Abstract

The lead applicant of this project is the George Washington University (GW) Libraries. We propose to enhance a prototype application developed at the GW Libraries in 2012 for collecting social media data from Twitter to meet diverse academic research, teaching, and library collection development needs. In this way, we propose to address the goal described in the 2012-2016 IMLS Strategic Plan to support “exemplary stewardship of museum and library collections” and to use “technology to facilitate discovery of knowledge and cultural heritage.”

Social media data, like that produced in Twitter, is a rich source for studying how we communicate and exchange ideas on a global scale. There is considerable interest in collecting this material for academic use; a keyword search for “twitter” in ProQuest Dissertations & Theses, for example, returns over 5,800 results, roughly 88% of which are from the year 2010 or later, thereby demonstrating the scholarly use of Twitter in a range of disciplines. Because most social science researchers and students are not trained in software development and working with application programming interfaces (APIs), however, collecting data can be a painstaking manual process for them. Additionally, like all Internet content, data from Twitter may disappear at any time. Further, although Twitter makes new data available for free from its API, acquiring older data requires purchases from third party vendors that can be cost-prohibitive. The prototype application developed at the GW Libraries automates the collection of data from Twitter’s API and makes it possible for scholars, students, and librarians to identify, select, collect, and preserve Twitter data for research purposes at little to no marginal cost.

During this project, we will develop our prototype Social Feed Manager application into a well-documented tool that can be implemented and used at diverse cultural heritage institutions. This application extends the domain of our stewardship, enabling implementation of innovative services to select, collect, and make locally available at-risk data from Twitter. To meet this objective and to broaden its potential impact, we will collaborate with colleagues at the University of North Texas, Yale University, the Center for Jewish History, and other cultural heritage organizations to ensure that the application can meet their diverse institutional needs.

The primary audience and potential users of this application will include staff in peer research libraries and archives, information technology, and academic support divisions of universities, who will use the application both to serve their local communities of scholars and students and to meet their own internal collection development objectives. By the end of the grant period, we will have a robust, reliable application, implemented and tested in multiple institutions, that is documented and available for use, study, copying, and modification under a free and open source software license. With this license, the Social Feed Manager software application will also be available for use by independent scholars to facilitate their own research data collection.

We anticipate that the project will extend from September 2013 to August 2014.