

Title: Data Science and the Data Management Lifecycle: A Library Perspective

Presented by: Karen Vagts

Authors: Martha Kelehan, Regina Raboin, and Karen Vagts, Tisch Library, Tufts University

Abstract:

The recent emergence of “Data Science” (or “Big Data”) within universities and other research institutions emphasizes the need for the proper curation of data sets. Such curation is necessary if data is to be preserved, transferred, and made understandable and discoverable by multiple parties. Recent mandates from the U.S. Office of Science and Technology Policy (OSTP) as well as requirements from the NSF and other funding agencies for data management and dissemination as a condition of grant awards underscore the importance of a “lifecycle” approach to working with data.

Effective data lifecycle curation is achieved by implementing a systematic data management plan (DMP) involving six critical steps: 1) the development of data products; 2) selection of appropriate data formats; 3) application of metadata; 4) procedures for dissemination; 5) archiving and preservation both during and after the research process; and 6) consideration of privacy, IP, and other potential restraints on data sharing.

Numerous Tufts’ organizations are involved in data science and data management, among them Tisch Library, which supports data management through a range of activities, including research support, consulting on data management plans, collections, and metadata/classification services as well as guidance on selecting data repositories, whether at Tufts or elsewhere.