



National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230

NSF 14-069

Dear Colleague Letter - Big Data For The Geosciences - Calling Attention to Big Data Challenges in the Geosciences

May 9, 2014

This Dear Colleague letter is intended to highlight the importance of Big Data challenges in the geosciences and identify an opportunity to address part of that challenge. NSF encourages active dialog between geoscientists and computational scientists to consider and develop community-oriented and broadly applicable solutions.

The geosciences, in the study of the Earth System, including atmosphere, geospace, earth, ocean and polar sciences, faces the growing need to manage, synthesize and understand unprecedented scales, rates, and diversity of data.

"Big Data" refers to data that challenge existing methods due to size, complexity, or rate of availability. The National Big Data Initiative is a coordinated set of activities involving multiple federal agencies that aims to address this challenge through:

- Advancing the fundamental techniques and technologies for data and knowledge management, data analytics, and data-enabled discovery;
- Accelerate scientific and engineering discovery and innovation;
- Enabling new fields of inquiry and new modes of discovery and innovation;
- Facilitate the development of new data management, data analytics, discovery algorithms and tools;
- Enabling scalable, accessible, and sustainable data infrastructure;
- Advancing understanding of natural, human and social processes and interactions;
- Supporting big data education and workforce development;
- Enabling multi-disciplinary collaborative teams and communities to address complex scientific, biomedical, and engineering grand challenges; and
- Promoting economic growth and improved health and quality of life.

The Geosciences Directorate (GEO) calls the attention of U.S. Academic geosciences researchers to the Critical Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) program, solicitation (NSF 14-543, http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767) organized by the Computer and Information Science and Engineering Directorate (CISE). The purpose of this solicitation is the development of fundamentally new techniques and technologies in Big Data knowledge management, data analytics challenges and computational scientific discovery, encompassing formalization, analysis, and algorithmic realization of all aspects of scientific discovery. These techniques and technologies may be attractive to the geosciences. In particular, GEO may find value in:

- Shared solutions to common geosciences Big Data problems
- Geosciences data issues that require the development of novel core technologies. Areas of interest can include but are not limited to, data mining; data assimilation; real time data management;

algorithm scaling; combining structured and unstructured data; new methods when scaling is not possible.

- Geosciences science problems for which the application of modern Big Data methods would make a sea change. The larger the community of use the better.

GEO is asking the community to organize and articulate common challenges in the Big Data arena to better shape future projects and collaborations. Key aspects the geosciences should consider are how outcomes from Big Data projects can be incorporated into geosciences research, broadly disseminated and shared across the discipline. The community should consider and express the definition and implications of Big Data for the geosciences or which aspects of Big Data require immediate attention for advancing geosciences research.

EarthCube provides resources to organize communities of scientists, to articulate common problems, find tools or concepts or application problems for tool development and to build partnerships between geoscientists, computational scientists, CI specialists and mathematicians (<http://earthcube.org>). This effort may be used by geoscientists to organize community-based Big Data requirements and potential solutions.

This year, we expect the geosciences community to begin development towards future Big Data projects and to identify the potential role of Big Data in the geosciences. Some projects that develop novel core technologies using geosciences data or use cases may be acceptable in the current solicitation. PIs are encouraged to contact their program directors to discuss their potential projects and ideas. Questions about GEO cyberinfrastructure may also be directed to Eva Zanzerkia, GEO/OAD, ezanzerk@nsf.gov.

Sincerely,

Roger Wakimoto, Assistant Director, Directorate for Geosciences (GEO)