

# *Digging into Data Challenge*

## **Request for Proposals**

**Version 1.0, Modified 16 January, 2009**

### **I. Program Description**

#### **General Overview of the Digging into Data Challenge**

The advent of what has been called “data-driven inquiry” or “cyberscholarship” 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, artworks, 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 organizations are organizing a joint grant competition 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.

In recognition of the international nature of cyberinfrastructure/e-science, the Digging into Data Challenge will bring together international research teams to advance research and to share their results openly, so that others may learn from them.

The Digging into Data Challenge competition is sponsored by four leading funders from three countries:

- The Canadian Social Sciences and Humanities Research Council (SSHRC);
- The U.K. Joint Information Systems Committee (JISC);
- The U.S. National Endowment for the Humanities (NEH); and
- The U.S. National Science Foundation (NSF).

This Request for Proposals (RFP) explains how to apply to the Digging into Data Challenge. Please note that each funder has also produced an RFP Addendum with information specific to that funder. Each funder will have different rules and requirements, depending on its mission and its policies and procedures. Therefore, before you apply you should carefully read this RFP as well as the RFP Addenda for any of the funders to which you wish to apply. All of these documents are available on the [Digging into Data Challenge Web site](http://www.diggingintodata.org) (www.diggingintodata.org).

This competition is open only to international partnerships involving at least two of the three participating countries: Canada, England or Wales in the United Kingdom, and the United States. Each project must be sponsored by at least two research teams from at least two different countries. (Please consult the RFP Addendum for each funder, as there may be eligibility differences in terms of what constitutes a "research team.") Each research team must have a project director. The partners will collaborate to write a single application package, which will be sent to (at least) two funding agencies.

Here are some hypothetical examples of eligible partnerships:

- 1) a research team in the U.S. and a team in England might apply to NEH and JISC for funding. Both NEH and JISC would review their application and, should it be chosen for funding, two awards would be made: NEH would fund the U.S. team and JISC the English team.
- 2) a research team in England and a team in Canada might apply to JISC and SSHRC for funding. Both JISC and SSHRC would review their application and, should it be chosen for funding, two awards would be made: JISC would fund the English research team and SSHRC the Canadian team.
- 3) Two teams in the U.S. and one in Canada, might apply to NSF, NEH, and SSHRC for funding. NSF, NEH, and SSHRC would review their application and, should it be chosen for funding, three awards would be made: NSF would fund one U.S. team, NEH would fund the second U.S. team, and SSHRC the Canadian team.

All potential applicants should carefully read both the RFP and the appropriate RFP Addenda.

Note that all applications are judged as an international team. All successful applications will receive two (or more) awards.

### **Providing Access to Grant Products**

The funders of the Digging into Data Challenge endeavor to make the products of this grant program available to the broadest possible audience. All funded projects will be expected to do the following:

- 1) prepare a final project white paper. Your white paper should describe in detail the results of your research. It should discuss how your project progressed over time, and how you managed it; document meetings and important milestones; describe lessons learned (both positive and negative); document any software, algorithms, or techniques that you developed; discuss your success in addressing your research question; and provide your candid opinions about the success of the project overall. The white paper will be posted on the Internet by one (or more) of the funders so that others may benefit from your research.
- 2) make freely available to the public any software or other products created as a result of the project.
- 3) attend a conference sponsored by the funders at the end of the grant period. Project directors will be asked to present the results of their research to the funders and to other attendees.

### **Choosing a Data Repository**

The Digging into Data Challenge seeks to demonstrate data analysis that draws from a very large set of data. To take the example of books, what can be learned by searching thousands or millions of books that can't be learned by a close reading of one? As such, it is important that your project make use of a large dataset. The four funding agencies have compiled a list of large data repositories that have expressed an interest in making their datasets available and have included technical support contacts. This list of data repositories can be found on the [Digging into Data Challenge Web site](#). You are not, however, required to use a dataset from this list of repositories for your project.

## **II. Award Information**

The grant period is for fifteen months, with projects being completed by March of 2011.

Each applicant must provide a separate budget for each funding partner. For example, if your team consists of an English research team and a Canadian research team, you would supply two budgets, one describing the Canadian budget and one describing the English budget. Award amounts are discussed in the individual RFP Addenda for the various funders. **Projects must include a project director and applicant organization from each country, so that funds may be distributed according to respective national laws and funding guidelines.**

### III. Eligibility

Applicants must apply as international research partnerships (a minimum of two teams, from two different countries, with two project directors, one from each country). Additional eligibility requirements for each of the funding agencies are explained in the RFP Addenda.

Ineligible applications will not be reviewed.

### IV. Application and Submission Information

In order to apply, interested potential applicants must first send a letter of intent by March 15, 2009. Final applications will be due July 15, 2009.

#### **Letter of Intent**

By March 15, 2009, the applicant must submit a letter of intent via the competition Web site, [www.diggingintodata.org](http://www.diggingintodata.org). The letter should be no more than three pages. It should be saved as a PDF file and submitted as per the instructions on the competition Web site. In the letter, the applicant should briefly address the following:

- 1) What are the names and countries of the research teams and the project directors?
  
- 2) Why is the project intellectually significant? What questions will be addressed through large-scale data analysis, or through the development of middleware or other software? How will the project benefit research? How will it serve as an exemplar for data-driven scholarship?
  
- 3) What dataset will you use, and how will you gain access to it? Is this dataset big enough to demonstrate what can be achieved using large amounts of data?

4) Who are the members of the project team, and why are they qualified to accomplish the project goals?

Funding agency representatives will read and respond to each letter of intent. We will provide advice and feedback about your project no later than May 1, 2009.

### **Final Application**

Final applications must be received by July 15, 2009. The final application should be submitted via the competition Web site, [www.diggingintodata.org](http://www.diggingintodata.org). Please read the RFP Addenda for more information.

### **Application Contents:**

**All of the following documents should be saved as PDF (Portable Document Format) files prior to uploading via the competition Web site:**

- 1) Statement of Significance. Provide a one-page abstract of the project, written for a general audience, that explains the significance of the project. **Please clearly indicate the names and countries of the project directors and their research teams.**
- 2) Table of Contents. List all parts of the application and, beginning with the narrative, number all pages consecutively.
- 3) List of Participants. On a separate page, list in alphabetical order, surnames first, all project participants and collaborators, and their institutional affiliations. This list should include—when relevant—advisory board members, consultants, and authors of letters of commitment and letters of support.
- 4) Narrative. The narrative should be a maximum of eight single-spaced pages. **Important: Review the RFP Addenda for the appropriate funders to ensure that your narrative covers any funder-specific requirements.**

In the narrative, please discuss the following:

a) Discuss how the project takes advantage of the large scale of the chosen digital dataset. How does the large scale effectively change the research paradigm? How does it allow for scholarship that could not be done on a small scale?

b) Describe the partnership. Explain why this project can successfully be undertaken by the research teams from the different countries that have formed a partnership. What strengths does each partner bring to the project?

c) How will the final product be used? Who would be its likely users?

d) Describe in detail the dataset chosen for the project. Describe what the dataset contains and how it is structured. Describe your means of accessing the data (e.g., via Application Programming Interface [API], Web services, etc). Are the data local or remote? Are the data freely accessible, or is there a charge to use the data? Discuss any intellectual property or privacy issues that might affect the availability of the materials. In an appendix, provide letters from the data repository, indicating permissions and proof of informed consent, if appropriate.

e) Provide a concise history of the project, including information about preliminary research or planning, financial support already received, and resources or research facilities available. If a project requires more than fifteen months to complete, describe the scope and duration of the entire project, but show clearly the specific accomplishments or products intended in the fifteen-month period for which funding is requested.

f) Describe the technology and methodologies used in the project, and make a case for your choices. Discuss your choice of technology. Explain if you are using new technology or repurposing existing tools or algorithms. Detail your development methodology.

g) Describe standards used. Project activities should conform to appropriate global standards and accepted professional practices. If the project methodology departs from usual standards and procedures, explain why the project's goals require this approach and how the results would be interoperable with other relevant resources that follow existing standards.

h) Environmental Scan. Provide a clear and concise summary of an environmental scan of the relevant field. The goal of an environmental scan is to call attention to similar work being done in the area of study. For example, if you are developing software, please discuss similar software developed for other projects and explain how the software proposed for this project differs. If there are existing software products that could be adapted and re-used for the proposed project,

please identify them and discuss the pros and cons of taking that approach. If there are existing projects that are similar in nature to your project, please describe them and discuss how they relate to the proposed project. The environmental scan should make it clear that you are aware of similar work being done; it should explain how the proposed project contributes to and advances the field.

i) Provide sample materials, when and as appropriate. In addition to or instead of sample materials, applications should include screen shots or reports that show the final or anticipated form of the project or illustrate the experience of the project's staff in doing comparable work.

5) Budget. Include the appropriate budget form for each of the individual funders to which you are applying, indicating the funds being requested from each of these funders. Please see the appropriate RFP Addenda to download the budget forms.

6) Résumés. Attach résumés for the principal investigators and major participants.

7) Letters of Support. Attach letters of support and/or letters of commitment.

8) Appendices. Attach any relevant samples or other materials critical to your project.

## V. Application Review

Reviewers will apply the following criteria in assessing applications:

**Project Aims:** What intellectual contribution will the project make? How will it increase understanding? How innovative is the project? Will it serve as a model for future work?

**Project Plans:** Is the project methodology sound? Does it adhere to accepted standards and professional practices? Is the work plan (including the ways in which the project staff and equipment will be employed) well thought out? Is the project staff well qualified? Is the project budget reasonable?

**Technology Plans:** Does the project make innovative use of technology? Are the chosen technologies and proposed development methodologies appropriate?

**Partnership:** Does the proposal describe an effective international partnership? Is the partnership likely to extend beyond the funding period?

**Open Access and Dissemination:** Will the project provide adequate access to grant products? Will it effectively disseminate the project outcomes?

### **Review and Selection Process**

Reviewers outside the funding agencies will read each application and advise the funding agencies about its merits. Final funding decisions will be made by each funding agency according to its own rules and procedures. Please see the individual RFP Addenda for more information.

Late applications will not be reviewed.

## **VI. Award Administration Information**

### **Award Notices**

Applicants will be notified by e-mail in December 2009. At that time grants administrators and project directors of successful applications will also receive award documents by mail. Applicants may obtain the evaluations of their applications by sending an e-mail to [did@neh.gov](mailto:did@neh.gov).

Other award administration information for specific funders may be found in each funder's RFP Addendum.

## **VII. Points of Contact**

General E-Mail Address for the Competition: [did@neh.gov](mailto:did@neh.gov)

Special Web site for the competition: [www.diggingintodata.org](http://www.diggingintodata.org)

Specific Contact Information for Each Funder is available in each funders's RFP Addendum.