



**Digging into Data Challenge**  
**Round Three Request for Proposals**

I. Program Description

**Overview of the Digging into Data Challenge**

The Digging into Data Challenge aims to address how "big data" changes the research landscape for the humanities and social sciences. Now that we have massive databases of materials available for research in the humanities and the social sciences—ranging from digitized books, newspapers, and music to information generated by Internet-based activities and mobile communications, administrative data from public agencies, and customer databases from private sector organizations—what new, computationally-based research methods might we apply? As the world becomes increasingly digital, new techniques will be needed to search, analyze, and understand these materials. Digging into Data challenges the research community to help create the new research infrastructure for 21st-century scholarship.

Now going into the third round of the competition, the Digging into Data Challenge has funded a wide variety of projects that explore how computationally intensive research methods can be used to ask new questions about and gain new insights into our world. To encourage innovative research from across the globe, Digging into Data is sponsored by ten international research funding organizations that are working together to focus the attention of the social sciences, humanities, library, archival, information, computer, mathematical, and statistical science communities on large-scale data analysis and its potential applications.

The four overarching goals of the Digging into Data are to

- promote the development and deployment of innovative research techniques in large-scale data analysis that focus on applications in the humanities and social sciences;
- foster interdisciplinary collaboration among researchers in the humanities, social sciences, computer sciences, library science, archival science, information sciences, mathematical and statistical sciences, engineering, and other fields, around questions of text and data analysis;
- promote international collaboration among both researchers and funders; and

- 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 ten 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) in collaboration with the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Foundation for Innovation (CFI);
- The Netherlands Organisation for Scientific Research – Humanities, Physical Sciences and Social Sciences (NWO) in collaboration with The Netherlands eScience Center (NLeSC)
- The UK Arts and Humanities Research Council (AHRC);
- The UK Economic and Social Research Council (ESRC);
- The UK Joint Information Systems Committee (Jisc);
- 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 its 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 will involve teams from two to four of the participating countries.** (See Section II “Eligibility” for more details.)

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.)

### **New Considerations for Round Three**

In this third round of the competition the funders encourage projects that address some of the findings from the earlier rounds. Potential applicants are therefore invited to take into account a special report—produced by the Council on Library and Information Resources (CLIR)—that focused on the outcomes from the first round. This report, [One Culture](#).

[Computationally Intensive Research in the Humanities and Social Sciences](#), contains specific recommendations to researchers, funders, libraries, and institutions. Applicants are asked to show how their projects address some of the recommendations from this report. In particular, please see the section entitled “Recommendations” found on pages 2-6 of the CLIR report.

### **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 do the following:

- Submit a final project “white paper.” 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; propose indicators to measure success; 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. Please bear in mind that the paper will be a public document.
- Deposit copies of any code developed under the grant into an appropriate repository. (How you preserve your code should be described in your Data Management Plan. Please see below for the instructions for the eighth attachment in Section IV. Application and Submission Information for more details on the content of your DMP.)
- Attend an end-of-grant Digging into Data Challenge conference. At the end of the grant period, the funders plan to host a major conference to highlight all the funded projects. Members from each project will be asked to participate and share the results of their team’s research.
- Submit a Project Management, Dissemination, and Communications (PMDC) plan that describes your management and public outreach activities. (Please see below for the instructions for the ninth attachment in Section IV. Application and Submission Information for more details on the content of your PMDC plan.)

### **Choosing a data repository**

The Digging into Data Challenge seeks to demonstrate data analysis methods that draw 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 ten 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 Netherlands, the US, and the UK). Each national team must be led by an eligible institution (for example, 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 receive grants only 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 (that is, 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 Dutch university. If the project were chosen for funding, the US team’s lead university would receive a grant from one of the US funders (that is, IMLS, NEH, or NSF), while the Dutch university would receive an award from the Dutch funder (that is, NWO).
- A project involves a US team, a Canadian team, and a UK team. The US team is led by an American library. 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 library would receive a grant from a US funder (that is, IMLS, NEH, or NSF), the Canadian university would receive an award from the Canadian funders (SSHRC, NSERC and CFI or

SSHRC only) and the UK team's lead university would receive a grant from Jisc (acting on behalf of the three UK funders, AHRC, ESRC, and Jisc).

Late, incomplete, or ineligible applications will not be reviewed.

### III. Award Information

The grant period will range between twelve and twenty-four months. Projects must be completed by January 31, 2016.

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 up to \$250,000 (SSHRC and NSERC) or up to \$750,000 (with CFI infrastructure funding).
- 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 €25,000 and €100,000.

When you apply, you need not stipulate to which funding agency you are applying. All funded teams will receive grants from one of their nation's funders. Applicants should consult the appropriate RFP addenda to familiarize themselves with each funder's eligibility rules. For example, a Dutch team must ensure that its application meets the requirements of the Dutch funder as stated in the addendum. By contrast, more than one funder might provide a grant to Canadian, US, and UK teams. In these cases, the nature of the work proposed—particularly the subject matter—will typically determine which funder makes a grant. (For example, 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.) Since some projects may be eligible for support from multiple funders (for example, a US team's linguistics project might receive a grant from either NEH or NSF, a Canadian team's project that advances both computer science and history might receive grants from NSERC, SSHRC and CFI), 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:

- CFI: CAN \$500,000
- NSERC: CAN \$400,000
- SSHRC: CAN \$1.0M
- NWO/NLeSC: EUR €50,000
- AHRC: GBP £500,000
- ESRC: GBP £500,000
- IMLS: US \$450,000
- NEH: US \$700,000
- NSF: US \$650,000

In the previous round (2011), 67 teams applied to the program and 14 received awards, for a funding ratio of 21 percent.

**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. Each funder will require that awardee institutions (those that receive funds from that agency) adhere to any special reporting requirements of that funding agency.**

#### IV. Application and Submission Information

##### **Application advice**

Applicants may wish to examine examples of projects funded in prior rounds of the competition. These examples can be found on the [Digging into Data Challenge website](#).

Please note that program officers are available to answer general questions about the grant program but are not able to read and respond to draft applications.

Final applications must be received by 23:59 (Greenwich Mean Time) on May 15, 2013 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 most applications will be submitted in English. Canadian teams that will be supported by SSHRC, NSERC and CFI and wish to submit an application in French should consult the SSHRC or NSERC or CFI addenda for further information.

##### **Application contents**

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

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

1) **Statement of significance** (one page, single spaced) 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) **Narrative** (maximum of ten 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:

3a. Delineate and discuss your research questions and objectives. Some projects may primarily focus on substance: answers to scholarly research questions in the humanities or social sciences. Others, however, may primarily focus on methodology: proposed advances in infrastructure, information science, or research techniques used for large data sets that may facilitate future scholarly research in the humanities or social sciences. Some projects will address both kinds of questions. Explain why your research questions are important and how they will advance knowledge and understanding in the humanities or social sciences.

3b. 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?

3c. Discuss how the project addresses the overarching goals of the Digging into Data Challenge as described in this RFP in the section entitled “General Overview of the Digging into Data Challenge” starting on page 1.

3d. Discuss how your project addresses one or more of the recommendations from the CLIR report, as described in this RFP in the section entitled “General Overview of the Digging into Data Challenge” starting on page 1. Please cite the page in the report on which the recommendation is mentioned.

3e. Describe the partnership and the importance of collaboration. 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? Discuss the interdisciplinary nature of your partnership and how this

particular collaboration will make for better research. What are the potential challenges, and how will they be addressed? For example, how will you collaborate? What model might you adopt to share resources and credit?

3f. 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 (for example, 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 attachment (see below for the instructions for the seventh attachment), provide letters from the data guardians, indicating permissions and proof of informed consent, if appropriate. In the Data Management Plan attachment (see below for the instructions for the eighth attachment), describe the project's plan for sustaining the data created during the grant period after the grant funding ends.

3g. 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. If a project has received previous Digging into Data awards, please detail the accomplishments during the previous grant period and indicate how this request for support builds on and extends the previous research questions explored.

3h. 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.

3i. 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.

3j. Describe how your research project will assist in the training of students and newer-to-the-field researchers on your team.

3k. 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.

3l. To support the narrative, provide sample materials in an appendix, when and as appropriate. Applications should also include—in the appendices attachment, see below for the instructions for the tenth attachment—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.

3m. 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.

4) **References cited** Please use this attachment to provide citations for all references in the narrative. This attachment should be no more than two pages.

5) **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 at least three budgets. (Note for Canadian teams: Depending on the work being performed, you may need to submit multiple budgets, particularly if you are requesting additional infrastructure funds from CFI. Please refer to the NSERC, SSHRC, and CFI addenda for details.) Budgets should be prepared using a spreadsheet program; the budget examples below are offered as guides. Prior to uploading budgets to the competition website, please convert them to an Adobe PDF format.

- Canadian budget example [on [Digging into Data website](#)]
- Dutch budget example [on [Digging into Data website](#)]
- UK budget example [on [Digging into Data website](#)]
- US 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 location has not yet been determined, we recommend assuming that 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.

6) **Résumés** Attach résumés for the principal investigators and major participants. You may use a format of your choosing. Regardless of format, please ensure that the résumé indicates all institutional affiliations, to ensure that peer reviewers face no conflicts of interest in any of the proposals that they evaluate. The maximum length for a résumé is two pages.

7) **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.

8) **Data Management Plan (DMP)** Prepare a data management plan for your project. The plan should describe how the project team will manage and disseminate data generated or collected by the project. For example, projects in this category may generate data such as software code, algorithms, digital tools, data sets, reports, articles, research notes, or websites. The DMP should be no more than two pages.

9) **Project Management, Dissemination, and Communications Plan (PMDC)**  
Prepare a Project Management, Dissemination, and Communications Plan to which all project partners agree. The PMDC should be no more than five pages. The PMDC should contain three sections:

9a. Roles and responsibilities. Briefly describe how the project will be managed, including reporting and decision making. List all members of the project team(s), their roles and contact details. Indicate the time allocated by those members to the project in days/hours over the course of the project.

9b. Work plan. Describe the project's key aims and objectives and your work products (for example, publications, software, etc). Include a month-by-month work plan that describes tasks and milestones. Discuss what "success" means for these milestones and how you plan to measure it. Also discuss possible risks or barriers to success, their likelihood, and how you plan to avoid or mitigate them. These risks might be related to time (for example, staff time, length of project), resources (for example, money, materials), assumptions/expectations, etc.

9c. Dissemination and communication. Identify the target audiences for your project's work products, how you will engage with them, and how they will benefit. Describe how you would communicate project results (for example, through web pages, social media outreach, training or mentoring opportunities, events, videos, or publications), should the project receive an award. Applicants should keep in mind that outreach is an important part of the Digging into Data Challenge, as one goal is to bring new research methods to the larger field. Please note that the funders encourage all resulting publications to be available via open

access (and any projects receiving UK funding must comply with the RCUK Policy on Access). Also note that, at a minimum, all awarded projects must have a project website.

10) **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 overarching goals of the Digging into Data Challenge?

**Consideration of recommendations from CLIR report:** Does the project appropriately address one or more of the recommendations described in the CLIR report on the Digging into Data Challenge in a way that might serve as a model for the field?

**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? Will it provide a resource that other researchers can build on?

**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) sufficiently detailed? Is the project staff well qualified? Is the project budget reasonable?

**Technology plans:** To what extent does the project make innovative use of technology and/or propose the development of new tools? 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? 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. Peer reviewers will recuse themselves from evaluating any proposals with respect to which they have conflicts of interest . 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 ten 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 online. This first stage will not eliminate any applications but will serve to calibrate the applications and identify those applications with the most potential. 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 use 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; in other words, panelists will be asked to assign equal weight to each criterion during their deliberations. 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 a slate of recommended applications.

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.

## VI. Award Administration Information

### **Award notices**

Applicants will be notified by e-mail in January, 2014. Grants administrators and project directors of successful applications will receive award documents by January 2014.

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

Digging into Data Challenge Program Officer Contacts:

### **The Canada Foundation for Innovation (CFI)**

Laurent Messier, [Laurent.Messier@Innovation.ca](mailto:Laurent.Messier@Innovation.ca), +1 613 943-1251

### **The Natural Sciences and Engineering Research Council of Canada (NSERC)**

Kristina Archibald, [Kristina.archibald@nserc-crsng.gc.ca](mailto:Kristina.archibald@nserc-crsng.gc.ca), +1 613 947 5636

### **The Social Sciences and Humanities Research Council of Canada (SSHRC)**

Crystal Sissons, [crystal.sissons@sshrc-crsh.gc.ca](mailto:crystal.sissons@sshrc-crsh.gc.ca), +1-613-947-4539

### **The Netherlands Organisation for Scientific Research (NWO/ NLeSC)**

Alice Dijkstra (Humanities), [a.dijkstra@nwo.nl](mailto:a.dijkstra@nwo.nl), (+31) 70-3440736

Marcus van Leeuwen (Social Sciences), [m.vanleeuwen@nwo.nl](mailto:m.vanleeuwen@nwo.nl), (+31) 70-3440931

Rosemarie van der Veen-Oei (Physical Sciences), [r.vanderveen@nwo.nl](mailto:r.vanderveen@nwo.nl), (+31) 70-3440851

René van Schaik (NLeSC), [r.vanschaik@esciencecenter.nl](mailto:r.vanschaik@esciencecenter.nl), (+31) 20 460 4770

### **The UK Arts and Humanities Research Council (AHRC)**

Pam Mason, [p.mason@ahrc.ac.uk](mailto:p.mason@ahrc.ac.uk), +44 (0)1793 416063

### **The UK Economic and Social Research Council (ESRC)**

Michael Bright, [michael.bright@esrc.ac.uk](mailto:michael.bright@esrc.ac.uk), +44 (0)1793 413042

Audrey Sharp, [audrey.sharp@esrc.ac.uk](mailto:audrey.sharp@esrc.ac.uk), +44 (0)1793 413150

### **The US Institute of Museum and Library Services (IMLS)**

Chuck Thomas, [cthomas@imls.gov](mailto:cthomas@imls.gov), +1-202-653-4663

### **The US National Endowment for the Humanities (NEH)**

Brett Boble, [bboble@neh.gov](mailto:bboble@neh.gov), +1-202-606-8401

Jennifer Serventi, [jserventi@neh.gov](mailto: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](mailto:etran@nsf.gov), +1-703-292-5338

The competition website is <http://www.diggingintodata.org/>.