Research Data Curator: the competencies gap

Anna Maria Tammaro: University of Parma, Information Technology Department, Via Usberti 181/a, 43125, Parma (Italy). annamaria.tammaro@unipr.it

Seamus Ross: University of Toronto, Faculty of Information Toronto, 140 St. George Street, Toronto, Ontario, Canada M5S 3G6, seamus.ross@utoronto.ca

Vittore Casarosa: Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" Area della Ricerca CNR di Pisa Via G. Moruzzi 156124 PISA - Italy, vittore.casarosa@isti.cnr.it

Abstract: *The increase of digital content in the broad areas of institutional and domain specific Repositories, Libraries, Archives and Museums and the increased interest in the sharing and preservation of "research data" have triggered the emergence of new competencies and skills such as Digital Curation. The paper refers about the ongoing investigation of current data curation education and training programs with regard to competencies gap and the role of information professionals in the research lifecycle. The investigation has been based on a series of workshops and events discussing the concerns of researchers and professors about digital library and digital curation. A first list of competencies and skills at technical and operational level that professionals should have, has been evidenced. The theoretical framework and structure of educational programmes should have sufficient flexibility to accommodate the needs of various groups of specialists.*

Keywords: Digital Library , Research Data Curation, Competencies

Introduction

Today practically all the research activities are based on digital sources, and therefore a particular aspect of Data Curation is the storage, management and preservation of digital research data. Digital research data can take many different aspects, such as previous publications, images, video, audio, data bases, email, web sites, etc., and most of the time those data are specific to the research field. Moreover, very often teams consist of members distributed across large geographic areas and researchers need better services that provide more sophisticated access control and collaboration tools, and that can support the generated, data, sometimes in big volumes. The increase of digital content in the broad areas of institutional and domain specific Repositories and the increased interest in sharing and preserving "research data" have triggered the emergence of new competencies and skills of data curators.

For the purpose of this paper, data curation is best described as life cycle data management; it encompasses a spectrum of activities ranging from: research data management planning at the project inception stage; through collection of data as part of the research process; through the identification, processing, and accession of data sets; and, finally, to the archival preservation and sharing of data in an appropriate repository. Those activities add value and knowledge to the collections, and the added value is usually given by the “data curator” and the participants in the research project. Different players can be taking the role of data curator: data managers, data creators, data librarians, data archivists, and data scientists. In other words, data curators include the convergent role of Library, Archives, Museums professionals, together with the creator of data, such as scientists involved in research projects. The convergence of education of information professional in archives, museums, and libraries will be the main topic of the presentation. While evidence of such convergence has been found in North America and in some other places internationally, clearly there is considerable additional research and discussion that is needed to determine the most effective way to proceed to realize a convergence, especially in Europe. The discussion will focus on competencies needed by

information professionals in these three fields, as they relate to the curation and management of digital collections.

Aims and objectives

The paper refers about the on-going investigation of current data curation education and training curricula, with regard to the competencies gap and the role of information professionals in the research lifecycle. The questions being investigated are:

1. More and more the professions of Librarian, Archivist and Museum curator seem to converge into an "Information professional". Is there a common core of knowledge/skills that could be taught ? What is it ? Should there be a Data Librarian, or a Data Archivist, or a Data Museum curator (Borgman 2010, 2012)? Or is this a new role to be invented from scratch ?
2. In addition to Librarian, Archivist and Museum curator curricula, what are the specific knowledge/skills that need to be taught in each discipline? Or should the responsibility of curating research data be given to the "data producers", i.e. the researchers themselves ?
3. Is the role/skills of a "Data Curator/Custodian" in the research/scientific domain the same as the role/skills of a Librarian/Archivist/Museum curator in the culture heritage domain ?

In the first phase of the survey, a literature review has been done, collecting information about competencies and skills of the data curator, and initiatives aimed at improving them. In the United States, the Institute of Museum and Library Services' (IMLS), within the Laura Bush 21st Century Librarian Program (IMLS 2013) has been supporting a significant range of projects that have positioned some Library and Information Science (LIS) schools as leaders in data curation education. At the University of North Carolina at Chapel Hill, a matrix of skills and functions has been developed (Lee 2009), describing 24 functional areas and 4 meta-level functions. These are broad, high-level categories, designed to address "digital curation 'know how,' as opposed to the conceptual, attitudinal or declarative knowledge." The Data Curation Education in Research Centers –(DCERC 2013) is a joint initiative of the University of Illinois at Champaign-Urbana, the University of Tennessee at Knoxville and the National Center for Atmospheric Research, aimed at developing a sustainable and transferable model for educating LIS masters and doctoral students in data curation through field experiences in research and data centers. Target research areas for the dissertation of doctoral students include:

• cross-disciplinary data sharing and reuse potentials
• ontology of datasets, formats, provenance, identity conditions
• metadata for description, discovery, interpretation, integration
• interoperability, provenance, preservation, and reuse
• research data in the scholarly communication continuum
• trust, security, confidentiality, ownership, quality, attribution.

Methodology

The investigation has been based on a series of workshops and events discussing the concerns about digital library and digital curation of researchers and professors. The co-authors of this presentation were among the organizers of the Workshop «iSchools and Education in Data Curation», held on September 26th in Malta in connection with the TPDL 2013 Conference (DataCur 2013). Earlier this year they were also among the organizer of the Workshop «iSchools Building on the Strengths Found in the Convergence of Librarianship, Archival, and Museum Studies to Improve the Education of Managing Digital Collections», held on

February 13th, during the iConference 2013,(WSiConf 2013). They have also been involved in similar events in the past, which started in 2005 with a workshop on "Information Technologies profiles and curricula for libraries" (held at the University of Parma), and has continued through five more events up to the last one in September 2013. For more details see (Tammaro et al. 2013).

Findings

Data curation development is not merely a technical/engineering issue, because data curation involves many different stakeholders, and data curators need to be aware of the organisational context and socio political issues. Another aspect to consider is the growing number of interdisciplinary projects, with the need to take discipline-specific differences into account to ensure the efficient and professional handling of research data. A broad distinction of the roles involved in research data curation is that of the "digital curator", more concerned and involved in the managerial and organizational aspects of data curation, and that of the "data curator", more concerned and involved in the actual creation, preservation and re-use of data.

A first list of required competencies and skills at technical and operational level that professionals have been evidenced appears in (Tammaro et al 2012), based on a Delphi study developed in the frame of a DILL Master thesis (Madrid 2011). The investigation has evidenced competency gaps in four areas.

Digital curator (managing organization)

Competencies include (but are not limited to these)

1. Plan, implement, and monitor digital curation projects.
2. Understand theory and practice of record keeping.
3. Formulate data curation policies.
4. Establish and maintain collaborative relationships with various stakeholders (e.g., IT specialist, information professionals inside and outside the institution, data creators, (re)users and other stakeholders like vendors, memory institutions and international partners) to facilitate the accomplishment of digital curation objectives.
5. Auditing and compliance assessment.
6. Be aware of the need of institutional organizational context
7. Understand and be able to communicate the risk of information loss or corruption of digital entities.
8. Manage the use of pre-requisite knowledge.
9. Be aware of advertising and promoting
10. Observe and adhere to all applicable legislation and regulations when making decisions about preservation, use and reuse of digital objects in collaboration with legal practitioners.

Digital curator (user services)

Competencies include (but are not limited to these)

1. Ensure access and re-use of preserved data
2. Appraise and select the collections
3. Allow discovery and retrieval
4. Do authors training
5. Be aware of disciplines related practice.

Data curator (managing data)
Competencies include (but are not limited to these)

1. Manage the workflow for long-term preservation.
2. Have an expert knowledge of different types of resources and big data.
3. Know data structure of different digital objects and determines appropriate support it needs.
4. Understand semantic ontologies, linked data.
5. Be aware of requirements for information infrastructure in order to ensure proper access, storage and data recovery.
6. Diagnose and solve problems to ensure continuous accessibility to digital objects, in collaboration with IT professionals.
7. Monitor the obsolescence of file formats, hardware and software and the development of new ones (e.g. using such tools as PRONOM registry)
8. Ensure the use of methods and tools that support interoperability of different applications and preservation technologies among users in different locations.
9. Verify collection definition and the provenance of the data to be preserved and ensures that it is properly documented.
10. Have the knowledge to assess the digital objects' authenticity, integrity and accuracy over time.

Data curator (technological skills)
Competencies include (but are not limited to these)

1. Understand software for curation of data
2. Be aware of data format obsolescence
3. Understand accessibility to data
4. Have an expert knowledge of system design and implementation
5. Know interoperability
6. Monitor digital repository
7. Ensure preservation.

Conclusions

The research workflow is evolving towards a global infrastructure providing preservation and access to research data and research results, and providing discipline-specific tools. In this model the role of the Information Professionals should evolve from one of holders and providers of knowledge resources to one of being an active partner in the research process. From the examination of existing education and training programs and the investigation results of the workshops, there is a clear need to make sense of the digital curator role as a whole. The theoretical framework and structure of educational programmes should have sufficient flexibility to accommodate the needs of the various groups of data specialists.

References

Borgman, C., (2010). "Why data matters to librarians – and how to educate the next generation", The Changing Role of Libraries in Support of Research Data Activities: A Public Symposium, National Academy of Science, Board on Research Data and Information, 12 February, Washington DC, USA, http://sites.nationalacademies.org/PGA/brdi/PGA_056901, (last visited Dec. 16th, 2013)

Borgman, C., (2012). IS289 "Data, Data Practices, and Data Curation", Graduate School of

Education and Information Studies, University of California, Los Angeles, http://polaris.gseis.ucla.edu/cborgman/Chriss_Site/Courses.html, (last visited Dec. 16th, 2013)

Choi, Y. & Rasmussen, E. (2009). What qualifications and skills are important for digital librarian positions in academic libraries: A job advertisement analysis. The Journal of Academic Librarianship, 35(5), 457-467. http://dx.doi.org/10.1016/j.acalib.2009.06.003

Choudhury, G.S. (2008). Case study in data curation at John Hopkins University. Library Trends, 57(2), 211-220. http://dx.doi.org/10.1353/lib.0.0028

Corrall, S. Roles and responsibilities: libraries, librarians and data. Managing Research Data. Ed. G. Pryor. London: Facet, 2012, pp. 105-133.

Cunningham, A. (2008). Digital curation/digital archiving: A view from the National Archives of Australia. American Archivist, 71(2), 530-534.

DataCur2013 (2013), Workshop "Moving beyond technology: iSchools and education in data curation. Is Data Curator a new role?"(in connection with TPDL 2013, 23-26 September, La Valletta, Malta) http://datacur.isti.cnr.it/index.html, (last visited Dec. 16th, 2013)

DCC Digital Curation Centre, "What is digital curation?" http://www.dcc.ac.uk/digital-curation/what-digital-curation (last visited Dec. 16th, 2013)

DCERC (2013), Data Curation Education in Research Centers, http://www.sis.utk.edu/dcerc, (last visited Dec. 16th, 2013) http://www.lis.illinois.edu/research/projects/dcerc, (last visited Dec. 16th, 2013)

DigiCurV (2010) Training Opportunities http://www.digcur-education.org/eng/Training-opportunities (last visited Dec. 16th, 2013)

Harris-Pierce, R.L. & Liu, Y.Q. (2012). Is data curation education at library and information science schools in North America adequate? New Library World, 113(11/12), 598-613 http://dx.doi.org/10.1108/03074801211282957

Heidorn, P.B. (2011). The emerging role of libraries in data curation and e-science. Journal of Library Administration, 51(7), 662-672. http://dx.doi.org/10.1080/01930826.2011.601269

IMLS (2013). Institute of Museum and Library Services', Laura Bush 21st Century Librarian Program http://www.imls.gov/applicants/detail.aspx?GrantId=9 (last visited Dec. 16th, 2013)

Jahnke, Lori, and Andrew Asher. (2012). The Problem of Data: Data Management and Curation Practices Among University Researchers. The Problem of Data, pp. 3–31. Washington, DC: Council on Library and Information Resources.

Kennan, M.A., Cole, F., Willard, P., Wilson, C. & Marion, L. (2006). Changing workplace demands: What job ads tell us. Aslib Proceedings, 58(3), 179-196. http://dx.doi.org/10.1108/00012530610677228

Keralis, S.D.C. (2012). Data curation education: A snapshot. In L. Jahnke, A. Asher & S.D.C. Keralis (Eds.) The problem of data. Washington, DC: Council on Library and Information Resources.

Lee, C.A. & Tibbo, H. (2011). Where's the archivist in digital curation? Exploring the possibilities through a matrix of knowledge and skills. Archivaria, 72, 123-168.

Madrid, M., (2011), A Study of Digital Curator Competences: A survey of experts; DILL Master Thesis http://dspace-unipr.cilea.it/handle/1889/1785 (last visited Dec. 16th, 2013)

Rusbridge, C. (2007). Create, curate, re-use: The expanding life course of digital research data. Paper presented at EDUCAUSE, Australasia.

Swan, A. & Brown, S. (2008). The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs. A report to the Joint Information Systems Committee (JISC). Retrieved from http://ie-repository.jisc.ac.uk/245/1/DataSkillsReport.doc (last visited Dec. 16th, 2013)

Tammaro, A.M., Madrid, M. Casarosa, V., (2012), Digital curators' education: professional identity vs. convergence of LAM (Libraries, Archives, Museums), IRCDL 2012, 9-10 February, Bari, Italy, in Communications in Computer and Information Science Volume 354, 2013, pp 184-194 http://link.springer.com/chapter/10.1007%2F978-3-642-35834-0_19, (last visited Dec. 16th, 2013)

Tammaro, A.M., Casarosa, V., Castelli, D., (2013), Closing the gap: interdisciplinary perspectives on research and education for digital libraries, IRCDL 2013, Rome, Italy, to appear in Communications in Computer and Information Science, Springer, 2013.

Walton, G. (2010). Data curation and the academic library. New Review of Academic Librarianship, 16(1), 1-3.

WS-iConf2013 (2013), Workshop "iSchools Building on the Strengths Found in the Convergence of Librarianship, Archival, and Museum Studies to Improve the Education of Managing Digital Collections" (in connection with iConference 2013, February 12-15, Fort Worth, TX, USA) https://www.ideals.illinois.edu/handle/2142/42495, (last visited Dec. 16th, 2013)