

“Digging into Data Challenge” Competition Winners December 2009

Digging into Image Data to Answer Authorship-Related Questions

(Peter Ainsworth, University of Sheffield, JISC; Peter Bajcsy, University of Illinois at Urbana-Champaign, NSF; *Dean Rehberger, Michigan State University, NEH*). This project will pursue research using advanced computational techniques to explore humanities themes related to the authorship of large collections of cultural heritage materials, namely 15th-century manuscripts, 17th- and 18th-century maps, and 19th- and 20th-century quilts.

Digging into the Enlightenment: Mapping the Republic of Letters

(*Dan Edelstein, Stanford University, NEH*; Chris Weaver, University of Oklahoma, NSF; Robert McNamee, University of Oxford, JISC). This project will focus on a body of 53,000 18th-century letters and analyze the degree to which the effects of the Enlightenment can be observed in the letters of people of various occupations.

Railroads and the Making of Modern America—Tools for Spatio-Temporal Correlation, Analysis, and Visualization

(*William Thomas, University of Nebraska-Lincoln, NEH*; Richard Healey, University of Portsmouth, JISC). This project will integrate a vast collection of textual, geographical, and numerical data about the railroad and its impact on society over the centuries, concentrating initially on the Great Plains and Northeast United States.

Towards Dynamic Variorum Editions

(*Gregory Crane, Tufts University, NEH*; John Darlington, Imperial College, London, JISC; Bruce Robertson, Mount Allison University, SSHRC). This project supports the creation of a framework to produce "dynamic variorum" editions of classic texts that enable the reader to automatically link not only to variant editions but also to relevant citations, quotations, people, and places that are found in a digital library of more than one million primary and secondary source texts.

Using Zotero and TAPoR on the Old Bailey Proceedings: Data Mining with Criminal Intent

(*Daniel Cohen, George Mason University, NEH*; Tim Hitchcock, University of Hertfordshire, JISC; Geoffrey Rockwell, University of Alberta, SSHRC). This project will develop tools and models for comparing, visualizing, and analyzing the history of crime, using the Old Bailey Online, which contains extensive court records of more than 197,000 individual trials held over a period of 240 years in Great Britain.

Harvesting Speech Datasets for Linguistic Research on the Web

(Mats Rooth, Cornell University, NSF; Michael Wagner, McGill University, SSHRC). This project will pull together audio and transcribed data from podcasts, news broadcasts, public and educational lectures, and other sources to create a comprehensive repository of everyday speech. This speech will be used by linguistic researchers to learn more about prosody—the study of rhythm, stress, and intonation—in everyday speech.

Mining a Year of Speech

(Mark Liberman, University of Pennsylvania, NSF; John Coleman, University of Oxford, JISC). This project focuses on large scale data analysis of audio—specifically the spoken word. The scholars will create tools to enable rapid and flexible access to more than 9,000 hours of spoken audio files, containing a wide variety of speech, drawn from some of the leading British and American spoken word corpora, allowing for new kinds of linguistic analysis.

Structural Analysis of Large Amounts of Music Information

(Stephen Downie, University of Illinois at Urbana-Champaign, NSF; David De Roure, University of Southampton, JISC; Ichiro Fujinaga, McGill University, SSHRC). This project will gather approximately 23,000 hours of digitized music representing a wide range of styles, regions, and time periods. The goal is to develop tools to tag and analyze the underlying structures of this music, resulting in a body of world music that will provide music scholars with interactive access to previously unavailable analysis and insights.