Selecting and Managing Content Captured From the Web: Expanding Curatorial Expertise and Skills in Building Library of Congress Web Archives

Abbie Grotke, Digital Media Project Coordinator, Office of Strategic Initiatives, Library of Congress, Washington, D.C. 20540-1310, abgr@loc.gov
and
Janice E. Ruth, Manuscript Specialist in American Women's History, Manuscript Division, Library of Congress, Washington, D.C. 20540-4680, jrut@loc.gov

Abstract

Since 2000, the Library of Congress, as part of its MINERVA program, has been capturing Web sites and developing thematic, event-driven Web archives on such topics as the terrorist attacks of September 11, 2001, national elections, and the Iraq War. A small number of Library staff worked on these projects and gained expertise in crawling technologies, tools, and workflow. In June 2005, Library managers sought to increase the number and expertise of staff involved and to extend the collecting scope beyond event-based Web captures. Staff from the Office of Strategic Initiatives (OSI) partnered with recommending officers, curators, specialists, and processing staff in various Library Services (LS) divisions to conduct a pilot project titled Selecting and Managing Content Captured from the Web (SMCCW). The initiative involved twenty-five staff members working on four distinct projects to develop selection criteria for Web content; examine how technical capabilities enable, affect, or prevent the building of Web site collections; and document the activities required to ensure the continued viability of the content. This paper describes the SMCCW initiative. Project manager Abbie Grotke summarizes the Library's pre-2005 Web archiving activities before presenting an overview of the goals, organization, training needs, and accomplishments of the SMCCW effort. As one of the four team leaders, Janice E. Ruth discusses the project's implementation in the Manuscript Division, describing the staff's appraisal and quality review of captured Web sites, how the work was integrated into existing division priorities and donor relationships, and the skills needed to complete the project.