Using and Developing with Open Source Forensics Software in Digital Archives Programs

MARK A. MATIENZO

Abstract: Archivists, librarians, curators, and academic researchers have begun investigating the application of digital forensics tools and methodologies to the processes of transfer, accessioning, analysis, and management of born-digital archival holdings (see, e.g., Kirschenbaum, Ovenden, and Redwine, 2010). The high cost and complexity of many proprietary digital forensics applications are a large concern for the cultural heritage sector. While the digital forensics community has had some interaction with the cultural heritage sector, in many ways that sector an emerging market to which the digital forensics community is still adjusting. These two factors have led a small but motivated group within libraries and archives to use and develop with open source software and frameworks for digital forensics (see, e.g., BitCurator Project, 2012).

This presentation is investigates the use of open source digital forensics software, including The Sleuth Kit, fiwalk, and bulk_extractor, in the context of digital archives programs at the Yale University Library (YUL). The YUL Manuscripts and Archives department and the Beinecke Rare Book and Manuscript Library have been collaborating semi-formally on developing shared infrastructure and workflows to support the accessioning, processing, and management of born-digital material, including legacy material received on physical media. In addition to analyzing how YUL has used these existing tools, we will also present the ways in which we have leveraged the ease of development on top of these open source applications, frameworks, and libraries. These tools include Gumshoe, a presentation interface for filename- and metadata-layer data based on Ruby on Rails and Solr.

About the author:

Mark A. Matienzo is a Digital Archivist, Manuscripts and Archives, Yale University Library, and is the Technical Architect of the ArchivesSpace project. Matienzo served as lead digital archivist for the Andrew W. Mellon Foundation-funded AIMS project ("Born-Digital Collections: An Inter-Institutional Model for Stewardship"), a partnership between Yale University, University of Virginia, Stanford University, and University of Hull. In addition, he is an adjunct professor at the iSchool at Drexel University, where he teaches digital preservation, and has taught at the Palmer School of Library and Information Science at Long Island University. He received a Master of Science in Information from the University of Michigan and a BA in Philosophy from the College of Wooster. More information about his projects can be found online at http://matienzo.org/