Identification and Redaction of Sensitive Information in Born-digital Archival Materials: Research and Development Directions

CHRISTOPHER (CAL) LEE

Abstract: Archivists must often identify information within their collections (either in their existing holdings or as part of new acquisitions) that is sensitive and should thus be removed, closed, restricted or filtered from public view. The process of manually identifying such information can be time-consuming and prone to error. Luckily, there are many automated methods that archivists can use to support this work.

Computer-assisted redaction is still a relatively undeveloped area of archival practice and there are numerous opportunities for further research and development. In this presentation, I will discuss several of the open research questions and propose future directions to address them.

Among other issues, I will discuss the potential impacts of block-level vs. file-level redaction of sensitive data. The former can be implemented much more efficiently and over a wider set of data types (including unallocated space on disks and disk images), but it has the potential to hinder the mounting and navigation of file systems, as well as the ability to use and render files. I will discuss potential implications for archival workflows. I will also place these research questions within the wider context of applying "more product less process" (MPLP) approaches to born-digital archival materials.

About the Author:

Christopher (Cal) Lee is Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches archival administration; records management; digital curation; understanding information technology for managing digital collections; and digital forensics. He is a lead organizer and instructor for the DigCCurr Professional Institute, and he teaches professional workshops on digital forensics methods and principles.

Cal's primary area of research is digital curation. Cal developed "A Framework for Contextual Information in Digital Collections," and edited and provided several chapters to I, Digital: Personal Collections in the Digital Era.

Cal is Principal Investigator (PI) of BitCurator Access and was PI of BitCurator. He was also PI of the Digital Acquisition Learning Laboratory (DALL) project, is Senior Personnel on the DataNet Federation Consortium, and has served as Co-PI on several digital curation education projects.