Chapter 26
Data Stewards, Curators, and Experts:
Library Data Engagement at Samuel J. Wood Library at Weil Cornell Medicine

Peter R. Oxley
Weill Cornell Medicine, USA

Sarah Ben Maamar
Weill Cornell Medicine, USA

Terrie Wheeler
Weill Cornell Medicine, USA

ABSTRACT

Research data management practices continue to evolve as data become more connected, distributed, granular, and continuous. While keeping their mandate to serve the academic community, libraries are seeking to foster knowledge creation, learning, and information access in an ever more data-driven and online environment. In this chapter, the authors describe their perspectives that shape the library's mandate, roles, and technology. They detail specific roles that have been created to foster data engagement: a secure enclave data core, an institutional data catalog, scientific software management, bioinformatics and research reproducibility support, data integrity, and community engagement. A constant focus on customer service has shaped each role, built the library's reputation, and contributed to high-impact services.

DOI: 10.4018/979-8-3693-3026-5.ch026

INTRODUCTION

As the 21st century marches on, academic libraries experience accelerating pressure to evolve beyond mere 'books and journals.' While keeping their mandate to serve the academic community, libraries are seeking to foster knowledge creation, learning, and information access in an ever more data-driven and online environment (Turner et al., 2013). Data management practices must adapt as data become ever 'bigger' - more connected, more distributed, more granular, and more continuous. Biomedical data has seen consistent development in the form of computable biomedical knowledge (Callahan et al., 2020; Walsh & Wroe, 2020) and Learning Healthcare Systems (Budrionis & Bellika, 2016). This chapter outlines the ways in which the Samuel J. Wood Library at Weill Cornell Medicine (henceforth Wood Library) has innovated services to engage with data science and researchers. The authors describe the perspectives that shape their mandate, roles developed, and technology employed.

Background

In 2014, the Samuel J. Wood Library offered traditional library services - primarily to medical and graduate students. A clinical librarian served one small population of clinicians. The budget had been cut by 15-20%, resulting in lost positions, a reduction in collections, and low morale. A new director was hired to lead a transformation to the library aligned with the tripartite mission of Weill Cornell Medicine (WCM): next-generation science, care, and education. The library director listened to every employee in the library, each faculty chairperson in the college, and key leaders in the research community. Incorporated into this listening tour was a scientific software survey, which helped the library director identify researcher software needs. These listening sessions revealed both the library's strengths in education and curriculum support and opportunities to expand its services to meet the needs of WCM researchers. One of the authors' key principles is to determine how the administrative burden of researchers can be reduced to benefit scientific inquiry. If the library can save researchers' time, it has provided a service of value (Wheeler & Holmes, 2017).

A vision emerged of the library as a resource not just for external information (books and journals), but also internal information (manuscripts, software, and scientific data). The concept of a Data Core as a protected environment for the computation of enormous sets of patient data, curated by data librarians, grew from these conversations with researchers. The Data Core was first conceived in 2015 and built on Information Technology Infrastructure Library (ITIL) principles, with security, redundancy, scalability, and customization in mind (Oxley et al., 2018). The Data Core provides library-curated projects for controlled access to computational resources and software for researchers to analyze sensitive data. These projects allow secure collaboration between researchers internally and outside our institution. Thus, the Data Core is a key component of the WCM data science ecosystem. From the beginning, the library partnered with WCM's Information Technology Service (ITS) to create a research environment that would enable better healthcare delivery practices. Initially, all tasks associated with the project lifecycle (onboarding, changes, and closure) were managed by librarians. As the Data Core service grew, an operational manager and data curation specialists were hired by the library to oversee established procedures and operational tasks. Thus today, librarians serve primarily as data custodians to actively de-identify data before export from the secure environment, or as architects to further develop the service.