



Cyberinfrastructure Initiative

Challenge 1 – Research Data Infrastructure

Expressions of Interest

March 27, 2015

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

The following provides information about the 60 Expressions of Interest submitted to the CFI on February 27. As mentioned in the Call for Proposals, the CFI challenges institutions and researchers to come together to form consortia and propose research data infrastructure projects that create tailored, shared and integrated data resources capable of enabling leading-edge research on significant scientific, social and economic questions. Since projects are expected to involve multi-institutional consortia of institutions and their researchers, data scientists, data analysts and software developers, this list should be used to identify opportunities and encourage collaboration between consortia in closely-related or identical domains prior to submission and merit-review of the Notices of Intent.

Following a careful review of the information provided in the EOIs, the CFI wishes to emphasize a few key points from the Call for Proposals and remind the applicants that projects submitted under this Initiative are expected to:

- **Involve a critical mass of leading Canadian researchers in the domain who are fully engaged in the project and who are capable of exploiting the full potential of the research data infrastructure.** It has been noted that many of the EOIs focus on a very limited number of researchers, often at a single institution.
- **Be realistically achievable in the expected timeframe and budget.** The scope of some of the EOIs appears to be so broad that the consortium may not be able to implement the project within the expected two- to three-year timeframe.
- **Be sufficiently mature, beyond the conceptual stage.** Some of the EOIs describe consortia and projects that are at the very earliest stages of discussion and are unlikely to be sufficiently developed in time for the first competition.

Institutions are reminded that there will be a second competition for Challenge 1 of the Cyberinfrastructure Initiative, with a Notice of Intent deadline of April 2016. We encourage applicants who believe their initiative would benefit from additional discussion and further development to consider submitting their notice of intent in the second competition.

The list below has been sorted by domains of research and can be easily accessible using the summary table below. Yet, the best way to use this document is likely using the search function of your PDF reader to look for relevant keywords.

HEALTH SCIENCES

[Online collaborative communities of researchers, clinicians, managers, patients and their caregivers for open-access research-based tools to guide research on aging and chronic diseases](#)

[Patient-centric research cyber-infrastructure \(PCR-CI\)](#)

[Amplifying the Patient Voice through Enhanced Process Orientation to Achieve Affordable Patient and Population Centered Care\(The Amplify Project\)](#)

[Data and information platform on Francophone health](#)

[Plateforme de données et d'information sur la santé des francophones](#)

[Pan-Canadian Real-world Health Data Network – Primary Care \(PRHDN-Primary Care\)](#)

[Ongoing Development and use of a national Primary Care Electronic Medical Record Network for Research](#)

[Developing a Canadian General Medicine Inpatient Clinical Research Platform \(GEMINI\) through automated data](#)

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

[retrieval from electronic hospital information systems](#)

[The Complementary and Alternative Medicine \(CAM\) Prospective e-Data Project \(CAMPED\)](#)

[A distributed infrastructure for Genomics Data Sharing and Analysis](#)

[Immune Receptor Repertoire Data Commons for Personalized Immunotherapy](#)

[CBRAIN: A national platform for brain research](#)

[National Resource for Sharing Normative Medical Imaging Data](#)

[EEGLab HPC integration: tools for large scale advanced neuroscience data processing/analysis, provenance, annotation and shared procedure development](#)

[Enabling Data-Driven Massive Mesoscale Models of the Brain](#)

[Medical image databases integration and image indexing in a multiuser environment for an enhanced medical diagnostics, knowledge extraction and improved health delivery system](#)

[A Platform to Integrate High-Throughput Biological and Clinical Data](#)

[Canadian Molecular Profiling in Cancer Trials \(CAMPACT\) Interchange](#)

[Heart-CODE - High Performance Computing, Data Collaborative Hub and Integrated Analytics for Arrhythmia Health](#)

[Dental Anatomy Software: Expansion and Mobile Upgrade](#)

[Maelstrom Research Canadian Harmonization Infrastructure \(CHaIn\)](#)

[DATA: Databases, Analytical pipelines, Tools and Approaches for “Big Data”](#)

[Parametric Human Platform for Biomedical and Ergonomic Research](#)

[Next General Signaling Biology Platform](#)

[A Pan-Canadian Data Collection and Analysis Platform for Patient Radiation Protection and Safety](#)

[Pan-Canadian Integrative Documentation Framework for Surgical Synoptic Reporting](#)

NATURAL SCIENCES AND ENGINEERING

[Unleashing Big Data Archives](#)

[Canadian Advanced Network for Astronomy Research](#)

[Synthesizing the Universe: a comprehensive framework to generate, manage and visualize high-resolution](#)

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

[astronomical simulation data](#)

[A Universe of Data and Complexity: The Convergence of Climate Physics, Astrophysics and Cosmology](#)

[Canadian Open Biodiversity Knowledge Platform](#)

[Barcode-based Rapid Assessment of Variability in Ecosystems \(BRAVE\)](#)

[Seamless Data Visualization from the Kilometre- to Nanometre-Scale for 12 Agronomically-important Plant Species: ePlant Apps for the BAR and other Plant Cyberinfrastructure Resources](#)

[Building integrated data analysis and archiving capacity for watersheds in the Lake Winnipeg Basin and Nelson River Watershed](#)

[Software Infrastructure for Deep Learning](#)

[A Cyberinfrastructure Framework to Automate Processing and Management of Mineral-Exploration Data](#)

[Platform for Integrated Cyber Tomography \(PICT\)](#)

[Cyberinfrastructure for Advanced Structured Materials Imaging](#)

[The Canadian Consortium for Arctic Data Interoperability – Data Sharing and Analysis for Arctic Research and Northern Communities](#)

[MERIDIAN: MarinE Research Data Infrastructure](#)

[Distributed data and cloud computing system for high energy physics](#)

[Cyberinfrastructure for Experimental Robotics Research \(CERR\)](#)

[Digital Soil Data Consortium](#)

[PREMO, a Web Server for the Prediction of Substrate Recognition Motifs for Enzymes Depositing Post-Translational Modifications](#)

[Joint Cyberinfrastructure for Applied Forest and Renewable Energy Research and Innovation](#)

SOCIAL SCIENCES AND HUMANITIES

[Inclusive Learning Analytics](#)

[Canadian Virtual Analytics Lab \(CVALU\)](#)

[Mise en place d'une infrastructure de collecte de données de recherche sur les usages et impacts des technologies mobiles en éducation pour créer des ressources de données vidéo adaptées, partagées et intégrées \(bases de données, dépôts de données\)](#)

[ExaMine: A Discovery Platform for New Knowledge from Cultural Texts](#)

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

[Multidisciplinary collaboration with heterogeneous cultural and textual heritage data](#)

[SSH Digital Research Data Project](#)

[Database of Realworld Corpora in Child Language Development Research](#)

[Linked Data Infrastructure for Cultural Enquiry](#)

[Building a Community Data Resource](#)

[The Canadian Research Data Centre Network \(CRDCN\): the Move to Thin-client Operation and Administrative Data Acquisition](#)

[Interdisciplinary, Integrated Data Platform for Oceans and Fisheries Research \(O-Fish D-Platform\)](#)

[Navigable Cyberinfrastructure for Live Events](#)

EOIs for proposals to be submitted in Competition 2

[Networked Environment for Media Archiving and Analysis: Creating a national network for preservation of and access to digital assets \(NEMAA\)](#)

[Research Data Infrastructure for Smart Animal Health and Production Management](#)

[A Cyber-framework for the Harmonization of Multiple Historical Data Sets and Linking to Each Other and to Modern Data](#)

[Precision Agri-food Technology Research & Innovation Consortium \(PATRIC\)](#)

HEALTH SCIENCES

Project title:	Online collaborative communities of researchers, clinicians, managers, patients and their caregivers for open-access research-based tools to guide research on aging and chronic diseases
Administrative Institution:	Jewish General Hospital
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	José Morais
Consortium members:	McGill University
Number of researchers:	11

Project summary:

CEVIMAC (Centre of excellence on aging and chronic disease) is a new knowledge brokering organization, created in 2012 by the Quebec Ministry of Health and Social Services, to address the challenges of aging and multi-chronicity, in the McGill integrated university health network, which covers 63% of the province's area and serves 1.8 million inhabitants.

Advances in information technologies and their greater acceptance from health providers and patients have created new opportunities to address these challenges in more innovative ways.

CEVIMAC intends to create real-time interactions between researchers, healthcare professionals, managers, patients and their natural caregivers through its digital platform to transfer knowledge more effectively and to promote the co-creation, standardization and sharing of evidence-based clinical tools adapted to multi-chronicity, frailty and clusters of patients' profiles. The ultimate goal is to improve healthcare coordination from tertiary to primary and community care.

These interactions will generate large volumes of data which once analyzed will be a source of creativity for research and innovation. The competition will allow CEVIMAC to develop methods and standards to connect its internal databases to large external ones, to produce intelligence and open-access research-based tools for its community of researchers and guide research in aging and chronic diseases.

Estimated total project cost: \$2 million

Amount of funding to be requested from CFI: \$800,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Patient-centric research cyber-infrastructure (PCR-CI)
Administrative Institution:	Queen’s University
CFI liaison at institution :	Mary Purcell (purcellm@queensu.ca)
Project Leader :	Harriet Feilotter
Consortium members:	Mount Sinai Hospital Sunnybrook Health Sciences Centre Ontario Institute for Cancer Research
Number of researchers:	20

Project summary:

Synergies have been identified among researchers, data scientists, and software developers who are engaged in supporting national and international consortia and projects including the NCIC Clinical Trials Group, the Canadian Open Genetics Repository, disease cohorts for breast and prostate cancer, and eight provincial cohorts.

We are launching a new consortium to support a range of next-generation clinical research projects. We propose to develop an integrated research data infrastructure that will facilitate: 1) the formation of a network of laboratories operating to defined standards, 2) development of common data elements for collection of clinical, imaging, pathologic and molecular data, 3) design of federated databases that support linkages between institutions and between clinical and biomarker datasets, 4) development of analytic pipelines designed to standardize data extraction and data interpretation and 5) development of regulatory compliant consent and data management strategies to support registration research for clinical use. The program builds upon bioinformatics platforms such as Biomatrix, Brain-CODE, the Canadian Data Integration Centre and the Cancer Genome Collaboratory.

This will ultimately accelerate recruitment into novel personalised medicine trials, translate into exportable processes and tools that can be used by research groups in Canada and beyond, and ultimately, by diagnostic laboratories and clinical organizations.

Estimated total project cost: Unstated

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Amplifying the Patient Voice through Enhanced Process Orientation to Achieve Affordable Patient and Population Centered Care(The Amplify Project)
Administrative Institution:	University of Ottawa
CFI liaison at institution :	Mohamad Nasser-Eddine (mnassere@uottawa.ca)
Project Leader :	Alan Forster
Consortium members:	McGill University Health Centre University of Toronto Institute for Clinical Evaluative Services
Number of researchers:	6

Project summary:

Acute care academic hospitals are expected to be the central care hub for patients with life threatening and complex illnesses including cancer, end stage organ failure, and trauma. Due to increasing patient expectations and advancing technologies, opportunities to improve care and associated costs are common. Health systems are struggling to optimize value. The Amplify Program is dedicated to advancing our knowledge in achieving the competing goals of improved quality and reduced costs for patients and populations. It will consist of the adaptation of well-established technologies from the business world to the healthcare environment. Specifically, we will adapt ‘Customer Relationship Management’ applications to capture individual patient reported health outcome data throughout the patient life cycle; ‘Business Process Modeling’ applications to enhance our ability to manage care processes across the care continuum; and ‘Knowledge Management’ applications to make data more available to support analyses. This will enable the creation of software solutions developed specifically for the health system. It will also enable researchers in Ontario and Montreal to access robust process level healthcare data. Collectively, this program will result in better decision making by patients, providers, and policy makers related to challenges at the clinic and population level.

Estimated total project cost: \$4 million

Amount of funding to be requested from CFI: \$1.6 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Data and information platform on Francophone health Plateforme de données et d'information sur la santé des francophones
Administrative Institution:	University of Ottawa
CFI liaison at institution :	Mohamad Nasser-Eddine (mnassere@uottawa.ca)
Project Leader :	Renaud Arnaud
Consortium members:	Institut de recherché de l'Hôpital Montfort (IRHM) Institute for Clinical Evaluative Services (ICES) Bruyère Research Institute (BRI) Manitoba Centre for Health Policy (MCHP)
Number of researchers:	5

Project summary:

In Canada, the lack of interoperability between health information systems prevents healthcare stakeholders (i.e. decision-makers, health professionals) from using up-to-date, evidence-based information to deliver services and adopt best practices that meet Francophone-minority community's needs.

IRHM-ICES-uOttawa-EBRI and MCHP will partner with organizations across Canada to create:

- A. A secure platform for mining and aggregating data and information on the health of Francophones and streamlining access to multiple health databases.
- B. A network of users (e.g., researchers, policymakers, service providers) to:
 - Develop and conduct research with large health databases.
 - Create research programs and advisory services on health and health services in French that address stakeholders' and policy maker's needs.
 - Delineate the availability/distribution of French services and French-speaking health professional (mapping).
 - Develop a health data dashboard and report cards on the health of Francophone and on the health system's efficiency to meet their needs.

By focusing on Francophone-communities in a minority context, this platform will enable:

1. Evidence-based decision making to implement innovating health services that deploy available resources to efficiently meet the needs of Francophones.
2. Mobilizing multi-disciplinary/multi-stakeholder teams to tackle key issues pertaining to: the health of Francophone-communities, French-language health services, and training of Francophone health professionals.

Estimated total project cost: \$1.6 million

Amount of funding to be requested from CFI: \$640,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Pan-Canadian Real-world Health Data Network – Primary Care (PRHDN-Primary Care)
Administrative Institution:	Sunnybrook Health Sciences Centre
CFI liaison at institution :	Kevin Hamilton (kevin.hamilton@sri.utoronto.ca)
Project Leader :	Alison Paprica
Consortium members:	Institute for Clinical Evaluative Sciences (ICES) Manitoba Centre for Health Policy (MCHP) Population Data British Columbia (PopData BC) Canadian Institute for Health Information (CIHI)
Number of researchers :	9

Project summary:

There is strong demand from health researchers and policymakers to advance from provincial studies to national-level analyses and comparative research that capitalize on Canada's unique population-wide individual-level health data holdings. In response, MCHP, ICES, PopData BC and CIHI have developed the Pan-Canadian Real-world Health Data Network (PRHDN) proposal to bring together diverse data and expertise and unite them through a Health & Social Data Council and National Coordinating Centre. As a first step toward the PRHDN, CFI Cyber-infrastructure project funding would establish two complementary primary health care distributed data networks: (i) common data model – a minimum common dataset of prepared primary care data within each province that would enable rapid multi-province queries (ii) common analytic protocol – validated tools and methods to be deployed independently within provinces (on data in unaltered provincial form) to produce findings that can be pooled for meta-analyses which provide multi-province estimates. PRHDN-Primary Care data infrastructure would be designed with input from researchers, practitioners and policymakers from across Canada, including members of the pan-Canadian CIHR SPOR PIHCl Network. One-time investment will thus create primary health care research data infrastructure with widespread utility that also serves as a cornerstone toward a fully national, whole health system, PRHDN.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Ongoing Development and use of a national Primary Care Electronic Medical Record Network for Research
Administrative Institution:	Queen’s University
CFI liaison at institution :	Mary Purcell (purcellm@queensu.ca)
Project Leader :	Richard Birtwhistle
Consortium members:	Memorial University Dalhousie University Université de Montréal University of Toronto Western University University of Manitoba University of Alberta University of Calgary University of British Columbia High Performance Computing Virtual Laboratory (HPCVL) IBM Provincial Centres for Health Data Ontario Genomics Institute Canadian Institute for Health Information Canada Health Infoway Health Canada Public Health Agency of Canada BrainCode/Indoc
Number of researchers:	Unstated

Project summary:

This application is to support research using a national primary care electronic medical record database. This data source is new and has not been available before for a wide range of health research and analytics. Examples include: the epidemiology and identification of novel risk factors of chronic disease, health service use in the population, post marketing pharmacosurveillance and phenotypic data for use in genomics. The database had been developed by the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) funded by the Public Health Agency of Canada since 2008 and is a network of primary care research networks at universities in Canada.

It currently has almost 1 million patient electronic medical records from >600 primary care practices. De-identified health data is extracted every 3 months for use in research, surveillance and practice audit and feedback. Much work has been done to establish this unique database and it has huge potential for research particularly linking to other databases.

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

CPCSSN data is stored at the High Performance Computing Virtual Laboratory and relationships with provincial and territorial health databases such as ICES and Manitoba Centre for Health Policy have been established.

Challenge 1 CFI funding is required to develop near real-time data extraction capability for acute disease surveillance and research and infrastructure support for ongoing development of cleaning algorithms, case definition, data analytics and refining linkage capabilities.

Estimated total project cost: Unstated

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Developing a Canadian General Medicine Inpatient Clinical Research Platform (GEMINI) through automated data retrieval from electronic hospital information systems
Administrative Institution:	St. Michael's Hospital
CFI liaison at institution :	Samar Saneinejad (saneinejads@smh.ca)
Project Leaders :	Andreas Laupacis, Muhammad Mamdani
Consortium members:	University Health Network University of Toronto Mount Sinai Hospital Sunnybrook Health Sciences Centre Trillium Health Partners McMaster University University of Ottawa McGill University University of Calgary University of Alberta University of British Columbia Harvard University
Number of researchers:	12

Project summary:

General medicine wards care for up to 50% of patients admitted to hospital from the emergency department. This represents the largest group of hospital patients. Yet little is known about general medicine patients and the quality of their care in hospital. Hospitals' electronic information systems store a wealth of untapped information about processes of care and patient outcomes. Data warehouse projects use this information to provide increasingly sophisticated analytics within individual institutions, across regions and nationally. Connecting electronic systems across institutions represents a significant opportunity to catalyze research that transforms the delivery of inpatient medical care. The GEMINI registry will be a scalable Canada-wide data collection platform for general medical inpatients using automated data retrieval from electronic hospital systems. This cyberinfrastructure will enable researchers to: study variations in processes of care and patient outcomes; identify opportunities to improve quality of care; study best practices for the care of patients with multiple medical conditions; implement and evaluate interventions to improve care; study knowledge translation strategies and their effects on outcomes; attract large multi-centre randomized controlled trials in general medical inpatients. GEMINI will become self-sustaining and job-creating by developing local research nodes and providing analytic support to hospitals, health authorities, and governments.

Estimated Total Project Cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	The Complementary and Alternative Medicine (CAM) Prospective e-Data Project (CAMPED)
Administrative Institution:	University of Toronto
CFI liaison at institution :	MayLiza Baak (m.baak@utoronto.ca)
Project Leader :	Lynda G. Balneaves
Consortium members:	University of Alberta Chinese University Hong Kong
Number of researchers :	8

Project summary:

Complementary and alternative medicine (CAM) is a burgeoning health issue, with 75% of Canadians using CAM at a cost of \$5 billion/year. However, there is limited data regarding which CAM therapies are most prevalent, their clinical benefits and adverse effects, and their impact on how other health services are utilized.

An international priority within the CAM research domain is the development of shared and integrated databases across primary and speciality clinical settings that will collect data on CAM use and its impact on patients and the health care system. Colleagues at Chinese University Hong Kong (CUHK), with industry and government partnerships, have developed a standardized, comprehensive electronic medical record (EMR) system that bridges conventional Western medicine and CAM to enable identification of CAM therapies worthy of clinical evaluation and those associated with adverse effects. Building on this existing infrastructure, a collaborative data initiative between the UofT, UofA, and CUHK will be undertaken that will collect and analyze large volumes of data on CAM therapies and related health and health care system outcomes using EMR and administrative data. The resulting data repository will provide an international, prospective database that will support innovative basic science, clinical and health services research on CAM.

Estimated total project cost: \$3 million

Amount of funding to be requested from CFI: \$1.2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: A distributed infrastructure for Genomics Data Sharing and Analysis
Administrative Institution: The Hospital for Sick Children
CFI liaison at institution : Belinda Potter (belinda.potter@sickkids.ca)
Project Leader : Michael Brudno
Consortium members: University of Toronto
McGill University and Genome Quebec Innovation Centre
University of British Columbia and BC Genome Sciences Centre
University Health Network
Université de Sherbrooke

Number of researchers : 5

Project summary:

Biological and health scientists require access to large-scale genomic datasets (including proteomic, metabolomic and other related data). These include both controlled datasets of cohorts with diseases such as cancers and autism, as well as public data. The exponential growth of such data, which can be hundreds of terabytes in size, has led to a significant cost in the computing infrastructure required for storage and analysis. Aggravating the problem is that such datasets are typically replicated across a wide number of sites, often with multiple copies within a single institution.

The Global Alliance for Genomics and Health (GA4GH) has been developing secure APIs for efficient storage and retrieval of such data. We propose to implement these APIs on Compute Canada infrastructure. We will utilize effective caching and sharing of large-scale data over the network of sites hosting diverse datasets, with appropriate privacy controls and role-based permissions, while taking advantage of both secure (HPC4Health and BCGSC) and open (Calcul Quebec, SciNet) data centres. Our consortium will also implement innovative tools to process, integrate and explore these private and public genomics data sets using Compute Canada resources. These developments will build on existing projects such as the Genetics and Genomics Analysis Platform (GenAP).

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Immune Receptor Repertoire Data Commons for Personalized Immunotherapy
Administrative Institution:	Simon Fraser University
CFI liaison at institution :	Sara Swenson (saras@sfu.ca)
Project Leaders :	Felix Breden and Jamie Scott
Consortium members:	University of Toronto Texas Advanced Computation Center, University Texas at Austin CNRS France VDJServer – University of Texas Southwestern Medical School HIPC (Human Immunology Project Consortium) ImmPort - Human Immunology Portal University College London and Wellcome Trust Sanger Institute Scripps Research Institute NIH Vaccine Research Centre (VRC) Duke Center for HIV/AIDS Vaccine Immunology (CHAVI)
Number of Researchers:	2

Project summary:

Our Challenge 1 project will integrate and provide access to large data sets from the immunogenetics research community, thereby increasing the value of these data for biomedical research and patient care. Over the past five years, next generation sequencing (**NGS**) has revolutionized our ability to monitor the immune system in exquisite detail; this is a critical advance for research in infectious and autoimmune diseases and cancer, and for the development of vaccines and therapeutic antibodies. The challenge is to integrate these data into one system that will allow the comparison and sharing of NGS immunogenetics data across research institutions.

This project will use Compute Canada resources to leverage and expand our current iReceptor project (funded by the CANARIE NEP program) to build an integrated data repository of immunogenetics. Our vision for this project derives from the consortium of 20 immuno-geneticists, industrial partners, funding agency representatives, ethicists, and legal experts, who identified the need for a common immunogenetic NGS data repository during a meeting at SFU in 2014, which will be further defined at a larger Community Meeting planned for May 2015. This initiative will greatly strengthen Canada's leadership position in the development, production, and dissemination of knowledge critical to the next generation of personalized immunotherapy.

Estimated total project cost: \$1.55 million

Amount of funding to be requested from CFI: \$620,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	CBRAIN: A national platform for brain research
Administrative Institution:	McGill University
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	Alan Evans
Consortium members:	University of Calgary University of Toronto Université de Montréal Western University University of British Columbia Simon Fraser University Concordia University Dalhousie University Université de Lyon, France UC Berkeley Neurospin, France Julich Research Centre, Germany Hanyang University, South Korea Yale University University of Vermont
Number of researchers:	31

Project summary:

The CBRAIN platform (Sherif et al., 2014 ; www.cbrain.mcgill.ca) gives Canada leadership in global neuroinformatics, serving 300+ users in 19 countries (70% Canadian). First funded by CANARIE (www.canarie.ca), CBRAIN provides researchers with web access to the high-performance computing infrastructure of Compute Canada (www.computecanada.ca). CBRAIN supports data storage, organization and dissemination through its LORIS web-database (Das et al., 2012).

We will enhance (i) CBRAIN service delivery and support for the research community, and (ii) CBRAIN functionality, e.g. cloud access, international interoperability, support for genetics and epigenetics research. Specifically, we will add developers to:

- Oversee the performance/load balancing of distributed computing.
- Leverage continuing Compute Canada and CANARIE IT developments.
- Oversee VM provisioning/deployment over heterogeneous clusters/clouds.
- Develop core services for user-deployed algorithms and tool packages.
- Develop GUIs for workflow and remote 3D visualization.
- Enhance LORIS functionality and interoperability with other major databases.

We will locate developers in Halifax, Montreal, Toronto and Vancouver, to maintain application software on CBRAIN from researchers in each Compute Canada region.

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Implementation of these goals will establish a world leading neuroinformatics platform for brain research. Coupled with disease-oriented networks (e.g. CCNA, NeuroDevNet), CBRAIN will provide the binding layer between Canada's IT infrastructure and receptor communities (clinicians, industry, government).

Estimated total project cost: \$4.5 million

Amount of funding to be requested from CFI: \$1.8 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	National Resource for Sharing Normative Medical Imaging Data
Administrative Institution:	University of Calgary
CFI liaison at institution :	Liz Scarratt (scarrate@ucalgary.ca)
Project Leader :	Richard Frayne
Consortium members:	None stated
Number of researchers:	3

Project summary:

The goal of this project is to share an actively managed, quality-assured, readily accessible database of brain medical images of healthy subjects. The database would contain volumetric data obtained from healthy Canadians over their lifespan and from both genders. The project will develop advanced software tools/strategies to:

- 1) Facilitate image uploading to populate and actively update imaging large databases,
- 2) Perform intrinsic quality assurance on incoming data as uploaded,
- 3) Aggregate data from different sites, different scanners and nominally equivalent protocols,
- 4) Generate normative image summaries, and
- 5) Create web-based tools with easy, interactive extraction tools for retrieval of customized summaries.

To achieve these objectives, we need to solve three major limitations: We need to 1) implement automated data- and image-management solutions, 2) realize sophisticated strategies that allows data to be aggregated using multiple criteria in real time, and 3) demonstrate the role of implicit quality assurance. Our proposed strategies will access high-performance processing and storage power resources and allow interactive access to individual Canadian researchers.

The project will initially focus on brain imaging using quantitative data acquired on magnetic resonance (MR) scanners; however, the framework that will be developed will be extensible to other body regions and imaging modalities.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	EEGLab HPC integration: tools for large scale advanced neuroscience data processing/analysis, provenance, annotation and shared procedure development
Administrative Institution:	Brock University
CFI liaison at institution :	Carrie Kelly (ckelly@brocku.ca)
Project Leaders :	Sid Segalowitz & James Desjardins
Consortium members:	University of California, San Diego Western University McMaster University Rotman Research Institute
Number of researchers:	21

Project summary:

Human brain electrophysiology (EEG/MEG) is expanding as a technology of choice for cognitive neuroscientists. Due to major advances over the last decade, EEG methods have opened up major new possibilities for the big data sets being gathered, but these require HPC for practical execution. EEGLab is the leading open source data analysis toolbox developed by the Swartz Center for Computational Neuroscience (at UCSD), with whom we have been working to implement HPC solutions. To fully integrate EEGLab with HPC, development is required for standardization of script generation and execution, flexible and intuitive job submission, automated job provenance, file management, as well as flexible data annotation. The Brock Lifespan Centre and SHARCNet have produced initial working prototypes of the needed EEGLab extensions. This project will support a team of programmers and analysts to develop these software extensions necessary to integrate the advanced EEG/MEG processing required for the big data capacity of the consortium. Further, this will establish a repository for the sharing and community development of optimized procedures. Completion of this technology will bring Compute Canada into the centre of human EEG/MEG neuroscience across Canada and eventually as a potential service provider internationally.

Estimated total project cost: \$2,275,000

Amount of funding to be requested from CFI: \$910,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: Enabling Data-Driven Massive Mesoscale Models of the Brain
Administrative Institution: Western University
CFI liaison at institution : Bryce Pickard (bpickard@uwo.ca)
Project Leader : Mark Daley
Consortium members: University of Toronto
University of Waterloo
Number of researchers: 3

Project summary:

The study of the brain has focused on the relationship between its structure and function. The functional capacity of the brain is unique, but the brain's structural architecture has several singular features that sculpt its computational capacity. Big data initiatives are now standard fare in neuroscience (e.g. CLSA, ADNI, MIND, etc) and have amassed significant structural, functional and behavioural datasets.

It has become clear that "big data" alone is not enough to carry us to the scientific insights we seek; indeed, our experimental data is only as valuable as the theoretical frameworks within which we can analyze it. A key distinguishing feature of 21st century science is that our theoretical frameworks can now be exponentially more comprehensive than in earlier times due to the availability of exponentially-growing compute power.

Our proposal is to bring these pieces together: "big data" and "big computational modelling" based on an international modeling platform, grounded in theoretical neuroscience, to integrate and organize this wealth of information and create new knowledge and understanding.

The platform will benefit enormously by starting from a strong existing leadership position established by two Canadian initiatives, The Virtual Brain (thevirtualbrain.org) and Nengo (nengo.ca).

Estimated total project cost: \$3 million

Amount of funding to be requested from CFI: \$1.2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Medical image databases integration and image indexing in a multiuser environment for an enhanced medical diagnostics, knowledge extraction and improved health delivery system
Administrative Institution:	University of British Columbia
CFI liaison at institution :	Sharon Wu (sharon.wu@ubc.ca)
Project Leader :	None stated
Consortium members:	None stated
Number of researchers:	Unstated

Project summary:

This project will focus mainly on medical data base images that are part of PAC systems and acquired during normal hospital/medical institutional operation. Considerable medical information is contained in such image databases and associated radiologist reports that can be used in response to numerous needs examples of which includes teaching and learning process of young physicians, data mining for improved diagnosis, quality medical services and management to name a few.

This proposal focuses on the coordination of existing databases and preparation and transformation of data into formats for multiuser application i.e. data usage by wide classes of users. Coordination and transforming data into suitable formats for multiuser application data normalization, cleaning, and anonymization will all be a major part of this project as well as data indexing, and image analysis. In this project, algorithms will be developed under which data transformation into useful format for multiuser application will be investigated. These will include application of advanced clustering techniques, which use image features combined with natural language processing of associated radiology/cardiology reports. We will use meta data available as DICOM header information (provided with each images) to categorize studies. Researchers can query this database based on criteria such as similarity with a given image or a given text reports. We note that medical image databases contain data that is gathered using alternative technologies. Data fusion may also be applied for a multi-mode data analysis and extraction of an improved query results. In this project, attempts will be made to identify criteria that reflect different needs of medical researchers for multiuser application and develop necessary algorithms

Estimated total project cost: \$1.2 million

Amount of funding to be requested from CFI: \$480,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	A Platform to Integrate High-Throughput Biological and Clinical Data
Administrative Institution:	Mount Sinai Hospital
CFI liaison at institution :	Gareth Taylor (grtaylor@lunenfeld.ca)
Project Leaders :	Laurence Pelletier and Jeff Wrana
Consortium members:	None Stated
Number of researchers:	Unstated

Project summary:

The Lunenfeld-Tanenbaum Research Institute (LTRI) at Mount Sinai Hospital has an international reputation for excellence in several key areas of biomedical and clinical science with particular foci in systems biology and modeling of complex biological systems. This includes genome-scale analyses of protein complexes using mass spectrometry, study of cellular events using high-resolution and high-throughput imaging and analysis of genomic alterations using advanced biostatistics. These key areas each require intricate analysis and processing schema. Furthermore, there is a pressing need to provide reliable infrastructure for data storage and retrieval and meta-analysis to mine linkages between cell biological and clinical datasets. There is also an urgent need to promote efficient sharing of data with the broader research community, essential for global scientific progress. Our goal with this proposal is to partner with CFI/Compute Canada and build cutting edge infrastructure to support systematic processing, analysis and dissemination of high complexity datasets. This includes, but is not restricted to, the establishment of GPU-based processing of Terabyte-scale image datasets (e.g. de-noising and deconvolution), mass spectrometry data analysis/repository and whole transcriptome analysis. This sophisticated infrastructure will be designed with data processing pipelines that enable availability to researchers across Canada using existing infrastructure and will build upon our partnering Genomics Innovation Network (GIN) platforms. We will play a leading role in the implementation and management of this infrastructure in close collaboration with other research centers in Ontario in our effort to link systems biology data to clinical outcome (e.g. HCP4Health and OICR). As world leaders in proteomics, functional genomics, populations genetics and cellular imaging with demonstrated experience in international data hosting and dissemination, we are uniquely poised to implement the proposed infrastructure.

Estimated total project cost: Unstated

Amount of funding to be requested from CFI: \$1.2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Canadian Molecular Profiling in Cancer Trials (CAMPACT) Interchange
Administrative Institution:	University Health Network
CFI liaison at institution :	Ian McDermott (mcdermot@uhnresearch.ca) Helen Chan (hchan@uhnresearch.ca)
Project Leader :	Trevor Pugh
Consortium members:	Jewish General Hospital BC Cancer Agency Centre of Genomics and Policy, Genome Quebec Innovation Centre British Columbia Cancer Agency
Number of researchers:	15

Project summary:

Our goal is to enable genome-guided clinical trials across Canada to target cancer patients with rare clinically actionable somatic mutations. The Canadian Molecular Profiling in Cancer Trials (CAMPACT) Interchange leverages significant investments that have created strong independent efforts at leading cancer hospitals across three Canadian provinces: the Princess Margaret Cancer Centre (Ontario), British Columbia Cancer Agency (BC), and Jewish General Hospital Segal Cancer Centre (Quebec). We have also engaged experts in health law and large-scale data access (GQIC). Through deployment on Compute Canada infrastructure, we will establish a data-sharing and analysis framework that can be extended to other centres within and outside Canada.

We will connect clinically-derived genomic, phenotypic, therapeutic, and outcome datasets currently maintained separately by each centre. Leveraging HPC4Health, a secure, cloud-based Compute Canada node, our respective clinical bioinformatics groups will adopt shared clinical cancer genome analysis and data federation tools, and adapt each for local hospital offerings and IT environments. All data exchange formats and protocols will be harmonized with similar data-sharing efforts (e.g. AACR Project GENIE and Global Alliance for Genomics and Health). All connectors and federation systems will enable secure and auditable querying of data without disclosing sensitive patient information outside of clinical care teams.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Heart-CODE - High Performance Computing, Data Collaborative Hub and Integrated Analytics for Arrhythmia Health
Administrative Institution:	Western University
CFI liaison at institution :	Bryce Pickard (bpickard@uwo.ca)
Project Leader :	Anthony Tang
Consortium members:	Canadian Arrhythmia Network (CANet) InDOC High Performance Computing Virtual Laboratory (HPCVL)
Number of researchers:	Unstated

Project summary:

The *IT plan* for CANet-ReCAR, a recently funded Networks of Centres of Excellence, is to provide research teams with access to powerful data management and high performance computing infrastructure to enable electronic data capture, centralized storage, secure data sharing, a collaboration hub, and integrated analytics. Members of the consortium include HPCVL which hosts the InDOC platform that developed *Brain-CODE*, an informatics platform managing the acquisition, storage, analysis, and sharing of multidimensional data. With this application, CANet-ReCAR will leverage this investment and adapt it for '*Heart-CODE*'. *Heart-CODE* will be an integrated data management system of all the data produced by CANet-ReCAR research programs. It will provide CANet-ReCAR with: **Infrastructure**-hardware, storage, software, combined with IT and high performance computing provided by HPCVL; **Security, Privacy Processes and Policies**; **User Portal**-customizable, secure, unified point of entry for researchers to access databases, reports, visualizations, collaboration tools and enable data sharing; **Data Management**-a suite of databases, custom data entry forms, quality control, site monitoring; integration and query tools supporting multiple data types (e.g. clinical assessments, ECG, imaging); **Analytics and Visualization**-data analysis and visualization software including automated pipelines, complex data mining, and visual analytics; **Standards**-common data elements and standards, case report forms, Delphi-based processes for consensus.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: Dental Anatomy Software: Expansion and Mobile Upgrade
Administrative Institution: University of Toronto
CFI liaison at institution : MayLiza Baak (m.baak@utoronto.ca)
Project Leader : Anuradha Prakki
Consortium members: Western University
University of Manitoba
North Dakota State University
Number of researchers : 7

Project summary:

First Year Dental Anatomy (FYDA) based on beta software version was developed in 2009, by U of T Dentistry IT support team and the Dental Anatomy course director. The proposed aim was to fill a void of educational, interactive software sought after by students and researchers. FYDA has been successfully recognized as an important educational tool based on educational research, using three-year controlled user's feedback and outcome. Moreover, FYDA has also been used and well accepted by international dental schools in China, Brazil and Peru. This proposal aims at expanding FYDA in content through upgrades and transition to HTML compliant version and mobile application formats using iOS and Android Frameworks. The final project deliverable is an upgraded, online and mobile version of FYDA. On a global level, professors and students from various departments (e.g. Anthropology, Biology, human and veterinary Medicine) and different disciplines throughout Dentistry will benefit with the newest content and teaching tools offered by FYDA. Researchers in Education field will use generated data to assess and improve student's perceptions of learning benefits and learning outcome stemming from the design and instructional use of such software.

Estimated total project cost: \$450,000

Amount of funding to be requested from CFI: \$180,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Maelstrom Research Canadian Harmonization Infrastructure (CHaIn)
Administrative Institution:	McGill University
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	Isabel Fortier
Consortium members:	Ontario Institute for Cancer Research McMaster University University of Victoria University of British Columbia Université de Sherbrooke Université de Montréal Alberta Health Services University of Bristol, UK Eindhoven University of Technology, The Netherlands University Medical Center Groningen, The Netherlands
Number of researchers:	15

Project summary:

CHaIn's goal is to offer the means to leverage international collaborative research in health and social sciences and foster better usage of Canadian data and samples. Building upon existing national and international expertise, networks (e.g. Maternal Infant & Youth Research Network and Integrative Analysis of Longitudinal Studies on Aging), and resources (specialized software, methods and data/samples collections), the CHaIn initiative will implement an infrastructure allowing efficient documentation, harmonization, integration and co-analysis of data across research resources. An interactive web-based catalogue will document designs and content of existing Canadian studies and databases available for research. The catalogue will permit identifying compatible data across research resources. A complementary software environment will support data harmonization and co-analysis. Over time, this infrastructure will act as a data hub to be exploited by networks of investigators looking to answer a broad range of research questions (e.g. investigation of gene-environment interaction, impacts of behavioral and social factors on health). To achieve its aims CHaIn brings together organizations producing and hosting health and social data, data scientist and software developers building the platform and researchers making use of this data to achieve multi-disciplinary research.

Estimated total project cost: \$3 million

Amount of funding to be requested from CFI: \$1.2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	DATA: Databases, Analytical pipelines, Tools and Approaches for “Big Data”
Administrative Institution:	University of British Columbia
CFI liaison at institution :	Sharon Wu (sharon.wu@ubc.ca)
Project Leader :	Elodie Portales-Casamar
Consortium members:	Arthritis Research Canada Child and Family Research Institute Population Data BC BC Cancer Agency BC Children’s Hospital Institute for Heart+Lung Health Providence Health Care Research Institute Simon Fraser University St. Paul’s Hospital University of Victoria Vancouver Coastal Health Research Institute Vancouver General Hospital Allergy, Genes and Environment Network (AllerGen) NCE Program Canadian Alliance of Pediatric Rheumatology Investigators Canadian Inflammatory Myopathy Study Canadian Network for Improved Outcomes in Systemic Lupus Erythematosus Canadian Primary Care Sentinel Surveillance Network Canadian Scleroderma Research Group Centre of Excellence for Prevention of Organ Failure Compute Canada Genetics-Environment-and-Therapies: Food Allergy Clinical Tolerance Studies Maternal Infant Child & Youth Research Network of Canada McGill University McMaster University NeuroDevNet NCE Program Plexia Electronic Medical Systems Inc. Queen’s University University of Alberta University of Calgary University of Manitoba University of Montreal University of Toronto University of Waterloo Channing Laboratories

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Childhood Arthritis and Rheumatology Research Alliance
Harvard University
Imperial College
Johns Hopkins University
Munich Allergy Research Center
Transnational Asthma Genetics Consortia
University of Groningen
University of Western Australia

Number of researchers: 5

Project summary:

We are an international, multidisciplinary consortium with common interests in inflammation in chronic diseases and commitment to building powerful data management, integration and analytic methods for research uses of “big data”. Management, integration and analysis of complex data from multiple sources and platforms (i.e., “big data”) is essential to accessing, harnessing and integrating vast, accumulating knowledge. Potential data resources are expanding rapidly, but their use for research suffers from their existence in discrete repositories, greatly reducing the potential for discovery and translational impact. This proposal has three broad goals:

1. Expanding data comprehensiveness and interoperability: Development of interoperable re-usable processes/procedures/tools that will enable the transformation of clinical data sets, including EMRs, cohorts and disease registries, into research-ready data for the scientific community.
2. Analytics: Producing a well-documented infrastructure of re-usable software and analytic/visualization approaches for high-dimensional information.
3. Ensuring privacy and confidentiality: Creation of infrastructure and processes for secure data flow, storage, linkage, and access, including development where possible of open-access data. Our initial focus will be on datasets including clinical, laboratory, disease registry, administrative, and publically available (e.g., NCBI, ENCODE, 1000 Genomes) information. While our consortium is interested in inflammation, the resulting infrastructure can be applied to any domain.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: Parametric Human Platform for Biomedical and Ergonomic Research
Administrative Institution: University of British Columbia
CFI liaison at institution : Sharon Wu (sharon.wu@ubc.ca)
Project Leader : Sidney Fels
Consortium members: University of Saskatchewan
University of Toronto
McMaster University
York University
Queen’s University
McGill University
Parametric Human Project

Number of researchers : 10

Project summary:

We seek to build an advanced platform for accelerating progress of biomedical and ergonomic research communities through collaborative development and dissemination of large-scale anatomical datasets and digital macro-level human models via an existing Canadian-based international consortium of researchers, data scientists, and software developers. The platform, utilizing Compute Canada’s capabilities, enables a large, growing network of researchers to browse, query, analyze, visualize and collaboratively author massive medical datasets and digital human models easily, while ensuring scientific reproducibility and compliance with established ethics and data privacy standards. Through a cloud-based infrastructure, our expanding database will be structured and integrated into advanced data-driven statistical and dynamic human musculo-skeletal models. The consortia will save researchers substantial time and resources spent collecting and processing their own human data and models. Advanced human macro-level modeling provides the next step to reduce costs and improve outcomes in Canadian healthcare and industrial design. The Parametric Human Platform enables hundreds of Canadian researchers to consolidate and coordinate the effort needed to make Canada the world’s leading authority in macro-scale human model-based research. Hence, this platform provides Canadian economic and quality of life benefits, and propels our human data and model intensive researchers to lead the world in this area.

Estimated total project cost: \$1,870,000

Amount of funding to be requested from CFI: \$784,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: Next General Signaling Biology Platform
Administrative Institution: University Health Network
Contacts at institution : Ian McDermott (mcdermot@uhnresearch.ca)
Helen Chan (hchan@uhnresearch.ca)
Project Leader : Igor Jurisica
Consortium members: Université de Montréal
University of Toronto
Queen's University
Ontario Institute for Cancer Research
Number of researchers: 5

Project summary:

Precision medicine needs to be data driven and corresponding analyses comprehensive and systematic. We will not find new treatments if only testing known targets and studying characterized pathways. Thousands of potentially important proteins remain drug, pathway or interactome "orphans". Computational biology methods can help fill this gap with accurate predictions, but biological validation and functional experiments are essential. Intertwining computational prediction and modeling with biological experiments will lead to the faster generation of more relevant findings while optimizing resources—leading to data-driven biology and medicine. We must reform literature curation to cope with increased volume/depth of information. This project will lead to improved accuracy and relevance of curated protein interaction databases, transforming the veracity and predictive power of biological interaction networks and disease:gene networks. Network-based drug design and use will be empowered by this initiative, as single target approaches are outdated and urgently need to be refined for clinical studies. No single interaction, pathway or drug database provides sufficient coverage and richness of information—intelligent integration is key. Building on our unique resources, this initiative will reduce research costs, increase quality, information content, and speed of discovery leading to precision medicine, new treatments, and eventually to proactive and preventive medicine.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	A Pan-Canadian Data Collection and Analysis Platform for Patient Radiation Protection and Safety
Administrative Institution:	McMaster University
CFI liaison at institution :	Sherisse Webb (webbsh@mcmaster.ca)
Project Leader :	David Koff
Consortium members:	University of Waterloo
Number of researchers:	2

Project summary:

The project will build innovative tools required to study the long-term effect of low dose medical radiation at the population level. By developing a platform to collate automatically the dose information acquired in hospitals and clinics, we will ultimately have created the first province-wide followed by nation-wide automatic collaboration solution to reconcile radiation dose with patient medical information. It will facilitate, for the first time in Ontario and Canada, an understanding of the long-term effects of low dose radiation exposure on populations along with benchmarking and real-time decision support for procedure justification. For example, we may want to understand what will be the impact of performing a chest X-ray every day on a neonate spending 45 days in Intensive Care Unit.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Pan-Canadian Integrative Documentation Framework for Surgical Synoptic Reporting
Administrative Institution:	University of Alberta
CFI liaison at institution :	Neil Sulakhe (neil.sulakhe@ualberta.ca)
Project Leader :	Ronald Moore
Consortium members:	University of Calgary Alberta Health Services Cancer Surgery Alberta
Number of researchers:	6

Project summary:

In partnership with Compute Canada, Alberta Health Services (Strategic Clinical Networks), the University of Calgary (Drs. Walley Temple, Joseph Dort, John Kortbeek), Alberta Health, the Softworks Group Inc. (Edmonton, Alberta), and the Royal College of Physicians and Surgeons of Canada, the University of Alberta (Drs. Ron Moore, Todd McMullen, Doug Hedden) proposes to lead the development of the Pan-Canadian Integrative Documentation Framework for Surgical Synoptic Reporting. This proposal builds on the successful initiatives by Cancer Surgery Alberta, Cancer Control Alberta, Alberta Health Services, to develop Alberta's WebSMR reporting tool, the Synoptec reporting system, and the Synoptic Interactive Patient Portal, the latter which was funded under Canada Health Infoway Innovation and Adoption investment program. These platforms are currently utilized in all cancer surgery sites in Alberta, integrating electronic health records, provincial patient registries, and electronic health records, and were successfully piloted across other Canadian provinces, including Quebec, Manitoba, Ontario, and Nova Scotia, as part of the Canadian Partnership Against Cancer Synoptic Reporting Tools Project. These initiatives have gathered patient outcomes and biorepository data for over 35,000 patients, and we propose to leverage this successful adoption towards integration of other patient care systems in all surgical disciplines.

Our proposal aims to address the prevalent need to improve access to, and quality of collected clinical data in our Canadian health care systems. Surgical care in Canada is uncoordinated and care practices vary widely by region and provider. Many patients do not receive the preventative measures necessary to avert suffering and clinical decisions often deviate from the latest scientific guidelines. While significant progresses have been made in the implementation of evidence-based clinical practices, further research is limited by the information captured at the point of care by physicians and health care providers from procedures and examinations. Synoptic reporting refers to a systematized method for capturing meaningful data during clinical documentation so that important information demonstrated to influence health outcomes through decision-making can be seamlessly captured, aggregated, and analyzed. A standardized framework for synoptic reporting which integrates clinical information systems collecting varied information, such as surgery outcomes, pathology, health records, and patient experience, across all Canadian health care providers presents the potential for widespread aggregation and study of multivariate data across varying care delivery systems. Access to complete, real-time, and

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

essential outcomes information can improve quality of surgical practice, facilitate assessment of new treatments, and inform resource planning. The primary goal of our proposal is to develop an unparalleled comprehensive and integrative Pan-Canadian clinical research data framework to facilitate synergistic collaborations and foster advancements in evidence-based medicine and health care delivery research.

The proposed project will be aligned with Compute Canada to define the computation resources needed to provide a Pan-Canadian research computing environment for secure surgical synoptic reporting. The project will be planned in partnership with Alberta Health to leverage on Canada's first provincial-wide integrated electronic health record system, Alberta ERHIS, and infrastructure developments which they have planned in the near long-term.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

NATURAL SCIENCES AND ENGINEERING

Project title:	Unleashing Big Data Archives
Administrative Institution:	University of Alberta
CFI liaison at institution :	Neil Sulakhe (neil.sulakhe@ualberta.ca)
Project Leader :	Erik Rosolowsky
Consortium members:	Calgary Scientific Harvard University The US National Radio Astronomy Observatory Square Kilometre Array South Africa
Number of researchers:	9

Project summary:

The Atacama Large Millimetre/submillimetre Array (ALMA) is Canada's newest telescope and is rapidly producing big data sets (TBs) that challenge traditional analysis methods. We have met this challenge by creating the Cube Analysis and Rendering Tool for Astronomy (CARTA), which uses a novel software architecture for high performance visualization of Big Data. Its key innovation is locating the visualization servers where the data are housed and connecting to these servers through an advanced web browser interface. This approach is used commercially for Earth mapping data and we have brought this approach to manage Big Science Data from current and next generation radio telescopes. Here, we propose two major innovations: first, we will broaden CARTA to work on different classes of data: seismic data from geophysical surveys and simulation data generated by astrophysical codes. Second, we will expand the CARTA analysis framework to interactively trigger high performance analyses of these datasets on Compute Canada resources. We will use these new interactive connections to Big Data to address key science challenges: how to recover robust information from incomplete imaging data (using both radio interferometry and seismic data as our test cases) and how statistical comparison between simulation and radio astronomical can unravel the complexities of star formation. Through integrating of seismic and simulation data visualization and advanced data analyses with CARTA and enabling model comparisons between big data archives, we will enable CARTA to be used by a broad community that extends beyond the astronomical and geophysical communities.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Canadian Advanced Network for Astronomy Research
Administrative Institution:	University of Victoria
CFI liaison at institution :	Joaquin Trapero (cfi@uvic.ca)
Project Leader :	Christopher Pritchett
Consortium members:	University of Alberta University of Toronto St. Mary's University McMaster University Western University Université de Montreal University of British Columbia National Research Council Canada Compute Canada CANARIE
Number of researchers:	Unstated

Project summary:

Effective data management has been a key to supporting Astronomy science based on large complex datasets and the support of large international collaborations. In 2009 the Canadian Advanced Network for Astronomy Research (CANFAR) was formed. This was a partnership of several universities, led by University of Victoria. CANFAR has been funded by CANARIE NEP and RPI programs and collaborates with the NRC Canadian Astronomy Data Centre (CADDC) in developing, implementing, and operating services that support data-intensive research. The current systems are fully available to the entire Canadian research community but is, in reality, unable to support the scaling necessary to do so. CANFAR is working with Compute Canada (CC) to migrate 100% of the CANFAR systems onto CC. Basic data archiving and distribution functions are well supported but the more advanced functions demanded by the CANFAR Science Management Committee remain in a state where users experience failures and lack of capacity. The CANFAR system needs a complete re-design during the transition to Compute Canada.

CANFAR is both a developer of, and a strong proponent of, the CANARIE-led concept of Research Platform Interfaces where web services that support primitive functions can be combined together to form components of a national platform to support research data management infrastructure.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Synthesizing the Universe: a comprehensive framework to generate, manage and visualize high-resolution astronomical simulation data
Administrative Institution:	McMaster University
CFI liaison at institution :	Sherisse Webb (webbsh@mcmaster.ca)
Project Leader :	Alison Sills
Consortium members:	Saint Mary's University National Research council – Herzberg Institute of Astrophysics University of Alberta
Number of researchers:	7

Project summary:

To understand how structure (planets, stars, galaxies) forms in the universe, astronomers rely on observed and simulated data sets. Increasingly, a direct comparison of simulations to observations provides the best route to physical explanations for the objects that we see in the sky. Thus, simulation and associated visualization techniques, must keep pace with the ever-increasing fidelity of observational data. Modelling structure formation requires high resolution in space and time combined with a complete treatment of gravitational forces, hydrodynamics, chemical evolution of dust and gas, and radiation. Such simulations generate prodigious quantities of high quality data. By combining our world-leading expertise in astrophysical simulations with a new software engineering team, we will develop a framework that will allow such multi-physics simulations. The framework will also provide critical management, visualization, and storage tools necessary to realize the full scientific value of a huge amount of model data. A key deliverable will be a pipeline that produces synthetic observations from simulated datasets - a prerequisite for confronting theory with observed cosmic structures and populations. This dedicated software team and the proposed framework will maximize Canada's investments in the ALMA, JWST, TMT and SKA observatories, and will catapult our simulation capability to international leadership.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	A Universe of Data and Complexity: The Convergence of Climate Physics, Astrophysics and Cosmology
Administrative Institution:	University of Toronto
CFI liaison at institution :	MayLiza Baak (m.baak@utoronto.ca)
Project Leader :	Bryan Gaensler
Consortium members:	Dalhousie University McGill University McMaster University Perimeter Institute University of British Columbia
Number of researchers :	17

Project summary:

All scientific fields are being forced to confront very large data sets. Although generation and analysis of “Big Data” is extremely demanding in itself, the new frontiers are simulations that are indistinguishable from experiments, and data sets that span an extreme dynamic range in both space and time.

While this frontier lies in the future for many fields, three intersecting disciplines face this challenge right now. In climate physics, astrophysics, and cosmology, a profound convergence is occurring, both in the scale of the observational data sets driving discovery, and in the sophistication and mathematical form of the theoretical models that extract meaning from these observations. Underpinning this are fundamental commonalities in the multi-process physics at play, including radiative transfer, turbulence and magnetohydrodynamics. In all three fields, new algorithms are required to solve massive sets of nonlinear partial differential equations at the accuracy and efficiency needed to reproduce and interpret observed reality.

We propose to create a shared software environment in which a team of software engineers and data analysts will work collaboratively with the core science group to solve a set of fundamental problems common to all three disciplines. We will develop methods of data analysis and simulation that extend and transcend the traditional Fourier approach, we will formulate novel ways of comparing experiments and simulations, and we will design and optimize new algorithms that can be scaled to the petaflop regime, all with the ultimate aim of extracting vital statistical signatures from massive multi-dimensional space-time-domain data sets. This work will drive cross-disciplinary discovery, will deliver a broad set of new tools and approaches to the wider community, and will position Canada for leadership in landmark upcoming projects ranging from climate physics (GRACE-FO, SWOT) to gravitational waves (Advanced LIGO) and cosmology (CHIME).

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Canadian Open Biodiversity Knowledge Platform
Administrative Institution:	Université de Montréal
CFI liaison at institution :	Anne Welcker (anne.welcker@umontreal.ca)
Project Leader :	Anne Bruneau
Consortium members:	McGill University University of Ottawa University of Alberta University of British Columbia York University Université du Québec à Rimouski Université du Québec à Montréal University of Guelph Concordia University University of Toronto University of Alberta Université de Sherbrooke Canadian Museum of Nature Canadian Food Inspection Agency Canadian Forest Service AAFC
Number of researchers:	37

Project summary:

The Canadian Open Biodiversity Knowledge Platform will be a viable and accessible information management system for data on species. It will establish the semantic interoperability among silos of highly valued biological data stores and will develop the analytical and visualization tools for cutting-edge research and the generation of new knowledge in biodiversity science.

The consortium will make significant use of data mining and high throughput data entry and pivot on three nodes with complementary and overlapping specializations on species: their discovery and characterization, their evolution, and their ecological relationships. The study of changes to species distributions and ecosystem services as a result of climate and environmental changes, inadvertent introductions and other disturbances will be galvanized by the development of the infrastructure. The data mobilized by the consortium and made interoperable by the Platform will include high resolution images, genetic and genomic information, geographic distribution, habitat and environmental information, occurrence observations, character traits, scientific articles, floras/faunas, keys and checklists, and information on associated taxa and inter-species relationships.

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Internationally-established standards and community norms, principles of public data access and open-source application development will be prioritized. The platform and associated projects will place Canada at the forefront of international biodiversity research and conservation efforts.

Estimated total project cost: \$2,880,000

Amount of funding to be requested from CFI: \$1,152,001

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title: Barcode-based Rapid Assessment of Variability in Ecosystems (BRAVE)
Administrative Institution: University of Guelph
CFI liaison at institution : Jill Rogers (jlrogers@uoguelph.ca)
Project Leader : Paul Hebert
Consortium members: University of New Brunswick
McGill University
Queen Mary University of London
Oxford University
University of Pennsylvania
University of Helsinki
Musée national d'histoire naturelle

Number of researchers: 13

Project summary:

The Biodiversity Institute of Ontario has developed BOLD, the informatics platform which supports international research efforts to develop a DNA barcode reference library. The present application seeks funds to develop BRAVE, a companion web-based platform for the deposition, analysis, visualization and publication of biodiversity data generated through the use of next generation sequencers (NGS). It will employ a client-server strategy for data submission, coupling a desktop tool for the validation, filtration and compression of datasets with interfaces for data submission. BRAVE will provide tools for assessing sequence quality, assembly, gene prediction and annotation. It will employ hardware-encoded algorithms using Field Programmable Gate Arrays (FPGAs) to accelerate species assignments for the large number (5M to 100M) of sequence records generated by each NGS run. Once such assignments are made, it will deliver metrics facilitating the comparison of biodiversity patterns among sites. BRAVE will also provide publication tools allowing researchers to share data and to deposit it in community archives, such as GenBank. By coupling a rich, distinctive data repository with specialized analytical tools, BRAVE will address many of the challenges associated with the use of NGS platforms to advance biodiversity science.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Seamless Data Visualization from the Kilometre- to Nanometre-Scale for 12 Agronomically-important Plant Species: ePlant Apps for the BAR and other Plant Cyberinfrastructure Resources
Administrative Institution:	University of Toronto
CFI liaison at institution :	MayLiza Baak (m.baak@utoronto.ca)
Project Leader :	Nicholas Provart
Consortium members:	University of Guelph University of British Columbia University of Victoria Agriculture and AgriFood Canada J. Craig Venter Institute Iowa State University University of Arizona Southern Illinois University University of Missouri Cold Spring Harbour Labs Brazilian Bioethanol Science and Technology Laboratory Brazilian Federal University of Rio Grande do Sul
Number of researchers	12

Project summary:

Being able to efficiently query and visualize the huge amount of data generated in the past 5 years for a dozen crop species for hypothesis generation will be key to improving crop yields to feed, shelter and power a world of 9 billion people by the year 2050. We will expand our ePlant framework for seamlessly exploring data from different biological levels to these 12 important crop species. The ePlant framework, built by the Provart Lab with funds from the Genome Canada Bioinformatics and Computational Biology competition, has been recently populated by us with data for *Arabidopsis thaliana*, see <http://bar.utoronto.ca/~dev/eplant/>. Its zoomable user interface permits exploration of sequence variation at the kilometre scale, through gene expression levels in tissues and cell types, subcellular localization of gene products, protein-protein interactors, to protein tertiary structure and gene sequences at the nanometre scale. We will develop 12 ePlants for crop species for which we have or will generate data. Some amount of data analysis is necessary to be able to map sequence variation data onto protein structures, to predict binding sites for transcription factors in genomes, or to compute interologs (predicted protein-protein interactors). The project would thus involve an analysis aspect plus a collational aspect of creating and populating databases to be able to power the various ePlants. These 12 ePlants (for Maize, Rice, Barley, Triticale, Poplar, *Medicago truncatula*, Soybean, Potato, Tomato, *Camelina sativa*, eucalyptus, and sugar cane) would be available on the widely used Bio-Analytic

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Resource at <http://bar.utoronto.ca> (run by the Provart Lab and used 60,000 times a month by plant researchers from around the world), and as “apps” on other plant cyberinfrastructure portals such as MaizeGDB.org, SoyKB.org, Gramene.org, CoGe, etc. in a manner similar to how the Arabidopsis ePlant will be available shortly on the new, U.S. National Science Foundation-funded portal for Arabidopsis information, <http://araport.org>.

Estimated total project cost: \$1 million

Amount of funding to be requested from CFI: \$400,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Building integrated data analysis and archiving capacity for watersheds in the Lake Winnipeg Basin and Nelson River Watershed
Administrative Institution:	University of Manitoba
CFI liaison at institution :	Tracy Mohr (tracy.mohr@umanitoba.ca)
Project Leader :	David Lobb
Consortium members:	University of Saskatchewan Water Security Agency International Institute for Sustainable Development Environment Canada Lake Winnipeg Basin Initiative Province of Manitoba Water Stewardship Lake Winnipeg Research Consortium Red River Basin Commission
Number of researchers:	3

Project summary:

The vision of the Lake Winnipeg Basin Information Network is to provide researchers within the Lake Winnipeg Basin and Nelson River Watershed with a consistent and central data management strategy to store, share, organize, access, mine and analyze digital resources from multiple research disciplines. This shared strategy enables integrated analyses of water, sediment and nutrient data for multiple watersheds across North America and the opportunity for analysis of data in new and innovative ways.

The LWBIN aims to create a shared and integrated research data infrastructure in two iterative phases:

Phase 1: Optimizing data storage and sharing capacity

- Compare currently used open access digital and spatial data repositories with other options
- Integrating existing national and international metadata standards (Government of Canada metadata standards, ISO 19115, Open data license) to ensure data consistency
- Refining existing database dimensional warehouse model to further dataset integration and archiving
- Integration with international datasets

Phase 2: Development of novel tools for data mining and analysis

- Near real-time mapping capabilities for remote sensing data
- Data access strategy to accommodate, e.g., embargo periods for current and on-going research
- Interactive data models

Estimated total project cost: \$1,237,000

Amount of funding to be requested from CFI: \$494,800

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Software Infrastructure for Deep Learning
Administrative Institution:	Université de Montréal
CFI liaison at institution :	Anne Welcker (anne.welcker@umontreal.ca)
Project Leader :	Yoshua Bengio
Consortium members:	Montreal Heart Institute École Polytechnique de Montréal McGill University Université Laval University of Toronto University of British Columbia
Number of researchers:	19

Project summary:

Based on the impetus of the Canadian Institute for Advanced Research (CIFAR), deep learning revolutionizes broad areas of Artificial Intelligence (AI) applications, including computer vision, speech recognition and natural language processing. These brain-inspired statistical learning methods are based on learning multiple levels of gradually more abstract and non-linear representations of data. The breakthroughs achieved recently have stimulated major investments (billions of dollars) by leading-edge information technology companies outside Canada (Google, Microsoft, Facebook, Amazon, Samsung, IBM, Baidu, etc.). Deep learning algorithms are already among the major uses of computing resources of Compute Canada, especially regarding high-performance computing and Graphics Processing Unit (GPU). Academics have been developing their own software for research in deep learning, which require specialized programming for GPUs. The proposal aims at putting together the talents of the CIFAR-centered consortium towards an open-source shared infrastructure for research and technology transfer of deep learning algorithms, with an initial focus on AI and registry-based medical research applications, and transfer to Canadian start-ups and users. Deep learning can help capture high-order interactions between large set of variables and take advantage of big data, enabling numerous applications, in particular, democratization of the research process, personalized medicine, and improvement in the health of Canadians.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	A Cyberinfrastructure Framework to Automate Processing and Management of Mineral-Exploration Data
Administrative Institution:	Western University
CFI liaison at institution :	Bryce Pickard (bpickard@uwo.ca)
Project Leader :	Neil Banerjee
Consortium members:	IBM Saskatchewan Geological Survey
Number of researchers:	3

Project summary:

A cyberinfrastructure can facilitate data acquisition, storage, management, integration, visualization and other information processing. The term is representative of a complex interaction of researchers, practitioners and computational facilities organized to serve a community, such as the mining and exploration industry. We use this example as this community accounts for a significant component of the economic growth of Canada (> than \$50 billion – nominal GDP), thus representing an important sector for the Canadian economy justifying the level of investment and research proposed.

In this context, the development of a cyberinfrastructure to support mineral exploration and mining is a reasonable venture. We argue that the process of mineral discovery needs to benefit from modern process automation and predictive analytics. Various disciplines of modern ore deposit science require a higher level of interconnection. Present research in the areas of 3D visualization and orebody modelling, GIS based mineral exploration, coupled thermal-mechanical-chemical, numerical simulations of ore forming systems are still fragmented fields in Economic Geology represented by an array of independent software tools that often are not capable of leveraging big data and supercomputing. Such efforts if properly regrouped into a computational framework that provides a linkage and a workflow oriented framework, could represent a unique/innovative approach to mineral discovery. Promoting a collaborative environment for academia and industry, the proposed Mineral Exploration Cyberinfrastructure will facilitate knowledge transfer, decision making and business sustainability and growth.

Estimated total project cost: \$2,208,000

Amount of funding to be requested from CFI: 883,200

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Platform for Integrated Cyber Tomography (PICT)
Administrative Institution:	University of Saskatchewan
CFI liaison at institution :	Susan Blum (susan.blum@usask.ca)
Project Leader :	David Cooper
Consortium members:	Saskatoon Health Region Commonwealth Scientific and Industrial Research Organisation (CSIRO) Canadian Light Source Monash University
Number of researchers:	17

Project summary:

The BioMedical Imaging and Therapy facility of the Canadian Light Source – our national synchrotron - is unique within Canada and supports a community of national and international scientists. The unparalleled capabilities for X-ray based imaging create data-related challenges – particularly for high resolution and high throughput tomographic imaging. Bottlenecks at many stages of the research pipeline (acquisition, processing, storage and sharing) limit the number of users that can be engaged and slow the path to dissemination of results. To address these challenges, we propose the creation of a unified platform which integrates and accelerates the analysis pipeline. It will consist of existing hardware for data acquisition, software for image reconstruction, new hardware for rapid reconstruction and storage, and a web-based platform for remote data sharing. With respect to software, our goal is to partner with international collaborators associated with the Australian Synchrotron to capitalize on a tomography reconstruction package they have already implemented (X-TRACT). This partnership will not only jump start the streamlining of processes locally, it will lead to a more integrated international community of imaging-focused scientists. Ultimately, the proposed platform would enable users to focus on tackling scientific problems rather than technological hurdles associated with large and complex datasets.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Cyberinfrastructure for Advanced Structured Materials Imaging
Administrative Institution:	University of Waterloo
CFI liaison at institution :	Andrew Barker (awbarker@uwaterloo.ca)
Project Leader :	Marios Ioannidis
Consortium members:	McMaster University Western University McGill University
Number of researchers:	7

Project summary:

Two- and three-dimensional imaging of advanced structured materials are of fundamental value to the understanding of their properties for scientific and engineering applications. Whether the materials are rocks hosting vital resources (water and hydrocarbons), or electrodes in fuel cells and batteries where energy is generated from electrochemical transformations, or polymeric scaffolds hosting living cells for tissue engineering applications, the ability to relate performance to structure is a key gap that stifles progress in many research and commercialization activities.

Significant portions of CFI funds have been directed towards the acquisition of advanced instruments for imaging of structured materials, thus there is abundant imaging data, however there is a gap in the unification of the collection, storage, and exchange of this data. The proposed cyberinfrastructure will provide a unified platform for structured materials imaging from a multitude of sources (MRI, x-CT, SEM, etc), resolutions (nanometre to centimetre), and dimensionalities (2D and 3D). The infrastructure will also enable the application of cutting-edge computational image fusion techniques for data of diverse origin to enable new insights into the relations between structure and physicochemical properties. Training, outreach and support will be provided for cyberinfrastructure usage through both online and in-person training activities and technical support.

Estimated total project cost: \$1 million

Amount of funding to be requested from CFI: \$400,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	The Canadian Consortium for Arctic Data Interoperability – Data Sharing and Analysis for Arctic Research and Northern Communities
Administrative Institution:	University of Calgary
CFI liaison at institution :	Liz Scarratt (scarrate@ucalgary.ca)
Project Leader :	Maribeth Murray
Consortium members:	Carleton University Université Laval University of Manitoba University of Waterloo Canadian High Arctic Research Station Cybera, Inc. International Arctic Science Committee and SAON Arctic Data Committee Inuit Tapiriit Kanatami
Number of researchers:	9

Project summary:

The Canadian Consortium for Arctic Data Interoperability (CCADI) connects Canada’s Arctic research centres to make data and information more accessible, and to provide analysis, visualization, and mapping tools to facilitate scientific discovery through data mining and integration. Partners will develop interoperable cyberinfrastructure among established institutes and online archives so that data generated from northern Canada is shared with stakeholder communities of researchers across all areas of Arctic science, with northern institutions, residents, organizations, decision makers, and international partners. Over 3 years, the CCADI will establish a national governance structure that facilitates: 1) discovery/access to Arctic data; 2) implementation of standards and policies for data interoperability/sharing; 3) creation of online analysis and mapping tools for organizing, translating, and integrating disparate data. System components will be optimized for use by stakeholders and consideration will be made for bandwidth restrictions in northern Canada, and for the ethical and legal requirements surrounding indigenous knowledge and confidential data. CCADI will engage researchers in data management, sharing and use requirements, facilitate data use by northern organizations, and partner with international institutions producing and managing Arctic data. The CCADI will enable Canada to lead in the development of data standards for Arctic regions and advance international interoperability.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	MERIDIAN: MarinE Research Data Infrastructure
Administrative Institution:	Dalhousie University
CFI liaison at institution :	Nancy Hayter (nancy.hayter@dal.ca)
Project Leader :	Stan Matwin
Consortium members:	Ocean Networks Canada Ocean Tracking Network University of British Columbia Simon Fraser University Université du Québec à Rimouski/Observatoire global du Saint-Laurent Memorial University
Number of researchers:	21

Project summary:

Canada's ocean scientists today face “Big Data” challenges, namely volume, variety and velocity of the data. To answer data intensive ocean science questions, scientists need to link and visualize existing massive oceanographic-sensor, geo-spatial and social data sets. We will develop a sustainable data infrastructure for Canada’s leading-edge ocean science research with a focus on marine soundscape data within a broader ocean data context. Key areas of focus are the impacts of sound for aquatic animal biology, and the analysis and visualization of the data from anthropogenic noise.

The proposed infrastructure will integrate state-of-the-art data management with data analytics and visualization solutions. The data management solutions will enable collection, quality-control, interoperability and federation of heterogeneous data sets. The data analytics infrastructure will adopt, adapt and develop tools for Canadian scientists to query existing marine life monitoring data for spatio-temporal movement, marine animal modeling, and ship movement patterns. The infrastructure, supported by Compute Canada nodes, will establish the hardware, software and human expertise platforms to support advanced predictive modeling and data visualization for ocean science and for the relevant social science issues. The project team will link the infrastructure with corresponding international ventures and an emerging, vibrant Canadian sensor industry sector.

Estimated total project cost: Not stated

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Distributed data and cloud computing system for high energy physics
Administrative Institution:	University of Victoria
CFI liaison at institution :	Joaquin Trapero (cfi@uvic.ca)
Project Leader :	Randall Sobie
Consortium members:	TRIUMF Laboratory Institute of Particle Physics
Number of researchers:	Unstated

Project summary:

High energy physics (HEP) experiments are evolving to a distributed computing model that employs cloud centres and data storage facilities around the world. Canadian researchers have been leading the international effort to use cloud computing resources for the HEP community. Currently, the system developed by the Canadian team seamlessly manages clouds in three continents for the generation of simulated particle collisions. The simulated data is used to design our experiments and develop software algorithms but the goal is to run data-intensive applications using the real particle collision data samples. The goal of the project is to develop context-aware cloud computing services so that the applications running on one cloud can dynamically locate and retrieve data from a remote site as required. The system will be used for the ATLAS and Belle-II experiments based at the CERN (Geneva) and KEK Laboratories (Japan); however, it will be developed so that other HEP experiments and researchers in other fields can use it. This project will develop new and innovative services, and provide a unique environment for the training of HEP.

Estimated total project cost: \$2.4 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Cyberinfrastructure for Experimental Robotics Research (CERR)
Administrative Institution:	McGill University
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	Gregory Dudek
Consortium members:	York University University of Toronto Simon Fraser University University of Alberta University of Waterloo Queen’s University Sherbrooke University Laval University Memorial University
Number of researchers:	18

Project summary:

We seek the development of shared software and datasets to support emerging research in experimental robotics, specifically targeting real-time acquisition, storage and analysis of large volumes of range and video data acquired in the field by mobile robots. The project is directly linked to activities of the NSERC Canadian Field Robotics Network (NCFRN), with additional researchers from across Canada. These researchers are conducting experimental robotics research that generates large amounts of data (video, 3D range sensing, audio, GPS, etc.) that is currently collected in many different formats, and stored locally by each lab. The goal of the proposed cyberinfrastructure is to develop the software tools necessary to share this data across the research community (both within the consortium and with the international community), thereby facilitating research into fundamental robotics algorithms and systems, including learning, localization, planning, mapping, human-robot interaction and multi-vehicle interaction. It is clear that the public sharing of datasets in other fields of applied computer science (e.g. machine learning, computer vision) has led to an accelerated pace of development throughout the last decade. The proposed infrastructure is crucial to allow Canadian researchers in experimental robotics to lead the way in terms of real-world deployment, validation and data-set sizes.

Estimated total project cost: \$2 million

Amount of funding to be requested from CFI: \$0.8 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Digital Soil Data Consortium
Administrative Institution:	University of Saskatchewan
CFI liaison at institution :	Susan Blum (susan.blum@usask.ca)
Project Leader :	Angela Bedard-Haughn
Consortium members:	Agriculture and Agri-Food Canada McGill University University of British Columbia University of Alberta University of Guelph Memorial University University of Manitoba Université Laval Atlantic Canada Conservation Data Centre Canadian Society of Soil Science The Reference Centre for Agriculture and Food of Quebec Alberta Agriculture and Rural Development LandMapper Environmental Solutions Department of Environment, Labour and Justice, PEI Department of Agriculture and Forestry, PEI
Number of researchers:	28

Project summary:

The Digital Soil Data Consortium will develop a next-generation database for integrating existing digital soil mapping information with spatially referenced soil properties, including both traditional biogeochemical data as well as novel spectroscopy and microbiome data. This will require a multi-dimensional system architecture that can incorporate different types of data and their spatio-temporal framework. This database could be integrated with other databases (geology, hydrology, climate, etc.), facilitating development of sophisticated models and unique land management decision tools. We have two specific objectives: Objective 1) Develop a repository for provincial and national soils data that ensures long-term access to existing soil survey data and allows for continual enhancement as new research data types and knowledge become available. This would build on existing initiatives in some regions and improve data integration across provincial borders resulting in infrastructure that maintains standards for openness, metadata, and common and interoperable schemas, protocols and formats. Objective 2) Develop specialized software and analytical tools, methods and standards that exploit Objective 1 infrastructure. One example might be a smart nutrient management tool for optimizing fertilizer use efficiency and minimizing environmental impact; this tool could in turn provide a platform for further digital decision making tools and expert system applications.

Estimated total project cost: \$1.5 million

Amount of funding to be requested from CFI: \$600,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	PREMO, a Web Server for the Prediction of Substrate Recognition Motifs for Enzymes Depositing Post-Translational Modifications
Administrative Institution:	University of Ottawa
CFI liaison at institution :	Mohamad Nasser-Eddine (mnassere@uottawa.ca)
Project Leader :	Roberto Chica
Consortium members:	None stated
Number of researchers:	3

Project summary:

The study of enzymes depositing post-translational modifications (PTMs) and the discovery of their substrates are essential to understand the biology of the cell. Yet, the discovery of substrates for PTM enzymes remains a significant challenge due to the limitations of current experimental techniques. To facilitate the discovery of substrates for PTM enzymes, we recently developed a rapid and inexpensive computational method that combines cutting-edge multistate computational protein design algorithms with k-means cluster analysis. Using our computational approach, we predicted the motif recognized by a protein lysine methyltransferase, leading to the discovery of four new substrates, which doubled in a single experiment the number of known targets for this PTM enzyme that has been difficult to characterize using conventional methods. With the help of the CFI, we will develop a web server (PREMO) to predict substrate recognition motifs for any enzyme depositing PTMs. PREMO will be freely accessible to the international scientific community and thus has the potential to represent a game-changing tool for systems biologists. Our unique and complementary set of skills coupled to the fact that we have exclusive access to this computational method put us in an excellent position to achieve this goal within the 3-year funding period.

Estimated total project cost: \$1 million

Amount of funding to be requested from CFI: \$400,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Joint Cyberinfrastructure for Applied Forest and Renewable Energy Research and Innovation
Administrative Institution:	Université de Moncton
CFI liaison at institution :	Lise Dubois (lise.dubois@umoncton.ca)
Project Leader :	Bedarul Alam
Consortium members:	Acadian Timber Corp. AV Group J.D. Irving Ltd. FP Innovations Groupe Savoie Remsoft FORUS Research Leading Edge Geomatics Precision Hawk Inc.
Number of researchers:	3

Project summary:

A leader in applied forest research and innovation, the Northern Hardwoods Research Institute (NHRI) performs leading edge research and innovation work to encourage a viable and sustainable development of hardwood resources, for the benefit of businesses and organizations in Canada. With the support of well established and emerging forest, ICT and aerospace-based companies, NHRI is proposing to establish a joint big-data center. Within its current research and innovation program, NHRI has been developing decision support systems (DSS) with optimization and simulation models using huge databases (terabytes), including fine-resolution LiDAR and other remote sensing information. However, in the development of LiDAR data intensive models and DSS, the NHRI and its partner organizations need a new cyberinfrastructure for big-data storage, management and computing in order to deliver concrete solutions to users. Further to the applications in hardwood resource research and innovation work, the cyberinfrastructure will be shared with its host institution (UMCE), notably in their research and innovation work on renewable energy resource mapping (biomass, wind, solar, micro-hydro) applied to Canada, and at the international level. This work in renewable energy resource mapping requires very large data-sets and computing capacity in order to analyse complex problems. Thus, the cyberinfrastructure proposed will assist in further developing cutting-edge research and innovation in precision forestry and in renewable energy in Canada, and with an international reach.

We wish to submit a proposal under the Cyberinfrastructure Initiative of the CFI program to create in-situ and/or cloud-based computing solutions to support our research and innovation work. The existing Technology Service Centre of the UMCE will be used to accommodate the equipment proposed by the NHRI, which will require planned improvements to its internal communications infrastructure and connectivity with outside networks such as the New Brunswick Optical Research and Educational Network.

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Estimated total project cost: \$1.5 million

Amount of funding to be requested from CFI: \$600,000

SOCIAL SCIENCES AND HUMANITIES

Project title:	Inclusive Learning Analytics
Administrative Institution:	OCAD University
CFI liaison at institution :	Helmut Reichenbacher (hreichenbacher@ocad.ca)
Project Leader :	Jutta Treviranus
Consortium members:	C21 Canada Canadian Education Association Canadian School Boards Associatio Independent Learning Centre Ontario Ministry of Education Simon Fraser University TakingITGlobal The William and Flora Hewlett Foundation Toronto District School Board University of Ontario Institute of Technology
Number of researchers:	Unstated

Project summary:

Learners learn differently. To respond to the complexity, fast pace of change, and unexpected shifts in the economic ecosystem, our education system should produce a diversity of life-long-learners. Learning analytics are a powerful tool to optimize learning and support inclusive education. Traditional Big Data research methods elide the outliers and minimize diversity. A partnership of organizations concerned with inclusive education, led by the Inclusive Design Research Centre at OCAD University, will develop and evaluate “Small” and “Thick” data research tools and inclusively designed personal learning analytics utilities or “dashboards” for educators, policymakers and students. Inclusive and multi-modal data representation methods will be refined and implemented to provide access to individuals using alternative access means.

This will provide formative and summative information to guide educators; enable learners to develop metacognition and self-awareness as life-long learners; and, equip policymakers with diversity-supportive metrics. The tools will also significantly advance the user experience design of data analytics, providing exemplars for personally monitoring health, improving financial literacy and other personalized data functions. The project will leverage the innovative cloud infrastructure developed through the Inclusive Design Institute. The partnership has established connections to dozens of school boards in Canada and over 40,000 educators in 149 countries.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Canadian Virtual Analytics Lab (CVALU)
Administrative Institution:	University of British Columbia
CFI liaison at institution :	Sharon Wu (sharon.wu@ubc.ca)
Project Leader :	Leonard Foster
Consortium members:	Simon Fraser University Université Laval Queen’s University University of Toronto University of Ottawa University of Victoria Bow Valley College, Calgary BC Ministry of Education Decoda Literacy Solutions Surrey School District Minister Ministry of Children and Youth Services, Ontario SRB Education Solutions Inc.
Number of researchers :	15

Project summary:

In British Columbia, the total 2014-15-budget estimate for education sector funding is \$11.89 billion. School Board expenditures in Canada in 2012 were \$54 billion. Post-Secondary Education Expenditures in Canada in 2010 were \$34 billion. Incorporating administrative or other statistical data “CVALU” software will create reports that use visualization to portray the mathematical trends it finds. We propose that increasing access to the tools of data analytics is essential to inform the design of public programs and policy and, especially, to ground practitioner decision-making across Canada and internationally. The dynamics of change in early childhood development, K-12 schooling, post-secondary education and non-formal learning settings need to be better understood beyond the boundaries of data scientists and statistics specialists. The CVALU Consortium will design and build an advanced research-computing infrastructure to help users to decipher change - trends and patterns in schooling and wellbeing. CVALU will promote discovery within data by providing tools that use open-source, secure enterprise-grade analytics, meta-data maps and custom connectors, libraries of predictive statistical functions, an intuitive interface, and interactive dashboards. These technologies will be architected for plug-and-play with additional data sources, extraction engines, and visualization widgets. CVALU will also support standard dynamic data flow diagrams and data mining features such as visualization, aggregate and breakout descriptive statistics, instance selection, feature selection, regression, clustering, classification and association rule learning.

Estimated total project cost: \$710,000

Amount of funding to be requested from CFI: \$280,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Mise en place d'une infrastructure de collecte de données de recherche sur les usages et impacts des technologies mobiles en éducation pour créer des ressources de données vidéo adaptées, partagées et intégrées (bases de données, dépôts de données)
Administrative Institution:	Université de Montréal
CFI liaison at institution :	Anne Welcker (anne.welcker@umontreal.ca)
Project Leader :	Thierry Karsenti
Consortium members:	Télé-Université du Québec Université du Québec à Montréal Université du Québec à Chicoutimi Université du Québec à Trois-Rivières Université du Québec en Outaouais École des Hautes-Études Commerciales Université Laval McGill University Université de Mons-Hainaut de Belgique Université de Cergy-Pontoise University of Patras, Greece
Number of researchers:	21

Project summary:

L'objectif de l'infrastructure de recherche — Mise en place d'un laboratoire de recherche sur les usages et impacts des technologies mobiles en éducation — est de contribuer à la pleine mise en œuvre d'une programmation scientifique rigoureuse et globale, articulée autour du domaine des technologies en éducation, lequel forme une des priorités majeures pour plusieurs des universités du Québec et la société en général. En effet, ce domaine est exceptionnellement novateur, dynamique et porteur à l'échelle mondiale. Cependant, la réalisation de recherches précises dans le domaine du numérique en éducation est indissociable d'une infrastructure technologique mobile, flexible, facilement adaptable à différents contextes éducatifs, à la fois pour comprendre les usages de ces technologies et mieux évaluer les impacts. Aussi, ce projet d'infrastructure de données de recherche créera des ressources de données vidéos adaptées, partagées et intégrées (bases de données, dépôts de données) qui favorisera une recherche de pointe dans le domaine des technologies en éducation, une thématique d'importance économique et sociale majeure. L'infrastructure permettra aussi, avec l'appui de partenaires majeurs, de situer ces activités scientifiques au cœur d'un réseau national et international de recherche. Sur le plan des retombées, elle contribuera de façon directe et indirecte à aider les quelque 340 000 enseignants et 6 millions d'élèves du Canada à faire un usage plus réfléchi et éducatif des technologies à l'école, dans un contexte de mondialisation où les compétences technologiques constituent des savoir-faire essentiels pour le marché du travail actuel et futur au Canada. Notre projet réunira un consortium multi-

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

établissements de chercheurs. Ce consortium mis en place permettra de garantir la pérennité et la pertinence à long terme ainsi que l'utilité de l'infrastructure de données de recherche.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	<i>ExaMine</i> : A Discovery Platform for New Knowledge from Cultural Texts
Administrative Institution:	McGill University
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	Stéfan Sinclair
Consortium members:	University of Guelph University of Lethbridge University of Alberta University of Victoria Université de Montréal
Number of researchers:	6

Project summary:

We live in a world of abundance of cultural texts that include millions of books, hundreds of millions of articles, and countless other potential sources from emails to tweets. Compiling, structuring and mining cultural data remains an enormous and ongoing challenge for cultural scholars as well as for many sectors of society, from governments and businesses to education and journalism. Various datasets and tools exist, but they remain disparate and require significant effort to use.

ExaMine will provide cyberinfrastructure in the form of an accessible data analytics platform that incorporates large cultural datasets and provides powerful analysis and visualization capabilities for discovering, exploring and analyzing textual content. *ExaMine* will leverage several existing initiatives in the digital humanities and focus on:

- federating multiple text collections in different formats and with varying metadata richness (e.g. Canadian Writing Research Collaboratory and Érudit.org)
- enabling navigation between primary and secondary text collections (e.g. works by Shakespeare and works on Shakespeare)
- leveraging visualization interfaces for browsing within very large full-text meta-collections (favouring fortuitous discovery)
- facilitating workset creation and transfer from query results to rich analytic interfaces (e.g. Voyant Tools)

Digital texts are out there, but so much useful knowledge remains unattainable because of the challenge of finding texts and associating relevant bits of information between them. *ExaMine* will help researchers explore culture in unprecedented ways.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Multidisciplinary collaboration with heterogeneous cultural and textual heritage data
Administrative Institution:	University of Lethbridge
CFI liaison at institution :	Brij Verma (brij.verma@uleth.ca)
Project Leader :	Daniel O'Donnell
Consortium members:	University of Saskatchewan University of Guelph University of Toronto Western University Dalhousie University Kislak Center and the Schoenberg Institute University of Pennsylvania Libraries Advanced Research Consortium (ARC) Texas A&M University (Texas) Visual Computing Lab ISTI-CNR (Pisa, Italy) University of Balamand (Lebanon) University of Turin (Italy) KU Leuven (Belgium) Net7 (Italy) Agile Humanities Agency (Canada)
Number of researchers:	11

Project summary:

Digital Cultural and Textual Heritage projects are at the forefront of public knowledge mobilization. They bring historical texts, monuments, artefacts, and research to the public and supply huge quantities of data to cultural research and education. Their interfaces, environments, and exhibits are the first point of contact with much cultural research. Such projects work collaboratively with heterogeneous datatypes: 2D, 3D, video, audio, GIS, and text. They use their data in various contexts: digital libraries and archives, semantic webs, crowdsourcing applications, gaming, immersive environments. Audiences include the public, entrepreneurs, government, and researchers. They are multidisciplinary and multisector, spanning University, GLAM, commercial, and policy sectors. Digital preservation and repatriation initiatives foster intercultural collaborations with First Nations, Métis, and Inuit communities.

The proposers are a strategic group of multidisciplinary partners with large existing datasets. They seek to extend, adapt, and develop tools and protocols for collaboration, visualisation, and mobilization. Canadian researchers play a leading role in this international endeavour.

Estimated total project cost: \$2.25 million

Amount of funding to be requested from CFI: \$900,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	SSH Digital Research Data Project
Administrative Institution:	Université de Montréal
CFI liaison at institution :	Anne Welcker (anne.welcker@umontreal.ca)
Project Leader :	Vincent Larivière
Consortium members:	Consortium Érudit Public Knowledge Project, Simon Fraser University Libraries Canadian Writing Research Collaboratory, University of Guelph Implementing New Knowledge Environments, University of Victoria
Number of researchers:	3

Project summary:

The project aims at creating a sustainable and open national research infrastructure that will give access to Canadian research in SSH (journals, monographs, data), foster their continuous and integrated growth in line with international standards, and allow their automated exploration from a high performance computing environment.

The infrastructure will build upon a set of existing initiatives dedicated to production, dissemination and preservation of SSH data research, a variety of research projects developing new contents and tools, and institutions involved in the establishment of structuration standards. The project is essential to the development of digital research in SSH, and to the evolution of the knowledge on Canadian society and culture, in the two national languages, through the interdisciplinary and historical perspectives it opens.

This infrastructure answers the strategic needs of relevant stakeholders: 1) normalization, interoperability and enrichment of Canadian social sciences and humanities datasets; 2) centralized access to full text and metadata of Canada's intellectual production; 3) enhancement of the data exploration capacities of large research collections from different sources and with different levels of structuration.

Estimated total project cost: \$5 million

Amount of funding to be requested from CFI: \$2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Database of Realworld Corpora in Child Language Development Research
Administrative Institution:	University of Manitoba
CFI liaison at institution :	Tracy Mohr (tracy.mohr@umanitoba.ca)
Project Leader :	Melanie Soderstrom
Consortium members:	University of Toronto Laboratoire de Sciences Cognitives et Psycholinguistiques, CNRS (France) Department of Psychology, University of California, Merced (USA) Brain and Cognitive Science Department, University of Rochester (USA)
Number of researchers:	5

Project summary:

With the rise of the e-data revolution, child language development researchers have begun using a device known as the LENA System and other tools to collect large-scale audio-recordings of the language environment experiences of infants and young children, together with automated analysis mechanisms. Additionally, the online sharing of audio files and transcripts among researchers has become a normative mode of research. The LENA System is currently being used internationally by a large number of researchers. This consortium seeks to allow for the sharing of large-scale audio corpora, together with associated data files and analytic tools, within a single platform. The proposed project would build basic infrastructure to house the database and share existing analytic tools developed by individual researchers. Experts in automated speech recognition will develop cutting edge new analytic tools to allow researchers to probe the data regarding important questions such as differentiating adult-directed from child-directed speech within the audio files, validity and norming for languages other than English, and automatic transcription. A significant focus would be on resolving practical and technical issues in maintaining confidentiality for the children and adults in the audio recordings while allowing for public access by researchers to the important data contained therein.

Estimated total project cost: \$1 million

Amount of funding to be requested from CFI: \$400,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Linked Data Infrastructure for Cultural Enquiry
Administrative Institution:	University of Guelph
CFI liaison at institution :	Jill Rogers (jlrogers@uoguelph.ca)
Project Leader :	Susan Brown
Consortium members:	Ryerson University Acadia University McGill University Université de Montréal University of Alberta University of British Columbia University of Lethbridge University of Ottawa University of Saskatchewan University of Victoria University of Waterloo Western University York University OCAD University Canadiana Texas A&M University, Applied Research Consortium
Number of researchers:	16

Project summary:

Computational research on human culture and history suffers from dearth of interoperable and accessible data. Canadian researchers in fields ranging from English to Computer Science and Information Science are investigating the potential of the semantic web, or linked, data to overcome both of these challenges and allow researchers to push the boundaries of knowledge while also making their data more broadly available and useful. A common linked data infrastructure would produce a step change, advancing the science of individual projects and creating a big dataset for cultural and historical inquiry that would open up new avenues of research and partnerships with the public, non-profit, and commercial sectors. This infrastructure would comprise: 1) a national semantic web resource produced from existing humanities data made newly accessible to researchers and, largely, to the public; 2) software for converting existing datasets to linked data form and making them interoperable with other datasets within and beyond the national database; 3) tools for exposing and investigating linked data in the form of query systems and visualization environments. The project will adopt and adapt existing technologies, and will train highly-qualified personnel to support innovation and secure Canada's leadership in the management of large linked cultural datasets.

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Estimated total project cost: \$3 million

Amount of funding to be requested from CFI: \$1.2 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Building a Community Data Resource
Administrative Institution:	McMaster University
CFI liaison at institution :	Sherisse Webb (webbsh@mcmaster.ca)
Project Leader :	Jeremiah Hurley
Consortium members:	None stated
Number of researchers:	8

Project summary:

The NSF recently described the frontier of social science research as data-intensive, multidisciplinary, collaborative, and problem-oriented. “Building a Community Data Resource” (BCDR) is an innovative research collaboration positioned precisely at this frontier that integrates an empirical, policy-oriented social science research program with a technology development program to achieve two goals: (1) to conduct innovative, multidisciplinary, multi-sectoral research on pressing social, health and economic challenges facing Canadian communities; and (2) to develop new data-cleaning and data-security tools needed to overcome data challenges researchers face in conducting research that exploits large stores of routinely collected community-level data. BCDR adopts a staged approach to creating a network of communities committed to supporting the use of their data for research and policy decision-making, beginning with a single community—Hamilton—to develop a partnership model, develop data cleaning/security tools and protocols for transforming raw data and ensuring security, demonstrate the benefits of such a university-community partnership, and then expand the network to other communities nationally, enabling powerful cross-community comparisons. The CFI-funded Public Economics Data Analysis Laboratory [PEDAL] at McMaster University will provide the test-bed for tool development, and the data tools developed will be made available through the national Compute Canada data platform.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	The Canadian Research Data Centre Network (CRDCN): the Move to Thin-client Operation and Administrative Data Acquisition
Administrative Institution:	McMaster University
CFI liaison at institution :	Sherisse Webb (webbsh@mcmaster.ca)
Project Leader :	Byron G. Spencer
Consortium members:	Dalhousie University Memorial University University of New Brunswick Université de Moncton Université Laval Université de Montréal McGill University/Concordia University (shared site) Université de Sherbrooke Université du Québec à Montréal/Institut national de la recherche scientifique (shared site) University of Ottawa/Carleton University (shared site) Queen’s University University of Toronto York University University of Waterloo University of Guelph Western University University of Windsor University of Manitoba University of Saskatchewan University of Alberta University of Calgary University of Lethbridge University of British Columbia University of Victoria Simon Fraser University Institute for Circumpolar Health Research (Yellowknife)
Number of researchers:	Unstated

Project summary:

By providing a national platform for research access to confidential Statistics Canada data the CRDCN, with its 27 Centres, has transformed quantitative social science research in Canada. But the research frontier now includes the analysis of large administrative data files (e.g., income, health, education, business operation), and linkages with survey files enhance the research potential. Work at that frontier

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

will require software development to bring our files to international metadata standards and thin-client technology to ensure security and enable 24/7 access. The software development will make metadata capacities interoperable with organizations such as the European Data without Boundaries initiative. One important research theme of the proposal will ask why the rate of productivity growth has declined in recent decades, thereby posing a significant threat to Canada's prosperity and to its fiscal and social sustainability. The answer may lie in understanding what is taking place at the firm level where productivity "happens." Firm-level data exist, but they need to be made research ready; related questions concerning the health of Canadians could be addressed in innovative ways by linkages with data from the firm-level pollution registry, records of employment, vital statistics, and local area health authorities.

Estimated total project cost: \$4 million

Amount of funding to be requested from CFI: \$1.6 million

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Interdisciplinary, Integrated Data Platform for Oceans and Fisheries Research (O-Fish D-Platform)
Administrative Institution:	University of British Columbia
CFI liaison at institution :	Sharon Wu (sharon.wu@ubc.ca)
Project Leaders:	Rashid Sumaila, William Cheung
Consortium members:	University of Winnipeg University of Victoria McMaster University Vancouver Aquarium Fisheries and Oceans Canada, Institute of Ocean Sciences Environment Canada Carleton University Université du Québec à Rimouski ARCTICconnexion University of Waterloo Dalhousie University Memorial University
Number of researchers:	14

Project summary:

This CFI request will build on and extend data collected by the SSHRC-funded OceanCanada Partnership, a project seeking to enhance sustainability across Canada’s oceans and coastal communities in the face of climate change, globalization and other marine ecosystem stressors. We propose the development of a new data management and integration platform that will link and harmonize existing datasets across geographical location, historical timescales, data type, and disciplinary domain. The research data infrastructure to be created will help unlock the full potential of research data on ocean and fisheries by organizing, structuring and integrating data sets and developing analytical tools to mine the data. We envision three key elements for this platform, which will involve shared computing and information processing services that only a consortium of institutions can provide. First, a password-protected database will be built to house existing data on Canadian oceans and fisheries, thereby making it easier for researchers to conduct complex analyses. This will require the development of new algorithms to harmonize disparate indicators and impute missing data where necessary. A second element will map existing quantitative and qualitative data alongside the historical development of law and public policies related to Canada’s oceans. Using harmonized datasets, a user/researcher will be able to use the database to uncover the impact or performance of a particular legal or policy decision over time and/or highlight areas where additional legal interventions may be needed.

Estimated total project cost: \$4,460,000

Amount of funding to be requested from CFI: \$1,856,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Navigable Cyberinfrastructure for Live Events
Administrative Institution:	York University
CFI liaison at institution :	Ryan McCarthy (rmccarth@yorku.ca)
Project Leader :	Doug Van Nort
Consortium members:	McMaster University University of Guelph OCAD University Concordia University McGill University Simon Fraser University University of Victoria Memorial University Stanford University Arizona State University Rensselaer Polytechnic Institute
Number of researchers:	21

Project summary:

The production and consumption of cultural events are becoming increasingly reliant on distributed digital media systems. However, the paradigms of interaction and exchange are not adequately keeping pace with these developments. The focus of this proposal is to develop a shared data platform that enables new ways of representing, navigating, processing and delivering multi-modal, time-based data for digital media applications, based on information extracted from live events that merges symbolic information (e.g. text, a musical score) with multi-modal sub-symbolic information (e.g. acoustic and kinetic sensor data streams). The project will enable new ways of fusing and interacting with this multi-layered data by merging it with higher-level information extracted from data analytics, including temporal patterns and correlations across databases. The project will allow researchers and practitioners from telematic performing arts, tele-education and data analytics to augment and coordinate their research. It will support a thriving and vibrant Canadian research community in collaboration with satellite partners, including major U.S. research universities and media production firms. By funding HQP for cyberinfrastructure development at core partner sites, this project will greatly strengthen Canada's position in performing arts and digital media research and Canada's industrial base in ICT through technology transfer and HQP training.

Estimated total project cost: \$2.5 million

Amount of funding to be requested from CFI: \$1 million

PROPOSALS TO BE SUBMITTED IN COMPETITION 2

Project title:	Networked Environment for Media Archiving and Analysis: Creating a national network for preservation of and access to digital assets (NEMAA)
Administrative Institution:	McGill University
CFI liaison at institution :	Nathalie Foisset (nathalie.foisset@mcgill.ca)
Project Leader :	Ichiro Fujinaga
Consortium members:	University of Waterloo Université de Montréal Western University DIAMM, UK Graduate School of Library and Information Science (iSchool at Illinois), US Library and Archive Canada
Number of researchers:	10

Project summary:

Canadian institutions produce and own intellectual, cultural, and scientific digital materials, including video, audio, images, text, and other formats. They are faced with the challenge of finding secure long-term preservation solutions. Preservation of data is not sufficient. Unless it can be searched, and the desired items retrieved, the data is useless. Large amounts of data that *can* be searched and analyzed provide a tremendous resource for research of many different kinds.

Solutions to the issues of secure archiving and sophisticated analysis of big data can be accomplished relatively easily using existing technologies, such as LOCKSS and Duraspace. By exploiting Compute Canada's unique network of distributed centres with high-speed computing and large storage capacity, we propose to design an infrastructure for an open repository where the data will be safe, well managed, distributed, easily deposited, accessible, and analyzable.

We will build on our own expertise by designing the system around music, in all its different formats, and by working with music libraries across Canada, as well as bringing in international experts with whom we already collaborate. The result will be a robust and flexible system that has the potential to provide Canada with a model for a national archiving solution.

Estimated total project cost: \$1.25 million

Amount of funding to be requested from CFI: \$500,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Research Data Infrastructure for Smart Animal Health and Production Management
Administrative Institution:	University of Guelph
CFI liaison at institution :	Jill Rogers (jlrogers@uoguelph.ca)
Project Leader :	Shayan Sharif
Consortium members:	University of Calgary University of Saskatchewan University of Prince Edward Island University of Montreal University of British Columbia University of Alberta University of Manitoba University of Waterloo McGill University Dalhousie University Sheridan College
Number of researchers:	Unstated

Project summary:

The global livestock industry is facing several challenges, including maintaining healthy animals, producing safe food for humans, improving animal welfare, increasing efficiency and sustainability of production while minimizing environmental impacts. With the advent and application of technologies for smart livestock farming, some of the above challenges are addressed. Using these technologies we have now generated very large bodies of data, including the data for animal genome sequences, genotypes, production and disease. However, we are not able to fully utilize these data, partly because of: 1-distributed and fragmented data sources with varying format and quality standards which can significantly hinder analysis, aggregation, and knowledge transfer and translation; 2- limited understanding of data quality and data processing standards and protocols for different animal health and production related applications. Here, we propose to establish a research data infrastructure to: 1- provide integrated infrastructure to enhance data integrity, privacy, accessibility, exchange, and interoperability among academia, industry and government; 2- build the analytics framework and tools for collecting, processing, and reporting accurate and actionable information to researchers, producers and other stakeholders; 3- create global standards for data governance best practices to address the need for protocols and policies for managing animal data.

Estimated total project cost: \$2,450,000

Amount of funding to be requested from CFI: \$980,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	A Cyber-framework for the Harmonization of Multiple Historical Data Sets and Linking to Each Other and to Modern Data
Administrative Institution:	University of Guelph
CFI liaison at institution :	Jill Rogers (jlrogers@uoguelph.ca)
Project Leader :	Kris Inwood
Consortium members:	University of Victoria University of British Columbia Simon Fraser University University of Calgary University of Alberta University of Saskatchewan University of Winnipeg Lakehead University Western University University of Waterloo McMaster University University of Toronto York University Trent University Queen’s University Carleton University University of Ottawa McGill University Université of Montreal Université Laval University of New Brunswick Dalhousie University University of Minnesota University of Michigan University of Essex London School of Economics (LSE)
Number of researchers:	unstated

Project summary:

The infrastructure identified in this application is a harmonized and integrated cyber-framework for historical data of the Canadian population and economy. The Dominion Bureau of Statistics and provincial Registrars provide comprehensive and consistent post-1921 data. For the earlier period, individual scholars and research teams have created hundreds of databases from the census and related historical sources. Collectively these data comprise an indispensable resource for the understanding of

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Canada's long-run evolution. These data currently are not coded or organized consistently to permit their joint use and survival. The organization, inter-operability and ongoing availability of these data is one function of the proposed infrastructure. A second function is the identification of entities that are common to two or more databases. Recent advances in supervised classification make it feasible to link databases using principles that differ depending on individual research needs. Entity identification across multiple historical and post-1921 databases, some of them comprising millions of records and dozens of features, is a significant computational challenge. Customized variation on demand for diverse research projects is only possible using the resources of Compute Canada. This infrastructure will be invaluable for ongoing research in the social sciences, history and in computing and information sciences.

Estimated total project cost: \$750,000

Amount of funding to be requested from CFI: \$300,000

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Project title:	Precision Agri-food Technology Research & Innovation Consortium (PATRIC)
Administrative Institution:	University of Guelph
CFI liaison at institution :	Jill Rogers (jlrogers@uoguelph.ca)
Project Leader :	Erin Skimson
Consortium members:	University of Waterloo Niagara College (IBM Industry Research Chair, Precision Agriculture) Conestoga College, Institute of Food Technology Processing Vineland Research & Innovation Centre Livestock Research Innovation Corporation Agricultural Research Institute of Ontario Agriculture and Agri-Food Canada Ontario Ministry of Agriculture, Food and Rural Affairs Southern Ontario Smart Computing Innovation Platform Dairy Farmers of Ontario Grain Farmers on Ontario IBM Communitech Innovation Guelph
Number of researchers:	Unstated

Project summary:

The convergence of agriculture and technology is providing new solutions to old problems. Managing the unprecedented streams of agri-food data being created through novel research stands to support more efficient sector management, and job creation. However, the fragmented data silos that currently support this field make analysis and data mining for insights about interrelationships between the billions of bits of genomic, environmental and phenomic data impossible and too complex to unravel.

In support of Canada's data-intensive research enterprise, the University of Guelph is proposing the **Precision Agri-food Technology Research and Innovation Consortium (PATRIC)**, the creation of a robust data platform to provide cyber-infrastructure, and tools to drive a collaborative and systematic approach to create, gather, organize, house, analyze and utilize agri-food big data. Project collaborators have identified research data management and infrastructure needs broadly into three applications that PATRIC will support:

- Data collection/edge processing tools

- Data storage, analysis and modelling systems

Cyberinfrastructure Initiative — Challenge 1 – Research Data Infrastructure

Data consumption applications/ decision support systems

This infrastructure will support data management to attract leading edge, data-intensive research, which will drive new insights, innovations and technologies, and secure Canada's position as a global leader in the ag-tech sector.

Estimated total project cost: \$4 million

Amount of funding to be requested from CFI: \$2 million



Research builds communities
La recherche au service des collectivités

450-230 Queen Street
Ottawa ON K1P 5E4
Tel 613.947.7260
Fax 613.943.0227

450-230 rue Queen
Ottawa ON K1P 5E4
Tél 613.947.7260
Télé 613.943.0227