

# Learning Hands-on and by Trial & Error with Data Curation Profiles

D. Scott Brandt

Purdue University Libraries  
West Lafayette, IN US  
[techman@purdue.edu](mailto:techman@purdue.edu)

**Abstract**— The Data Curation Profiles Toolkit can be used in several ways to capture requirements for data sets, as articulated by researchers. As a flexible instrument it can facilitate dialog between librarians and researchers to discuss data concerns, current data workflow, and possible outcomes for the future. As a structured tool it can help identify areas of concern and need, to begin to making informed decisions about the data. Published Profiles offer insight into similarities and variations in data and data workflow, across multiple research areas or sub-disciplines. Designed as a tool for practitioners, it can help build knowledge and skill through application. Librarians who have completed Profiles have found the process to improve their comfort, increase their confidence, and build competencies in working with researchers. Mapping use of the Profiles Toolkit to the DigCurV Curriculum Framework can help identify strengths and weaknesses in what is currently a hands-on, trial and error self-learning approach.

**Keywords:** *data curation, competencies, learning outcomes, workshop, self-learning, Curriculum Framework, Data Curation Profiles Toolkit*

## I. INTRODUCTION

The Data Curation Profiles (DCP), and its associated Toolkit, is an outcome of a research investigation launched in 2007 to understand data sharing, particularly amongst single or small groups of investigators [1]. The instrument probes the kinds of data researchers are working with, what they are currently doing with it, and what they would like to do with it. Librarians and other information professionals use the DCP Toolkit to interview researchers, and the result can produce a stand-alone publication describing a researcher's intent and needs related to a data set—the Data Curation Profiles Directory currently has published 30 Profiles [2]. In effect, the DCP Toolkit supports the role of librarian or information professional that finds herself or himself in the position of mediator between the researcher and the archivist.

There is no explicit curriculum for learning how to use the DCP Toolkit, which consists of a User Guide, Interviewer Manual, Interviewee Worksheet and Profile Template. The US Institute of Museum and Library Services, who funded the

original research project that produced the Toolkit, supported a short project to teach librarians how to use it. Over 300 librarians attended a series of twelve workshops taught between 2011-2012. The workshops were developed to guide attendees in how to use the DCP Toolkit and what to expect when doing so. The goal was for participants to gain comfort using a tool, so that they could gain confidence in exploring researcher needs and concerns, and later gain competence in working with researchers on data management planning and depositing data in a repository [3].

To this point training has consisted of a one-day session focused on instruction in using the DCP Toolkit. It has covered concepts of data curation as they relate to researchers (identified in the original DCP research), how the components of the Toolkit were designed to probe them, and detailed information on working through two main sections, data lifecycle and data sharing [4]. Probing about the research data lifecycle is seen as important for both researcher and librarian because it uncovers what might otherwise be unrecognized: that the research process can be seen as stages of data collection and analysis that produce tangible products (data sets) in addition to intellectual findings disseminated in articles. Probing about data sharing also is seen as important because it helps identify researcher concerns in doing so.

An instructional design approach, ADDIE, was used in developing the workshop. It required identification of users' needs (how to use the Toolkit), developing specific learning outcomes (application of skills), designing learning objectives to meet the outcomes through lecture, discussion, exercises and additional resources, to scaffold learning (scenarios) [5]. Assessing the workshop with the Framework would be helpful in building and expanding training. Initially it is assumed that training will continue to be generic in regard to digital curation overall, but specific in focusing on the particular use of the Toolkit. It is also hoped that the framework will help us describe the value of the training in a meaningful way for a broader range of people.

## II. CURRICULUM FRAMEWORK AS ASSESSMENT TOOL

The Framework is meant to be useful to those building new training courses. Depending on the user's aims, the Framework can assist in providing a structure for a generic training program for the role of digital curator, or it can suggest which subjects should be covered in shorter, more specialized courses addressing one particular area of professional digital curation practice. The Framework may also supply a common language to allow those building and developing training to meaningfully describe the value of their training offerings.

Utilizing the Framework to assess the DCP Toolkit and associated learning initially seems somewhat problematic. First, training is focused on emulating use of the DCP Toolkit. The workshop gives context (lecture), offers instruction (reading), provides modeling (videos) and employs problem solving (exercises). Basically, the learning objectives are meant to facilitate self-learning through scenarios of how others employed the DCP Toolkit, and providing expert feedback by workshop leader during discussion. Second, users of the DCP Toolkit are not learning about data curation per se (i.e., digital preservation), they seek to learn what researchers are doing with their research (i.e., data management). This is because the DCP Toolkit supports the role of librarian or information professional in the position of mediator between the researcher and the archivist. However, the Framework can still be helpful in assessing learning goals and outcomes.

## III. ASSESSING WITH THE FRAMEWORK

The Framework looks first at the knowledge and principles to be learned. The DCP Toolkit outcomes map especially well with understanding research data management. That is, the elements under the Framework map with the goal of understanding the concepts and terminology employed in research data collection and analysis, organizing and describing or documenting it, and meeting funder or institutional requirements. However, beyond a general data lifecycle, it is hard to anticipate what someone will encounter when meeting with a researcher—methodology differs between physical, life and social science, not to mention the humanities. What has been key, is to represent a data lifecycle in a data table depicting various stages of collecting data, processing or anonymizing it, analyzing it, and depositing and publishing it. Classifying data stages can help clarify what will or won't or can't be shared.

Assessment with the Framework at this level raises the question of how much knowledge about digital curation a librarian needs to interview a researcher to get "the story" around his research data and workflow. Up to now, the goal of using the DCP Toolkit is to gain insight and to gather many Profiles so they can be studied to understand research data in broader terms, in order to give greater context to digital curation. However, to fulfill a mediating role and provide

specific services would demand more knowledge and application of related principles. For instance, to make suggestions about depositing a mediator would need to know more about requirements related to file formats (i.e., which are more appropriate for preservation).

In assessing skills and competences, it can be argued in this case that there are three perspectives: those related to the researcher, to the librarian or information professional (as mediator) and to the archivist. Subject knowledge in the discipline relates to the researcher, and anyone who uses the DCP Toolkit to interview should have some familiarity with the discipline (i.e., background, or should review the researchers work to become familiar with it). Skills for the librarian or information professional include first those that help in interviewing the researcher, which are enumerated in the Framework: creativity (inquiring mind), professional conduct (ethics regarding confidentiality, familiarity with institutional policies), and communication (articulating and clarifying needs). Skills for the librarian or information professional also include those related to creating Profiles to publish: creativity (an inquiring mind to pursue the interview), personal qualities (able to engage in deep conversation), self-management (self-initiative) professional conduct (clear and accurate reporting of information synthesized), research management (project planning/delivery) and, obviously, communication. Competence is required, even if these are considered "soft skills;" although it is not clear how to teach them. This is an area in particular for which the workshop employs hands-on practice, and practicing through trial and error is important.

## IV. AMBIGUITIES IN DESIGN

Understanding how the DCP Toolkit fits with a digital curation lifecycle depends mostly on where or how the research lifecycle is seen as having overlap. As with skills and competences, learning here may depend on three perspectives, the researcher, the librarian or informational professional and the archivist. Researchers may not see themselves directly involved in digital curation. However, they conceptualize the research project and associated data, and the intersection with data management planning. Obviously they create or receive data to analyze, they perform a kind of appraisal and selection determining what they will share or publish of a given project, and they allow access (e.g., usually to peers). Librarians and information professionals need to understand how and where the research lifecycle intersects the digital curation cycle, and when or where to work with archivists and preservation.

We have argued up to now that using the DCP Toolkit can be learned through hands-on application and trial and error. The Toolkit includes interview questions and a worksheet, along with suggestions for how to ask questions, what to focus on, and how to stay on track. By following general instructions on how to probe data needs and concerns, a practitioner can

demonstrate success at performing a set of associated tasks, which evidences learning. As there are no “right” answers, practitioners can (must) use trial and error to work through the interview process, collecting information and distilling information into a Profile. Trial and error is important because “[W]hen the problem space is too large to explore completely, a learning agent must have the ability to guess about new situations based on experience with similar situations” [6]. In previous workshops, several teaching methods were used: lecture, small group exercise, manual, one-to-one training, and workshops. The Framework helps understand how training will likely require using an online approach that integrates webinars, readings and videos and allowing learners to move at their own pace. Without funding for more workshops, it will be critical to use the Framework as a guide to create a tutorial in which there will be no expert leading the lessons.

#### V. CONCLUSION

There is a paradox that is highlighted by using the Framework in assessing this situation. The knowledge to be gained is about the researcher’s data (workflow, etc.), yet the skill is about learning to explore and understand the needs before being able to attend to them. As the gatekeeper, some researchers hold onto their data for reasons associated with lack of time to do more and lack of understanding how to organize and disseminate it. For the researcher to relinquish the gatekeeper goal, librarians must understand the researchers’ perspective, context, situation and needs. They must, in effect, learn to use a tool to learn about the researcher data, and then learn what can be done with it [7].

Trying to map the use of the DCP Toolkit to the Framework reveals some weaknesses and strengths in relying on the Toolkit itself to facilitate “self-training,” and a traditional training approach. The Framework is meant to be useful to those building new training courses, but can give insight into assessing for changes or using different approaches, which will be a next step for us. While the Framework doesn’t assist in providing a structure for a generic training program, it has helped identify the need to clarify perspectives, clarify which subjects should be covered, and where more specialized coursework addressing professional digital curation practice would be helpful. With further application, the Framework will help articulate better the value of learning how to use, and use, the DCP Toolkit.

#### REFERENCES

- [1] Data Curation Profiles Toolkit. <http://datacurationprofiles.org>
- [2] Data Curation Profiles Directory. <http://docs.lib.purdue.edu/dcp/>
- [3] [Data Curation Profiles Symposium, September 24, 2012, Purdue University, West Lafayette IN <http://docs.lib.purdue.edu/dcpsymposium/>
- [4] Carlson, J. “Demystifying the data interview: developing a foundation for reference librarians to talk with researchers about their data” *Reference Services Review* 40(1): 2012. 7-23. [http://docs.lib.purdue.edu/lib\\_research/153/](http://docs.lib.purdue.edu/lib_research/153/)
- [5] Brandt, D. S. Teaching Technology: A How-To-Do-It Manual for Librarians. How-To-Do-It Manuals for Librarians. Neal-Schuman Publishers, Inc., 100 Varick St., New York, NY, 2002
- [6] Lin, L. J. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning*, 8(3-4), 1992, 293-321.
- [7] Brandt, D. S. “Disambiguating the role of data lifecycle gatekeeper,” Workshop on Research Data Lifecycle Management, July 18-20, 2011, Princeton University, Princeton NJ.