The Role of Librarians in Data Science: A Call to Action

Elaine R. Martin
University of Massachusetts Medical School

Let us know how access to this document benefits you.

UMass Chan
MEDICAL SCHOOL

Journal of eScience Librarianship
putting the pieces together: theory and practice

Editorial

The Role of Librarians in Data Science: A Call to Action

Elaine R. Martin, DA, Director of Library Services

University of Massachusetts Medical School, Worcester, MA, USA

Focus

Many academic institutions and their libraries have developed research data services, but sometimes institutional objectives, professional organizations, and librarians' current and future roles aren't always in sync. In this issue of the Journal of eScience Librarianship, librarians report on moving forward with various services, but frequently face institutional and professional obstacles.

After recently attending a meeting of librarians where the discussion about their roles within the field of data science was put on "hold," the focus of this editorial had to be the topic. A lesson for me based on this experience, where I was the lone voice advocating for leadership, is my library and its librarians are part of an exclusive group of health sciences librarians contributing to data science initiatives. The numbers of health sciences librarians won't expand without some intentionality or strategy, so I want to talk about the issues of data science, management and practice from a library science perspective (regardless of specialization), although my bias is in the health sciences. One more piece of context, there was some question raised in this meeting as to what "data" are. I would argue that there are many accepted definitions of data (Hey 2009), the data life-cycle (Humphrey 2006) and many recent books on the topic (Corti 2014) that define data better than I, so I will refer readers there.

All of this hesitancy on the part of librarians to participate in the data movement is happening at a time when we have seen an increase in the money and involvement in data initiatives from a range of other professions and academic disciplines (e.g. computer science, informatics, etc.). For me, this is an especially critical moment for librarians to talk about data and actively plan and implement our strategies collectively.

I want to share with you a proposed framework for the librarian's role in data science. I come to the discussion with the fear that data science is an evolving academic discipline being defined solely by computer science and that the field of library and information science is being left behind. I would argue that the principles and values of the field of library and information science that form the core of our profession need to be part of this new discipline and that we can add unique perspectives and roles. I hope this proposed framework illustrates these points.
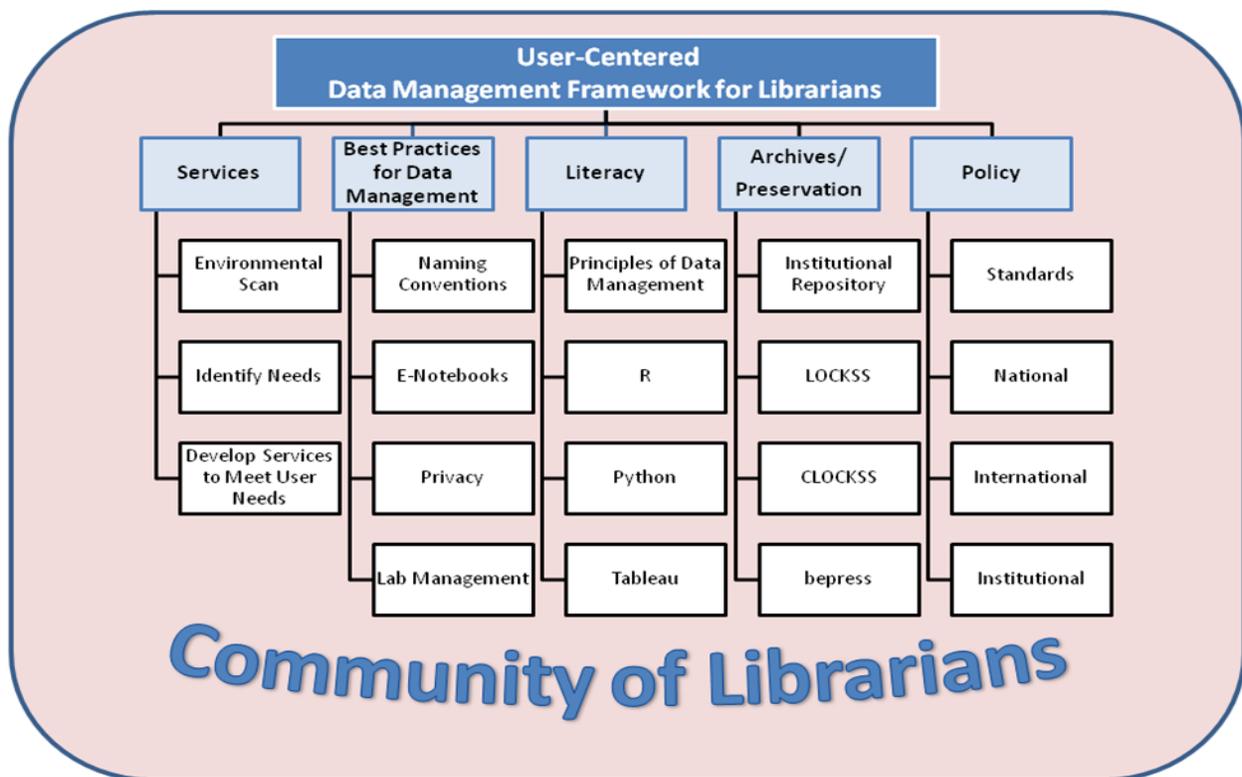
Figure 1: User-Centered Data Management Framework for Librarians. Building a community of data science librarians to evolve the data science discipline.

Sitting outside the framework are the values the library science profession brings to the data science table. It is these values which help us bring a unique and necessary perspective to the evolving data science discipline. These values include: focus on the user, user needs and user behavior, an ethical base, a penchant for collaboration and equal access to all types of information.

First, let me point out that data science is the realm of different communities (e.g. computer science, informatics, etc.), including the library science community. The outside circle illustrates the community of data science librarians who interact within and among these different communities as well as interact within the center of the framework. The center of the framework is the **User**, not the data. The User-centered framework consists of buckets. These buckets include: data services, data management practices, data literacy, data archiving and preservation, and data policy.

The **Data Services** bucket may include the following activities: assessing researcher needs, performing an institutional data environmental scan, conducting the research interview, designing a suite of services such as assistance with DMPs based on user needs, etc. The **Data Management Practices** bucket may include: recommending file naming conventions, helping a lab institute an electronic lab notebook program, providing assistance with HIPAA guidelines for clinical research data, etc. The **Data Literacy** bucket may include offering in-person and online classes on topics such as: Data Management Best Practices (e.g. NECDMC) and data tools (e.g. R and Python) and data visualization software (e.g. Tableau). The **Archives/Preservation** bucket focuses on data repositories, institutional and subject, and their corresponding preservation options such as LOCKSS, CLOCKSS, etc. The **Data Policy** bucket involves working in collaboration with local, institutional, regional, national and international standards organizations to develop data sharing standards.

Journal of eScience Librarianship readers can help flesh out this proposed framework and its corresponding buckets. Certainly the articles in this issue and earlier issues illustrate how librarians can play roles in the activities suggested by the framework.

This way of organizing the conversation around a common framework will lead to clarity of discussion and strategic thinking around the question so that we move beyond "hold" to "action." Perhaps this framework can help us deliberate about and act upon what we want to, or should, focus our collaborative efforts on regarding the different buckets at any given time.

On that note, I would like to encourage *JeSLIB* readers to think about the framework and the buckets, and add to the discussion.

Please send your comments to me at Elaine.Martin@umassmed.edu.

References

Hey, Tony, Stewart Tansley, and Kristin Michele Tolle, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft research. http://research.microsoft.com/en-us/collaboration/fourthparadigm

Humphrey, Charles. 2006. e-*Science and the Life Cycle of Research*. University of Alberta. http://www.usit.uio.no/om/organisasjon/uav/itf/saker/forskningsdata/bakgrunn/life-cycle.pdf

Corti, Louise, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard. 2014. *Managing and Sharing Research Data: A Guide to Good Practice*. London: Sage. https://us.sagepub.com/en-us/nam/managing-and-sharing-research-data/book240297