National Science Foundation
4201 Wilson Boulevard
Arlington, Virginia 22230

NSF 12-094

Dear Colleague Letter - Data Infrastructure in Mathematical and Physical Sciences

DATE: June 22, 2012

Science and engineering are rapidly and profoundly becoming data-intensive. Nowhere is this felt more acutely than in the mathematical and physical sciences (MPS), as national facilities, campus, and laboratory instruments collectively generate petabytes of data that can be shared across communities. As such, the creation, storage, dissemination, and curation of data play a key role in continuing to advance the MPS disciplines to new frontiers. The opportunities that are afforded by the wealth of data that is currently available to the scientific community represent a critical component of the NSF vision for the *Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)*. Through CIF21, NSF aims "to accelerate research and education and new functional capabilities in computational and data-intensive science and engineering". The CIF21 structure is not a monolithic entity but an integrated set of activities that are designed to work in conjunction with each other to ensure that the scientific community has the infrastructure tools that will enable the solution of the major grand challenges that are faced by 21st century science and engineering.

One component of this portfolio is the recently-announced crosscutting program announcement *Data Infrastructure Building Blocks (DIBBs)* that addresses the need to "develop, implement and support the new methods, management structures and technologies to store and manage the diversity, size, and complexity of current and future data sets and data streams." The goals of the program, issued through the Office of Cyberinfrastructure, touch upon every scientific research area in NSF. The term "building blocks" has been carefully chosen to reflect the fact that the intent of the investments made through the program is to create infrastructure components that have potential long-term utility beyond just the immediate use. Properly viewed, these building blocks can contribute fundamentally to a broader data-intensive science capability that will serve the entire national science community.

The Mathematical and Physical Sciences (MPS) Directorate views this activity as being a critical component of CIF21 that can enhance and complement its own efforts in addressing the data needs for the MPS community. These include areas such as: managing large data sets, collecting data from distributed and/or heterogeneous sources, accessing and integrating multi-source data from multiple groups or communities developing standards and data certification, and ensuring access to data at the appropriate stages and in useful ways. The three components of DIBBs (conceptualization, implementation, and interoperability) offer opportunities to address one of more of these core data needs at various stages of their development. Scientists engaged in DIBBs projects will thus be able to directly participate in developing, and thereby helping to shape, the resulting infrastructure.

Members of the MPS community are encouraged to take advantage of the opportunities offered through the DIBBs activity wherever appropriate. Feel free to contact any of the program officers in the MPS Divisions listed below to obtain a better understanding of the MPS goals expressed in the program announcement and how they relate to the overall goals of the infrastructure development.

Tom Statler (tstatler@nsf.gov) Evelyn Goldfield (egoldfie@nsf.gov)
Daryl Hess (dhess@nsf.gov) Christopher Stark (cstark@nsf.gov)
Jean Cottam Allen (jcallen@nsf.gov) Dean Evasius (devasius@nsf.gov)

With best regards,

H. Edward Seidel
AD/MPS