



EScience in Action

**Data Management and Curation for Qualitative
Research: Collaborative Curriculum Development
and Implementation**

Kendall Roark

Purdue University, West Lafayette, IN, USA

Abstract

Objective: This eScience in Action article describes the collaborative development process and outputs for a qualitative data curation curriculum initiative led by a library faculty (research data specialist) at an R1 research university.

Methods: The collaborative curriculum development activities described in this article took place between 2015-2020 and included 1) a college-wide “call out” meeting with graduate methods instructors and additional one-on-one conversations, 2) a year-long training series for disciplinary faculty teaching graduate-level qualitative research methods courses, 3) guest lectures and co-curricular workshops, and 4) the development of a credit-bearing graduate-level course.

Correspondence: Kendall Roark: roark6@purdue.edu

Received: February 29, 2020 **Accepted:** May 28, 2020 **Published:** October 9, 2020

Copyright: © 2020 Roark. This is an open access article licensed under the terms of the [Creative Commons Attribution-Share Alike License](https://creativecommons.org/licenses/by-sa/4.0/).

Data Availability: All non-identifying data referenced included in tables.

Disclosures: The author reports no conflict of interest.

Abstract Continued

Results: This practice-based article includes a reflection on the collaborative curriculum development process and impacts, including the development of networks between the Library and qualitative researchers across campus. The article provides a proof-of-concept example for developing relevant and trustworthy library data services for humanities and qualitative social-science researchers.

Conclusions: Curriculum development activities focused predominately upon researcher-centered perspectives and identified needs. However, changes in institutional expectations for library faculty (i.e. requirement to teach credit-bearing courses) played a major role in how the curriculum was implemented, its impact and continued sustainability of outputs going forward.

Introduction

This article describes the collaborative development process for a qualitative data curation curriculum initiative led by a library faculty research data specialist (author) at an R1 research university. The goals of the initiative were to 1) build networks and collaboration opportunities between the Libraries and interdisciplinary humanities and social science researchers across campus, 2) provide professional development opportunities for disciplinary faculty engaged in qualitative research and graduate-level qualitative methods education, 3) collaboratively develop and implement data management and curation curriculum and opportunities on campus, and 4) develop a formal data management and curation curriculum within existing and/or new credit-bearing course options for graduate students.

Collaborative Curriculum Development Process

The data curation curriculum development and implementation process is described in four overlapping parts or “legs” of the initiative. The first leg of the initiative began with discussions between the author and the Associate Dean Research and Graduate Education for the College of Liberal Arts (CLA) in 2016. This discussion led to a Fall 2016 Data Management Call Out for all CLA faculty teaching graduate research methods courses. Call Outs are common at this particular campus and are used to gauge faculty and student interest in new research, learning and engagement opportunities.

The Call Out began with both the author and the organizer of the Library’s graduate research instruction project describing current library data management and curation services and training opportunities. An open discussion period followed where the nine CLA faculty present were asked to share data management and curation topics they felt should be prioritized and the extent to which they currently incorporated these topics within existing graduate-level research methods courses. Faculty then brainstormed about potential gaps in graduate methods education surrounding data management and curation and identified preferred next steps in developing curriculum and training opportunities. Table 1 provides a summary of the needs, curricular gaps and next steps suggested by CLA faculty who attended the Call Out.

Data Management & Graduate Methods Training – Faculty Series

The second leg of the collaborative curriculum development process, a methods faculty training series was informed by needs identified during the initial CLA Call Out (Table 1) and the model developed by the *Digital Data Management for a New Generation* project (AAA 2016). The *Digital Data Management* project brought together faculty from across the four-fields of anthropology to develop data management modules that could be incorporated into existing research methods classes in archaeology, cultural

Table 1: Data Management Training Gaps – CLA Call Out

Topic	Description
Faculty professional development	Need for faculty and student specific training and professional development opportunities around data management and data literacy so that faculty can have a better idea of how to incorporate into their training of graduate students.
	Faculty currently learn from students as well in regard to data management and digital methodologies and tools.
	Need for more outreach in CLA around data management plan writing requirements and available resources and tools in grant writing phase.
Graduate Certificate in Data Management	Explore credit and non-credit graduate certificates.
Graduate training by cohort or research area focus	Ex. gerontology program. More information for students earlier in career or at beginning of research projects.
Institutional Repository (IR) oversight & long-term storage	Need for more involvement of faculty advisor/supervisor in IR project creation and publication to provide more stable contact as students graduate.
	Related to this issue there were several questions about what happens to data published in IR after the 10-year commitment window and if the digital object identifier or DOI (persistent identifier) can be transferred back to the researcher or another repository if it is not maintained by the Libraries.

Topic	Description
Ethical and privacy implications of data collection, access and reuse	The need to embed research ethics and privacy awareness into data management and data literacy education, especially in regard to emerging technologies and methodologies (such as research involving social media and other quasi-private data available online and elsewhere, indigenous knowledge and community data ownership).
	Restricted access alternatives like data enclaves and virtual research environments to meet ethical and legal obligations.
	Need to connect with IRB and responsible conduct of research.
Interdisciplinary research and secondary data analysis	Need for data management and data literacy education for those engaged in interdisciplinary research and/or those engaged in secondary data analysis of data collected in other disciplines or by nonacademic data producers.
	Need for ways to connect researchers across campus, build on Digital Humanities Directory and other initiatives in Libraries and CLA, existing STEM collaborations, who are working in similar interdisciplinary space or dealing with similar issues around secondary data analysis / reuse.
Undergraduate students	Discussed the possibility of also targeting undergraduates with a similar program to graduate school workshop series.
	Potential for working with various undergraduate learning communities and directed research internship groups on campus

anthropology, physical anthropology and sociolinguistics.

In order to address the needs of CLA graduate methods faculty (see Table 1), the author invited twelve faculty to participate in a three-part training series. The first and second two-hour sessions took place in the Fall of 2016 and the third two-hour session took place in the Spring of 2017 (see Table 2). The goal of the series was to explore how data management and curation principles might best be incorporated into the faculty members' existing graduate methods courses. Seven of the twelve invited faculty members agreed to participate in the series, with five faculty attending at least two of the three sessions. Departments represented included Anthropology, American Studies, Communication, Linguistics and Political Science.

Table 2: Data Management & Graduate Methods Education Faculty Series

Session Name	Semester	Year
Data literacy competencies & frameworks	Fall	2016
Assessment, active learning & co-curricular resources	Fall	2016
Learning Objectives & Course Planning	Spring	2017

Prior to each session, faculty participants were asked to read a selection of articles related to the session topic and completed a series of data management and curation-related pedagogical assignments related to their own research methods courses. The author facilitated these hands-on work sessions complemented by an additional ten-fifteen-minute presentation by a library faculty member with expertise embedding data and information literacy (session one) and active learning (session two) assignments into credit-bearing courses.

The first session focused on a review of different data literacy and data management competencies (AAA 2016, Sapp Nelson 2016, Calzada Prado and Marzal 2013, Carlson et al. 2011). Faculty were asked to identify data management skills desired or already covered in their existing research methods courses. The second session focused on a review of active learning approaches to data and information literacy. Faculty were asked to identify specific data management topics and skills that might best be taught using active learning and creative strategies (Whitmire 2015, Maybee et al. 2016, Bhargava et al. 2016). Prior to meeting, the group were encouraged to review open educational resources related to data management and curation (AAA 2016, Corti 2014, DataONE 2012, n.d., History DMT 2014, NECDMC 2016, MANTRA 2014, RDMRose 2015).

By the end of the second session, faculty determined that it would be desirable for the author to create a for-credit data management course for the humanities and social sciences, as it would be difficult to cover a full range of data management and curation topics in existing courses. For the third and final session, the group was joined by the CLA Associate Dean for Research and Graduate Education and asked to identify potential learning objectives for a credit-bearing course.

Data Management Workshops and Guest Lectures

The third leg of the collaborative curriculum development process was informed by needs and gaps identified in Table 1 and priorities identified during the ongoing faculty series described above. During the 2016 – 2017 academic year, the author was asked to propose a series of workshops for a Graduate School Data Management Series. The author offered workshops on data management for human participant data, qualitative data management, data management with NVivo (2012/2019) and REDCap (2019) workshops. At the same time the author coordinated with the CLA Associate Dean for Research and Graduate Education to offer a series of data management related workshops targeted specifically at social science and humanities faculty and students in the college. Workshops continue to be offered upon request. The initial CLA series included workshops focused on data management for humanities and social sciences, data management and sharing for community-engaged research, data management with NVivo, and accessing and preparing data deposits for ICPSR and the university's Institutional Repository. See Table 3 for a full list of workshop titles.

Initial requests for data management guest lectures (mini-workshops) included data sharing with collaborators and research participants, legal and ethical issues, data management for specific disciplines, data management with NVivo and other computer assisted qualitative data analysis software, data cleaning and documentation for Big Data Projects (e.g. social media). Guest lectures continue to be offered upon request. See Table 4 for a full list of guest lecture topics.

The post-workshop and guest lecture evaluations included an open-ended question about topics the attendees would have liked to have seen covered either during the workshop/guest lecture and/or in subsequent offerings. Table 5 provides a summary of the topics and descriptions gathered from post-workshop/lecture evaluations.

Table 3: Workshops

	Workshop Name	Venue	Year
W-01	<i>Data Management and Sharing Human Participant Research</i>	Graduate School	2016
W-02	<i>Data Management and the Ethical Implications of Open Science</i>	American Studies Dissertation Symposium	2016
W-03	<i>Data Management for Qualitative Researchers</i>	Graduate School	2016
W-04	<i>Data Management for the Arts, Humanities & Social Sciences</i>	College of Liberal Arts	2016
W-05	<i>Open Ethical Data Sharing</i>	College of Liberal Arts	2017
W-06	<i>Data Management with NVivo</i>	Graduate School	2017
W-07	<i>Data Management with REDCap</i>	Graduate School	2017
W-08	<i>ICPSR data deposit & access</i>	College of Liberal Arts	2017
W-09	<i>Data Management for NVivo Team Projects</i>	Forestry (Lab)	2018
W-10	<i>Data Management for Literature Reviews</i>	Department of Anthropology	2020

Table 4: Guest Lectures

	Lecture Name	Class	Year
GL-01	<i>Scalar Digital Exhibits</i>	American Studies / Research Methods	2016
GL-02	<i>Data Management & Sharing for Reciprocal Anthropology</i>	Anthropology / Qualitative Methods	2016
GL-03	<i>Big Data Management: Documentation and Data Cleaning</i>	Communication / Research Methods	2016
GL-04	<i>Data Management & Sharing for Reciprocal Anthropology</i>	Anthropology/ Qualitative Methods	2017
GL-05	<i>Data Management for Linguistics</i>	Linguistics/ Experimental Methods	2018
GL-06	<i>Data Management for Lit Review with NVivo</i>	Anthropology/ Public Engagement	2018
GL-07	<i>Coding & Data Management with NVivo</i>	Anthropology/ Qualitative Methods	2018
GL-08	<i>Data Management for Lit Review with NVivo</i>	Anthropology/ Political Ecology	2019
GL-09	<i>Data Management for Community Collaborations</i>	Anthropology / Applied	2019
GL-10	<i>Data Management for Qualitative & Mixed Method Research</i>	Anthropology/ Qualitative Methods	2019
GL-11	<i>Qualitative & Multi-modal Data Analysis with NVivo</i>	Anthropology/ Qualitative Methods	2019

Table 5: Suggested Future Topics – Workshop and Guest Lecture Feedback

Topic	Description
Collaborative Data Management and Access	Data management with collaborator.
	More about accessing once data is stored (ease of access), especially in a team.
Best Practices and Tools	Best practices and software that could help with managing data.
	Best practices in managing data (in terms of how do I have to store, compress data!).
	What are different programs/scripts that work best with different data types?
Case Study Examples	More case study examples.
	Examples of various data management plans with an array of cultural contexts and differing research would be helpful
	Applications to anthropology
	Applications to social science
Data Cleaning and Messy Data	More on what to do with data that you already had that wasn't managed well.
	More specifics of the data cleaning.
Data Ethics	More discussion of IRB/Ethics.
Data Storage	Opportunities for data storage.
	Where else can you find storage if you aren't affiliated with an institution?
Focus on participant research	Perhaps a narrower focus to not cover quite so much? I like the idea of getting us to reflect and discuss our own projects.
	Some integration/discussion of our own research interests.
More Hands-on or Interactive Exercises	More interactive stuff would be helpful.
	More hands-on practice. More examples of each phase in the data lifecycle.
	Practice using one of these [data cleaning and documentation] tools.
More Time	More time than one class period-more time for questions.

Graduate Data Management & Curation for Qualitative Research Course

The fourth leg of the collaborative curriculum development process was informed by needs discussed in Table 1 and Table 5, a review of literature on critical perspectives on qualitative data curation and literacy (Antes et al. 2018, Bishop 2009, Cligget 2013, Corti 2011, D'Ignazio 2017, Hardy 2016, Rawson 2016, Yardly 2014) and developing learning outcomes and assessing active learning in graduate-level data management courses (Whitmire 2015). During the Spring of 2017 the author circulated a draft course proposal for the three-credit data management and curation course. CLA Graduate Directors and faculty involved in teaching graduate methods courses and graduate certificates (qualitative research, innovative methods and digital humanities) reviewed and offered feedback on the draft proposal and sample syllabus.

During this time, the author learned that a newly hired CLA faculty member planned to develop a course in data management for quantitative and computational social science. Given the changing landscape and underrepresentation of qualitative research within planned course offerings, the author began to focus specifically on data management and curation for interdisciplinary qualitative research. The author developed the course throughout 2017 in consultation with CLA faculty. The cross-listed (LIS/Anthropology) course proposal was approved in variable title format by Libraries and School of Information Studies, the Department of Anthropology and the Graduate School in 2018 and offered for credit during Spring 2019 and 2020. The author plans to submit a permanent course proposal to the School of Information Studies for 2021.

Table 6 describes topics covered in the course during the first (2019) and second (2020) iteration. Open educational resources such as Data Carpentry (Becker et al. 2019), HILT (Taylor and Tilton 2019) and Programming Historian (van Hooland et al. 2013, Posner and Brett 2016) were consulted when developing some lab assignments. The Spring 2020 syllabus and labs required some modification due to the COVID-19 global pandemic. Some of these changes, such as utilizing a university-based virtual research environment instead of using in-person lab versions of software, provided a more standardized and accessible working environment and will be utilized in the third iteration of the course in the Spring of 2021.

The first two iterations of the course were structured as a two-hour-and-fifty-minute weekly evening class that met in a computer lab. In the Spring of 2020, the last six weeks of class took place in a virtual asynchronous environment. The first hour-and-a-half of class was devoted to lecture (instructor or guest lecturer), student presentations, and student and faculty-led discussion. After a 15-20-minute break in which the instructor was available for questions, the class resumed with an hour-long lab related

Table 6: Data Management & Curation for Qualitative Research Course Topics **additional content introduced in second iteration of course (2020)*

Topic
Archival Information Packages (Bag-it*, Bagger)
Content Management Systems (Scalar, Omeka*)
Data Cleaning (Excel, Open Refine, Python*, command line*)
Data Ethics
Data Management and Sharing Plans
Data Licensing and Use Agreements
Data Management Systems – (CAQDAS - NVivo)
Data Repository Software & Deposit
Data Security (confidential and sensitive data handling)
Documentation (Data Dictionaries, Code Books, Readme files, Jupyter Notebooks*)
File Organization and Naming Conventions
Metadata
Preservation Planning
Storage, Distributed Computing*
Workflows

to the topic of discussion for the day. All students were expected to come to class able to lead discussion on an assigned reading. In the online portion of the 2020 hybrid class, students were asked to post critical summaries and discussion questions to the learning system discussion board. Lab work was in most cases completed and saved to the student's class directory by the end of class, but students had the option to turn in lab assignments at the end of week if the student needed to work on the assignment at a slower pace.

Throughout the semester students developed a novel data curation project related to qualitative research and/or digital scholarship. Students had the opportunity to turn in draft proposals, data management and sharing plans, and project workflows. Students gave project presentations and received feedback before the final semester project was due at the end of the semester. Final data curation projects ranged from preparing dissertation research materials for a data publication or deposit into data repository,

creating archival and dissemination information packages, creating an Omeka digital archive and exhibit, and using the Scalar platform to create a research data portal.

Conclusion

The most significant impact of the collaborative curriculum development and implementation process focused on building trust and sustainable networks between the research data management unit of the Libraries and those engaged in humanities, social sciences and interdisciplinary qualitative research across campus. This initiative allowed the author to better leverage institutional priorities for underserved disciplines by providing a use case and set the groundwork for subsequent sustainable curriculum development and service initiatives.

An unexpected outcome was the interest for qualitative data management and curation training among students and faculty outside of the College of Liberal Arts. This is reflected in the departmental affiliation of graduate students participating in the workshops and first two iterations of the credit-bearing course, which in addition to CLA disciplines included students from the College of Science, Agriculture, Forestry, Education and Health Sciences. For this reason, the author continues to work with faculty to make sure that co-curricular workshops are promoted across campus, and that the graduate-level course will count toward existing and proposed cross-college graduate certificates and cross-listed with a CLA department to satisfy graduate degree methods requirements.

Acknowledgments

The author would like to thank the faculty and students who offered feedback, shared time and resources, and participated in the initiatives described in this article.

Data Availability

All non-identifying data referenced included in tables.

References

- American Anthropological Association. 2016. Digital Data Management for a New Generation. <https://www.americananthro.org/LearnAndTeach/Landing.aspx?ItemNumber=20641&navItemNumber=20708>
- Antes, Alison L., Heidi A. Walsh, Michelle Strait, Cynthia R. Hudson-Vitale, and James M. DuBois. 2018. "Examining data repository guidelines for qualitative data sharing." *Journal of Empirical Research on Human Research Ethics* 13(1): 61-73. <https://doi.org/10.1177/1556264617744121>
- Arnold, Taylor and Lauren Tilton. 2019. "A Brief Introduction to Python," in *Image Analysis with Deep Learning* (HILT 2019). <https://github.com/statsmaths/hilt2019-image-analysis/blob/master/nb/session02-brief-intro-python.ipynb>

Becker, Erin Alison, Anita Schürch, Tracy Teal, Sheldon John McKay, Jessica and Elizabeth Mizzi, François Michonneau. 2019. "Introducing the Shell," in *Introducing the Command Line for Genomics. Data Carpentry*. June 2019 (Version v2019.06.1).

<https://datacarpentry.org/shell-genomics/01-introduction/index.html>

Bhargava, Rahul, Ricardo Kadouaki, Emily Bhargava, Guilherme Castro, and Catherine D'Ignazio. 2016. "Data murals: Using the arts to build data literacy." *The Journal of Community Informatics* 12(3). <http://ci-journal.net/index.php/ciej/article/view/1276>

Bishop, Libby. 2009. "Ethical sharing and reuse of qualitative data." *Australian Journal of Social Issues* 44(3): 255-272. <https://doi.org/10.1002/j.1839-4655.2009.tb00145.x>

Calzada Prado, Javier, and Miguel Ángel Marzal. 2013. "Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents." *Libri* 63(2): 123-134.

<https://doi.org/10.1515/libri-2013-0010>

Carlson, Jake, Michael Fosmire, C.C. Miller, and Megan Sapp Nelson. 2011. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *Libraries and the Academy* 11(2): 629-657. <https://doi.org/10.1353/pla.2011.0022>

Cliggett, Lisa. 2013. "Qualitative data archiving in the digital age: Strategies for data preservation and sharing." *The Qualitative Report* 18(24): 1-11. https://uknowledge.uky.edu/anthro_facpub/1

Corti, Louise, and Arofan Gregory. 2011. "CAQDAS comparability. What about CAQDAS data exchange?" *Forum: Qualitative Social Research* 12(1). <http://dx.doi.org/10.17169/fqs-12.1.1634>

Corti, Louise. 2014/2020. *UK Data Service Companion website for Managing and Sharing Research Data Handbook: A guide to good practice*. <https://www.ukdataservice.ac.uk/manage-data/handbook>

DataONE. 2016. DataONE Education Modules. <https://www.dataone.org/education-modules>

DataONE. n.d. Data Stories. <https://www.dataone.org/data-stories>

D'Ignazio, Catherine. 2017. "Creative data literacy: Bridging the gap between the data-haves and data-have nots." *Information Design Journal* 23(1): 6-18. <https://doi.org/10.1075/idj.23.1.03dig>

Hardy, Lisa J., Amy Hughes, Elizabeth Hulen, and Anna L. Schwartz. 2016. "Implementing qualitative data management plans to ensure ethical standards in multi-partner centers." *Journal of Empirical Research on Human Research Ethics* 11(2): 191-198.

<https://doi.org/10.1177/1556264616636233>

History DMT. 2014. Institute of Historical Research.

<https://port.sas.ac.uk/mod/page/view.php?id=1363>

Maybee, Clarence, Tomalee Doan, and Michael Flierl. 2016. "Information literacy in the active learning classroom." *The Journal of Academic Librarianship* 6(42): 705-711.

<https://doi.org/10.1016/j.acalib.2016.07.005>

MANTRA. 2014. University of Edinburgh (EDINA and Data Library). <http://datalib.edina.ac.uk/mantra>

NECDMC (New England Collaborative Data Management Curriculum). 2016/2017. Lamar Soutter Library, University of Massachusetts Medical School.

<https://library.umassmed.edu/resources/necdmc/index>

NVivo (software). 2012/2019. QSR International. Version 10-12.

<https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home>

- Piorun, Mary E., Donna Kafel, Tracey Leger-Hornby, Slamak Najafi, Elaine R. Martin, Paul Colombo, Nancy R. LaPelle. 2012. "Teaching Research Data Management: An Undergraduate / Graduate Curriculum." *Journal of eScience Librarianship* 1(1): e1003.
<http://dx.doi.org/10.7191/jeslib.2012.1003>
- Posner, Miriam and Megan R. Brett. 2016. "Creating an Omeka Exhibit." *The Programming Historian* 5. <https://doi.org/10.46430/phen0049>
- Qin, Jian and John D'Ignazio. 2010. "The Central Role of Metadata in a Science Data Literacy Course." *Journal of Library Metadata* 10(2-3): 188-204.
<https://doi.org/10.1080/19386389.2010.506379>
- Rawson, Katherine. 2016. "Curating Menus: Digesting Data for Critical Humanistic Inquiry" in *Laying the Foundations: Digital Humanities in Academic Libraries*. <http://dx.doi.org/10.2307/j.ctt163t7kq.7>
- REDCap - Research Electronic Data Capture (software). 2019. REDCap Consortium.
<https://projectredcap.org>
- RDMRose. 2015. University of Sheffield (Information School). <http://rdmrose.group.shef.ac.uk>
- Sapp Nelson, Megan. 2016. "Scaffolding for data management skills: From undergraduate education through postgraduate training and beyond." *Purdue University Research Repository*.
<https://doi.org/10.4231/R7QJ7F9R>
- van Hooland, Setj. Ruben Verborgh and Max De Wilde. 2013. "Cleaning Data with OpenRefine." *The Programming Historian* 2. <https://doi.org/10.46430/phen0023>
- Whitmire, Amanda L. 2015. "Implementing a graduate-level research data management course: Approach, outcomes, and lessons learned." *Journal of Librarianship and Scholarly Communication* 3(2): p.eP1246. <https://doi.org/10.7710/2162-3309.1246>
- Yardley, Sarah J., Kate M. Watts, Jennifer Pearson, and Jane C. Richardson. 2014. "Ethical issues in the reuse of qualitative data: Perspectives from literature, practice, and participants." *Qualitative Health Research* 24(1): 102-113. <https://doi.org/10.1177/1049732313518373>