



Digging into Data Challenge

Second Year

Request for Proposals

Final Version, 16 March 2011

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, 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 have led to remarkable new discoveries in genetics, astronomy, and other fields, and—importantly—connections between different academic disciplines. The Digging into Data Challenge seeks to discover how these new research techniques might also be applied to questions in the humanities and social sciences.

New techniques of large-scale data analysis allow researchers to discover relationships, detect discrepancies, and perform computations on so-called “big data” sets that are so large that they can be processed only by using computing resources and computational methods that were developed and made economically affordable within the past few years. This “data deluge” has arisen not just from the capture and storage of data on everyday transactions such as Internet searches, consumer purchases, cell phone records, “smart” metering systems and sensors, but also from the digitization of all types of media, with books, newspapers, journals, films, artworks, and sound recordings being digitized on a massive scale. It is possible to apply data linkage and analysis techniques to large and diverse data collections, including survey data, economic data, digitized newspapers, books, music, and other scholarly and scientific resources. How might these techniques help researchers use these materials to ask new questions about and gain new insights into our world? To encourage innovative approaches to this question, eight international research organizations are organizing a joint grant competition to focus the attention of the social sciences, humanities, library, archival, and information sciences communities on large-scale data analysis and its potential applications.

The four goals of the initiative are

- to promote the development and deployment of innovative research techniques in large-scale data analysis that focus on applications for the humanities and social sciences;
- to foster interdisciplinary collaboration among researchers in the humanities, social sciences, computer sciences, library, archive, information sciences, and other fields, around questions of text and data analysis;
- to promote international collaboration among both researchers and funders; and
- to ensure efficient access to and sharing of the materials for research by working with data repositories that hold large digital collections.

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 eight leading funders from four countries (Canada, the Netherlands, the United Kingdom and the United States):

- The Social Sciences and Humanities Research Council of Canada (SSHRC);
- The Netherlands Organisation for Scientific Research (NWO);
- The UK Joint Information Systems Committee (JISC);
- The UK Arts and Humanities Research Council (AHRC);
- The UK Economic and Social Research Council (ESRC);
- The US Institute of Museum and Library Services (IMLS);
- The US National Endowment for the Humanities (NEH); and
- The US 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 their respective rules, requirements, funding mandates, policies, and procedures. Please consult the appropriate addenda prior to applying. All of these documents are available on the [Digging into Data Challenge website](#).

This competition is open only to international research projects. Each project represents a collaboration among 2-4 teams, each team representing one of **the competition's** participating countries. (See Section II "Eligibility" for more details.)

Once received, the Digging into Data applications will be reviewed by an international peer review committee. (See Section V "Application Review and Adjudication" for more information on the peer review process.)

The DiD Challenge is an open competition, soliciting applications from researchers in the information, library, archival, and computational sciences as well as the humanities and the social sciences. A successful application is likely to be one which addresses the goals of the DiD initiative (innovative research applied to large scale datasets, effective

interdisciplinary collaboration, and improving access to and sharing of data for work in the humanities and/or social sciences). Examples of projects funded in the first year of the competition can be found on the [Digging into Data Challenge website](#).

Providing access to grant products and research outcomes

The funders of the Digging into Data Challenge endeavor to make the products and research results of this grant program available to the broadest possible audience. All funded projects will be expected to prepare a final project white paper and make freely available to the public any software or other products created as a result of the project:

- The 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 Digging into Data Challenge website, so that others may benefit from your research. The white paper will be due ninety days after the end of the grant period. The white paper should be about as long as a typical academic paper in your discipline. A white paper could in fact be a pre-press version (or an early draft) of an academic paper. More details will be provided at a later date. The text of your white paper may also be used to satisfy the reporting requirements of your funding agency, so there is no extra burden of time or resources for reporting the results of your project. Please consult the Addenda for individual funders for details. However, please bear in mind that the paper will be a public document.
- Copies of any code developed under the grant should be placed in an appropriate repository.
- At the end of the grant period, the funders plan to sponsor a major conference to highlight all the funded projects. Each project will be asked to participate and share the results of their research.
- Each funder will require that awardee institutions (those that receive funds from that agency) adhere to any special reporting requirements of that funding agency.

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 cannot be learned by a close reading of one? As such, it is important that your research project make use of a large dataset. The eight 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 website](#). You are not, however, required to use a dataset from this list of repositories for your project.

II. Eligibility

Applicants must apply as an international research project partnership. Each project is a partnership among two to four national teams. Each team represents one of the four nations participating in the Digging into Data Challenge (Canada, the US, the UK, or the Netherlands). Each national team must be led by an eligible institution (e.g. a university) with a designated principal investigator.

If more than one institution from the same country is participating, they must work together and designate one of the institutions as the team “lead.” Teams can only receive grants from a funder from their own country. However, please note that each funder has its own restrictions on the eligibility of potential applicants. Please read the appropriate RFP Addenda or contact us if you have questions about eligibility.

Here are some hypothetical examples of eligible project partnerships:

- A project involves a US team and a UK team. The US team is led by an American university and the UK team is led by a UK university. If the project were chosen for funding, the US university would receive an award from a US funder (i.e., IMLS, NEH, or NSF), while the UK university would receive an award from JISC (acting on behalf of the three UK funders, AHRC, ESRC, and JISC).
- A project involves a US Team and a Dutch team. The US team consists of two American universities, one of which is designated as the “lead.” The Dutch team is led by a university. If the project were chosen for funding, the US team’s lead university would receive a grant from one of the US funders (i.e., IMLS, NEH, or NSF), while the Dutch university would receive an award from the Dutch funder (i.e. NWO).
- A project involves a US team, a Canadian team, and a UK team. The US team is led by an American university. The Canadian team is led by a Canadian university. The UK team consists of two UK-based universities, one of which is designated as the “lead.” If the project were chosen for funding, the US university would receive a grant from a US funder (i.e., IMLS, NEH, or NSF), the Canadian university would receive an award from the Canadian funder (i.e. SSHRC) and the UK team’s lead university would received a grant from JISC (acting on behalf of the three UK funders, AHRC, ESRC, and JISC).

Applications that are late, incomplete, and/or ineligible will not be reviewed.

III. Award Information

The grant period will range between six and twenty-four months. Projects must be completed by January 31, 2014.

Each project is a partnership among two or more national teams, from at least two of the four participating countries. When a project is selected for funding, each of the teams will receive a grant.

- For Canadian teams, the award amount will range between CAN \$25,000 and \$125,000.
- For UK teams, the award amount will range between GBP £15,000 and £100,000. If the UK team consists of two or more institutions, the maximum award is increased to £150,000.
- For US teams, the award amount will range between US \$25,000 and \$125,000. If the US team consists of two or more institutions, the maximum award is increased to \$175,000.
- For Dutch teams, the award amount will range between EUR €17,000 and €100,000.

Each project can be awarded up to a maximum of four grants.

When you apply, you need not stipulate to which funding agency you are applying. All funded teams will receive grants from one of their own nation's funders. It is recommended that applicants consult the appropriate RFP Addenda to familiarize themselves with each funder's eligibility rules. For example, a Canadian team must ensure that its application meets the requirements of SSHRC, the only Canadian funder. By contrast, more than one funder might provide a grant to US and UK teams. In these cases, the nature of the work proposed will typically determine which funder makes a grant (e.g., if a US team's project is in philosophy, it is more likely to be funded by NEH, whereas if the project is in economics, it is more likely to be funded by NSF). But since some projects may be eligible for support from multiple funders (e.g., a US team's linguistics project might receive a grant from either NEH or NSF), applicants should consult all the relevant Addenda, to ensure that they are aware of eligibility rules and any other funder-specific requirements. Applicants with questions may wish to consult with program staff (please see Section VII Points of Contact).

In order to give applicants a better notion of the funding levels, below is an estimate of the amount of money that each participating funder is planning to make available for this competition:

- SSHRC: CAN \$1,500,000
- NWO: EUR €200,000
- JISC: GBP £500,000
- AHRC: GBP £250,000
- ESRC: GBP £250,000
- IMLS: US \$800,000
- NEH: US \$1,000,000
- NSF: US \$650,000

Based on these levels, and assuming an average award of US\$125,000 and an average of 2.5 teams undertaking each project, this would mean approximately 50 total grants being made to 18 projects. This estimate is subject to the quality of the applications submitted as well as the total grant amount requested.

Funds will be distributed to each awardee according to respective national laws and **each funder's internal policies and procedures. Please see appropriate addenda for more details.**

IV. Application and Submission Information

Final applications must be received by 23:59 (Greenwich Mean Time) on June 16, 2011 and must be submitted via the competition website, <http://www.diggingintodata.org/>.

Applications that are late, incomplete, and/or ineligible will not be reviewed.

It is assumed that the majority of applications will be submitted in English due to the language of each of the participating funding organizations. However, for Canadian teams who will be supported by SSHRC and who wish to submit the application in French on behalf of their international team, please consult the SSHRC Addendum for further information on steps for submission.

Application contents:

The application consists of a cover sheet and nine separate documents that will be uploaded by the applicant to the competition website.

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

- 1) **Statement of significance.** (one page) Provide an abstract of the project, written for a general audience, which explains the significance of the project. Please clearly indicate the names and countries of the principal investigators and their research teams.
- 2) **Table of contents.** List all parts of the application and number all pages consecutively.
- 3) **List of participants.** 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.
- 4) **Narrative.** (maximum of eight single-spaced pages). All pages should have one-inch margins, and the font size should be no smaller than eleven point. In the narrative, please discuss the following:

4a. Delineate and discuss your research questions and objectives. Some projects will primarily address substantive scholarly research questions. Others will primarily address infrastructure, information science, or methodological research questions that have the potential to create new avenues for future scholarly research. Some projects will address both kinds of questions. Explain why these research questions are important and will advance knowledge and understanding in the humanities or social sciences.

4b. 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?

4c. 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?

4d. Describe in detail the data chosen for the project. Describe what these data contain and how they are 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 the Letters of Commitment section below, provide letters from the data guardians, indicating permissions and proof of informed consent, if appropriate.

4e. Provide a concise history of the project, including information about preliminary research or planning, financial support and/or in-kind contributions already received, and resources or research facilities available. If a project would not be completed during the grant period, describe the scope and duration of the entire project, but show clearly the specific accomplishments that would be completed during the grant period.

4f. 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.

4g. 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.

4h. Describe how your research project will assist in the training of graduate students and newer-to-the-field researchers on your team.

4i. 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 reused 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.

4j. To support the narrative, provide sample materials in an appendix, when and as appropriate. In addition to or instead of sample materials, applications should include in an appendix 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. Since the applications will be read by peer reviewers online, we encourage you to include URLs when possible.

4k. Innovative research often involves elements of risk. Discuss potential risks to your project and risk-mitigation strategies that you would employ.

4l. The Digging into Data funders require that all funded research be conducted in accordance with relevant ethical principles and be approved by the relevant ethical authorities. Please consult individual RFP addenda for more information on each funder's ethics requirements.

5) **References cited.** Please use this attachment for all references cited.

6) **Budget.** Each national team that is applying must submit a budget. For example, if your project has a US team, a UK team, and a Canadian team, you would submit three budgets. The budget should be prepared using a spreadsheet program, using the example budgets listed below as a guide. Prior to uploading your budget to the competition website, please convert it to an Adobe PDF format.

- Canadian Budget Example [on [Digging into Data website](#)]
- US Budget Example [on [Digging into Data website](#)]
- Dutch Budget Example [on [Digging into Data website](#)]
- UK Budget Example [on [Digging into Data website](#)]

At the end of the grant period, the funders plan to sponsor a major conference to highlight all the funded projects. In your budget(s), please include funds for your principal investigators to travel to this conference. As the exact location has not been set yet, we recommend assuming it will be an international trip of approximately three nights.

Please note that if your project is selected for funding, you may be asked to provide additional funding documentation to clarify items in the budget.

7) **Résumés.** Attach résumés for the principal investigators and major participants. You may use the format of your choosing. Regardless of format, please ensure that the résumé indicates all institutional affiliations, as this is required to ensure there are no conflicts of interest with peer reviewers. The maximum length for a résumé is two pages.

8) **Letters of commitment.** Attach letters of commitment, as appropriate. A letter of commitment is typically written by a person or organization that is committing something to your project: for example, giving you access to a collection of materials for your research or agreeing to make some kind of contribution to your project.

9) **Appendices.** (Maximum of five single-spaced pages). Attach any relevant samples or other materials critical to your project.

V. Application Review and Adjudication

Reviewers will apply the following criteria in assessing applications:

Relevance to the challenge: Does the project promise to meet the goals of the Digging into Data Challenge?

Project aims: What intellectual contribution will the research 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 research 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) sufficiently outlined? Is the project staff well qualified? Is the project budget reasonable?

Technology plans: Does the research 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? Does the partnership have an appropriate management or governance plan?

Open access and dissemination: Will the project provide adequate access to grant products? Where appropriate, does it bring outside knowledge into the project? Will it effectively disseminate the project outcomes?

Value for money: Is the project likely to make effective and efficient use of the requested funds? Is there an institutional commitment to the project, beyond the funds requested from the funding bodies?

Review and selection process

Once applications have been received, they will be distributed to peer review panels based on disciplines and areas of research represented as well as methods used in each application. The funders will then jointly agree on a group of peer reviewers to be assigned to each panel, chosen from the scholarly and scientific community and other experts, where appropriate. Only peer reviewers who are free from conflicts of interest with the senior personnel on the proposals will be used. The total number of peer reviewers will depend on the number of applications received. However, the funders will choose the reviewers with the goal of forming a group drawn equally from the eight funders. Each panel will be co-chaired by a member of the scholarly community, along with a program officer from one of the funders. The panel chairs will not participate in the review. Rather, their job will be to answer questions about the process and ensure that each application is thoroughly discussed.

Prior to a scheduled face-to-face meeting, the peer reviewers will initially read and rate each application via the Internet. This first stage will not eliminate any applications but will serve to calibrate the applications and identify those applications with the most potential. Then, at a later date, the peer reviewers will meet face-to-face for final panel sessions in which they will discuss the applications with one another, focusing their deliberations on the applications that received high ratings during the first stage. At the end of these sessions, the peer reviewers will provide final ratings and comments to the funders. The ratings will be a common five-point scale (Excellent, Very Good, Good, Some Merit, Not Competitive). The peer reviewers will use the review criteria described above in assigning their ratings. The criteria are not weighted -- panelists will be asked to take each into consideration during their deliberations before arriving at a final rating. The panelists will sign a confidentiality document prior to the start of the panel meeting to ensure that they understand and will comply with the competition's rules regarding confidentiality.

At the conclusion of all the face-to-face peer review meetings, the funders will meet to put together the slate of recommended applications. In finalizing this slate, the guiding principle will be scientific or scholarly merit; that is, proposals with the highest ratings from the peer reviewers will be recommended for funding. In some cases, notably when two or more proposals receive identical ratings, the funders will also take into account the entire funding portfolio, with an eye toward creating a "balanced portfolio," addressing such matters as value for money, diversity of what is being funded, and institutional, disciplinary, and geographical balance.

Final funding decisions will be made by each funding agency, according to its own rules and procedures. In some cases, funders may require teams to submit a copy of the proposal or other materials directly to the funder prior to making an award.

Please see the individual RFP Addenda for more information.

Applications that are late, incomplete, and/or ineligible will not be reviewed.

VI. Award Administration Information

Award notices

Applicants will be notified by e-mail in December, 2011. Grants administrators and project directors of successful applications will receive award documents by December 2011.

Applicants may obtain the evaluations of their applications by sending an e-mail message to 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

Digging into Data Challenge Program Officer Contacts:

The Social Sciences and Humanities Research Council of Canada (SSHRC)
Gail Zboch, Gail.Zboch@sshrc-crsh.gc.ca, +1-613-943-1148

The Netherlands Organisation for Scientific Research (NWO)
Rosemarie van der Veen-Oei, catch@nwo.nl, (+31) 70-3440851

The UK Joint Information Systems Committee (JISC)
Alastair Dunning, a.dunning@jisc.ac.uk, +44 (0)203 006 6065

The UK Arts and Humanities Research Council (AHRC)
Pam Mason, p.mason@ahrc.ac.uk, +44 (0)1793 416063

The UK Economic and Social Research Council (ESRC)
Michael Bright, michael.bright@esrc.ac.uk, +44 (0)1793 413042
Audrey Sharp, audrey.sharp@esrc.ac.uk +44 (0)1793 413150

The US Institute of Museum and Library Services (IMLS)
Joyce Ray, jray@imls.gov, +1-202-653-4660
Chuck Thomas, cthomas@imls.gov, +1-202-653-4663

The US National Endowment for the Humanities (NEH)
Brett Bobley, bbobley@neh.gov, +1-202-606-8401
Jennifer Serventi, jserventi@neh.gov, +1-202-606-8395
Hearing-impaired applicants can contact NEH via TDD at 1-866-372-2930.

The US National Science Foundation (NSF)
Elizabeth Tran, etran@nsf.gov, +1-703-292-5338

Special website for the competition: <http://www.diggingintodata.org/>

Please note: While program officers are available to answer general questions about the grant program, they aren't available to read and respond to draft applications.