Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)

PROGRAM SOLICITATION
NSF 12-499

National Science Foundation

Directorate for Biological Sciences

Directorate for Computer & Information Science & Engineering

Directorate for Education & Human Resources

Directorate for Engineering

Directorate for Geosciences

Directorate for Mathematical & Physical Sciences

Directorate for Social, Behavioral & Economic Sciences

Office of Cyberinfrastructure

Office of Polar Programs

National Institutes of Health

National Cancer Institute

National Institute of Biomedical Imaging and Bioengineering

National Institute on Drug Abuse

National Institute of General Medical Sciences

National Institute of Neurological Disorders and Stroke

National Library of Medicine

National Human Genome Research Institute

Full Proposal Deadline(s) (due by 5 p.m. proposer's local time):

June 13, 2012

Mid-Scale Projects

July 11, 2012

Small Projects

IMPORTANT INFORMATION AND REVISION NOTES

A revised version of the *NSF Proposal & Award Policies & Procedures Guide* (PAPPG), NSF 11-1, was issued on October 1, 2010 and is effective for proposals submitted, or due, on or after January 18, 2011. Please be advised that the guidelines contained in NSF 11-1 apply to proposals submitted in response to this funding opportunity.

Cost Sharing: The PAPPG has been revised to implement the National Science Board's recommendations regarding cost sharing. Inclusion of voluntary committed cost sharing is prohibited. In order to assess the scope of the project, all organizational resources necessary for the project must be described in the Facilities, Equipment and Other Resources section of the proposal. The description should be narrative in nature and must not include any quantifiable financial information. Mandatory cost sharing will only be required when explicitly authorized by the NSF Director. See the PAPP Guide Part I: *Grant Proposal Guide (GPG)* Chapter II.C.2.g(xi) for further information about the implementation of these recommendations.

Data Management Plan: The PAPPG contains a clarification of NSF's long-standing data policy. All proposals must describe plans for data management and sharing of the products of research, or assert the absence of the need for such plans. FastLane will not permit submission of a proposal that is missing a Data Management Plan. The Data Management Plan will be reviewed as part of the intellectual merit or broader impacts of the proposal, or both, as appropriate. Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units are available on the NSF website at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp. See Chapter II.C.2.j of the GPG for further information about the implementation of this requirement.

Postdoctoral Researcher Mentoring Plan: As a reminder, each proposal that requests funding to support postdoctoral researchers must include, as a supplementary document, a description of the mentoring activities that will be provided for such individuals. Please be advised that if required, FastLane will not permit submission of a proposal that is missing a Postdoctoral Researcher

Mentoring Plan. See Chapter II.C.2.j of the GPG for further information about the implementation of this requirement.

SUMMARY OF PROGRAM REQUIREMENTS

General Information

Program Title:

Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)

Synopsis of Program:

The Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) solicitation aims to advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large, diverse, distributed and heterogeneous data sets so as to: accelerate the progress of scientific discovery and innovation; lead to new fields of inquiry that would not otherwise be possible; encourage the development of new data analytic tools and algorithms; facilitate scalable, accessible, and sustainable data infrastructure; increase understanding of human and social processes and interactions; and promote economic growth and improved health and quality of life. The new knowledge, tools, practices, and infrastructures produced will enable breakthrough discoveries and innovation in science, engineering, medicine, commerce, education, and national security -- laying the foundations for US competitiveness for many decades to come.

The phrase "big data" in this solicitation refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.

This solicitation is one component in a long-term strategy to address national big data challenges, which include advances in core techniques and technologies; big data infrastructure projects in various science, biomedical research, health and engineering communities; education and workforce development; and a comprehensive integrative program to support collaborations of multi-disciplinary teams and communities to make advances in the complex grand challenge science, biomedical research, and engineering problems of a computational- and data-intensive world.

Today, US government agencies recognize that the scientific, biomedical and engineering research communities are undergoing a profound transformation with the use of large-scale, diverse, and high-resolution data sets that allow for data-intensive decision-making, including clinical decision making, at a level never before imagined. New statistical and mathematical algorithms, prediction techniques, and modeling methods, as well as multidisciplinary approaches to data collection, data analysis and new technologies for sharing data and information are enabling a paradigm shift in scientific and biomedical investigation. Advances in machine learning, data mining, and visualization are enabling new ways of extracting useful information in a timely fashion from massive data sets, which complement and extend existing methods of hypothesis testing and statistical inference. As a result, a number of agencies are developing big data strategies to align with their missions. This solicitation focuses on common interests in big data research across the National Institutes of Health (NIH) and the National Science Foundation (NSF).

This initiative will build new capabilities to create actionable information that leads to timely and more informed decisions. It will both help to accelerate discovery and innovation, as well as support their transition into practice to benefit society. As the recent President's Council of Advisors on Science and Technology (PCAST) 2010 review of the Networking Information Technology Research and Development (NITRD) [http://www.nitrd.gov/pcast-2010/report/nitrd-program/pcast-nitrd-report-2010.pdf] program notes, the pipeline of data to knowledge to action has tremendous potential in transforming all areas of national priority. This initiative will also lay the foundations for complementary big data activities -- big data infrastructure projects, workforce development, and progress in addressing complex, multi-disciplinary grand challenge problems in science and engineering.

Cognizant Program Officer(s):

Please note that the following information is current at the time of publishing. See program website for any updates to the points of contact.

- Vasant G. Honavar, NSF - CISE, telephone: (703) 292-7129, email: vhonavar@nsf.gov
- Jia Li, NSF - MPS, telephone: (703) 292-4870, email: jli@nsf.gov
- Dane Skow, NSF - OCI, telephone: (703) 292-4551, email: dskow@nsf.gov
- Peter H. McCartney, NSF - BIO, telephone: (703) 292-8470, email: pmccartn@nsf.gov
- Doris L. Carver, NSF - EHR, telephone: (703) 292-5038, email: dcarver@nsf.gov
- Eduardo A. Misawa, NSF - ENG, telephone: (703) 292-5353, email: emisawa@nsf.gov
- Eva Zanzerkia, NSF - GEO, telephone: (703) 292-8556, email: ezanzerk@nsf.gov
- Peter Muhlberger, NSF - SBE, telephone: (703) 292-7848, email: pmuhlber@nsf.gov
- Vladimir Papitashvili, NSF - OPP, telephone: (703) 292-7425, email: vpapita@nsf.gov
- Tandy Warnow, NSF - CISE, telephone: (703) 292-8491, email: twarnow@nsf.gov
- Karin A. Remington, National Institute of General Medical Sciences, NIH, telephone: (301) 451-6446, email: remingka@nigms.nih.gov

- Jerry Li, National Cancer Institute, NIH, telephone: (301) 435-5226, email: jiayinli@mail.nih.gov
- Vinay M. Pai, National Institute of Biomedical Imaging & Bioengineering, NIH, telephone: (301) 451-4781, email: paiv@mail.nih.gov
- Karen Skinner, National Institute on Drug Abuse, NIH, telephone: (301) 443-1887, email: kskinner@nida.nih.gov
- Yuan Liu, National Institute of Neurological Disorders and Stroke, NIH, telephone: (301) 496-1917, email: liuyuan@ninds.nih.gov
- Valerie Florance, National Library of Medicine, NIH, telephone: (301) 496-4621, email: florancev@mail.nih.gov
- Vivien Bonazzi, National Human Genome Research Institute (NHGRI), NIH, telephone: (301) 451-8276, email: bonazziv@mail.nih.gov

Applicable Catalog of Federal Domestic Assistance (CFDA) Number(s):

- 47.041 --- Engineering
- 47.049 --- Mathematical and Physical Sciences
- 47.050 --- Geosciences
- 47.070 --- Computer and Information Science and Engineering
- 47.074 --- Biological Sciences
- 47.075 --- Social Behavioral and Economic Sciences
- 47.076 --- Education and Human Resources
- 47.078 --- Office of Polar Programs
- 47.080 --- Office of Cyberinfrastructure
- 93.172 --- National Human Genome Research Institute
- 93.279 --- National Institute on Drug Abuse
- 93.286 --- National Institute of Biomedical Imaging and Bioengineering
- 93.396 --- National Cancer Institute
- 93.853 --- National Institute of Neurological Disorders and Stroke
- 93.859 --- National Institute of General Medical Sciences
- 93.879 --- National Library of Medicine

Award Information

Anticipated Type of Award: Standard Grant or Continuing Grant or Cooperative Agreement or Contracts (depending on the needs of the particular awarding agency)

Estimated Number of Awards: 15 to 20 Fifteen to twenty projects will be funded, subject to availability of funds.

Anticipated Funding Amount: $25,000,000 Up to $25,000,000 will be invested in proposals submitted to this solicitation, subject to availability of funds.

Eligibility Information

Organization Limit:

The categories of proposers eligible to submit proposals to the National Science Foundation are identified in the Grant Proposal Guide, Chapter I, Section E.

PI Limit:

None Specified

Limit on Number of Proposals per Organization:

None Specified

Limit on Number of Proposals per PI: 2

An investigator may participate as PI, co-PI, senior personnel or consultant in no more than two proposals submitted in response to this solicitation.

In the event that an individual exceeds this limit, proposals will be accepted based on earliest date and time of proposal submission. That is, the first two proposals received will be accepted and the remainder will be returned without review. No exceptions will be made.

Proposals submitted in response to this solicitation may not duplicate or be substantially similar to other proposals concurrently under consideration by NSF, NIH, or other agencies' programs or study sections.

Proposal Preparation and Submission Instructions

A. Proposal Preparation Instructions

- **Letters of Intent:** Not Applicable
- **Preliminary Proposal Submission:** Not Applicable
- **Full Proposals:**
  - Full Proposals submitted via FastLane: NSF Proposal and Award Policies and Procedures Guide, Part I: Grant Proposal Guide (GPG) Guidelines apply. The complete text of the GPG is available electronically on the NSF

website at: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg.
- Full Proposals submitted via Grants.gov: NSF Grants.gov Application Guide: A Guide for the Preparation and Submission of NSF Applications via Grants.gov Guidelines apply (Note: The NSF Grants.gov Application Guide is available on the Grants.gov website and on the NSF website at: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=grantsgovguide)

B. Budgetary Information

- **Cost Sharing Requirements:** Inclusion of voluntary committed cost sharing is prohibited.
- **Indirect Cost (F&A) Limitations:** Not Applicable
- **Other Budgetary Limitations:** Not Applicable

C. Due Dates

- **Full Proposal Deadline(s)** (due by 5 p.m. proposer's local time):

June 13, 2012

Mid-Scale Projects

July 11, 2012

Small Projects

Proposal Review Information Criteria

Merit Review Criteria: National Science Board approved criteria. Additional merit review considerations apply. Please see the full text of this solicitation for further information.

Award Administration Information

Award Conditions: Standard NSF award conditions apply.

Reporting Requirements: Additional reporting requirements apply. Please see the full text of this solicitation for further information.

TABLE OF CONTENTS

Summary of Program Requirements

I. Introduction

II. Program Description

III. Award Information

IV. Eligibility Information

V. Proposal Preparation and Submission Instructions
- A. Proposal Preparation Instructions
- B. Budgetary Information
- C. Due Dates
- D. FastLane/Grants.gov Requirements

VI. NSF Proposal Processing and Review Procedures
- A. NSF Merit Review Criteria
- B. Review and Selection Process

VII. Award Administration Information
- A. Notification of the Award
- B. Award Conditions
- C. Reporting Requirements

VIII. Agency Contacts

IX. Other Information

I. INTRODUCTION

The Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) solicitation aims to advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large, diverse, distributed and heterogeneous data sets so as to: accelerate the progress of scientific discovery and innovation; lead to new fields

of inquiry that would not otherwise be possible; encourage the development of new data analytic tools and algorithms; facilitate scalable, accessible, and sustainable data infrastructure; increase understanding of human and social processes and interactions; and promote economic growth and improved health and quality of life. The new knowledge, tools, practices, and infrastructures produced will enable breakthrough discoveries and innovation in science, engineering, medicine, commerce, education, and national security-laying the foundations for US competitiveness for many decades to come.

The phrase "big data" in this solicitation refers to large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.

This solicitation is one component in a long-term strategy to address national big data challenges, which include advances in core techniques and technologies; big data infrastructure projects in various science, biomedical research, health and engineering communities; education and workforce development; and a comprehensive integrative program to support collaborations of multi-disciplinary teams and communities to make advances in the complex grand challenge science, biomedical research, and engineering problems of a computational- and data-intensive world.

Today, US government agencies recognize that the scientific, biomedical and engineering research communities are undergoing a profound transformation with the use of large-scale, diverse, and high-resolution data sets that allow for data-intensive decision-making, including clinical decision making, at a level never before imagined. New statistical and mathematical algorithms, prediction techniques, and modeling methods, as well as multidisciplinary approaches to data collection, data analysis and new technologies for sharing data and information are enabling a paradigm shift in scientific and biomedical investigation. Advances in machine learning, data mining, and visualization are enabling new ways of extracting useful information in a timely fashion from massive data sets, which complement and extend existing methods of hypothesis testing and statistical inference. As a result, a number of agencies are developing big data strategies to align with their missions. This solicitation focuses on common interests in big data research across the NIH and the NSF.

This initiative will build new capabilities to create actionable information that leads to timely and more informed decisions. It will both help to accelerate discovery and innovation, as well as support their transition into practice to benefit society. As the recent President's Council of Advisors on Science and Technology (PCAST) 2010 review of the Networking Information Technology Research and Development (NITRD) [http://www.nitrd.gov/pcast-2010/report/nitrd-program/pcast-nitrd-report-2010.pdf] program notes, the pipeline of data to knowledge to action has tremendous potential in transforming all areas of national priority. This initiative will also lay the foundations for complementary big data activities -- big data infrastructure projects, workforce development, and progress in addressing complex, multi-disciplinary grand challenge problems in science and engineering.

II. PROGRAM DESCRIPTION

Pervasive sensing and computing across natural, built, and social environments is generating heterogeneous data at unprecedented scale and complexity. Today, scientists, biomedical researchers, engineers, educators, citizens and decision-makers live in an era of observation: data come from many disparate sources, such as sensor networks; scientific instruments, such as medical equipment, telescopes, colliders, satellites, environmental networks, and scanners; video, audio, and click streams; financial transaction data; email, weblogs, twitter feeds, and picture archives; spatial graphs and maps; and scientific simulations and models. This plethora of data sources has given rise to a phenomenal diversity in data types; data can be temporal, spatial, or dynamic and can be derived from both structured and unstructured sources. Data may have different representation types, media formats, and levels of granularity, and may be used across multiple scientific disciplines. These new sources of data and their increasing complexity contribute to an explosion of information.

A. Broader Research Goals of BIGDATA

The potential for transformational science and engineering for all disciplines is enormous, but realizing the next frontier depends on effectively managing, using, and exploiting these heterogeneous data sources. It is now possible to extract knowledge and useful information in ways that were previously impossible, and to gain new insights in a timely manner. To understand the full spectrum of what advances in big data might mean, imagine a world where:

- Responses to disaster recovery empower rescue workers and individuals to make timely and effective decisions and provide resources where they are most needed;
- Complete health/disease/genome/environmental knowledge bases enable biomedical discovery and patient-centered therapy;
- The full complement of health and medical information is available at the point of care for clinical decision-making;
- Accurate high-resolution models support forecasting and management of increasingly stressed watersheds and eco-systems;
- Access to data and software in an easy-to-use format are available to everyone around the globe;
- Consumers can purchase wearable products using materials with novel and unique properties that prevent injuries;
- The transition to use of sustainable chemistry and manufacturing materials has been accelerated to the point that the US leads in advanced manufacturing;
- Consumers have the information they need to make optimal energy consumption decisions in their homes and cars;
- Civil engineers can continuously monitor and identify at-risk man-made structures like bridges, moderate the impact of failures, and avoid disaster;
- Students and researchers have intuitive real-time tools to view, understand, and learn from publicly available large scientific data sets on everything from genome sequences to astronomical star surveys, from public health databases to particle accelerator simulations and their teachers and professors use student performance analytics to improve that learning; and
- Accurate predictions of natural disasters, such as earthquakes, hurricanes, and tornadoes, enable life-saving and cost-saving preventative actions.

Opportunities abound for learning from large-scale data sets, which can provide researchers, educators, and decision makers with information of enhanced range, quality, and depth. To jump-start a national initiative in big data discovery, this solicitation focuses on the shared research interests across NIH and NSF, and has four related objectives:

- Promote new science, address key science questions, and accelerate the progress of discovery by harnessing the value of large, heterogeneous data.
- Exploit the unique value of big data to address areas of national need, agency missions and societal and economic challenges in all parts of society.
- Support responsible stewardship and sustainability of data resulting from federally- funded research.
- Develop and sustain educational resources, a competent, knowledgeable workforce and the infrastructure needed to

advance data-enabled sciences and broaden participation in data-enabled inquiry and action.

BIGDATA seeks proposals that develop and evaluate core technologies and tools that take advantage of available collections of large data sets to accelerate progress in science, biomedical research, and engineering. Each proposal should include an evaluation plan. (See details in the Proposal Preparation and Submission Instructions section). Proposals can focus on one or more of the following three perspectives:

1. Data collection and management (DCM). Dealing with massive amounts of often heterogeneous and complex data coming from multiple sources -- such as those generated by observational systems across many scientific fields, as well as those created in transactional and longitudinal data systems across social and commercial domains -- will require the development of new approaches and tools. Potential research areas include, but are not limited to:

- New data storage, I/O systems, and architectures for continuously generated data, as well as shared and widely-distributed static and real-time data;
- Effective utilization and optimization of computing, storage, and communications resources;
- Streaming, filtering, compressed sensing, and sufficient statistics -- potentially in real time and allowing for reduction of data sizes as data are generated;
- Fault-tolerant systems that continuously aggregate and process data accurately and reliably, while ensuring integrity;
- Novel means of automatically annotating data with semantic and contextual information (that both machines and humans can read), including curation;
- Model discovery techniques that can summarize and annotate data as they are generated;
- Tracking how, when and where data are created and modified, including provenance, allowing long-lived data to provide insights in the future;
- New designs for advanced data architectures, including clouds, addressing extreme capacity, power management, and real-time control while providing for extensibility and accessibility;
- New architectures that reflect the structure and hierarchy of data as well as access techniques enabling efficient parallelism in operations across a data structure or database schema;
- Next generation multi-core processor architectures and the next generation software libraries that take maximum advantage of such architectures;
- Tools for efficient archiving, querying, retrieval and data recovery of richly structured, semi-structured, and unstructured data sets, in particular those for large transactional-intensive databases;
- Research in software development to enable correct and effective programming of big data applications, including new programming languages, methodologies, and environments; and
- New approaches to improve data quality, validity, integrity, and consistency, as well as methods to account for and quantify uncertainty in large data sets, including the development of data assurance processes, formal methods, and algorithms.

2. Data analytics (DA). Significant impacts will result from advances in analysis, simulation, modeling, and interpretation to facilitate discovery of phenomena, to realize causality of events, to enable prediction, and to recommend action. Advances will allow, for example, modeling of social networks and learning communities, reliable prediction of consumer behaviors and preferences, and the surfacing of communication patterns among unknown groups at a larger, global scale; extraction of meaning from textual data; more effective correlation of events; enhanced ability to extract knowledge from large-scale experimental and observational datasets; and extracting useful information from incomplete data. Potential research areas include, but are not limited to:

- Development of new algorithms, programming languages, data structures, and data prediction tools;
- Computational models and the underlying mathematical and statistical theory needed to capture important performance characteristics of computing over massive data sets;
- Data-driven high fidelity modeling and simulations and/or reduced-order models enabling improved designs and/or processes for engineering industries, and direct interfacing with measurements and equipment;
- Novel algorithmic techniques with the capability to scale to handle the largest, most complex data sets being created now and in the future;
- Real-time processing techniques addressing the scale of continuously generated data sets, as well as real-time visualization and analysis tools that allow for more responsive and intuitive study of data;
- Computational, mathematical and statistical techniques for modeling physical, engineering, social or other processes that produce massive data sets;
- Novel applications of inverse methods to big data problems;
- Mining techniques that involve novelty and anomaly detection, trend detection and/or taxonomy creation as well as predictive models, hypothesis generation and automated discovery, including fundamentally new statistical, mathematical and computational methods for identifying changes in massive datasets;
- Development of data extraction techniques (e.g. natural language processing) to unlock the vast amounts of information currently stored as unstructured data (e.g. text);
- New scalable data visualization techniques and tools, which are able to illustrate the correlation of events in multidimensional data, synthesize information to provide new insights, and allow users to drill down for more refined information;
- Techniques to integrate disparate data and translate data into knowledge to enable on-the-fly decision-making;
- Development of usable state-of-the-art tools and theory in statistical inference and statistical learning for knowledge discovery from massive, complex, and dynamic data sets; and
- Consideration to potential limitations, e.g., the number of possible passes over the data, energy conservation, new communication architectures, and their implications for solution accuracy.

3. E-science collaboration environments (ESCE). A comprehensive "big data" cyberinfrastructure is necessary to allow for broad communities of scientists and engineers to have access to diverse data and to the best and most usable inferential and visualization tools. Potential research areas include, but are not limited to:

- Novel collaboration environments for diverse and distant groups of researchers and students to coordinate their work (e.g., through data and model sharing and software reuse, tele-presence capability, crowdsourcing, social networking capabilities) with greatly enhanced efficiency and effectiveness for the scientific collaboration;
- Automation of the discovery process (e.g., through machine learning, data mining, and automated inference);
- Automated modeling tools to provide multiple views of massive data sets that are useful to diverse disciplines;
- New data curation techniques for managing the complex and large flow of scientific output in a multitude of disciplines;
- Development of systems and processes that efficiently incorporate autonomous anomaly and trend detection with human interaction, response, and reaction;
- End-to-end systems that facilitate the development and use of scientific workflows and new applications;
- New approaches to development of research questions that might be pursued in light of access to heterogeneous, diverse, big data;
- New models for cross-disciplinary information fusion and knowledge sharing;
- New approaches for effective data, knowledge, and model sharing and collaboration across multiple domains and disciplines;

Securing access to data using innovative techniques to prevent excessive replication of data to external entities;
- Providing secure and controlled role-based access to centrally managed data environments;
- Remote operation, scheduling, and real-time access to distant instruments and data resources;
- Protection of privacy and maintenance of security in aggregated personal and proprietary data (e.g., de-identification);
- Generation of aggregated or summarized data sets for sharing and analyses across jurisdictional and other end user boundaries; and
- E-publishing tools that provide unique access, learning, and development opportunities.

In addition to the three science and engineering perspectives on big data described above, all proposals must also include a description of how the project will build capacity:

Capacity-building Requirement (CB). CB activities are critical to the growth and health of this emerging area of research and education. There are three broad types of CB activities: 1) appropriate models, policies and technologies to support responsible and sustainable big data stewardship; 2) training and communication strategies, targeted to the various research communities and/or the public; and 3) sustainable, cost-effective infrastructure for data storage, access and shared services.

To develop a coherent set of stewardship, outreach and education activities in big data discovery, each research proposal must focus on at least one capacity-building activity. Examples include, but are not limited to:

- Novel, effective frameworks of roles and responsibilities for various big data stakeholders (i.e., researchers, collaborators, research communities, research institutions, funding agencies);
- Efficient and effective models for data management, considering issues such as structure and formatting of data, standardization of terminology, metadata and provenance, persistent identifiers, data quality, etc.;
- Development of accurate cost models and structures;
- Establishing appropriate cyberinfrastructure models, prototypes and facilities for long-term sustainable data;
- Policies and processes for evaluating data value and balancing cost with value in an environment of limited resources;
- Policies and procedures to ensure appropriate access and use of data resources
- Economic sustainability models;
- Community standards, provenance tracking, privacy, and security;
- Communication strategies for public outreach and engagement;
- Education and workforce development; and
- Broadening participation in big data activities.

It is expected that at least one PI from each funded project will attend a BIGDATA Principal Investigator (PI) meeting in year two of the initiative to present project research findings and capacity building or community outreach activities. Requested budgets should include funds for travel to this event. An overarching goal is to leverage all the BIGDATA investments to build a successful science and engineering community that is well trained in dealing with and analyzing big data from various sources.

Finally, a project may choose to focus its science and engineering big data project in an area of national priority, but this is optional:

National Priority Domain Area Option. In addition to the research areas described above, to fully exploit the value of the investments made in large-scale data collection, BIGDATA would also like to support research in particular domain areas, especially areas of national priority, including health IT, emergency response and preparedness, clean energy, cyberlearning, material genome, national security, and advanced manufacturing. Research projects may focus on the science and engineering of big data in one or more of these domain areas while simultaneously engaging in the foundational research necessary to make general advances in "big data."

B. Sponsoring Agency Mission Specific Research Goals

1. NATIONAL SCIENCE FOUNDATION

NSF intends to support excellent research in the three areas mentioned above in this solicitation. It is important to note that this solicitation represents the start of a multi-year, multi-agency initiative, which at NSF is part of the Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21).

Innovative information technologies are transforming the fabric of society and data is the new currency for science, education, government and commerce. High performance computing (HPC) has played a central role in establishing the importance of simulation and modeling as the third pillar of science (theory and experiment being the first two), and the growing importance of data is creating the fourth pillar. Science and engineering researchers are pushing beyond the current boundaries of knowledge by addressing increasingly complex questions, which often require sophisticated integration of massive amounts of highly heterogeneous data from theoretical, experimental, observational, and simulation and modeling research programs. These efforts, which rely heavily on teams of researchers, observing and sensor platforms and other data collection efforts, computing facilities, software, advanced networking, analytics, visualization, and models, lead to critical breakthroughs in all areas of science and engineering and lay the foundation for a comprehensive, research requirements-based approach to the development of NSF's Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21).

Finally, NSF is interested in integrating foundational computing research with the domain sciences to make advances in big data challenges related to national priorities, such as health IT, emergency response and preparedness, clean energy, cyberlearning, material genome, national security, and advanced manufacturing.

2. NATIONAL INSTITUTES OF HEALTH

NIH's mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce the burdens of illness and disability. Toward this end, NIH provides leadership and direction to programs designed to improve the health of the Nation by conducting and supporting research in the causes, diagnosis, prevention, and cure of human diseases; in the processes of human growth and development; in the biological effects of environmental contaminants; in the understanding of mental, addictive and physical disorders; and in directing programs for the collection, dissemination, and exchange of information in medicine and health, including the development and support of medical libraries and the training of medical librarians and other health information specialists.

To support these goals, NIH seeks proposals of core technologies and tools in the areas described in Part II.A. that take advantage of imaging, molecular, cellular, electrophysiological, chemical, behavioral, epidemiological, clinical, and/or other data sets applicable to understanding health, and to preventing and treating the various diseases and conditions of relevance to the participating Institutes, Centers, and Offices, including, but not limited to:

- Long-term, economically and technically self-sustaining storage solutions that enable archiving, mining, retrieving, and analyzing diverse data sets relevant to biomedical and/or behavioral research in an environment of dynamic provenance, including evolving formats, standards and ontologies;

- Approaches that facilitate examining comparisons of genetic association data from across the globe to identify possible epistatic and/or gene-environment interactions;
- Computational and informatics infrastructure and tools for the achieving and analysis of "structural and functional connectomes" of large, complex, and dynamic networks, including those relevant to neuroscience and addiction research;
- Development of better tools and computational approaches to mine and store emergent crowd-sourced datasets (e.g., those produced by Twitter, Facebook, and other social media platforms), where de facto datasets or data points often emerge spontaneously, and can reveal insights into economic, social and technical factors affecting local, regional, national and global health, including drug use and drug toxicity;
- Computational and informatics infrastructure and tools for conducting landscape analyses of biomedical research areas (e.g., combinations of ontology, text mining, search engine tools to integrate and analyze heterogeneous data and literature to capture the stage of a research topic or the current knowledge of a disease);
- Efforts in collection, de-identification, validation, archiving, and dissemination of large volumes of imaging and associated genetic, pathological, and clinical data (e.g., cancer or neurological disorders) generated from clinical research projects and clinical trials to enable meaningful coalescing and organization of the data that are complex and diverse, and to enable collective, collaborative, and comprehensive analyses of the enhanced data collections for extraction of valuable information that would otherwise be obscured by the limited trial size or the diversity of data and disparities in data collection methodologies;
- Development of novel data validation, normalization, and analysis approaches 1) for the identification and development of biomarkers for (e.g., cancer, or neurological disorders) diagnosis and therapy monitoring, 2) for the development of clinical and imaging data modeling that is purposeful and predictive, and directed at elucidating primary biological and pathological driving factors and processes underlying disease and treatment response, and 3) for the development of effective clinical decision support criteria and tools that can be consistently and uniformly adopted;
- Development of informatics tools and infrastructure for creation of National Registries with the intent of improving the care of patients (e.g., cancer or neurology patients) by capturing real-time, real-world quality assured information on treatment delivery and health outcomes through a prospective electronic registry infrastructure; and
- Development of predictive modeling techniques, based on a large volume of patient data, to provide real-time individualized and optimal diagnosis and treatment plans "at the bedside".

Also of specific interest are novel tools and techniques for interacting with and managing very large and/or heterogeneous data sets, including, but not limited to:

- Approaches for minimizing human intervention in the organization and management of large biomedical knowledge resources, such as automated annotators and intelligent agents to handle updates and quality control in health knowledge repositories;
- Approaches for *in silico* science using published knowledge to generate or test hypotheses;
- Approaches, technical and cultural, to share and compare data among research groups and patient advocacy groups;
- Intelligent agents that can read a biomedical article and explain its contents to a layperson;
- Interactive publications that incorporate access to data/knowledge resources, along with tools and approaches for adding data and reanalyzing findings;
- Accessible data infrastructures that strongly facilitate a culture of data sharing among biomedical researchers; and
- Data infrastructures for benchmarked biomedical data that can be reused for validation and verification purposes.

NIH encourages applicants to address (where relevant) the sustainability of any software and data sharing plans after funding from this BIGDATA program would cease.

Because of the varied interests and priorities of the Institutes, Centers and Offices participating in this announcement, prospective applicants are strongly encouraged to contact the scientific officer of any targeted Institute, Center or Office prior to application.

III. AWARD INFORMATION

Estimated program budget, number of awards and average award size/duration are subject to the availability of funds. An estimated fifteen to twenty projects will be funded, subject to availability of funds. Up to $25,000,000 will be invested in proposals submitted to this solicitation, subject to availability of funds.

All awards under this solicitation made by NIH and / or NSF will be as grants or cooperative agreements or other contract vehicles as determined by the supporting agency. Two sizes of projects are expected to be funded under this solicitation:

1. Small projects: One or two investigators can ask for up to $250,000 per year for up to three years.
2. Mid-scale projects: Three or more investigators can ask for funding between $250,001 and $1,000,000 per year for up to five years.

For both types of projects, we encourage scientists from all disciplines to participate. Projects will be awarded depending on the availability of funds and with consideration for creating a balanced overall portfolio, from foundational big data science and engineering to areas of national priority, including health IT, emergency response and preparedness, clean energy, cyberlearning, material genome, national security, and advanced manufacturing.

IV. ELIGIBILITY INFORMATION

Organization Limit:

The categories of proposers eligible to submit proposals to the National Science Foundation are identified in the Grant Proposal Guide, Chapter I, Section E.

PI Limit:

None Specified

Limit on Number of Proposals per Organization:

None Specified

Limit on Number of Proposals per PI: 2

An investigator may participate as PI, co-PI, senior personnel or consultant in no more than two proposals submitted in response to this solicitation.

In the event that an individual exceeds this limit, proposals will be accepted based on earliest date and time of proposal submission. That is, the first two proposals received will be accepted and the remainder will be returned without review. No exceptions will be made.

Proposals submitted in response to this solicitation may not duplicate or be substantially similar to other proposals concurrently under consideration by NSF, NIH, or other agencies' programs or study sections.

V. PROPOSAL PREPARATION AND SUBMISSION INSTRUCTIONS

A. Proposal Preparation Instructions

Full Proposal Preparation Instructions: Proposers may opt to submit proposals in response to this Program Solicitation via Grants.gov or via the NSF FastLane system.

- Full proposals submitted via FastLane: Proposals submitted in response to this program solicitation should be prepared and submitted in accordance with the general guidelines contained in the NSF Grant Proposal Guide (GPG). The complete text of the GPG is available electronically on the NSF website at: http://www.nsf.gov/publications/pub_summ.jsp?ods_key=gpg. Paper copies of the GPG may be obtained from the NSF Publications Clearinghouse, telephone (703) 292-7827 or by e-mail from nsfpubs@nsf.gov. Proposers are reminded to identify this program solicitation number in the program solicitation block on the NSF Cover Sheet For Proposal to the National Science Foundation. Compliance with this requirement is critical to determining the relevant proposal processing guidelines. Failure to submit this information may delay processing.

- Full proposals submitted via Grants.gov: Proposals submitted in response to this program solicitation via Grants.gov should be prepared and submitted in accordance with the NSF Grants.gov Application Guide: A Guide for the Preparation and Submission of NSF Applications via Grants.gov. The complete text of the NSF Grants.gov Application Guide is available on the Grants.gov website and on the NSF website at: (http://www.nsf.gov/publications/pub_summ.jsp?ods_key=grantsgovguide). To obtain copies of the Application Guide and Application Forms Package, click on the Apply tab on the Grants.gov site, then click on the Apply Step 1: Download a Grant Application Package and Application Instructions link and enter the funding opportunity number, (the program solicitation number without the NSF prefix) and press the Download Package button. Paper copies of the Grants.gov Application Guide also may be obtained from the NSF Publications Clearinghouse, telephone (703) 292-7827 or by e-mail from nsfpubs@nsf.gov.

In determining which method to utilize in the electronic preparation and submission of the proposal, please note the following:

Collaborative Proposals. All collaborative proposals submitted as separate submissions from multiple organizations must be submitted via the NSF FastLane system. Chapter II, Section D.4 of the Grant Proposal Guide provides additional information on collaborative proposals.

Additional Instructions include the following:

Proposal Titles: Proposal titles must indicate the **BIGDATA** program, followed by a colon, then the project size, followed by a colon, then the research perspective, followed by the title of the project. For example, a Small data collection and management project proposal title would be **BIGDATA: Small: DCM: Title.** If all perspectives are involved, they should be ordered by their importance in the research being proposed. For example, **BIGDATA: Small: DA: ESCE: DCM: Title.** Titles of collaborative Mid-Scale proposals should be prepared as above, but should also include "Collaborative Research" followed by a colon before the title of the project. For example, the title of each proposal for a collaborative set of proposals for a Mid-Scale Project might be **BIGDATA: Mid-Scale: DA: Collaborative Research: Title.**

Project Summary (1-page limit): At the top of the Project Summary page enter the title of the BIGDATA project, the name of the PI and the lead institution. Provide a summary description of the BIGDATA project, including its transformative research and education goals, and the community (communities) that will be impacted by its results. In separate statements, provide a succinct summary of the intellectual merit and broader impacts of the proposed project. Those proposals that are targeting a specific agency sponsorship should indicate so in the last line of their 1-page Project Summary, e.g., "Requested funding agency:" followed by that agency's abbreviated name, "NSF" or "NIH," but only if the PIs have previously communicated with a program officer from that agency and received permission or instruction to do so. Those not so designated will be considered for funding by either of the joint sponsoring agencies.

Full proposals that do not address the intellectual merit and broader impacts of the proposed project in separate statements will be returned without review.

Project Description. There is a 15 page limit for small proposals and non-collaborative mid-scale proposals**. Mid-scale** collaborative proposals must include a Coordination Plan. **Up to two additional pages are permitted in the Project Description for this purpose only, allowing a maximum of 17 pages.** ***Proposals that do not comply with this requirement will be returned without review***.

Proposers are reminded that each proposal must have a separate section that deals explicitly with Capacity Building activities and goals, as described in II.A.

The Coordination Plan. It must include 1) the specific roles of the collaborating PIs, Co-PIs, other Senior Personnel and paid consultants at all organizations involved; 2) how the project will be managed across institutions and disciplines; 3) identification of the specific coordination mechanisms that will enable cross-institution and/or cross-discipline scientific integration (e.g., workshops, graduate student exchange, project meetings at conferences, use of videoconferencing and other communication tools, software repositories, etc.); and 4) specific references to the budget line items that support these coordination mechanisms.

The Evaluation Plan. Within the project description, include a plan to evaluate the technologies developed, which could include results from applications of that technology to specific domains, efficacy studies, assessments of learning and engagement, and other such activities. The evaluation plan should be appropriate for the size and scope of the project.

Proposal Budget. It is expected that at least one PI from each funded project will attend a BIGDATA Principal Investigator (PI) meeting in year two of the initiative to present project research findings and capacity building or community outreach activities. Requested budgets should include funds for travel to this event.

Supplementary Documents. Supplementary documents are limited to the specific types of documentation listed in the GPG, with the following exceptions:

Human Subjects Protection. Proposals involving human subjects should include a supplementary document of no more than two pages in length summarizing potential risks to human subjects; plans for recruitment and informed consent; inclusion of women, minorities, and children; and planned procedures to protect against or minimize potential risks.

Vertebrate Animals. Proposals involving vertebrate animals should include a supplementary document of no more than two pages in length that addresses the following points:

1. Detailed description of the proposed use of the animals, including species, strains, ages, sex, and number to be used;
2. Justification for the use of animals, choice of species, and numbers to be used;
3. Information on the veterinary care of the animals;
4. Description of procedures for minimizing discomfort, distress, pain, and injury; and
5. Method of euthanasia and the reasons for its selection.

Data Management Plan. Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplement should describe how the proposal will manage its data and share research results and may include:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

2. the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);

3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;

4. policies and provisions for re-use, re-distribution, and the production of derivatives; and

5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Data management requirements and plans specific to the Directorate, Office, Division, Program, or other NSF unit, relevant to a proposal are available at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp. If guidance specific to the program is not available, then the requirements established in this section apply.

Simultaneously submitted collaborative proposals and proposals that include subawards are a single unified project and should include only one supplemental combined Data Management Plan, regardless of the number of non-lead collaborative proposals or subawards included. Fastlane will not permit submission of a proposal that is missing a Data Management Plan. Proposals for supplementary support to an existing award are not required to include a Data Management Plan.

All BIGDATA proposals must have a Data Management Plan. Proposers who feel that the plan cannot fit within the supplement limit of two pages may use part of the 15-page Project Description for additional data management information. Proposers are advised that the Data Management Plan may not be used to circumvent the 15-page Project Description limitation. The Data Management Plan will be reviewed as an integral part of the proposal, coming under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.

Software Sharing Plan. A brief software dissemination plan (with appropriate timelines) must accompany the Data Management Plan. It can be part of the 2-page Data Management Plan Supplementary Document. If two pages are insufficient for Data Management and Software Sharing Plans, then the Software Sharing Plan can be included under a separate heading in the Project Description. There is no prescribed single license for software produced through grants responding to this announcement. However, the Program does have goals for software dissemination, and reviewers will be instructed to evaluate the dissemination plan relative to these goals:

1. The software should be freely available to science, biomedical and engineering researchers and educators in the non-profit sector, such as institutions of education, research institutions, and government laboratories.

2. The terms of software availability should permit the dissemination and commercialization of enhanced or customized versions of the software, or incorporation of the software or pieces of it into other software packages.

3. To preserve utility to the community, the software should be transferable such that another individual or team can continue development in the event that the original investigators are unwilling or unable to do so.

4. The terms of software availability should include the ability of researchers to modify the source code and to share modifications with other colleagues. An applicant should take responsibility for creating the original and subsequent official versions of a piece of software.

5. To further enhance the potential impact of their software, applicants may consider proposing a plan to manage and disseminate the improvements or customizations of their tools and resources by others. This proposal may include a plan to incorporate the enhancements into the official core software, may involve the creation of an infrastructure for plug-ins, or may describe some other solution.

Proposals that do not comply with this requirement will be returned without review.

B. Budgetary Information

Cost Sharing: Inclusion of voluntary committed cost sharing is prohibited

C. Due Dates

- **Full Proposal Deadline(s)** (due by 5 p.m. proposer's local time):

  June 13, 2012

  Mid-Scale Projects

  July 11, 2012

  Small Projects

D. FastLane / Grants.gov Requirements

- **For Proposals Submitted Via FastLane:**

  Detailed technical instructions regarding the technical aspects of preparation and submission via FastLane are available at: https://www.fastlane.nsf.gov/a1/newstan.htm. For FastLane user support, call the FastLane Help Desk at 1-800-673-6188 or e-mail fastlane@nsf.gov. The FastLane Help Desk answers general technical questions related to the use of the FastLane system. Specific questions related to this program solicitation should be referred to the NSF program staff contact(s) listed in Section VIII of this funding opportunity.

  Submission of Electronically Signed Cover Sheets. The Authorized Organizational Representative (AOR) must electronically sign the proposal Cover Sheet to submit the required proposal certifications (see Chapter II, Section C of the Grant Proposal Guide for a listing of the certifications). The AOR must provide the required electronic certifications within five working days following the electronic submission of the proposal. Further instructions regarding this process are available on the FastLane Website at: https://www.fastlane.nsf.gov/fastlane.jsp.

- **For Proposals Submitted Via Grants.gov:**

  Before using Grants.gov for the first time, each organization must register to create an institutional profile. Once registered, the applicant's organization can then apply for any federal grant on the Grants.gov website. Comprehensive information about using Grants.gov is available on the Grants.gov Applicant Resources webpage: http://www07.grants.gov/applicants/app_help_reso.jsp. In addition, the NSF Grants.gov Application Guide provides additional technical guidance regarding preparation of proposals via Grants.gov. For Grants.gov user support, contact the Grants.gov Contact Center at 1-800-518-4726 or by email: support@grants.gov. The Grants.gov Contact Center answers general technical questions related to the use of Grants.gov. Specific questions related to this program solicitation should be referred to the NSF program staff contact(s) listed in Section VIII of this solicitation.

  Submitting the Proposal: Once all documents have been completed, the Authorized Organizational Representative (AOR) must submit the application to Grants.gov and verify the desired funding opportunity and agency to which the application is submitted. The AOR must then sign and submit the application to Grants.gov. The completed application will be transferred to the NSF FastLane system for further processing.

VI. NSF PROPOSAL PROCESSING AND REVIEW PROCEDURES

Proposals received by NSF are assigned to the appropriate NSF program where they will be reviewed if they meet NSF proposal preparation requirements. All proposals are carefully reviewed by a scientist, engineer, or educator serving as an NSF Program Officer, and usually by three to ten other persons outside NSF who are experts in the particular fields represented by the proposal. These reviewers are selected by Program Officers charged with the oversight of the review process. Proposers are invited to suggest names of persons they believe are especially well qualified to review the proposal and/or persons they would prefer not review the proposal. These suggestions may serve as one source in the reviewer selection process at the Program Officer's discretion. Submission of such names, however, is optional. Care is taken to ensure that reviewers have no conflicts of interest with the proposal.

While NSF allows proposers to recommend reviewers, NIH does not. Any proposers that are considering NIH funding should therefore follow this guideline: *Proposers are invited to suggest areas of research (e.g., conferences) they believe are a close fit with the work outlined in this proposal, but should not name individual potential reviewers. Proposers may suggest persons they would prefer not review the proposal.*

A. NSF Merit Review Criteria

All NSF proposals are evaluated through use of the two National Science Board (NSB)-approved merit review criteria: intellectual merit and the broader impacts of the proposed effort. In some instances, however, NSF will employ additional criteria as required to highlight the specific objectives of certain programs and activities.

The two NSB-approved merit review criteria are listed below. The criteria include considerations that help define them. These considerations are suggestions and not all will apply to any given proposal. While proposers must address both merit review criteria, reviewers will be asked to address only those considerations that are relevant to the proposal being considered and for which the reviewer is qualified to make judgments.

What is the intellectual merit of the proposed activity?
How important is the proposed activity to advancing knowledge and understanding within its own field or across different fields? How well qualified is the proposer (individual or team) to conduct the project? (If appropriate, the reviewer will comment on the quality of the prior work.) To what extent does the proposed activity suggest and

explore creative, original, or potentially transformative concepts? How well conceived and organized is the proposed activity? Is there sufficient access to resources?

What are the broader impacts of the proposed activity?
How well does the activity advance discovery and understanding while promoting teaching, training, and learning? How well does the proposed activity broaden the participation of underrepresented groups (e.g., gender, ethnicity, disability, geographic, etc.)? To what extent will it enhance the infrastructure for research and education, such as facilities, instrumentation, networks, and partnerships? Will the results be disseminated broadly to enhance scientific and technological understanding? What may be the benefits of the proposed activity to society?

Examples illustrating activities likely to demonstrate broader impacts are available electronically on the NSF website at: http://www.nsf.gov/pubs/gpg/broaderimpacts.pdf.

Mentoring activities provided to postdoctoral researchers supported on the project, as described in a one-page supplementary document, will be evaluated under the Broader Impacts criterion.

Additional Solicitation Specific Review Criteria

Additional NIH Review Criteria

The mission of the NIH is to support science in pursuit of knowledge about the biology and behavior of living systems and to apply that knowledge to extend healthy life and reduce the burdens of illness and disability. In their evaluations of intellectual merit, reviewers will be asked to consider the following criteria that are used by NIH:

Overall Impact. Reviewers will provide an overall impact/priority score to reflect their assessment of the likelihood for the project to exert a sustained, powerful influence on the research field(s) involved, in consideration of the following five core review criteria, and additional review criteria (as applicable for the project proposed).

Significance. Does the project address an important problem or a critical barrier to progress in the field? If the aims of the project are achieved, how will scientific knowledge, technical capability, and/or clinical practice be improved? How will successful completion of the aims change the concepts, methods, technologies, treatments, services, or preventative interventions that drive this field?

Investigator(s). Are the PD/PIs, collaborators, and other researchers well suited to the project? If Early Stage Investigators or New Investigators, do they have appropriate experience and training? If established, have they demonstrated an ongoing record of accomplishments that have advanced their field(s)? If the project is collaborative or multi-PD/PI, do the investigators have complementary and integrated expertise; are their leadership approach, governance and organizational structure appropriate for the project?

Innovation. Does the application challenge and seek to shift current research or clinical practice paradigms by utilizing novel theoretical concepts, approaches or methodologies, instrumentation, or interventions? Are the concepts, approaches or methodologies, instrumentation, or interventions novel to one field of research or novel in a broad sense? Is a refinement, improvement, or new application of theoretical concepts, approaches or methodologies, instrumentation, or interventions proposed?

Approach. Are the overall strategy, methodology, and analyses well-reasoned and appropriate to accomplish the specific aims of the project? Are potential problems, alternative strategies, and benchmarks for success presented? If the project is in the early stages of development, will the strategy establish feasibility and will particularly risky aspects be managed? If the project involves clinical research, are the plans for 1) protection of human subjects from research risks, and 2) inclusion of minorities and members of both sexes/genders, as well as the inclusion of children, justified in terms of the scientific goals and research strategy proposed?

Environment. Will the scientific environment in which the work will be done contribute to the probability of success? Are the institutional support, equipment and other physical resources available to the investigators adequate for the project proposed? Will the project benefit from unique features of the scientific environment, subject populations, or collaborative arrangements?

Where applicable, the following items will also be considered:

Protections for Human Subjects. For research that involves human subjects but does not involve one of the six categories of research that are exempt under 45 CFR Part 46, the committee will evaluate the justification for involvement of human subjects and the proposed protections from research risk relating to their participation according to the following five review criteria: 1) risk to subjects, 2) adequacy of protection against risks, 3) potential benefits to the subjects and others, 4) importance of the knowledge to be gained, and 5) data and safety monitoring for clinical trials.

For research that involves human subjects and meets the criteria for one or more of the six categories of research that are exempt under 45 CFR Part 46, the committee will evaluate: 1) the justification for the exemption, 2) human subjects involvement and characteristics, and 3) sources of materials.

Inclusion of Women, Minorities, and Children. When the proposed project involves clinical research, the committee will evaluate the proposed plans for inclusion of minorities and members of both genders, as well as the inclusion of children.

Vertebrate Animals. The committee will evaluate the involvement of live vertebrate animals as part of the scientific assessment according to the following five points: 1) proposed use of the animals, and species, strains, ages, sex, and numbers to be used; 2) justifications for the use of animals and for the appropriateness of the species and numbers proposed; 3) adequacy of veterinary care; 4) procedures for limiting discomfort, distress, pain and injury to that which is unavoidable in the conduct of scientifically sound research including the use of analgesic, anesthetic, and tranquilizing drugs and/or comfortable restraining devices; and 5) methods of euthanasia and reason for selection if not consistent with the AVMA Guidelines on Euthanasia.

Biohazards. Reviewers will assess whether materials or procedures proposed are potentially hazardous to research personnel and/or the environment, and if needed, determine whether adequate protection is proposed.

Budget. The reasonableness of the proposed budget and the requested period of support in relation to the proposed research will be assessed.

For those proposals that are selected for funding consideration by participating NIH Institutes, the NIH will ask the applicant(s) to resubmit the proposal in an NIH-approved format directly to the Center for Scientific Review (CSR) at the NIH. Each of these NIH applications will be accompanied by a cover letter that associates the application with BIGDATA. Applicants will not be allowed to increase the proposed budget or change the scientific content of the application in the resubmission to the NIH. These NIH applications, along with the summary statements generated based on the review, will be entered into the NIH IMPAC-II system.

Additional Joint NSF and NIH Review Criteria:

NSF and NIH Plans for Data Management and Sharing of the Products of Research.

The required Data Management and Sharing of Products of Research Plans that are described below will be evaluated as part of the proposal review process.

Data Management Plan. All BIGDATA proposals must have a Data Management Plan. Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplement should describe how the proposal will manage its data and share research results and may include:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

2. the standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);

3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;

4. policies and provisions for re-use, re-distribution, and the production of derivatives; and

5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Software Sharing Plan. A brief software dissemination plan (with appropriate timelines) must accompany the Data Management Plan. It can be part of the 2-page Data Management Plan Supplementary Document. If two pages are insufficient for Data Management and Software Sharing Plans, then the Software Sharing Plan can be included under a separate heading in the Project Description. There is no prescribed single license for software produced through grants responding to this announcement. However, the Program does have goals for software dissemination, and reviewers will be instructed to evaluate the dissemination plan relative to these goals:

1. The software should be freely available to science, biomedical and engineering researchers and educators in the non-profit sector, such as institutions of education, research institutions, and government laboratories.

2. The terms of software availability should permit the dissemination and commercialization of enhanced or customized versions of the software, or incorporation of the software or pieces of it into other software packages.

3. To preserve utility to the community, the software should be transferable such that another individual or team can continue development in the event that the original investigators are unwilling or unable to do so.

4. The terms of software availability should include the ability of researchers to modify the source code and to share modifications with other colleagues. An applicant should take responsibility for creating the original and subsequent official versions of a piece of software.

5. To further enhance the potential impact of their software, applicants may consider proposing a plan to manage and disseminate the improvements or customizations of their tools and resources by others. This proposal may include a plan to incorporate the enhancements into the official core software, may involve the creation of an infrastructure for plug-ins, or may describe some other solution.

All proposals **must** include a section on Capacity Building that will be evaluated by the reviewers.

NSF staff also will give careful consideration to the following in making funding decisions:

Integration of Research and Education
One of the principal strategies in support of NSF's goals is to foster integration of research and education through the programs, projects, and activities it supports at academic and research institutions. These institutions provide abundant opportunities where individuals may concurrently assume responsibilities as researchers, educators, and students and where all can engage in joint efforts that infuse education with the excitement of discovery and enrich research through the diversity of learning perspectives.

Integrating Diversity into NSF Programs, Projects, and Activities
Broadening opportunities and enabling the participation of all citizens -- women and men, underrepresented minorities, and persons with disabilities -- is essential to the health and vitality of science and engineering. NSF is committed to this principle of diversity and deems it central to the programs, projects, and activities it considers and supports.

B. Review and Selection Process

Proposals submitted in response to this program solicitation will be reviewed by Ad hoc Review and/or Panel Review.

Reviewers will be asked to formulate a recommendation to either support or decline each proposal. The Program Officer assigned to manage the proposal's review will consider the advice of reviewers and will formulate a recommendation.

After scientific, technical and programmatic review and consideration of appropriate factors, the NSF Program Officer recommends to the cognizant Division Director whether the proposal should be declined or recommended for award. NSF is striving to be able to tell applicants whether their proposals have been declined or recommended for funding within six months. The time interval begins on the deadline or target date, or receipt date, whichever is later. The interval ends when the Division Director accepts the Program Officer's recommendation.

A summary rating and accompanying narrative will be completed and submitted by each reviewer. In all cases, reviews are treated as confidential documents. Verbatim copies of reviews, excluding the names of the reviewers, are sent to the Principal Investigator/Project Director by the Program Officer. In addition, the proposer will receive an explanation of the decision to award or decline funding.

In all cases, after programmatic approval has been obtained, the proposals recommended for funding will be forwarded to the Division of Grants and Agreements for review of business, financial, and policy implications and the processing and issuance of a grant or other agreement. Proposers are cautioned that only a Grants and Agreements Officer may make commitments, obligations

or awards on behalf of NSF or authorize the expenditure of funds. No commitment on the part of NSF should be inferred from technical or budgetary discussions with a NSF Program Officer. A Principal Investigator or organization that makes financial or personnel commitments in the absence of a grant or cooperative agreement signed by the NSF Grants and Agreements Officer does so at their own risk.

NIH: For those proposals that are selected for potential funding by participating NIH Institutes, the PI will be required to resubmit the proposal in an NIH-approved format directly to the Center for Scientific Review (http://www.csr.nih.gov/) of the NIH. PIs invited to resubmit to NIH will receive further information on resubmission procedures from NIH. An applicant will not be allowed to increase the proposed budget or change the scientific content of the proposal in the resubmission to the NIH as an NIH application. Indirect costs on any foreign subawards/subcontracts will be limited to eight (8) percent. These NIH applications will be entered into the NIH IMPAC II system. The results of the review will be presented to the involved Institutes' National Advisory Councils for the second level of review. Subsequent to the Council reviews, NIH Institutes will make their funding determinations and selected awards will be made. Subsequent grant administration procedures for NIH awardees, including those related to New and Early Stage Investigators (http://www.niaid.nih.gov/researchfunding/grant/Pages/aag.aspx), will be in accordance with the policies of NIH. Applications selected for NIH funding will use the NIH funding mechanisms.

Proposals that are funded by NIH are expected to be renewed as competing continuing applications. PIs should contact their NIH Program Officer for additional information. For information purposes, NIH PIs may wish to consult the NIAID web site, "All about Grants," which provides excellent generic information about all aspects of NIH grantsmanship, including competitive renewals (http://www.niaid.nih.gov/researchfunding/grant/Pages/aag.aspx).

VII. AWARD ADMINISTRATION INFORMATION

A. Notification of the Award

Notification of the award is made to *the submitting organization* by a Grants Officer in the Division of Grants and Agreements. Organizations whose proposals are declined will be advised as promptly as possible by the cognizant NSF Program administering the program. Verbatim copies of reviews, not including the identity of the reviewer, will be provided automatically to the Principal Investigator. (See Section VI.B. for additional information on the review process.)

B. Award Conditions

NSF Award Conditions: An NSF award consists of: (1) the award letter, which includes any special provisions applicable to the award and any numbered amendments thereto; (2) the budget, which indicates the amounts, by categories of expense, on which NSF has based its support (or otherwise communicates any specific approvals or disapprovals of proposed expenditures); (3) the proposal referenced in the award letter; (4) the applicable award conditions, such as Grant General Conditions (GC-1); * or Research Terms and Conditions * and (5) any announcement or other NSF issuance that may be incorporated by reference in the award letter. Cooperative agreements also are administered in accordance with NSF Cooperative Agreement Financial and Administrative Terms and Conditions (CA-FATC) and the applicable Programmatic Terms and Conditions. NSF awards are electronically signed by an NSF Grants and Agreements Officer and transmitted electronically to the organization via e-mail.

*These documents may be accessed electronically on NSF's Website at http://www.nsf.gov/awards/managing/award_conditions.jsp?org=NSF. Paper copies may be obtained from the NSF Publications Clearinghouse, telephone (703) 292-7827 or by e-mail from nsfpubs@nsf.gov.

More comprehensive information on NSF Award Conditions and other important information on the administration of NSF awards is contained in the NSF *Award & Administration Guide* (AAG) Chapter II, available electronically on the NSF Website at http://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag.

NIH Award Conditions: Contact the cognizant NIH organization program officer for additional information.

C. Reporting Requirements

For all multi-year grants (including both standard and continuing grants), the Principal Investigator must submit an annual project report to the cognizant Program Officer at least 90 days before the end of the current budget period. (Some programs or awards require more frequent project reports). Within 90 days after expiration of a grant, the PI also is required to submit a final project report, and a project outcomes report for the general public.

Failure to provide the required annual or final project reports, or the project outcomes report will delay NSF review and processing of any future funding increments as well as any pending proposals for that PI. PIs should examine the formats of the required reports in advance to assure availability of required data.

PIs are required to use NSF's electronic project-reporting system, available through FastLane, for preparation and submission of annual and final project reports. Such reports provide information on activities and findings, project participants (individual and organizational), publications, and other specific products and contributions. PIs will not be required to re-enter information previously provided, either with a proposal or in earlier updates using the electronic system. Submission of the report via FastLane constitutes certification by the PI that the contents of the report are accurate and complete. The project outcomes report must be prepared and submitted using Research.gov. This report serves as a brief summary, prepared specifically for the public, of the nature and outcomes of the project. This report will be posted on the NSF website exactly as it is submitted by the PI.

More comprehensive information on NSF Reporting Requirements and other important information on the administration of NSF awards is contained in the NSF *Award & Administration Guide* (AAG) Chapter II, available electronically on the NSF Website at http://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag.

NIH: Contact the cognizant NIH organization Program Officer for additional information.

VIII. AGENCY CONTACTS

Please note that the program contact information is current at the time of publishing. See program website for any updates to the points of contact.

General inquiries regarding this program should be made to:

- Vasant G. Honavar, NSF - CISE, telephone: (703) 292-7129, email: vhonavar@nsf.gov
- Jia Li, NSF - MPS, telephone: (703) 292-4870, email: jli@nsf.gov
- Dane Skow, NSF - OCI, telephone: (703) 292-4551, email: dskow@nsf.gov
- Peter H. McCartney, NSF - BIO, telephone: (703) 292-8470, email: pmccartn@nsf.gov
- Doris L. Carver, NSF - EHR, telephone: (703) 292-5038, email: dcarver@nsf.gov
- Eduardo A. Misawa, NSF - ENG, telephone: (703) 292-5353, email: emisawa@nsf.gov
- Eva Zanzerkia, NSF - GEO, telephone: (703) 292-8556, email: ezanzerk@nsf.gov
- Peter Muhlberger, NSF - SBE, telephone: (703) 292-7848, email: pmuhlber@nsf.gov
- Vladimir Papitashvili, NSF - OPP, telephone: (703) 292-7425, email: vpapita@nsf.gov
- Tandy Warnow, NSF - CISE, telephone: (703) 292-8491, email: twarnow@nsf.gov
- Karin A. Remington, National Institute of General Medical Sciences, NIH, telephone: (301) 451-6446, email: remingka@nigms.nih.gov
- Jerry Li, National Cancer Institute, NIH, telephone: (301) 435-5226, email: jiayinli@mail.nih.gov
- Vinay M. Pai, National Institute of Biomedical Imaging & Bioengineering, NIH, telephone: (301) 451-4781, email: paiv@mail.nih.gov
- Karen Skinner, National Institute on Drug Abuse, NIH, telephone: (301) 443-1887, email: kskinner@nida.nih.gov
- Yuan Liu, National Institute of Neurological Disorders and Stroke, NIH, telephone: (301) 496-1917, email: liuyuan@ninds.nih.gov
- Valerie Florance, National Library of Medicine, NIH, telephone: (301) 496-4621, email: florancev@mail.nih.gov
- Vivien Bonazzi, National Human Genome Research Institute (NHGRI), NIH, telephone: (301) 451-8276, email: bonazziv@mail.nih.gov

For questions related to the use of FastLane, contact:

- FastLane Help Desk, telephone: 1-800-673-6188; e-mail: fastlane@nsf.gov.

For questions relating to Grants.gov contact:

- Grants.gov Contact Center: If the Authorized Organizational Representatives (AOR) has not received a confirmation message from Grants.gov within 48 hours of submission of application, please contact via telephone: 1-800-518-4726; e-mail: support@grants.gov.

General Correspondence email

For general correspondence, please reply to bigdata@nsf.gov.

IX. OTHER INFORMATION

The NSF Website provides the most comprehensive source of information on NSF Directorates (including contact information), programs and funding opportunities. Use of this Website by potential proposers is strongly encouraged. In addition, National Science Foundation Update is a free e-mail subscription service designed to keep potential proposers and other interested parties apprised of new NSF funding opportunities and publications, important changes in proposal and award policies and procedures, and upcoming NSF Regional Grants Conferences. Subscribers are informed through e-mail when new publications are issued that match their identified interests. Users can subscribe to this service by clicking the "Get NSF Updates by Email" link on the NSF web site.

Grants.gov provides an additional electronic capability to search for Federal government-wide grant opportunities. NSF funding opportunities may be accessed via this new mechanism. Further information on Grants.gov may be obtained at http://www.grants.gov.

National Science Foundation

http://www.nsf.gov

National Institutes of Health

http://nih.gov

ABOUT THE NATIONAL SCIENCE FOUNDATION

The National Science Foundation (NSF) is an independent Federal agency created by the National Science Foundation Act of 1950, as amended (42 USC 1861-75). The Act states the purpose of the NSF is "to promote the progress of science; [and] to advance the national health, prosperity, and welfare by supporting research and education in all fields of science and engineering."

NSF funds research and education in most fields of science and engineering. It does this through grants and cooperative agreements to more than 2,000 colleges, universities, K-12 school systems, businesses, informal science organizations and other research organizations throughout the US. The Foundation accounts for about one-fourth of Federal support to academic institutions for basic research.

NSF receives approximately 55,000 proposals each year for research, education and training projects, of which approximately 11,000 are funded. In addition, the Foundation receives several thousand applications for graduate and postdoctoral fellowships. The agency operates no laboratories itself but does support National Research Centers, user facilities, certain oceanographic vessels and Arctic and Antarctic research stations. The Foundation also supports cooperative research between universities and industry, US participation in international scientific and engineering efforts, and educational activities at every academic level.

Facilitation Awards for Scientists and Engineers with Disabilities provide funding for special assistance or equipment to enable persons with disabilities to work on NSF-supported projects. See Grant Proposal Guide Chapter II, Section D.2 for instructions regarding preparation of these types of proposals.

The National Science Foundation has Telephonic Device for the Deaf (TDD) and Federal Information Relay Service (FIRS) capabilities that enable individuals with hearing impairments to communicate with the Foundation about NSF programs, employment or general information. TDD may be accessed at (703) 292-5090 and (800) 281-8749, FIRS at (800) 877-8339.

The National Science Foundation Information Center may be reached at (703) 292-5111.

The National Science Foundation promotes and advances scientific progress in the United States by competitively awarding grants and cooperative agreements for research and education in the sciences, mathematics, and engineering.

To get the latest information about program deadlines, to download copies of NSF publications, and to access abstracts of awards, visit the NSF Website at http://www.nsf.gov

- **Location:** 4201 Wilson Blvd. Arlington, VA 22230
- **For General Information** (NSF Information Center)**:** (703) 292-5111
- **TDD (for the hearing-impaired):** (703) 292-5090
- **To Order Publications or Forms:**

  Send an e-mail to: nsfpubs@nsf.gov

  or telephone: (703) 292-7827
- **To Locate NSF Employees:** (703) 292-5111

PRIVACY ACT AND PUBLIC BURDEN STATEMENTS

The information requested on proposal forms and project reports is solicited under the authority of the National Science Foundation Act of 1950, as amended. The information on proposal forms will be used in connection with the selection of qualified proposals; and project reports submitted by awardees will be used for program evaluation and reporting within the Executive Branch and to Congress. The information requested may be disclosed to qualified reviewers and staff assistants as part of the proposal review process; to proposer institutions/grantees to provide or obtain data regarding the proposal review process, award decisions, or the administration of awards; to government contractors, experts, volunteers and researchers and educators as necessary to complete assigned work; to other government agencies or other entities needing information regarding applicants or nominees as part of a joint application review process, or in order to coordinate programs or policy; and to another Federal agency, court, or party in a court or Federal administrative proceeding if the government is a party. Information about Principal Investigators may be added to the Reviewer file and used to select potential candidates to serve as peer reviewers or advisory committee members. See Systems of Records, NSF-50, "Principal Investigator/Proposal File and Associated Records," 69 Federal Register 26410 (May 12, 2004), and NSF-51, "Reviewer/Proposal File and Associated Records," 69 Federal Register 26410 (May 12, 2004). Submission of the information is voluntary. Failure to provide full and complete information, however, may reduce the possibility of receiving an award.

An agency may not conduct or sponsor, and a person is not required to respond to, an information collection unless it displays a valid Office of Management and Budget (OMB) control number. The OMB control number for this collection is 3145-0058. Public reporting burden for this collection of information is estimated to average 120 hours per response, including the time for reviewing instructions. Send comments regarding the burden estimate and any other aspect of this collection of information, including suggestions for reducing this burden, to:

Suzanne H. Plimpton
Reports Clearance Officer
Division of Administrative Services

National Science Foundation
Arlington, VA 22230