



Evidence Summary

Researchers May Need Additional Data Curation Support

A Review of:

Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., & Stewart, C. (2018). How important are data curation activities to researchers? Gaps and opportunities for academic libraries. *Journal of Librarianship and Scholarly Communication*, 6(1), 1-24.
<https://doi.org/10.7710/2162-3309.2198>

Reviewed by:

Robin E. Miller
Associate Professor and Research & Instruction Librarian
McIntyre Library
University of Wisconsin-Eau Claire
Eau Claire, Wisconsin, United States of America
Email: millerob@uwec.edu

Received: 3 Dec. 2018

Accepted: 18 Feb. 2019

© 2019 Miller. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

DOI: [10.18438/eblip29539](https://doi.org/10.18438/eblip29539)

Abstract

Objective – To identify the data curation activities most valued by researchers at universities.

Design – Focus group and survey instrument.

Setting – Six R1: Doctoral Universities in the United States of America that are part of a Data Curation Network (DCN) project to design a shared data curation service.

Subjects – 91 researchers, librarians, and support staff.

Methods – The authors used focus group methodology to collect data about valued data curation activities, current practices, and satisfaction with existing services or activities. Six focus groups were conducted at participants' places of employment. Participants reviewed a list of 35 possible data curation activities, including documentation, data visualization, and rights management. A card-swapping exercise enabled subjects to rank the most important issues on a scale of 1-

5, with “most important” activities becoming the subject of a facilitated discussion. In a short paper-based survey, participants also noted whether a data curation practice is in place at their institution, and their satisfaction with the practice.

Main Results – Twelve data curation activities were identified as “highly rated” services that academic institutions could focus on providing to researchers. Documentation, Secure Storage, Quality Assurance, and Persistent Identifier were the data curation activities that the majority of participants rated as “most important.” Participants identified the data curation practices in place at their institutions, including documentation (80%), secure storage (75%), chain of custody (64%), metadata (63%), file inventory or manifest (58%), data visualization (58%), versioning (56%), file format transformations (55%), and quality assurance (52%). Participants reported low levels of satisfaction with their institutions’ data curation activities.

Conclusion – Academic libraries have an opportunity to develop or improve existing data curation services by focusing on the twelve data curation activities that researchers, staff, and librarians value but that could be implemented in a more satisfactory way. The authors conclude that their organization, the Data Curation Network, has an opportunity to improve data curation services or to offer new or expanded services.

Commentary

The strength of this research is in the methods employed to gather data from employees of the nine Data Curation Network member institutions. While focus groups can often be conducted with a rigid set of questions, in this study each focus group’s facilitator used rating and card-swapping to direct the inquiry to the primary interests of the participants at each institution. Quantitative data collected aided in the interpretation of verbal comments about the challenges and barriers to curating data. Card-swapping exercises demonstrated that the participants valued 12 data curation practices in particular. A subsequent

questionnaire about data curation efforts actually in practice revealed that most participants were relatively unsatisfied with institutional practices, even among data curation services they valued. For example, “documentation” was rated as “most important” and 80% of participants indicated this is in place at their institutions. However, only 46.2% of participants were “somewhat” satisfied, and 9.9% were not satisfied with documentation of data curation processes. “Secure storage” received the highest satisfaction ratings among participants, with 38.3% expressing satisfaction, although 40% of participants responded “N/A.”

The authors identify self-selection as a limitation of their research. In light of the study’s specialized topic, another view is that participant knowledge and experience enhanced the focus group’s outcomes. Additional granularity in reporting participant views would improve the results. For example, the authors do not indicate whether the data show differences of opinion between the researchers, librarians, and staff who participated in the study. Researchers, particularly those who require data curation services in order to fulfill contractual obligations, may have different expectations for the outcome of data curation activities at their institutions than the support staff or librarians developing data curation services.

Most of the study’s participants agreed on a set of highly valued data curation activities, which may form the basis of any academic library’s data curation program. While the authors do not directly suggest that the study’s results are generalizable, the title of the article implies that “academic libraries” have opportunities to invest in, develop, and market data curation services. However, the authors repeatedly use the phrase “research libraries,” implying that the results of this research are more likely directed to practitioners at large research universities, if not exclusively at DCN member institutions.

The article does not indicate that the researchers coded the qualitative data collected during the six focus groups. “Case studies” of

two of the six focus groups are presented, highlighting problems with the data curation process, like limited time, de-identification of sensitive data, and a desire for standardized data curation practices. The authors also point out that the literature about data curation raises themes similar to those that emerged during facilitated focus group discussions, including limited time and staffing, and pointing to a need for greater support in the form of documentation, templates, and standards. Coding the qualitative data collected during focus group discussions would improve the authors' communication about the prevalence and frequency of the issues raised by participants in this study.

While the results of the study cannot be generalized to all universities or libraries, library practitioners building a data curation service may find that this research serves as a reference point for the data curation services that researchers value or need.