documentation of curation process: What to capture from curation process
<ul style="list-style-type: none"> • Exporting tables to CSV: did curator do it or were they provided by the researcher? • Documenting the presence of linked tables or queries that reference external servers • Version of MS Access needed to open file • Any changes to the internal metadata. E.g., if the curator added descriptions to fields or modules using MS Access built-in functionality. This would affect what is produced by the Database Documenter. • Document relationship among any files or parts exported from the database, e.g., individual tables and the keyfields that relate them, spatial data, reports. Document whether the curator or the researcher generated them. • Stating whether the curator generated the Database Documenter report vs being generated by the researcher. This implies differences in the quality of the documentation - especially if the curator-generated report was not vetted by the researcher. (A researcher-generated report does not necessarily imply quality either.)

Future Primer Development

Primers have evolved from an original template that was focused on curation for a single software format into a variety of curatorial functions, disciplines, and research methods. For example, there are primer creation groups working on topics in curation, such as: human subjects, oral histories and neuroimaging data. These shifts from the purpose outlined in the original template have modified the way primers are organized and formatted, and this

illustrates how primers can be expected to develop with changing needs and participation of the community.

Community participation is necessary for these resources to thrive. Our group has made an effort to grow a community of collaborative authorship from the specialized data curation workshop participants. We have engaged primer authors by assigning mentorship from professionals within the Data Curation Network to assist with their projects. Additionally, webinars that showcase each primer groups' efforts to create these resources are provided. We also encourage others to participate by adding their expertise to existing data curation primers on GitHub.

Peer review is an essential aspect of primer creation expected to continue beyond the original project. Our review process examines each submission for accuracy and currency of the content, usefulness of the resource, and keeping to a defined scope. Professionals who have attended our workshops or those who are members of the Data Curation Network may currently participate in the peer reviews. Primer authorship is also presently limited to these same groups. However, both activities may eventually include others who are interested in joining this opportunity for collaboration of expertise in creating resources for data curators.

Other ideas are presently being discussed that may affect the future of data curation primers. One idea that is being significantly explored is a researcher-focused primer. This type of data curation primer would help an individual who wants to reuse data found on a data repository know how to open the file, how to evaluate it for completeness, and otherwise evaluate the quality of the documentation.

Primers may also inspire new learning experiences. We have discussed their potential as training resources, and primer content may help to shape new curriculum ideas. We have a glimpse of what is important to our communities through the primer topic selections and the experiences of the group members in investigating these topics. This information may provide insight for better serving our research communities.

Conclusion

The creation of the primers as a community resource and a mechanism for community development has been a success. Assessment data from our educational workshops indicates that creating primers, contributing expertise, and building community data curation resources is what attendees are most excited about taking away from the training. This excitement can be felt throughout the six months of the primer development and into the final share-out webinar in which each group presents their primer to the community.

To date, eighteen primers have been published through this process and program and an additional eight are currently in development. A list of the primer topics is below.

- Geodatabases
- Microsoft Excel
- Jupyter Notebooks
- Microsoft Access
- netCDF files
- Wordpress
- SPSS
- Atlas.ti
- Confocal microscopy

- GeoJSON
- Google Docs
- Lidar Point Clouds
- NVivo
- PDF
- R
- STL files
- Tableau
- Text/character encoding
- Shape files
- ISO Images
- GEOTiff
- NVIVO
- Oral Histories
- SAS
- DarwinCore
- NIFTI/BIDs

Future work on the primers will focus on: 1) assessing how they are used within the curation workflow, and 2) evaluating the usefulness of the primers for curation. Additionally, we are exploring mechanisms and workflows for curators who have not attended our trainings to publish a primer while still upholding our quality measures. Workflows for this broader community contribution will be developed within the next few months.

Finally, we are exploring where these primers will have the biggest impact and be most useful for community discoverability. Existing preservation focused API's such as PRESQT and PRONOM are being explored for possible integration and linking. All in all the data curation community have indicated their continued interest and support of these community resources.

References

- Cox, Andrew M., Dr, Mary Anne Kennan, Dr, Elizabeth Josephine Lyon, Dr, Stephen Pinfield, Dr, Laura Sbaffi, and Dr. (2019). "Progress in Research Data Services." *International Journal of Digital Curation* 14 (September): 126–35. <https://doi.org/10.2218/ijdc.v14i1.595>
- DCN Hosts Data Curation Workshop. <https://datacurationnetwork.org/2018/01/31/dcn-hosts-data-curation-workshop/>

Data Curation Network (2019). Data Curation Network Primers. Retrieved from University of Minnesota Digital Conservancy. <http://hdl.handle.net/11299/202810>

Hedstrom, M., et. al. (2015). "Preparing the Workforce for Digital Curation". The National Academies Press. Accessed December 22, 2016 from:
<http://www.nap.edu/catalog/18590/preparing-the-workforce-for-digital-curation>

Hudson-Vitale, Cynthia, Heidi Imker, Lisa R. Johnston, Jake Carlson, Wendy Kozlowski, Robert Olendorf, and Claire Stewart. Data Curation. SPEC Kit 354. Washington, DC: Association of Research Libraries, May 2017.

Johnston, L.R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R. and Stewart, C., 2018. How Important is Data Curation? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6(1), p.eP2198. DOI:
<http://doi.org/10.7710/2162-3309.2198>

Johnston, Lisa R., Jake R. Carlson, Patricia Hswe, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert K. Olendorf, and Claire 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_

Sciolla, S. and Borda, S. (2018) Data Curation Format Profile: netCDF. Deep Blue. University of Michigan <https://deepblue.lib.umich.edu/handle/2027.42/145724>

Update from the April 2019 Data Curation Workshop.
<https://datacurationnetwork.org/2019/08/07/jhu-workshop-an-update/>

Update from the November 2019 Specialized Data Curation Workshop.
<https://datacurationnetwork.org/2019/11/26/update-from-the-november-2019-specialized-data-curation-workshop/>