National Snow and Ice Data Center

449 UCB
University of Colorado
Boulder, CO 80309-0449
Phone: +1.303.492.6199
Fax: +1.303.492.2468
Email: nsidc@nsidc.org
http://nsidc.org

NSIDC Response to the National Big Data R and D Initiative:

Dr. Mark Serreze, Director NSIDC
National Snow and Ice Data Center
Cooperative Institute for Research in Environmental Science
University of Colorado at Boulder
449 UCB Boulder, CO 80309

Our Experience Working with Big Data and Our Role in the Big Data Innovation Ecosystem

For nearly 40 years, the National Snow and Ice Data Center (NSIDC) has been at the forefront of cryospheric data management. We make data and information about the cryosphere and Polar Regions available and useful to researchers, policymakers, the media, industry, and the public at the behest of many agencies, including NASA, NOAA, NSF, and USAID. As leaders in data management, we are well versed in making data available across complex domains and disciplines, including remote sensing, traditional and local knowledge, and the social sciences. As a data center, we have contributed to education and training in data management at all levels through a number of activities, including the Federation of Earth Science Information Partners data management short course for scientists. We are also committed to the education of the next generation of data curators, through classes at the Graduate School of Library and Information Systems at UIUC and through the inculcation of undergraduate, graduate, and early career scientists from a variety of fields and programs into NSIDC activities.

For millions of people, NSIDC is the trusted source of information about Arctic sea ice and the Polar Regions. This trust was gained through our efforts to both provide thoroughly documented data and value-added higher-level data and information products and tools, all accessible to other disciplines and communities. Because our staff includes domain scientists, who are noted researchers in their own right, a strong informatics/data management research team, talented technical writers, and software developers, we can provide a number of valuable services, including expert analyses of the data for multiple audiences, with a superb user services office to help people understand how to use the data. As such, NSIDC and other domain repositories fill a critical role within the overall data ecosystem.

Our Comments and Suggestions:

What are the gaps that are not addressed in the Visions and Priority Actions document?

- Recognize that a Big Data innovation ecosystem will be built upon existing resources, such as domain repositories and other initiatives, and that a primary initial effort should be to identify these resources and determine how to integrate them into the emerging ecosystem.
- Understand that there is a major gap between how business and science see Big Data. The language and definitions used by these communities varies. Additionally, the scope and types of questions from each of these perspectives diverge greatly; however, each community has much to learn from each other.

From an interagency perspective, what do you think are the most high impact ideas at the frontiers of big data research and development?

- Provide incentives and funding that allow for full documentation and metadata with the goal of increasing attention to data usability. This is a high impact area that is currently underfunded/undervalued. Our experience has shown us that without attention to data reuse, none

NSIDC National Snow and Ice Data Center

449 UCB
University of Colorado
Boulder, CO 80309-0449
Phone: +1.303.492.6199
Fax: +1.303.492.2468
Email: nsidc@nsidc.org
http://nsidc.org

National Snow and Ice Data Center
University of Colorado Boulder

of the NITRD values are obtainable. We advise making documentation, metadata, and data usability a funded part of all future data generation.

What new research, education, and/or infrastructure investments do you think will be game-changing for the big data innovation ecosystem?

- Support funding models for domain repositories that allows cross-agency and cross-sector data stewardship activities over the long-term rather than focusing on maintaining short term non-data stewardship-focused grants.
- Create cloud-based test beds for R&D with sample datasets, community supported algorithm libraries, with cross-agency access and support.
- Integrate the library and information science communities with the existing suite of domain repositories so that the experiences gained over decades within repositories becomes visible and is used within those communities.
- Establish rewards throughout the research enterprise for researchers who engage in good data management practices.

How can the federal government most effectively enable new partnerships, particularly those that cross sectors or domains?

- Facilitate activities that build upon the missions of multiple agencies to ensure better cross-agency funding coordination for data activities.
- Develop funding models for infrastructure and data management projects that provide funding for extended periods of time.
- Continue funding planning workshops and Research Coordination Networks to ensure that communities have the opportunity to connect and build collaborations around Big Data.
- Create new fellowships and programs for graduate students to enable both scientific research and the development of technical/data systems (perhaps something similar to the IGERT fellowships that produced researchers fluent in interdisciplinary research)

Why do you feel your contribution/ideas should be included in the strategic plan?

NSIDC covers the gamut of types of Big Data - from large volume satellite and airborne sensors (the MODIS L2 data set includes more than 7 million data, metadata, and browse files), to extremely diverse long-tail data sets (the ACADIS system has more than 2800 data sets), to observations made by Indigenous Arctic residents. Our data are extremely diverse, both qualitative and quantitative, and most of the data are openly available except where ethical or legal considerations prohibit it. We have experience tying widely disparate types of data and information together, including experience working with and connecting Local and Traditional Knowledge (LTK) and community data. While these data are not as large as current standards, they are diverse ranging from instrument-based point samples, maps, text narratives, photographs, video, audio narratives and knowledge models. Increasingly, these data are being linked to other data, such as remote sensing and sensor measurements, held by NSIDC and other organizations. Additionally, we handle the whole range of data from the lowest level raw data through the development of higher level data products, such as interpretive products useful to the public and policy makers.

Examples, where appropriate

IceBridge example: IceBridge collects data from airborne sensors, ties it with ICESat, Cryosat, ICESat-2 big data, but also with data collected from diverse measurements, the Navy is making on the ground in selected locations.