Integrating Digital Forensics Metadata into Archival Management Systems

PORTER OLSEN, KAM WOODS, AND ALEX CHASSANOFF

Abstract: Over the last two years, members of the BitCurator project have been developing digital forensics tools for use by libraries, archives and museums as they process born-digital collections. A key outcome of this work has been the BitCurator Environment, a suite of open source digital forensics tools capable of analyzing both raw bitstreams and electronic files to extract items of interest and file-level metadata from disk images. The data and metadata exported by these digital forensics tools (e.g. bulk_extractor, fiwalk) includes detailed PII analysis, file name, file type, MAC records (modified, accessed, created times), notation of deleted files, and cryptographic hashes for each individual file. This rich collection of metadata is originally output as Digital Forensics XML (DFXML) and then transformed by the BitCurator Reporting tool into editable file formats including Excel (.xlsx).

For archivists to fully utilize file system metadata encoded as DFXML, we are developing crosswalks (full and partial) to metadata standards commonly used in archival managements systems. The addition of this information is intended to simplify accession of digital collections, along with improving finding aids.

This poster first describes the metadata produced by the digital forensics tools incorporated into the BitCurator Environment, and then explores how that metadata can be transformed and packaged for use in digital repository systems. We expect that archivists will use facets of DFXML and other BitCurator produced metadata in many interesting ways, and we hope to begin this much-needed conversation by offering our research and insight into digital forensics metadata and its potential integration into archival management systems.

About the authors:

Porter Olsen is a research faculty member at the Maryland Institute for Technology in the Humanities (MITH) where he is the Community Lead on the BitCurator project, a Mellon funded project to bring digital forensics tools and techniques to collecting institutions working with born-digital material. Simultaneously, he is a Ph.D. candidate in the English Department at the University of Maryland studying digital humanities and postcolonial literature. His dissertation, under the direction of Dr. Matthew Kirschenbaum, is titled "Hacking the Empire: Reading the Digital in Twentieth and Twenty-First Century Postcolonial Literature" and is a study of how hacker culture is represented in later 20th and early 21st century postcolonial fiction. Porter teaches classes on electronic literature and globalization and has been recognized for his contributions to the growing online teaching program in the university's English department. Before returning to graduate school, Porter worked as product manager on the United Linux initiative, an effort to create a single Linux platform shared among distributors from Germany, Brazil, the U.S., and Japan.

Kam Woods is a Postdoctoral Research Associate in the School of Information and Library Science at the University of North Carolina at Chapel Hill. He is currently Technical Lead on the BitCurator project, and works with Dr. Cal Lee developing techniques and tools to assist in long-term archiving of born-digital data. Kam's research focuses on long-term preservation of born-digital materials. He is interested in interdisciplinary approaches that combine technologies and expertise in the areas of archiving, computer science, and digital forensics for the purpose of enabling and maintaining access to digital objects that are at risk due to obsolescence. Prior to his current work at UNC, Kam worked with Cal Lee on the development of educational materials to support the use of realistic forensic datasets in professional training and to identify and explore novel uses of forensic data and tools in the context of digital archives.

Alex Chassanoff is a Ph.D. candidate in the School of Information and Library Science (SILS) at the University of North Carolina at Chapel Hill. Her research interests include the information behavior of digital archive users, digital materiality, and curation and preservation environments. Her current research looks at how historians use digitized photographs as evidence. Alex received her MSIS in 2009 and was a Digital Curation Fellow at SILS from 2008-2009. Prior to graduate school, she worked as a database programmer, IT consultant, and digital asset manager.