

FOR IMMEDIATE RELEASE

News Release

Leading International Research Agencies Announce Winners of Prestigious New “Digging into Data Challenge” Competition

(Ottawa, December 3, 2009) – The Honourable Gary Goodyear, Minister of State (Science and Technology), announced today the winners of a new international competition, the Digging into Data Challenge, which is a partnership between four leading research agencies: the Social Sciences and Humanities Research Council (SSHRC), of Canada; the National Endowment for the Humanities (NEH), of the United States, the National Science Foundation (NSF), of the United States, and the Joint Information Systems Committee (JISC), of the United Kingdom. The Digging into Data Challenge promotes innovative humanities and social science research using large-scale data analysis.

“Our government supports science and technology to create jobs, improve the quality of life of Canadians and strengthen the economy,” said Minister Goodyear. “Supporting international research partnerships helps our universities develop, attract and retain the world’s best research talent here in Canada.”

“This exciting joint initiative with NEH, NSF and JISC enables Canadian researchers to further develop sophisticated text and image mining and data visualization technologies while forging international research partnerships,” said SSHRC President Dr. Chad Gaffield. “The results of these projects will build new knowledge that crosses disciplines from the vast digital resources now available to researchers.”

“Trying to manage a deluge of data and turn bits of information into useful knowledge is a problem that affects almost everyone in today’s digital age,” said U.S. National Endowment for the Humanities Chairman Jim Leach. “With this international grant program, NEH is hoping to seed projects that will not only benefit researchers in the humanities, but also lead to shared cultural understanding.”

The 22 successful candidates make up eight international research teams, composed of scientists and scholars using advanced computational techniques for humanities and social sciences research. Each team includes researchers from at least two of the participating countries. The teams will demonstrate how data mining and data analysis tools currently used in the sciences can improve scholarship in the humanities and social sciences. Total project funding by all four agencies amounts to C\$2,082,926.



Detailed descriptions of the eight winning projects are attached.

-30-

SSHRC is the federal agency that promotes and supports university-based research and training in the humanities and social sciences. Through its programs and policies, the Council enables the highest levels of research excellence in Canada and facilitates knowledge-sharing and collaboration across research disciplines, universities and all sectors of society. More information is available at www.sshrc-crsh.gc.ca.

Created in 1965 as an independent federal agency, the NEH supports learning in history, literature, philosophy and other areas of the humanities. NEH grants enrich classroom learning, create and preserve knowledge, and bring ideas to life through public television, radio, new technologies, museum exhibitions, and programs in libraries and other community places. Additional information is available at www.neh.gov.

NSF is an independent federal agency that supports fundamental research and education across all fields of science and engineering, with an annual budget of \$6.06 billion. NSF funds reach all 50 states through grants to over 1,900 universities and institutions. Each year, NSF receives about 45,000 competitive requests for funding, and makes over 11,500 new funding awards. NSF also awards more than \$400 million in professional and service contracts yearly. More information is available at www.nsf.gov.

JISC is a joint committee of the U.K. further and higher education funding bodies and is responsible for supporting the innovative use of information and communication technology to support learning, teaching, and research. It is best known for providing a U.K. national infrastructure network, a range of support, content and advisory services, and a portfolio of high-quality resources. More information is available at www.jisc.ac.uk.

Ce document est également disponible en français.

For additional information on this release and other SSHRC research projects, please contact:

Trevor Lynn
Communications Manager
Tel.: 613-992-7302
Cell: 613-302-9879
Email: Trevor.Lynn@sshrc-crsh.gc.ca



Backgrounder

The Digging into Data Challenge is an international grant competition launched in January 2009 by four leading research agencies: the Joint Information Systems Committee (JISC) from the United Kingdom, the National Endowment for the Humanities (NEH) from the United States, the National Science Foundation (NSF) from the United States, and the Social Sciences and Humanities Research Council (SSHRC) from Canada. This event is to announce the winners of the first competition who have been selected by rigorous peer review.

The advent of what has been called “data-driven inquiry” or “cyber-scholarship” has changed the nature of inquiry across many disciplines, including the sciences and humanities, revealing new opportunities for interdisciplinary collaboration on problems of common interest. The creation of vast quantities of Internet accessible digital data and the development of techniques for large-scale data analysis and visualization have led to remarkable new discoveries in genetics, astronomy, and other fields, and – importantly – connections between academic disciplinary areas. New techniques of large-scale data analysis allow researchers to discover relationships, detect discrepancies, and perform computations on data sets that are so large that they can be processed only using computing resources and computational methods developed and made economically affordable within the past few years. With books, newspapers, journals, films, artwork, and sound recordings being digitized on a massive scale, it is possible to apply data analysis techniques to large collections of diverse cultural heritage resources as well as scientific data. How might these techniques help scholars use these materials to ask new questions about and gain new insights into our world?

To encourage innovative approaches to this question, four international research agencies launched this joint grant competition in January 2009 to focus the attention of the social science and humanities research communities on large-scale data analysis and its potential application to a wide range of scholarly resources.

The goals of the initiative are:

- to promote the development and deployment of innovative research techniques in large-scale data analysis;
- to foster interdisciplinary collaboration among scholars in the humanities, social sciences, computer sciences, information sciences, and other fields, around questions of text and data analysis;
- to promote international collaboration; and
- to work with data repositories that hold large digital collections to ensure efficient access to these materials for research.

Applicants were required to form international teams from at least two of the participating countries. Winning teams will receive grants from two or more of the funding agencies and will be invited to present their work at a special conference. These teams, which may be composed of scholars and scientists, will be asked to demonstrate how data mining and data analysis tools currently used in the sciences can improve humanities and social science scholarship. The hope of this competition is that these projects will serve as exemplars to the field and encourage new, international partnerships among scholars, computer scientists, information scientists, librarians, and others.



The eight winning teams include:

- 1. 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.
- 2. 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.
- 3. 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 compare and visualize instances of crime, analyzing them for criminal and non-criminal violence.
- 4. Towards Dynamic Variorum Editions** (Gregory Crane, Tufts University, NEH; John Darlington, Imperial College, London, JISC; Bruce Robertson, Mount Allison University, SSHRC). This project will establish an important resource for classical scholars by developing a range of tools for making dynamic comparisons, generating lexicons, identifying topics and extracting quotes from over 10,000 Greek and Roman texts.
- 5. Digging into Image Data to Answer Authorship Related Questions** (Dean Rehberger, Michigan State University, NEH; Peter Bajcsy, University of Illinois at Urbana-Champaign, NSF; Peter Ainsworth, University of Sheffield, JISC). This project will take three specific resources and develop tools to analyze and identify the authorship of visual images.
- 6. 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 speech. Tools will then be developed to analyze this communication.
- 7. 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 over the centuries, concentrating initially on the Great Plains and Northeast United States.
- 8. Mining a Year of Speech** (Mark Liberman, University of Pennsylvania, NSF; John Coleman, University of Oxford, JISC). This project will create tools to enable rapid and flexible access to over 9,000 hours of spoken audio files, drawn from some of the leading British and American spoken word corpora.

Further information can be found at www.diggingintodata.org.