Levels of Representation in Digital Collections: A Framework and Implications for Archival Research

CHRISTOPHER LEE

Abstract: Any use of digital resources is highly mediated. It involves the interaction of various hardware and software components. Computer professionals frequently describe hardware and software as a stack of independent but interacting layers. Archivists who are responsible for digital collections need similar ways to coordinate their work, channel their actions and reduce the complexity of their professional decisions.

I will present a framework of digital information that is based on eight levels of representation—from the bitstream on a physical medium up to aggregations of digital objects. I will argue that each level reflects specific and important implications for professional decisions and actions by archivists who are responsible for digital resources. The properties of information at a given level of representation are directly based upon, but are not fully reducible to, properties of information at the level immediately below it. Each level has emergent properties, which convey potential meaning that is not available through any of the other layers. This is because moving between layers always involves a process of translation that both adds and removes information.

I will offer a proposed set of implications for future research related to digital archives. Some of these implications relate to further elaborating fundamental archival concepts. Others relate to archival description and metadata conventions. I believe that the framework also raises significant considerations related to digital preservation strategies and professional education. Perhaps most importantly, I will argue that there are major, unavoidable issues of professional ethics.

About the author:

Christopher (Cal) Lee is Associate Professor at the School of Information and Library Science at the University of North Carolina, Chapel Hill. He teaches courses on archival administration; records management; digital curation; understanding information technology for managing digital collections; resource selection and evaluation; and acquiring information from digital storage media. He is a lead organizer and instructor for the DigCCurr Professional Institute, a week-long continuing education workshop on digital curation, and he teaches professional workshops on the application of digital forensics methods and principles to digital acquisitions.

Cal's primary area of research is the long-term curation of digital collections. He is particularly interested in the professionalization of this work and the diffusion of existing tools and methods (e.g. digital forensics, web archiving, automated implementation of policies) into professional practice. Cal recently developed "A Framework for Contextual Information in Digital Collections" (Journal of Documentation). He also edited and provided several chapters to I, Digital: Personal

Collections in the Digital Era published by the Society of American Archivists.

Cal was Principal Investigator of the Digital Acquisition Learning Laboratory (DALL) project, which investigated and tested the incorporation of digital forensics tools and methods into digital curation education. He also served as Co-PI on several projects focused on the preparation of professions to take on digital curation responsibilities: Preserving Access to Our Digital Future: Building an International Digital Curation Curriculum (DigCCurr), DigCCurr II: Extending an International Digital Curation Curriculum to Doctoral Students and Practitioners; Educating Stewards of Public Information for the 21st Century (ESOPI-21), Educating Stewards of the Public Information Infrastructure (ESOPI2), and Closing the Digital Curation Gap (CDCG). In a project called Curation of a Forensic Data Collection for Education, Cal is investigating and developing resources to enhance access and use of disk images created to support digital forensics education.