

Digital Curation at the Library of Congress: Lessons Learned from American Memory and the Archive Ingest and Handling Test

Liz Madden, Library of Congress, OSI/IM/DCS, 101 Independence Ave SW, Washington DC 20540, emad@loc.gov

Abstract

The Library of Congress (LC) created the American Memory (AM) website in 1995 to provide public access to digitized versions of materials from the LC collections. The technical architecture, programming, and bulk of AM content were created before the development or widespread application of digital library metadata and data standards. The staff that produced much of this digital content did so without tools designed specifically for this type of work and with no existing model of a web-based digital library for guidance. Twelve years later the data and metadata driving AM are managed in the same framework in which they were created, and the effects of digital-curation decisions made early on are coming to light in efforts to sustain, use, and share this valuable body of content. The retrospective view provides an opportunity to identify approaches to data that contributed to the long-term sustainability of the contents, and approaches that were detrimental.

In 2003, LC staff applied many of these lessons learned to the analysis of a donated digital archive. The archive comprised diverse formats and metadata collected on the web from different donors. The LC analysis provided further confirmation of good and bad practices with regard to the maintenance of digital materials, and helped refine the list of useful skills and practices for digital curation staff “in the trenches.” This paper discusses the findings and conclusions of analyses of AM and the donated digital archive as they pertain to digital curation. It also suggests opportunities for digital curation curriculum development.