Mark A. Parsons
National Snow and Ice Data Center/World Data Center for Glaciology, Boulder

Data and Scientists in a Sustained Arctic Observing Network

The central requirement of a Sustained Arctic Observing Network (SAON) is to provide data to scientists. This then raises the questions: what data, which scientists, and how do they interact? Answering these questions can help us define how best to develop systems and processes to meet the fundamental requirements of SAON. The National Science Board (NSB 2005) defines three basic categories of digital data—research data, resource or community data, and reference data—and show how these different categories of data create different policy implications. Research data are typically collected by focused research projects and are intended to serve a particular group of people. They may be useful to other researchers, but that is not the initial intent, so the data often do not adhere to common standards (metadata, formats, policies) or have well-defined archive and distribution systems. Community data serve a broader, but still defined, single scientific or engineering community. They are more likely to adhere to community standards and have defined archive and distribution systems, but these systems are subject to shifting agency priorities and may not be maintained. Reference data serve large and diverse communities. The standards used for these collections often define standards for broader use. The budgets supporting these data are typically large and the expectation is that the data will be maintained indefinitely. Ballagh, et al. (2005) provide examples of how different polar data can be categorized this way and how the categorization may evolve over time. The National Research Council (NRC 2006) provides a good list of "key variables" that need to be monitored in the Arctic, existing activities to collect and share data on these variables, and major gaps in these observing activities. It would be useful to document the status of these variables in terms of the NSB categories and how or whether certain data collections should evolve to a higher category. In doing this analysis, it is important to consider what the Open Archival Information System Reference Model calls the "designated community" (i.e., which scientists) for a given collection, because this, in turn, defines many of the archival and access requirements for the data (CCSDS 2002). This is especially important when we consider the NRC's recommendation to explicitly involve Arctic residents in the design of an AON system (NRC 2006, p. 4) and the fact that user communities can change over time, often in unanticipated ways (Parsons and Duerr 2005). We should also consider how these user communities think. For example, David Fulkner, in a keynote presentation[1] to the principle investigators of the U.S. National Science Foundation's (NSF) AON projects, showed how scientists have two worldviews. One view sees the world as a collection of features arranged in space (e.g., GIS users), while the other view sees the world as a set of parameters that vary over time (e.g., climate modelers). While Fulkner emphasizes that this is an over-simplified dichotomy, it illustrates how the two basic approaches to data integration (i.e., integration through time or space) may be relevant in different situations. More importantly, it also illustrates how consideration of the human element in the network is essential to developing a good system to provide data to scientists. In developing SAON, we must think beyond the technical problems to develop what Van House et al. call a sociotechnical system—a "network of technology, information, documents, *people, and practices"* (2003, p. 1 my emphasis). Three recent workshops have helped define some of the practices required to develop such a sociotechnical system. The related themes of building trust and understanding quality were persistent in these workshops and should guide the practices that underpin an effective network. One workshop explored how researchers search for and understand data outside their expertise. The ability to communicate with data experts in order to assess the quality of data in question was viewed as a critical piece of an interdisciplinary data discovery system (Parsons and Wilson 2007). Another workshop of

Canadian investigators working on International Polar Year projects revealed the tensions created by the IPY Data Policy's[2] requirement for timely data release in that some investigators do not trust "outsiders" to use their data fairly or appropriately. Both themes emerged in a NSF workshop on Arctic system science, which recommended the formation of an "Arctic Synthesis Collaboratory" to support the Arctic science community by providing "(1) a Community Network and Synthesis 'Meeting Grounds,' (2) Data and Modeling Support, (3) Education, Outreach, and Policy [resources], and (4) Scientist Training and Development" (Vörösmarty et al. 2007). The last point on educating scientists in data management is particularly important, and is also emphasized by the International Council of Science (ICSU 2004). Finally, we must consider how best to extend existing data systems to enable broad discovery and use of diverse data types. The NRC (2006, Table 3A.4) provides an initial inventory. This inventory should be updated[3] and the systems assessed in light of the themes identified here and the requirements identified in the SAON and other workshops. SAON can then move effectively forward to the next step of determining how these systems and activities can be coordinated and sustained over the long-term.

[1] http://www.eol.ucar.edu/projects/aon-cadis/meetings/200703/misc/Fulker/
[2] http://classic.ipy.org/Subcommittees/final_ipy_data_policy.pdf

References Cited

Ballagh, L. M. et al. 2005. "'Long-Lived Digital Data' for NSF Scientists." Poster presented to the National Science Board Meeting at the University of Colorado. See http://nsidc.org/cgi-bin/publications/pub_list.pl.

CCSDS (Consultative Committee for Space Data Systems). 2002. *Reference Model for An Open Archival Information System (OAIS) CCSDS 650.0-B-1 Issue 1* Washington, DC: CCSDS Secretariat.

ICSU (International Council for Science). 2004. *ICSU Report of the CSPR Assessment Panel on Scientific Data and Information*

NRC (National Research Council). 2006. *Toward An Integrated Arctic Observing Network* Washington, DC: National Acadamies Press.

NSB (National Science Board). 2005. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21St Century* National Science Foundation. 87 pp.

Parsons, MA, and BE Wilson. 2007. User-driven design of a data system for the International Polar Year. *Eos, Transactions of the American Geophysical Union. 88.*

Parsons, MA, and R Duerr. 2005. Designating user communities for scientific data: challenges and solutions. *Data Science Journal. 4*:31-38.

Van House, NA, AP Bishop, and BP Buttenfield. 2003. Introduction: Digital Libraries as Sociotechnical Systems. In AP Bishop, NA Van House, and BP Buttenfield (ed.). *Digital Library Use* Cambridge, Mass.: The MIT Press.

Vörösmarty, CJ, et al.. 2007. New Perspectives Through Data Discovery and Modeling. *Eos, Transactions of the AGU. 88.*

[3] For example, elements of the IPY Data and Information Service (http://ipydis.org) are now operational including the Cooperative Arctic Data and Information Service which supports the Arctic Observing Network (AON)(http://www.eol.ucar.edu/projects/aon-cadis/)