Emerging opportunities in data curation, bioinformatics and data linkage provide new opportunities for data librarians and eResearch intermediaries. The role of academic libraries in supporting research is evolving, with novel library services under development and new collaborations between researchers and library staff emerging [1]. These changes have been prompted by new technology, growth in the quantity of digital research data being generated and collected, increasing interdisciplinary research and increasing recognition of data as an important research output. New positions and skills are required, for example in biocuration, big data analytics, high performance computing and data linkage. Increasingly, experienced staff are sought for roles as data librarians, data curators, data archivists and eResearch intermediaries. While formal training and qualifications are important, job rotation, mentoring and temporary secondments can enable library staff to develop practical experience and skills in these new and emerging areas [2]. We discuss a case study from UNSW Australia in which a data librarian was assigned to a multidisciplinary team within the Faculty of Medicine to help curate colorectal cancer data and make it publicly accessible via the tranSMART bioinformatics platform [3].

Case Study – Data Curator Role

Over the past decade, the University Library at UNSW has been involved in a range of repository projects and research data management initiatives designed to support digital curation and collaborative research [4]. In mid 2014, a data librarian from the Library was seconded to the Translational Cancer Research Network (TCRN) tranSMART project [5]. This project involved curating and loading colorectal cancer data from the Molecular and Cellular Oncology (MCO) study into an open-source bioinformatics platform. The project team consisted of an international group of biomedical researchers, software developers and bioinformaticians. The data librarian was involved with business analysis, developing use cases, and designing metadata models, as well as assisting with data curation, user acceptance testing and reporting. Skill sets required for these tasks were varied, encompassing both data curation and software development (See Figure 1). Familiarity with university data management policies and practices was key.

Challenges and Opportunities

In the TCRN tranSMART project, embedding the data librarian within the faculty enabled rapid immersion into the research environment. While developing the discipline and project-specific expertise necessary to contribute meaningfully to a data management project may be challenging, specialist subject librarians may have domain-specific knowledge that can be leveraged [1]. In this case, the data librarian had relevant previous experience from roles at the NSW Central Cancer Registry and the Department of Epidemiology and Preventive Medicine, Monash University. Project time-lines were also adapted to ensure sufficient time for background research and investigations. Other challenges included data harmonization, data access and communication across disciplines.

Specialised data librarians can offer a range of generic skills in eResearch, software development, and research data management, and typically bring an understanding of local policies and services. Acting as an eResearch intermediary - a translator between research and technical staff – is another valuable role. In addition, the traditional expertise of librarians in literature searching, bibliography management and bibliometrics remain highly valued by research staff [6]. Practicing data librarians have drawn attention to sometimes unrealistic expectations and mission creep [7, 8]. Yet many of the core abilities of information professionals – such as advanced metadata skills – remain pivotal across different domains.

Conclusion

Emerging roles in data curation and linkage offer new opportunities for specialized data librarians, working in multidisciplinary teams to deliver eResearch projects. Data librarians can offer a number of important skill sets that are readily transferrable across research disciplines, such as metadata, copyright, digital preservation and software development. A research project can benefit from the novel perspectives and skill sets of a data librarian. Embedding data librarians with researchers for specific eResearch projects can also develop tacit skills and knowledge around research data management.

Acknowledgements

This study is supported by the Australian National Data Service (ANDS) through the National Collaborative Research Infrastructure Strategy Program, as well as through the Cancer Institute NSW and UNSW Australia. We also wish to acknowledge the MCO Study Group, the TCGA Research Network and the participants in the MCO Study.

References