University of Denver

Digital Commons @ DU

2016

Understanding Roles and Responsibilities of Data Curators: An International Perspective

Anna Maria Tammaro
University of Parma, annamaria.tammaro@unipr.it

Krystyna K. Matusiak
University of Denver, krystyna.matusiak@du.edu

Frank Andreas Sposito
University of Denver, frank.sposito@du.edu

Ana Pervan
Library and Information Management International Labour Organization, ana.pervana@gmail.com

Vittore Casarosa
The Institute of Information Science and Technologies of the Italian National Research Council (ISTI-CNR), casarosa@isti.cnr.it

Follow this and additional works at: https://digitalcommons.du.edu/lis_facpub

Part of the Scholarly Communication Commons

Recommended Citation

Understanding Roles and Responsibilities of Data Curators: An International Perspective

Comments

Originally published in *Libellarium* at

Understanding roles and responsibilities of data curators: an international perspective

Anna Maria Tammaro, annamaria.tammaro@unipr.it

University of Parma, Italy

Krystyna K. Matusiak, krystyna.matusiak@du.edu

University of Denver, USA

Frank Andreas Sposito, frank.sposito@du.edu

University of Denver, USA

Ana Pervan, ana.pervana@gmail.com

Library and Information Management International Labour Organization, Switzerland

Vittore Casarosa, casarosa@isti.cnr.it

The Institute of Information Science and Technologies of the Italian National Research Council (ISTI-CNR), Italy

Abstract

Data curation has emerged as a new area of responsibility for researchers, librarians, and information professionals in the digital library environment. This paper presents the preliminary findings of a large research project sponsored by the International Federation of Library Associations (IFLA), under the auspices of its Library Theory and Research (LTR) Section. The primary objective of the project is to identify the characteristic tasks and responsibilities of data curators in both international and interdisciplinary contexts. The ultimate objective, however, is to develop a "data curation ontology" that will better define the profession and support the development of educational curricula to train future data curators.

KEYWORDS: data curator, IFLA, digital library

Introduction

The variety and sheer volume of research data that must be processed, preserved, and made available to the scholarly community and public at large is creating new challenges, both technical and theoretical, for researchers and the librarians and information professionals who support them. In the past few years, national funding agencies have begun to require researchers to prepare data management plans and to make data sets available in open access

repositories. In addition, the demands for open data publishing are coming from scientific publishers. The concept of data curation has roots in the management of scientific information, but its usage has branched out to other disciplines, including digital humanities.

The IFLA Section Library Theory and Research (LTR) international team has conducted preliminary research using multiple data collection techniques, including literature review, questionnaires, in-depth interviews with professional data curators, as well as content analysis of data curation job announcements.

The research questions addressed in this study are:

- How is data curation currently defined by its practitioners working in the field?
- What specific professional vocabularies are used to describe the data curator's core competencies?
- What are the data curator's primary roles and responsibilities?
- What are the key educational qualifications and areas of expertise required of successful data curators?
- How do these vocabularies, roles, skills, and responsibilities vary by country and region around the world?

First, the team has conducted a broad, comprehensive, and systematic survey of the literature related to the area of digital curation. Next, a subgroup conducted content analysis of job descriptions, disseminated a questionnaire, and conducted interviews with data curators working in the field. The first workshop during BOBCATSSS 2016 conference presented the initial findings and was conducted with the active participation of the audience. Since then the scope of the research project has expanded and more data has been collected using a combination of quantitative and qualitative data collection techniques. The preliminary findings of this study provide a starting point for what we hope will be an in-depth description of the roles and responsibilities of data curators.

Terminology

Data curation has emerged as a new area of responsibility for researchers, librarians, and information professionals in the digital library environment (Heidorn 2011, Witt. 2008). In a 2011 article published in the International Journal of Digital Curation, Higgins (2011) proposes that digital curation has made great strides in establishing itself as a new discipline. However, identifying digital curation as a newly emerged discipline may be premature. Although there is a visible growth in educational programs, it is still unclear how the education for this knowledge and skills set fits within the educational landscape. Digital curation is instead mostly embedded in practice.

A number of alternative vocabularies have been deployed to describe the same or similar practices, reflecting the diverse environments in which data is now archived, as well as an evolving understanding of what the practice means in the field.

An exhaustive list of search terms was used to conduct the literature review:

- data curator
- digital curator
- data curation
- data management
- data curation curriculum
- data curation activity
- data curation infrastructure
- role data curator
- role librarian
- e-science.

The databases consulted were:

- Proquest
- Elsevier
- Emerald
- International Journal of Digital Curation
- Bailey's Digital Curation and Preservation Bibliography
- College & Research Libraries
- Journal of empirical research on human research ethics
- Libraries and the Academy

For this subset, 100 items were retrieved and a more detailed analysis was conducted.

Terms such as digital curation, digital archiving, or digital preservation are often used interchangeably (Beagrie 2008). The term digital curation is increasingly being used for the actions needed to add value to and maintain these digital assets over time for current and future generations of users. Digital curation refers to the actions people take to maintain and add value to digital information over its lifecycle, including the processes used when creating digital content. Digital preservation focuses on the series of managed activities necessary to ensure continued access to digital materials for as long as necessary (Walters and Skinner 2011).

Diverse vocabularies are also invoked to describe the various roles for professionals working with organizing, managing, preserving, and disseminating data. Moreover, there are important differences in understanding and usage of basic data curation terminology among countries and regions around the world.

Multidisciplinary engagement

The digital curation is an area of inter-disciplinary research and practice, and different disciplinary trends are influencing its development (Beagrie 2008). The traditional roles in library and information science (LIS) relating to work with digital assets are in transition. Information professionals don't only require skills and knowledge from the LIS field, but also skills in collaboration and professional approach to interdisciplinary work.

Poole (2013) considers and evaluates digital curation work undertaken in the sciences and in the humanities. In theory and in practice digital curation has benefited substantially from practices developed and tested first in the natural sciences and subsequently adapted for and extended in the humanities. The roles of libraries and data centres are not easy to define. Traditionally positioned at opposite ends of the research lifecycle, the convergence of data and publications and interdependencies between both has modified this traditional scope of duties. Both libraries and data centres are in a transition process. Today the tasks of research libraries and data centres are starting to partially overlap, but generally have complementary roles (Reilly et al. 2011).

Pryor (2012) predicts a doomsday scenario for librarians where computer scientists are doing library work with data and are limited to published materials. The new digital environment poses new challenges for librarians: digital deluge, life cycle of research, data policy, discipline differences, data curation, confidentiality of data. Librarians need subject knowledge to manage data and there is a culture of self-sufficiency among scholars. In alternative, national Data Centres are built (e.g. UK and Australia).

Education and training

There are two Frameworks providing a common language and helping to define the skills, knowledge and abilities that are necessary for the development of digital curation training and for promoting the continuing production, improvement and refinement of digital curation training programmes.

The DigCCurr (Digital Curation Curriculum) project (Lee, Tibbo, and Schaefer 2007) has developed a graduate level curricular framework, course modules, and experiential components to prepare students for digital curation in various environments.

In 2013, the DigCurV collaborative network (Molloy et al. 2014) completed the development of a Curriculum Framework for digital curation skills in the European cultural heritage sector. DigCurV synthesised a variety of established skills and competence models in the digital curation and LIS sectors with expertise from digital curation professionals, in order to develop a new Curriculum Framework.

Librarians need opportunities to learn more about these services either on campus or through attendance at workshops and professional conferences (Tenopir et al. 2014).

Preliminary findings

The role of data curator has become an important information management responsibility worldwide, stimulated by the growing need to organize and preserve large volumes of data generated by scholars, governments, and research institutions. A number of research studies have examined the roles of librarians and information professionals in research data management, but have focused primarily on U.S. libraries and research institutions (Akers et al. 2014, Kim, Warga and Moen 2012, 2013, Palmer et al. 2014). Our study builds upon this prior research and expands it by providing an inclusive international perspective.

The empirical part of this study included two phases of data collection and was designed using a mixed-method strategy. The first phase focused on quantitative content analysis of job announcements derived from a variety of library and information science job posting sites, including American Library Association (ALA) Job List, International Association for Social Science Information Services and Technology (IASSIST), and Code4Lib. IASSIST and Code4Lib provided international coverage and were the main sources of data. In the second phase of the study, interviews were conducted with data curators working at academic libraries and research centres. The following section presents the initial findings from the two phases of the study completed as of May 2016. As more data collection activities are scheduled for summer 2016, all results reported here are preliminary.

Phase I: Content analysis of job announcements

Phase one of the IFLA Data Curator research project consisted of a quantitative content analysis of job listings for data curators and related positions. The source data for this analysis were extracted from multiple sources, with a majority (more than 98%) scraped from the Code4Lib and IASSIST websites. The initial data set included 6051 job announcements. From there the database was winnowed to 441 data curator positions using both automated and hand coding, with 54% classified as composed of "primary" data curation responsibilities and 46% as "secondary". This final database is preliminary in nature as in likely to change as our research deepens.

Because of their source these results showed a selection bias towards positions located in the United States, though 34 countries in all were included in the database, and 12 countries were represented among the positions coded as data curators.

Perhaps the most interesting result of the analysis was a comparison of data curator and non-data curator positions by their location. Only 11.9% of non-data curator job listings were located in countries outside of the United States. Yet 26.4% of data curator jobs (including both "primary" and "secondary" positions) were located in other countries. These data suggest that data curation, as well as the data curator job market, is a more international practice than other kinds of work in the library and information science sector.

Other findings from the analysis show that the title "Data Curator" is a poor indicator of data curation responsibilities. There were 280 unique titles across the 441 positions coded as "data curator." No single title covered more than 3.5% of the total jobs. The most common titles were "Data Librarian", "Data Services Librarian", "Data Curator", and "Digital Scholarship Librarian". The prevalence of "librarian" in these titles indicates that typical responsibilities of librarianship, including reference, instruction, and outreach, are common for data curators as well.

Among the types of organizations employing data curators, universities (and university departments) were by far the most common (74.6%), which possibly reflects the North American bias in the source dataset. After universities, university libraries were second most frequent (9.8%), followed by research centres (6.1%) and government agencies (4.8%).

Phase II: Interviews with data curators

The second phase of the study provided an opportunity to gain insight into the practice of data curation and look at the roles and responsibilities from the perspective of information professionals working in the field. Interviews were a primary source of data and were supplemented by questionnaires and documentary evidence. Interviews followed a semi-structured protocol and focused on the participant's' job functions, work activities, distribution of responsibilities, and skills and competencies. Participants were recruited from academic libraries, research centres, and data curation centres in Australia, Canada, and the United States. Convenience and snowball sampling were used in the recruitment. Nine participants were interviewed as of May 2016. The interviews were conducted over Skype or phone and lasted between 30 and 45 minutes. All interviews were recorded and transcribed. The research team plans to conduct additional interviews in summer 2016.

The participants recruited for the study held different position titles, including coordinator of data curation and scholarly communications, data curation librarian, data librarian, data curation scientist, digital curation coordinator,

e-research project officer, project scientist, research data management librarian, and research services coordinator. A phrase "data curation" and a word "librarian" appear in several titles. However, all titles are unique, which confirms the findings from the first phase about a wide variety of titles for positions with data curation responsibilities. Six participants in the sample worked at the university libraries, one at the research centre, one at the university department, and one at the newly funded, campus-wide research services unit. The participants came from a wide variety of backgrounds with Bachelor education in computer science, engineering, health sciences, and history. All participants held Masters in library and information science (MLIS) and three had PhDs.

The importance of outreach and training responsibilities emerged as a pattern in all interviews, regardless of the disciplinary or international contexts. A number of participants discussed a mismatch between perceptions of data curation and their actual work activities. As it turned out, most professionals interviewed for this project have not been involved in managing data directly. Instead, their efforts have been focused on outreach, consultation, and training of researchers. As indicated by Participant E, "data curation is more about providing information about good data curation practices to the people who need to curate their data or could be curating data." While more time is devoted to reaching out to researchers in the early stages of data curation programs, outreach remains a key responsibility for data curators working for more than 2 years. Outreach efforts are focused on:

- Informing researchers about new data curation services available their institutions
- Educating about data life cycle and good data management practices
- Promoting open access and data sharing through data repositories
- Consulting on metadata standards, data formats, and citation standards.

Data curators are involved in teaching workshops and providing one-on-one consultations to faculty and graduate students. Workshops cover a wide spectrum of topics from data management plans to data citation and sharing. In addition to expertise in data management practices and standards, data curators need to have good communication and instruction skills to deliver effective presentations. Although data curators may not work directly with research data, their work requires some technical skills and knowledge of new technological solutions since they often make recommendations to researchers and lead data curation initiatives at their institutions.

Data curation, or perhaps more appropriately research data services, emerges as a new specialty that combines technical and public services skills. As one of the study participants emphasizes, "It's almost like multiple jobs in one really. I'm a technical services librarian, as well as an outreach librarian" (Participant C). The emerging specialty requires expertise in multiple areas of data creation,

organization, publishing, and preservation. It also involves collaboration - often a cross-campus collaboration - and building bridges between a library, IT unit, campus departments, or specialized centres. In a traditional library environment, technical and public services librarians tend to work in separate departments and have been educated using different tracks in LIS programs. Data curation as a new area of responsibility for librarians and information professionals requires a combination of technical, reference, and instruction skills and poses new challenges to the development of LIS curricula.

Conclusion

The IFLA Data Curator research project is a work in progress. Moving forward, we hope to find additional sources for non-North American data curator positions, engage a natural language analysis of the entire position database, conduct additional interviews, and perform cross-analysis of job descriptions, questionnaires, and interview data. The ultimate goal of the project is to identify key responsibilities of data curators and to develop a glossary that should help to better define the profession and develop appropriate educational curricula. The preliminary findings of this research study indicate that different terms are used to describe various roles for professionals working with organizing, managing, and preserving data. Data curation is an emergent field and precise vocabulary has yet to mature.

References

Akers, K. G., F. C. Sferdean, N. H. Nicholls, and J. A. Green. 2014. "Building support for research data management: Biographies of eight research universities." International Journal of Digital Curation 9, 2: 171-191. https://doi.org/10.2218/ijdc.v9i2.327

Beagrie, N. 2008. "Digital curation for science, digital libraries, and individuals." International Journal of Digital Curation 1, 1: 3-16. https://doi.org/10.2218/ijdc.v1i1.2

Heidorn, P. B. 2011. "The emerging role of libraries in data curation and e-science." Journal of Library Administration 51, 7-8: 662-672.

Kim, J., E. Warga, and W. E. Moen. 2012. "Digital curation in the academic library job market." Proceedings of the American Society for Information Science and Technology 49, 1. https://doi.org/10.1002/meet.14504901267

Kim, J., E. Warga, and W. E. Moen. 2013. "Competencies required for digital curation: An analysis of job advertisements." International Journal of Digital Curation 8, 1: 66-83. https://doi.org/10.2218/ijdc.v8i1.242

Lee, C. A., H. R. Tibbo, and J. C. Schaefer. 2007. "Defining what digital curators do and what they need to know: The Digccurr project." In Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, 49–50. New York, NY, USA: ACM. http://doi.org/10.1145/1255175.1255183 https://doi.org/10.1145/1255175.1255183

Molloy, L., A. Gow, and L. Konstantelos. 2014. "The DigCurV curriculum framework for digital curation in the cultural heritage sector." International Journal of Digital Curation 9, 1: 231–241. http://doi.org/10.2218/ijdc.v9i1.314 https://doi.org/10.2218/ijdc.v9i1.314

Palmer, C. L., C. A. Thompson, K. S. Baker, and M. Senseney. 2014. "Meeting data workforce needs: Indicators based on recent data curation placements." In iConference 2014 Proceedings. http://hdl.handle.net/2142/47308

Poole, A. H. 2013. "Now is the future now? The urgency of digital curation in the digital humanities". Digital Humanities Quarterly 7, 2. http://www.digitalhumanities.org/dhq/vol/7/2/000163/000163.html

Pryor, G. 2012. "Why manage research data?" In Managing Research Data, G. Pryor (ed). London: Facet Publishing.

Tenopir, C., R. J. Sandusky, S. Allard, and B. Birch. 2014. "Research data management services in academic research libraries and perceptions of librarians." Library %26 Information Science Research 36, 2: 84–90. http://doi.org/10.1016/j.lisr.2013.11.003

Walters, T., and K. Skinner. 2011. New roles for new times: Digital curation for preservation. Washington, DC: Association of Research Libraries. http://www.arl.org/storage/documents/publications/nrnt_digital_curation17mar11.pdf

Witt, M. 2008. "Institutional repositories and research data curation in a distributed environment." Library Trends 57, 2: 191-201. https://doi.org/10.1353/lib.0.0029