

Five Repositories, One Dataset: Using Exploratory Data Analysis Techniques to Track Patterns of Use.

MARK CUSTER, NOAH HUFFMAN, JENNIE LEVINE KNIES, KYLE RIMKUS, and SARA SNYDER

Abstract: On November 14, 2005, Google Inc. announced that its newly-released web analytics service, Google Analytics, would be provided free of charge. Since that time it has been implemented by websites of all sorts and sizes, making it one of the most popular methods for measuring web traffic currently available. In the archival community, for instance, users of Google Analytics range from state-wide consortium websites, like the Online Archive of California, to historical society websites, like Emma, “the catablog of Archives, Manuscripts & Special Collections at the Brooklyn Historical Society.”

An auspicious consequence of Google Analytics’ popularity is that many archives are now in the possession of uniform datasets. In hopes of taking advantage of this, Mark Custer sent an email to the EAD listserv on October 28, 2010, asking for volunteers to share their Google Analytics’ reports with one another. That inaugural call has resulted in a total of five repositories sharing read-only access to their online finding aids datasets. In alphabetical order, those five repositories are: Archives of American Art, Duke University, East Carolina University, University of Maryland, and the University of Miami.

We will present the approach that we have taken to investigate this multi-institutional dataset, which has been guided by Exploratory Data Analysis (EDA) principles. The presentation will include: visualizations of the use patterns discovered; indications about whether these patterns exist in the aggregate or repository-specific sense; and recommendations about which metrics reported by Google Analytics are the most important for archivists to analyze.

About the authors:

Mark Custer is the EAD Manager at the Smithsonian Institution. Previously, he worked at East Carolina University’s Joyner Library as Text and Markup Coordinator and Digital Archivist. He earned a BA in English Literature from Indiana University, Bloomington, and an MLIS from Syracuse University.

Noah Huffman is the Archivist for Metadata and Encoding at Duke University’s David M. Rubenstein Rare Book & Manuscript Library. Previously, he worked in the Southern Historical Collection at UNC-Chapel Hill and at the Filson Historical Society in Louisville, Kentucky. He earned a BA in History from Furman University, an MA in History from the University of Louisville, and an MSLS from UNC-Chapel Hill.

Jennie Levine Knies is the Manger of Digital Collections at the University of Maryland, College Park. She has a BA in Russian Area Studies and an MLS in Library Science/Archival Management, both from the University of Maryland, College Park.

Kyle Rimkus is the Preservation Librarian at the University of Illinois at Urbana-Champaign. Prior to that, he was Head of Digital Scholarship and Programs at the University of Miami Libraries. He has Master's degrees in Library and Information Science as well as in French Literature from the University of Illinois at Urbana-Champaign, and a Bachelor's in Germanic Studies from the University of Illinois at Chicago.

Sara Snyder has been webmaster at the Archives of American Art since 2007. She holds an MLS in archival studies and an MA in American history from the University of Maryland, College Park.