

Conducting a Data Management Audit of a Scientific Research Group

Introduction

As research-intensive science produces increasingly higher volumes of digital data, managing it with long term preservation in mind is more important than ever. Data management has become an often overlooked practice, taking a backseat to the science it supports (Heidorn, 2011; Carlson, Fosmire, Miller, & Nelson, 2011). Libraries and information science professionals can offer valuable assistance in managing that data, as they, historically, have strong skills related to the access and preservation of information and knowledge, such as classifying, describing, distributing, and storing (Pryor, 2012; Jones, Ball, & Ekmekcioglu, 2008).

One such way information science professionals can offer help to their scientific colleagues is by conducting an audit of their data management practices. By conducting an audit, the auditor can identify areas of strength, areas needing improvement, redundant activities, and useful tools. The goal of the data management audit is to increase efficiency and improve data management practices within a research group or project, which will, in turn, support the long term preservation and access of data sets. A data management audit supports the entire lifecycle of data, including Project Planning, Data Collection, Data Analysis, Data Archiving, and Data Distribution and Sharing.

This poster will highlight the skills an information professional needs and demonstrate a practical approach to performing a data management audit. Based on the Digital Curation Centre's Data Asset Framework (2009) and framed within the data lifecycle, this poster will explain the data management audit process and provide examples of questions asked, data workflow diagrams, and suggestions for improvements. Below are descriptions of the tasks involved in each phase of the audit.

Project Planning

During the project planning phase of the data management audit, the information science professional holds introductory meetings with science project managers. The purpose of these meetings is to determine which groups will be audited and be introduced to project personnel who will be participating in the data management audit. The auditor must obtain Internal Review Board approval, if necessary, before conducting any data gathering interviews.

Data Collection

During the data collection phase, the data collected includes, but is not limited to, surveys of project personnel; internal documents, such as reports and project logs; external documents, such as brochures and websites; and demonstrations of tools.

Data Analysis

The data analysis phase of the data management audit includes processing and analyzing the data collected from the previous phase. Tasks within this phase include transcribing interviews, studying each group's workflow, and extracting trends and finding gaps in those workflows. Additionally, the auditor should develop a workflow diagram that visualizes and describes the research process in each group audited. Color-coding each phase of the workflow process on the diagram facilitates easy comparison across research groups. Finally, during data analysis, the auditor must consistently consult with project personnel to determine if the workflow diagrams accurately represent the groups' processes.

Data Distribution & Sharing

Once the analysis is complete, the auditor must produce a final report that outlines key findings and lays out recommendations for improvements. This report is then shared with the project leaders and other project personnel who will determine how to proceed.

Implementation

Implementation of the recommendations contained within the data management audit report is the goal of the data management audit. The recommendations must be useful, reasonable suggestions to improve efficiency, reduce redundancy, and increase quality. Important areas where recommendations may be made include data management planning, metadata, archiving and preservation, and project tracking.

References

- Carlson, J., Fosmire, M., Miller, C., & Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries and the Academy* 11(2), 629-657. The Johns Hopkins University Press.
- Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51(7-8), 662-672. doi: 10.1080/01930826.2011.601269
- Jones, S., Ball, A., & Ekmekcioglu, Ç. (2008). The data management audit framework: A first step in the data management challenge. *International Journal of Digital Curation*, 3(2).
- Pryor, G. (2012). Why manage research data?. In G. Pryor (Ed.), *Managing research data* (p. 15). London: Facet Publishing.