A Competency Framework for Digital Curation and Data Science

Sheila Corrall
Professor and Chair, Library & Information Sciences Program
School of Information Sciences, University of Pittsburgh

Defining the human infrastructure needed for data-driven science and digital stewardship is a continuing challenge for the information and research communities because of the many different players involved in the scholarly knowledge cycle and the unsettled nature of the field, which have made it difficult to reach consensus on their roles, responsibilities, and relationships, their training and education requirements, and who should deliver programs and courses for all the potential recipients identified (including archive, library, and other information professionals; undergraduate, graduate, and postdoctoral researchers; research managers and citizen end-users). The educational preparedness of both specialists and non-specialists for dealing with digital data has been a recurring theme of high-level government publications, as well as papers and reports specifically concerned with the information disciplines.

Library and information science schools have responded to the needs identified with significant curricular developments, including individual courses, specialized concentrations and post-masters certificates – particularly in the USA, where the Institute of Museum and Library Services has invested around $10M in education and capacity building over a five-year period. However, educational development has not been uniform, with diversity in syllabus content, also reflected in the variety of course titles, and much of the data-related content across the sector has come from revision of existing digital library courses, rather than from new course development. Data curation education for future researchers in other disciplines has received less attention, but recent US initiatives point to opportunities for both information schools and library practitioners to extend their activities in this direction and confirm the desirability of aligning approaches.

The ANADP conference proceedings cite numerous projects across the globe that are working to improve education and training in data curation for students and practitioners, but also show how the multiplicity of efforts and resulting products have contributed to the problem and created the necessity to join up efforts and bring more coherence to such endeavors, to help stakeholders make sense of both needs and offerings. The proceedings list eight opportunities for alignment related to Education (including one under Standards and two added by Cliff Lynch). Two recommendations specifically call for definition and standardization of skillsets, and others identify the need to make it easier for employers and practitioners to match course offerings and learning outcomes with workplace activities and competency requirements.

The present proposal is to build on and consolidate existing valuable work (for example, the DCC Curation Lifecycle Model, DigCCurr Matrix of Knowledge and Competencies, DigCurV Evaluation Framework, DaMSSI Career Profiles, Vitae Researcher Development Framework), in order to develop a comprehensive competency framework for digital curation and data science. Competency frameworks have been developed for many organizations and some professions (e.g. information specialists in UK government agencies) and can be used to support recruitment, performance management and career planning, in addition to staff training and organization development. The proposed framework would specify the types and levels of skills, knowledge and understanding needed for particular roles, tasks, activities, etc, performed by information specialists, technologists, researchers, administrators and others across the whole scholarly knowledge lifecycle (the entire data, information and knowledge continuum). Including non-specialist roles would also support the need to reach out to other disciplines and the public.