

Collaborating to Catalog: Addressing a Digital Backlog

JESSE JOHNSTON and GRACE THOMAS

Abstract: This presentation will discuss current collaborative efforts to increase description, access, and discovery of web archives at the Library of Congress (LC). The work poses particular challenges due to the LC's selective approach to creating web archives, with a broad cohort of collecting units that create collections around varied themes, events, and creator communities. LC has maintained a Web Archiving program since 2000, and varying practices and priorities have shaped the initiative. Over time, a backlog of under-described resources has accrued. Aside from a standard one-year embargo period, most of these sites are available for onsite LC users to browse, but only a few are discoverable through the Library's search platforms. As LC aims to grow its Web archiving effort, staff have explored strategic collaboration and automation to describe web archives. The major collaborators in this effort have been the Web Archiving Team (WAT), within the newly created Digital Collections Management section (DCM), and catalogers from the Acquisitions and Bibliographic Access (ABA) division. Since February 2018, these entities have forged new intra-institutional partnerships to describe web archives: an ABA metadata librarian sits with the WAT, processing records for the backlog and providing guidance on best practices for a future automated cataloging workflow. In this presentation, DCM and WAT representatives will report on the pilot collaboration, position it in relation to the broader discussion on descriptive metadata for web archives, describe the possibility to improve discovery layers for web archives, and comment on the potential to shape new and semi-automated descriptive records for digital content.

About the Authors:

Grace Thomas became a Digital Collections Specialist for the Library of Congress Web Archiving Team in August 2016. Currently, many of her tasks revolve around strategizing digital means through which to open up and understand the various formats and content within the Library's 1.3 petabyte web archive. Thomas has a background in Digital Libraries through her masters work at Indiana University and prior experience with Digital Humanities projects at the University of Nebraska-Lincoln.

Jesse Johnston is a Senior Digital Collections Specialist at the Library of Congress and an adjunct faculty member at the University of Maryland iSchool. He works on policy for digital content management at the Library of Congress in the Digital Collection Management section, which intersects with the Library's special collections units, resource description, and the Web Archiving Team. Previously, he was an Archives Specialist in the Ralph Rinzler Folklife Archives and Collections at the Smithsonian Center for Folklife and Cultural Heritage and worked on metadata and digital description for archival audio collections. From 2013 to 2018, he

was a Senior Program Officer for preservation and access and Acting Records Officer at the National Endowment for the Humanities (NEH). Web: <http://www.jessejohnston.net/>