National Snow and Ice Data Center

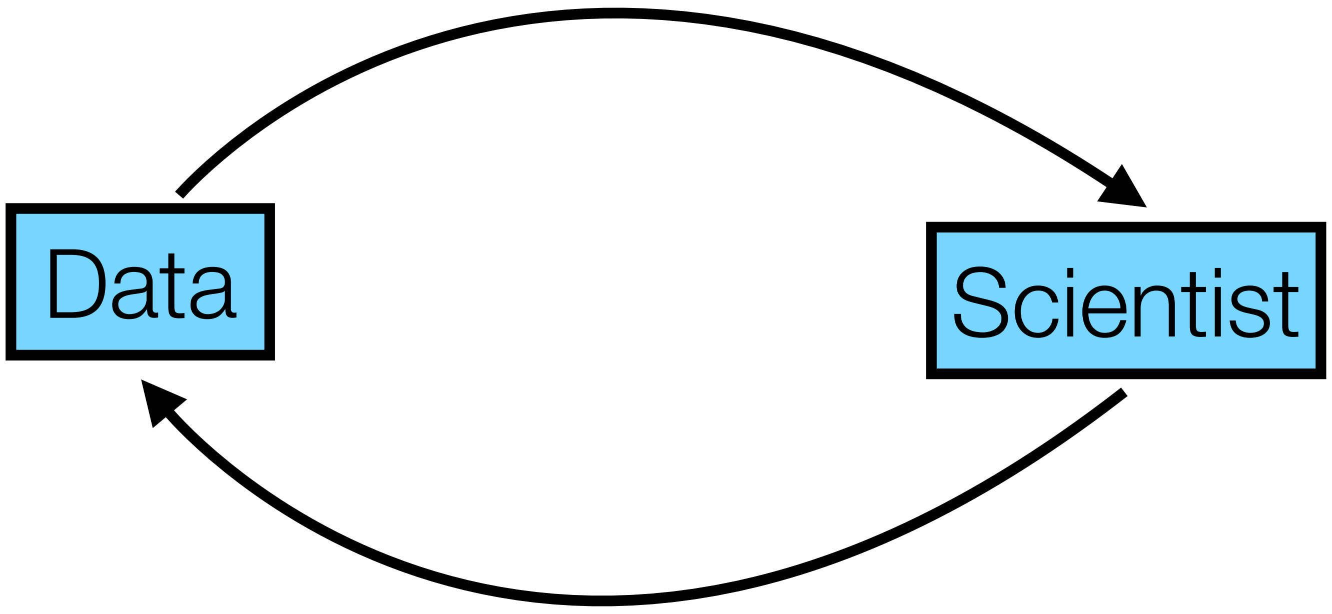
Data and Scientists in a Sustained Arctic Observing Network

Mark A. Parsons

SAON Workshop
Stockholm, Sweden
13 November 2007

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 Need



like all models, it's wrong but it may be useful

emphasize the blue as "communication interfaces" IT can help that communication but it's broader and includes a large data and human component--planned data and etadata, social networks, professional connection and development, mechanisms for collaboration, means of communications

So I will present
- vision
- context of that vision and how it might guide a way forward
- thoughts on IT
- best practices
- some possible approaches

My vision is...

A Data Utility

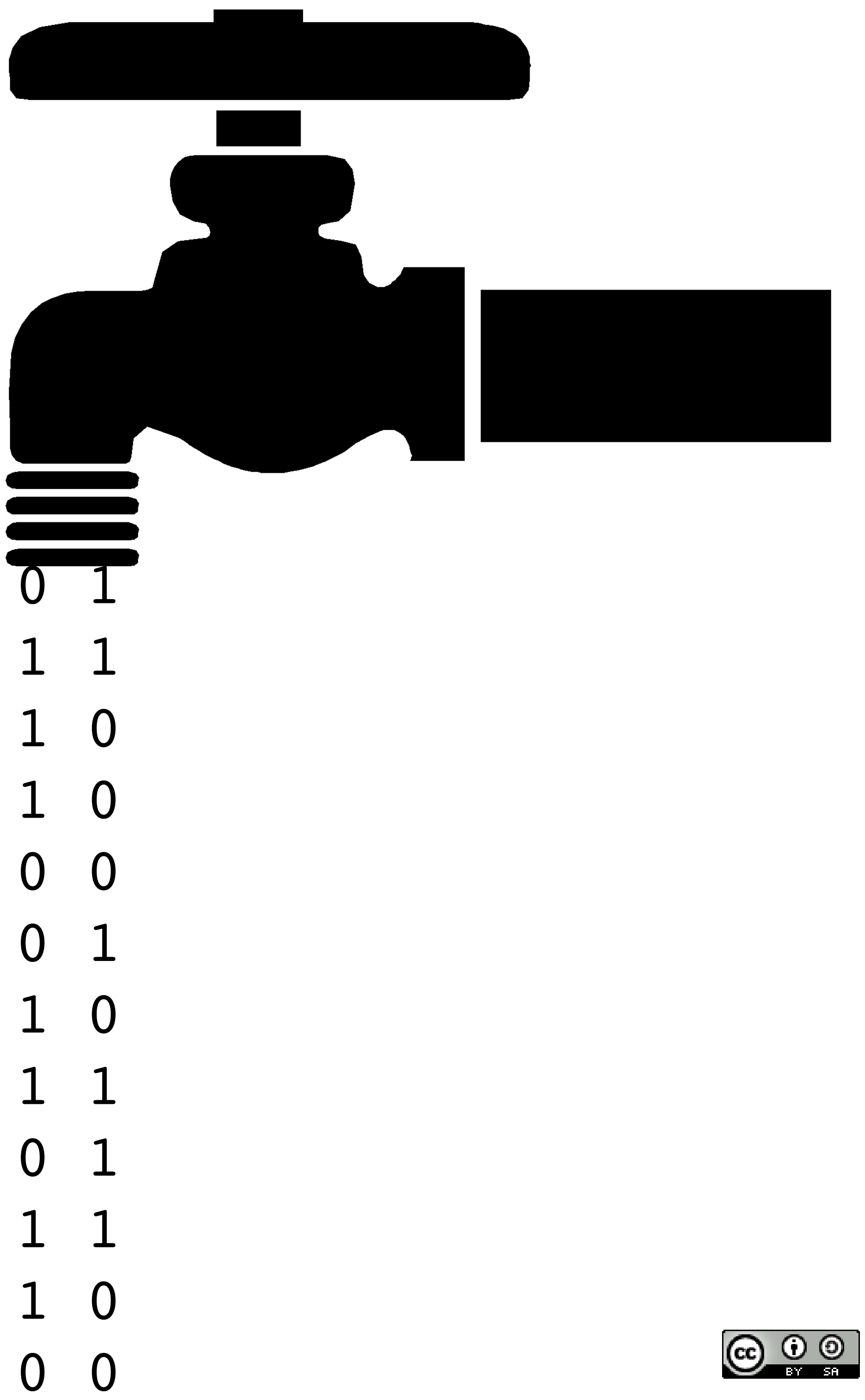
Simple

Predictable

Reliable

Extensible

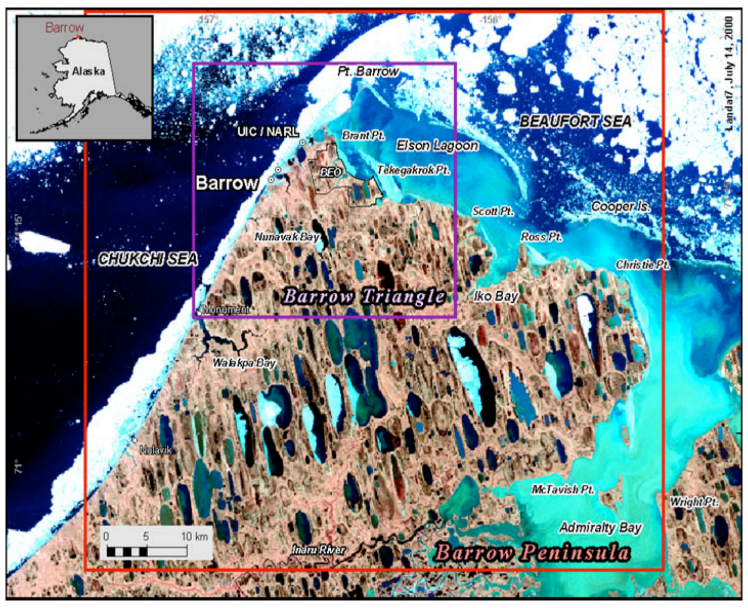
Durable



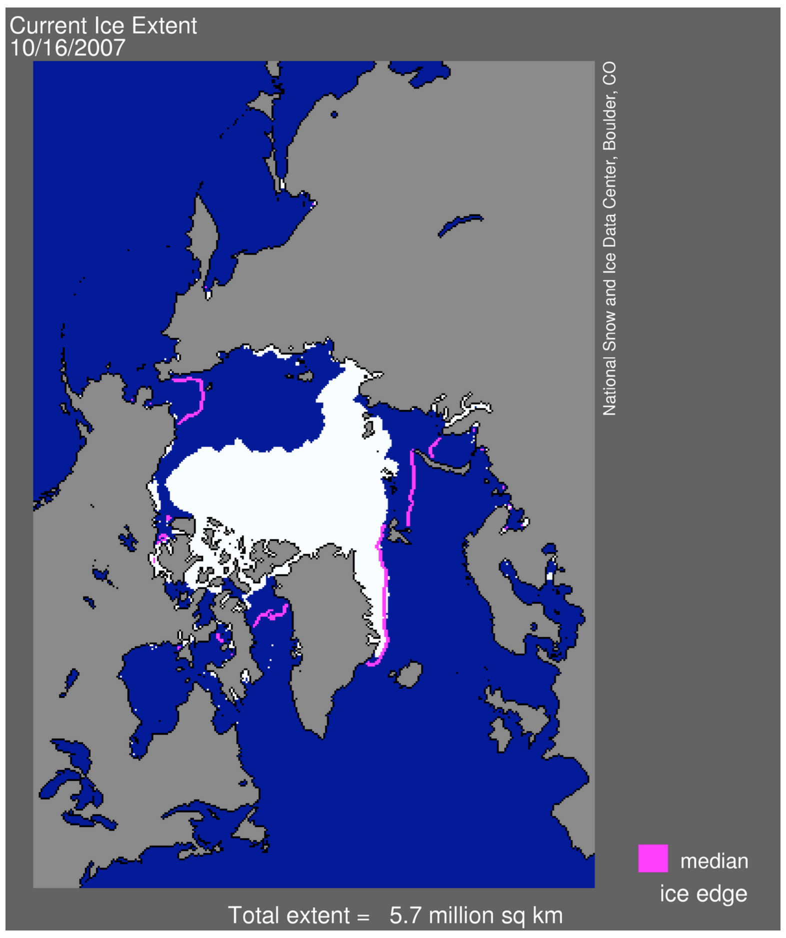
Well wouldn’t it be nice if access to scientific data was as simple as access to water? In other words data access was a basic utility

What are the Data

- National Science Board 2005:
  - Research collections
  - Community or Resource collections
  - Reference collections

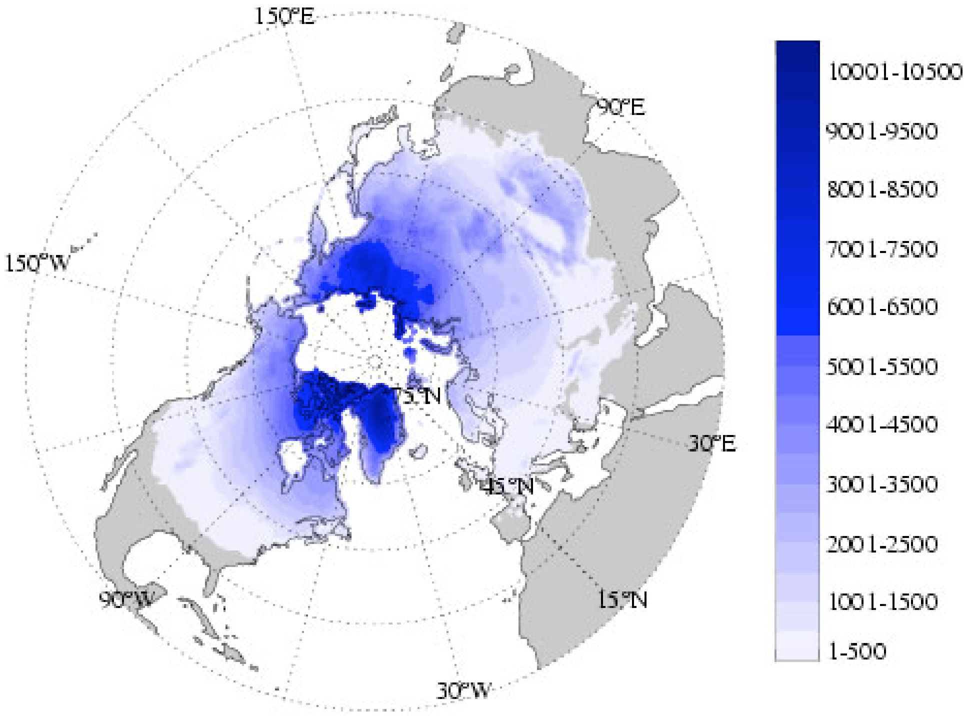


Manley, W. F. et al. 2005. Reduced-Resolution Radar Imagery, Digital Elevation Models, and Related GIS Layers for Barrow, Alaska, USA. nsidc.org/data/arcss303.html



Current Ice Extent
10/16/2007
National Snow and Ice Data Center, Boulder, CO
median ice edge
Total extent = 5.7 million sq km

Fetterer and Knowles. 2004. Sea Ice Index. nsidc.org/data/seaice_index/



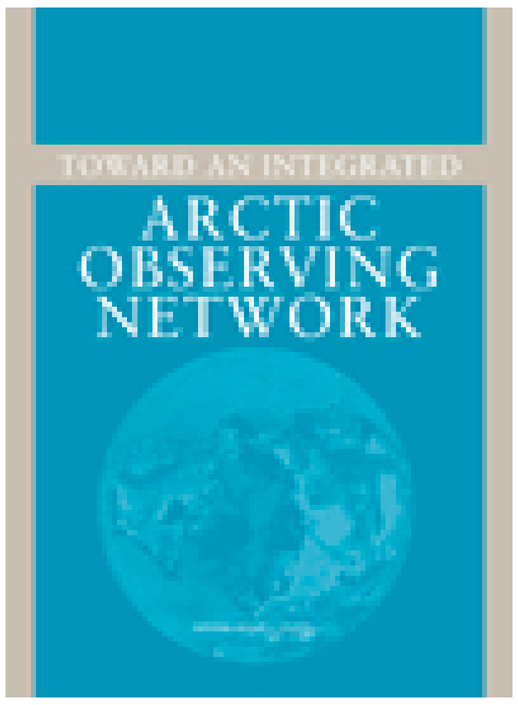
Zhang, T. et al. 2005. Northern Hemisphere EASE-Grid Annual Freezing and Thawing Indices, 1901 - 2002. nsidc.org/data/ggd649.html

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.

Sea ice show evolution in categories not because of its fame but because of the increased recognition of sea ice’s role on climate, ecology, and society. The index may not be the reference collection per se but it integrates the information for broader consumption.

The Data

- Toward an Integrated Arctic Observing Network
  Committee on Designing an Arctic Observing Network, National Research Council
  http://www.nap.edu/catalog/11607.html
  - Table 2.1 Key Variables
  - Table 3A.1 Current and Planned Networks

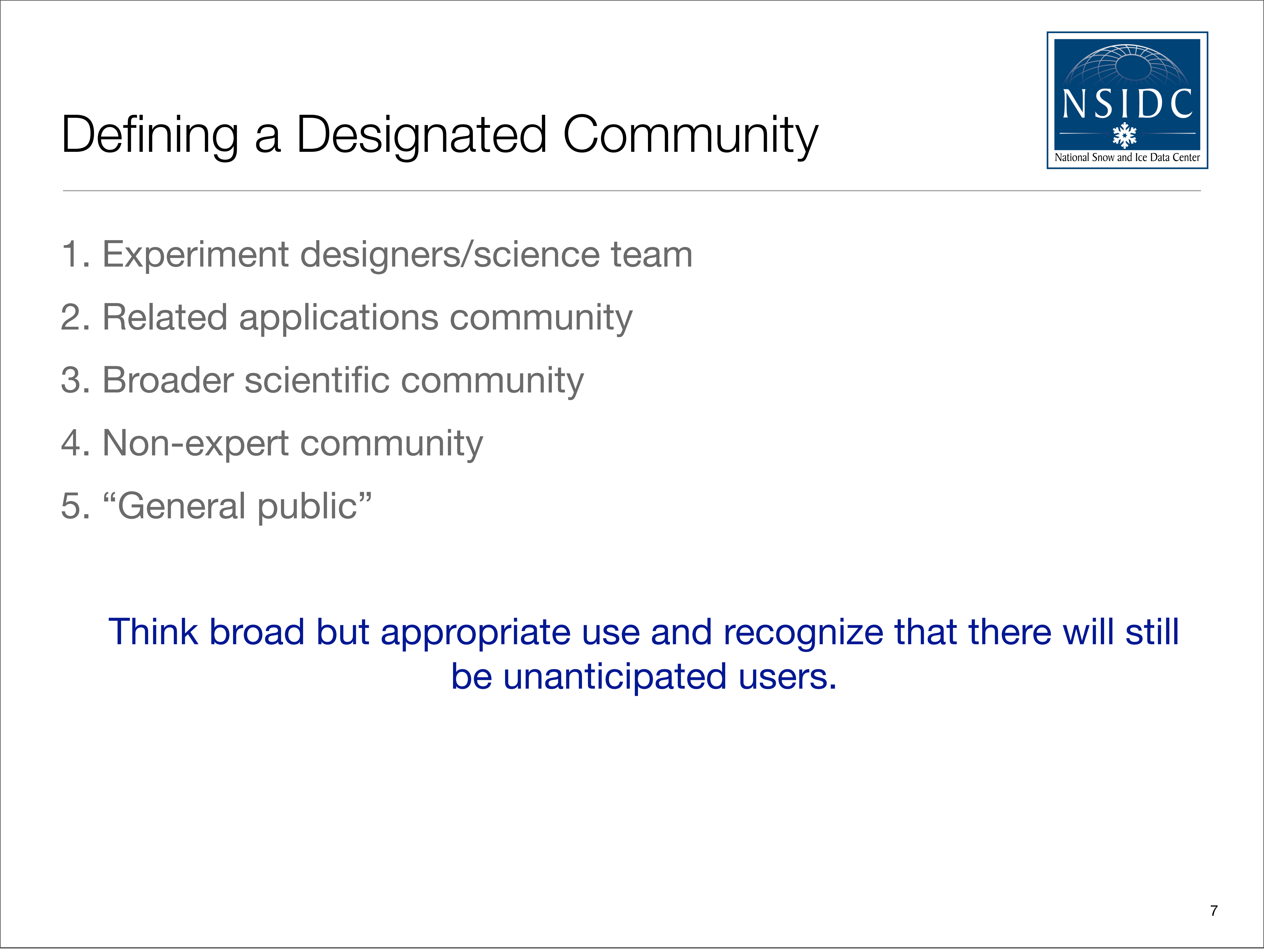


TOWARD AN INTEGRATED
ARCTIC
OBSERVING
NETWORK

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).

The Scientists

- Open Archival Information System Reference Model
  - An “organization that intends to preserve information for access and use by a Designated Community”
  - The data need to be “independently understandable” by the designated community.

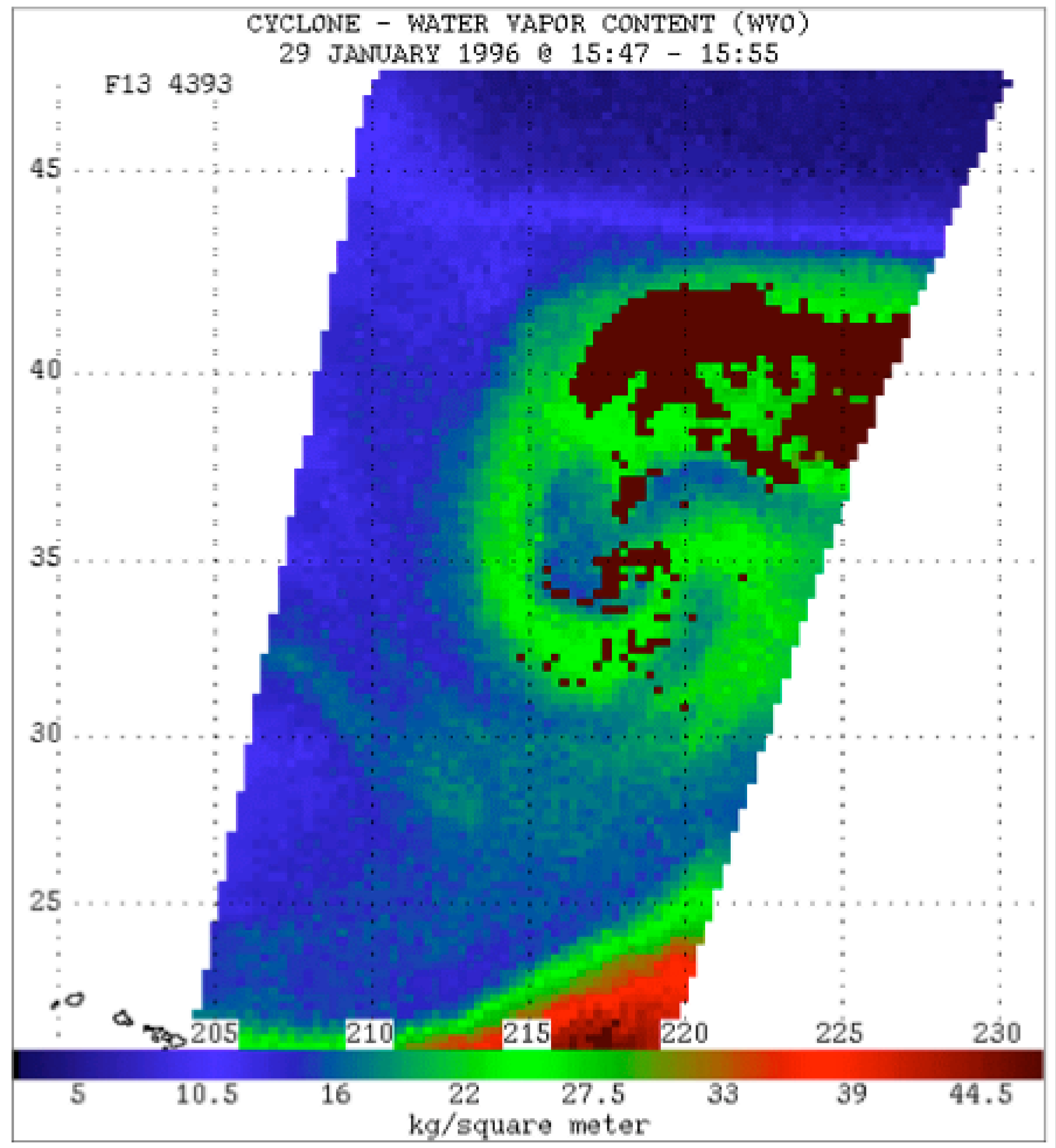


Defining a Designated Community

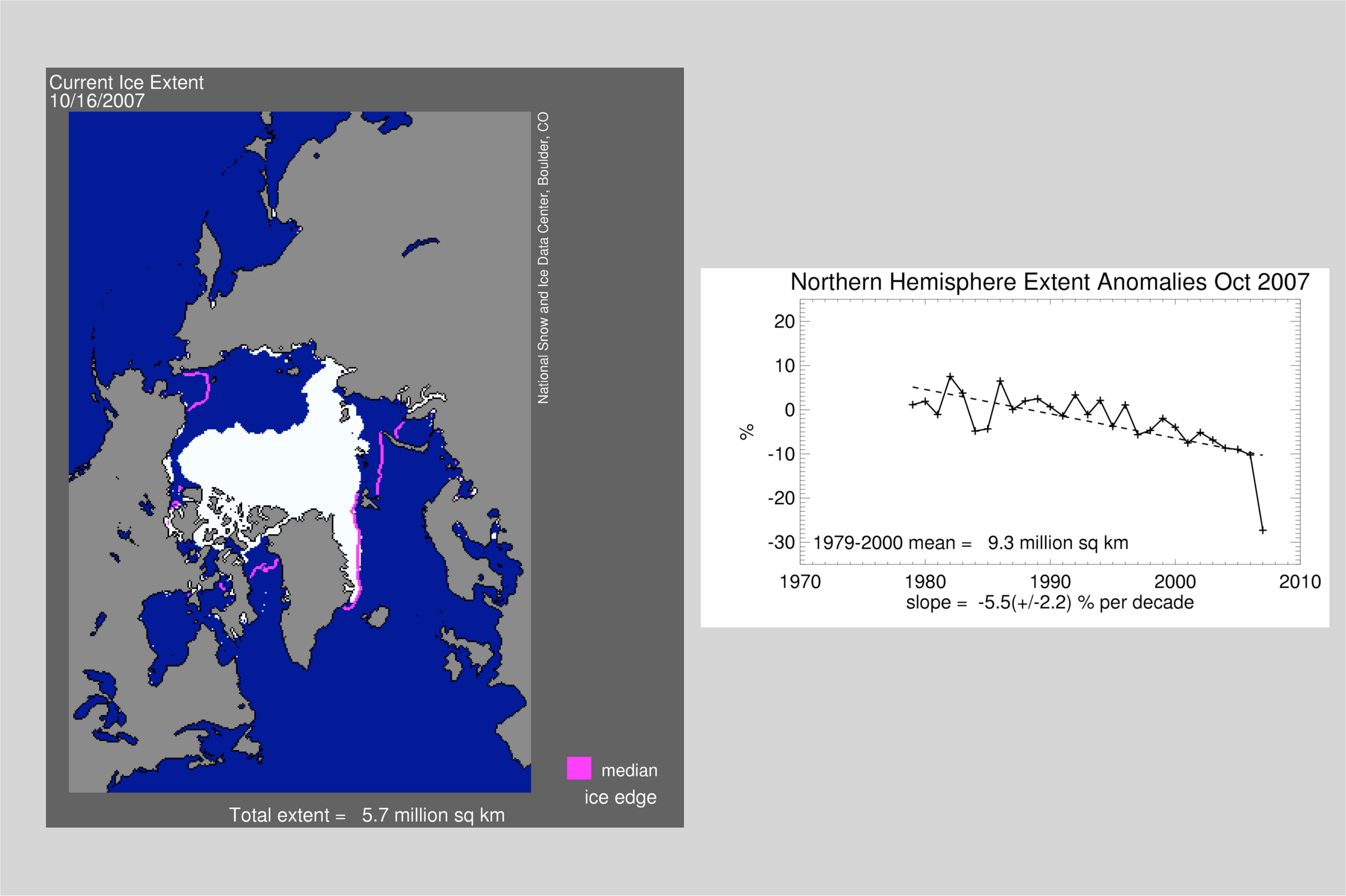
1. Experiment designers/science team
2. Related applications community
3. Broader scientific community
4. Non-expert community
5. “General public”

Think broad but appropriate use and recognize that there will still be unanticipated users.

Communities evolve. Consider SSM/I again

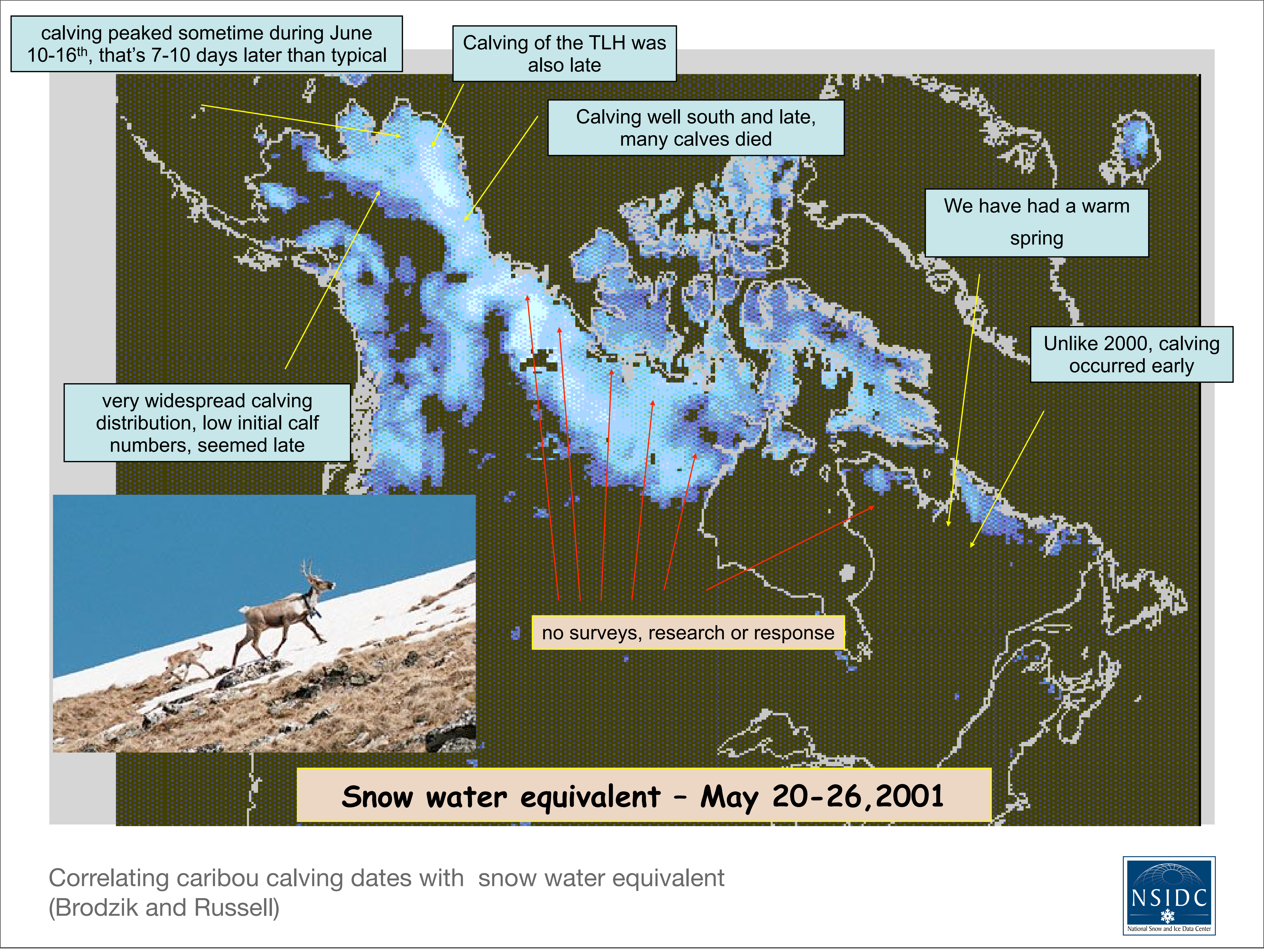


Operational Image of Water Vapor Content EDR values from the Special Sensor Microwave Imager (SSM/I). 29 January 1996 (FNMOC)



Sept. 2007 Northern Hemisphere gridded sea ice extent compared with the 1979-2000 median September extent and 28-year trend in extent anomalies

NSIDC
National Snow and Ice Data Center



calving peaked sometime during June 10-16th, that's 7-10 days later than typical

Calving of the TLH was also late

Calving well south and late, many calves died

We have had a warm spring

Unlike 2000, calving occurred early

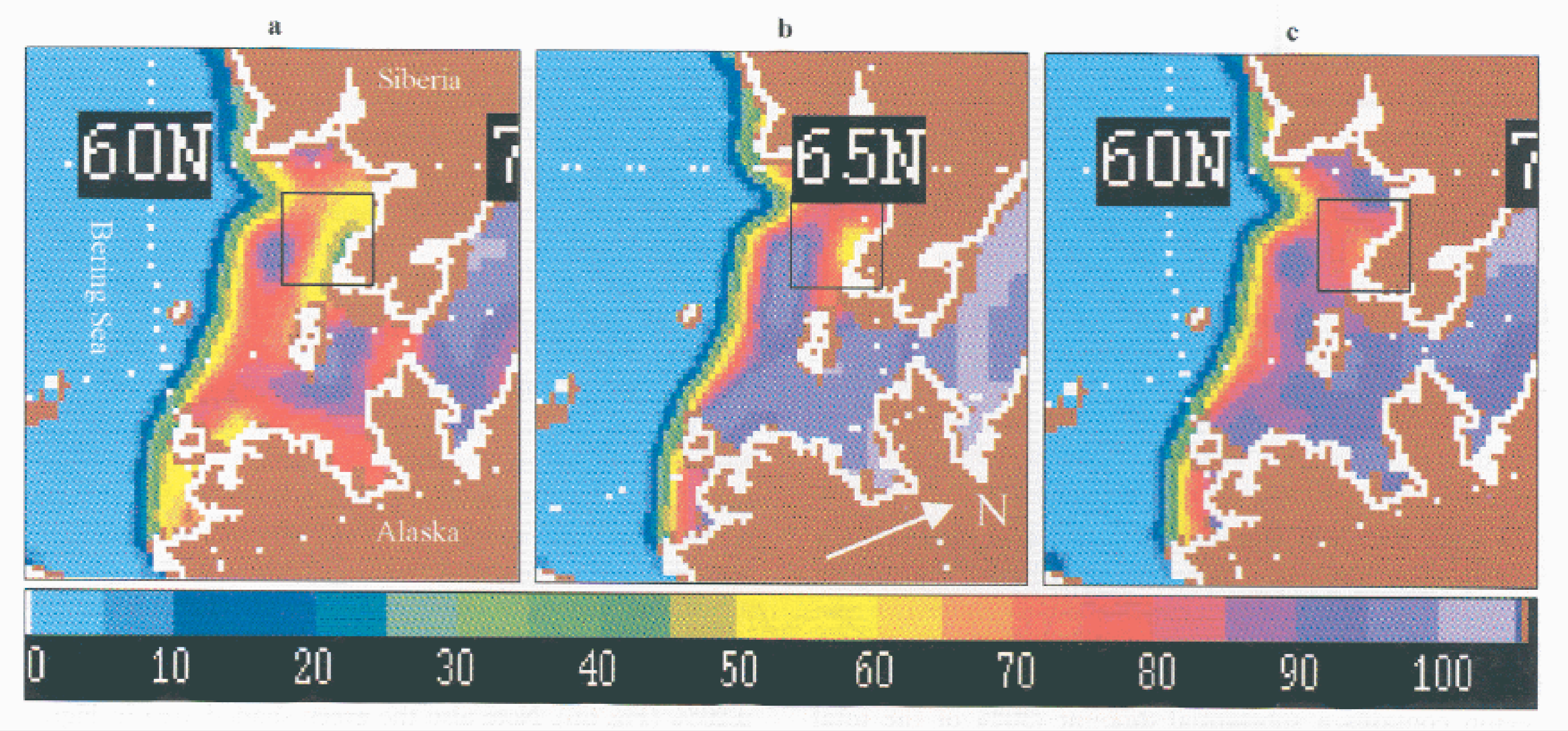
very widespread calving distribution, low initial calf numbers, seemed late

no surveys, research or response

Snow water equivalent – May 20-26,2001

Correlating caribou calving dates with snow water equivalent (Brodzik and Russell)

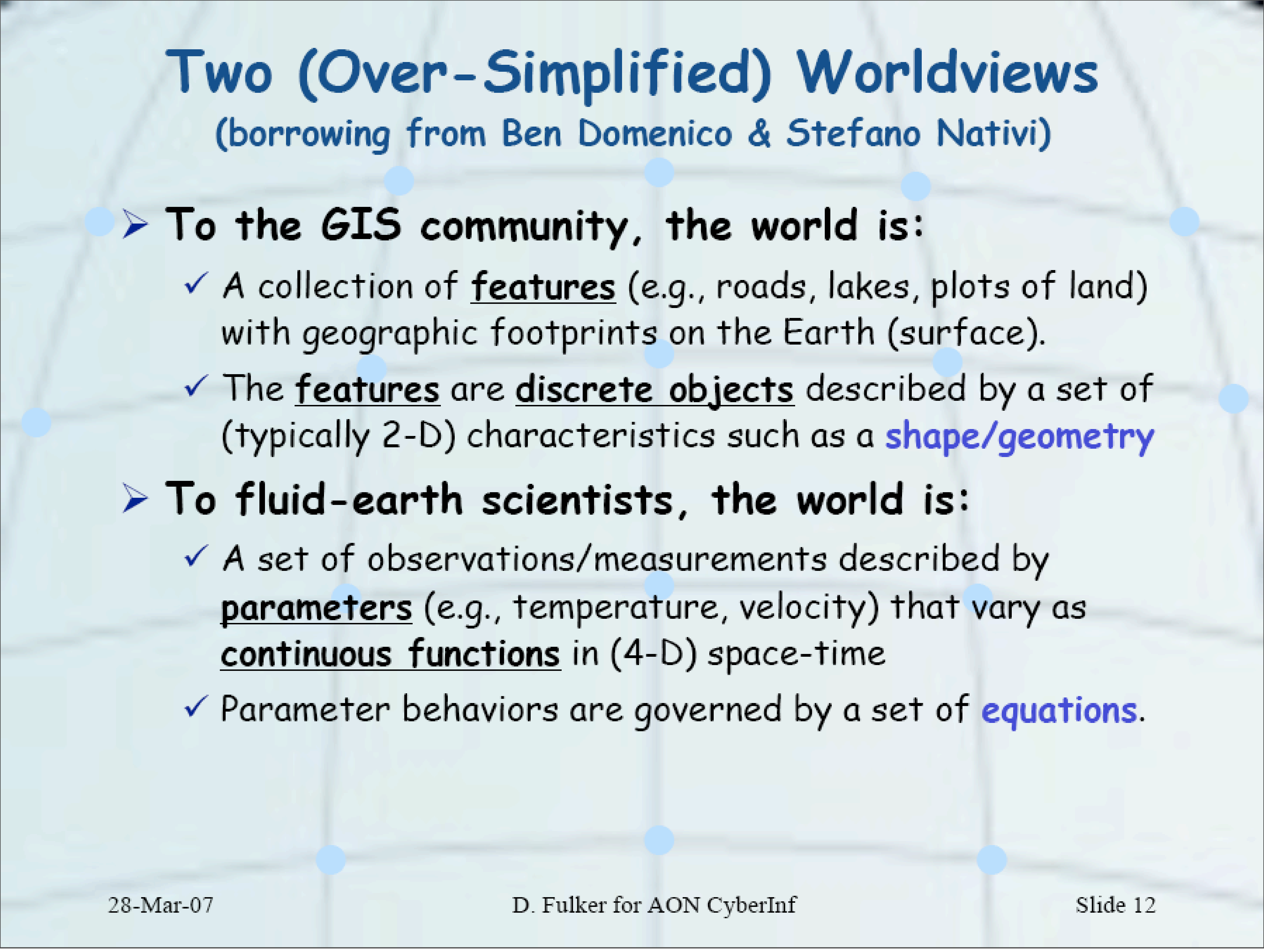
From a presentation by Don Russel of the Canadian Wildlife Service



Example of differences in SSM/I derived ice concentration values calculated with three different passive microwave algorithms (Meier et al. 2001)

Designating a user community

- Arctic residents
- Policy makers and government agencies.
- Educators
- Applied scientists and engineers
- Research scientists—different disciplines creates the core challenge



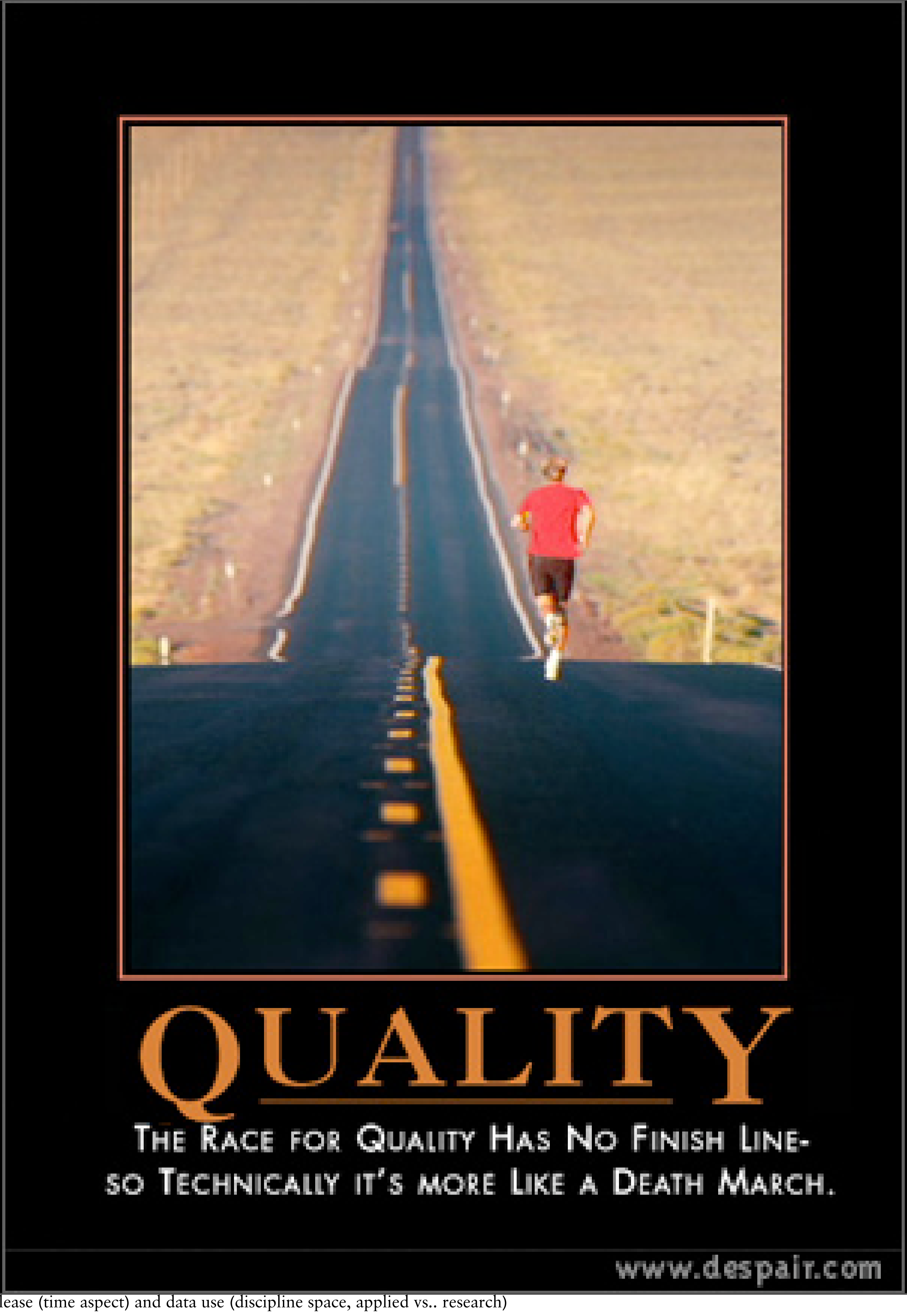
Two (Over-Simplified) Worldviews

(borrowing from Ben Domenico & Stefano Nativi)

- To the GIS community, the world is:
  - ✓ A collection of **features** (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).
  - ✓ The **features** are **discrete objects** described by a set of (typically 2-D) characteristics such as a **shape/geometry**
- To fluid-earth scientists, the world is:
  - ✓ A set of observations/measurements described by **parameters** (e.g., temperature, velocity) that vary as **continuous functions** in (4-D) space-time
  - ✓ Parameter behaviors are governed by a set of **equations**.

28-Mar-07 D. Fulker for AON CyberInf Slide 12

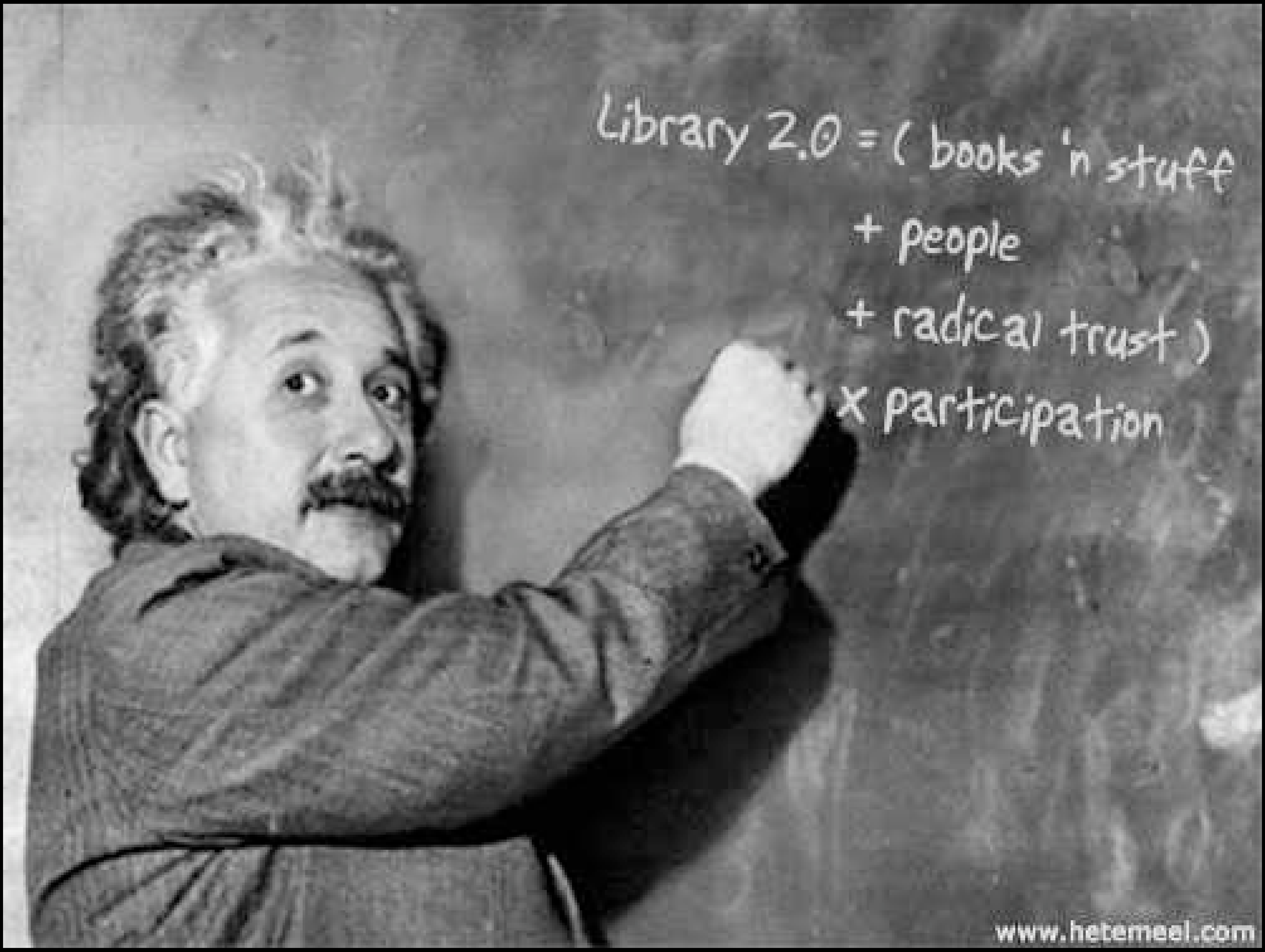
We should also consider how these user communities think. For example, David Fulkner, in a keynote presentation to the principle investigators of the U.S. National Science Foundation's (NSF) AON projects, showed how scientists have two worldviews. One view sees the world as a collection of features arranged 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.


QUALITY
The Race for Quality Has No Finish Line-
so Technically it's more Like a Death March.
www.despair.com

how do we define quality for data release (time aspect) and data use (discipline space, applied vs.. research)

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.



SOME RIGHTS RESERVED Darlene Fichter

Sociotechnical systems--“network of technology, information, documents, *people, and practices*” (Van House et al. 2003, p. 1 my emphasis)

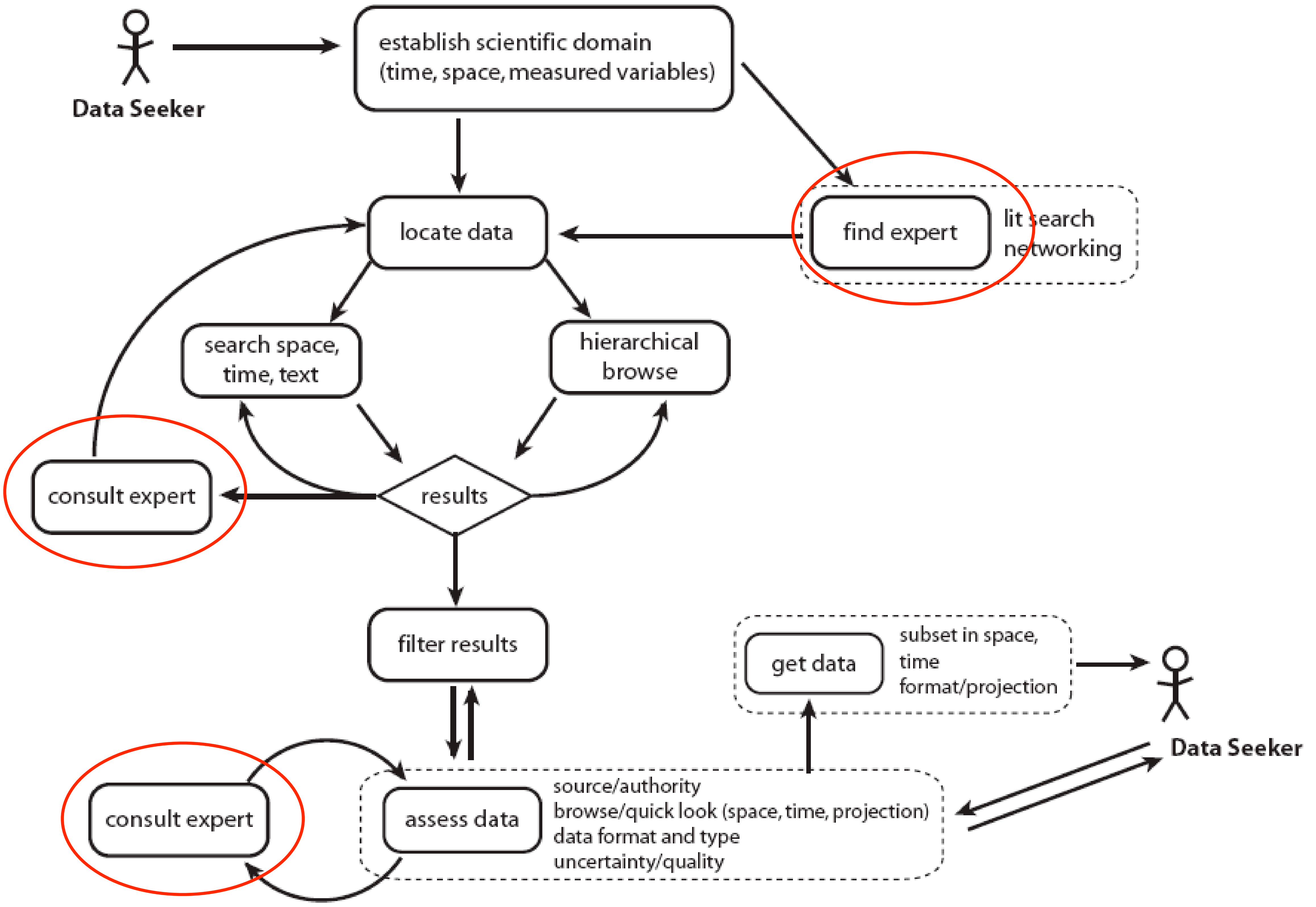
“Radical trust”
Hobbes says trust is key to social transactions and that this is maintained through autocratic authority. Liberal market theory says competition and transparency solve the problem. Neither really applies here, so we begin to explore deeper dimensions of the social--the “embededness” of human action within social context (Granovetter 1985)--the creation of radical(?) trust in an open scientific society.

Does IPY create a new social context? If it does, it will be because of our active “participation”.

—

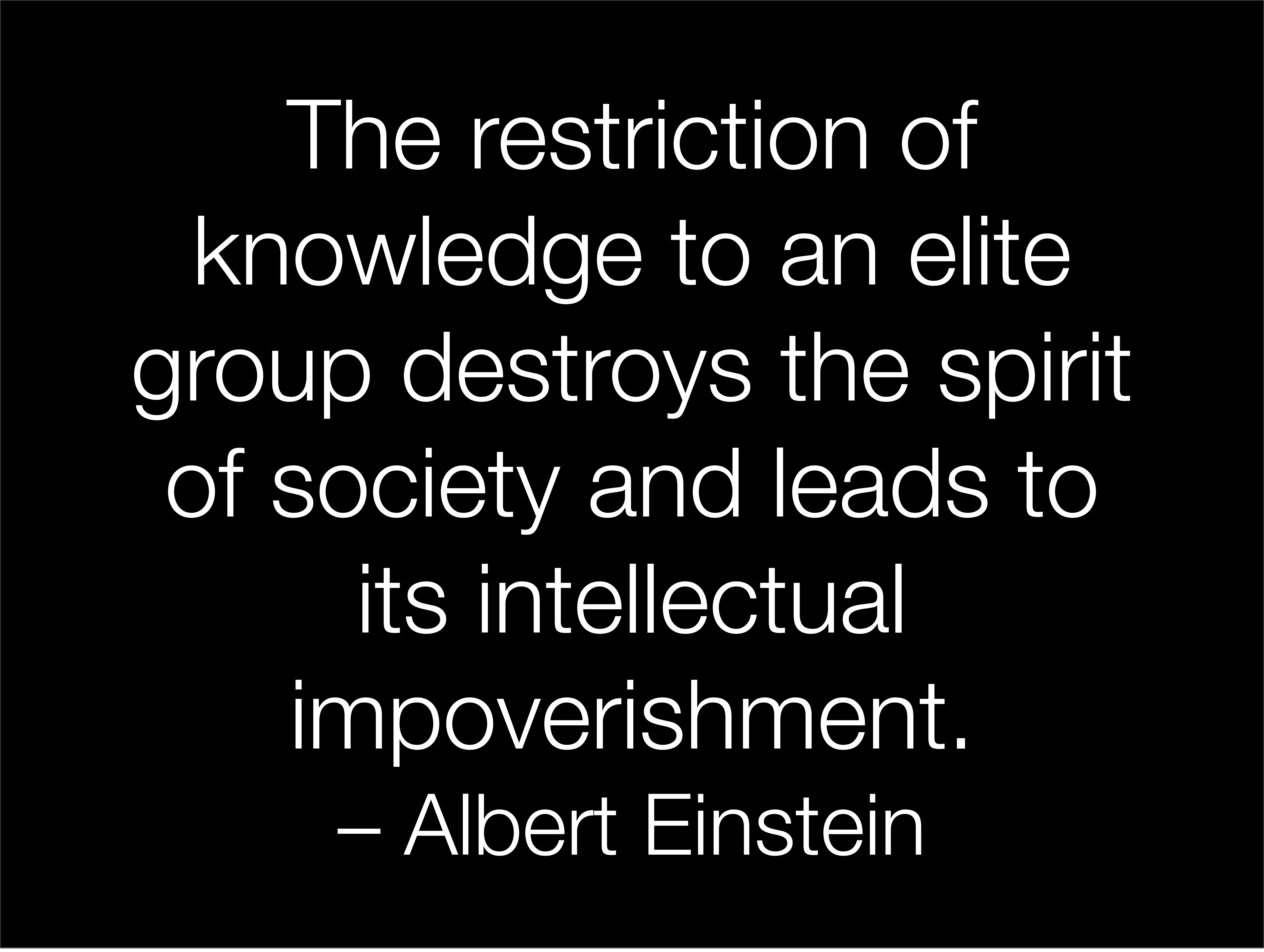
GRANOVETTER, M. 1985. ECONOMIC-ACTION AND SOCIAL-STRUCTURE - THE PROBLEM OF EMBEDDEDNESS. *AM J SOCIOL. 91*:481-510.

Search use-case diagram



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).

This is not the model
highlight
    search vs. explore
    role of expert consultation
    filter



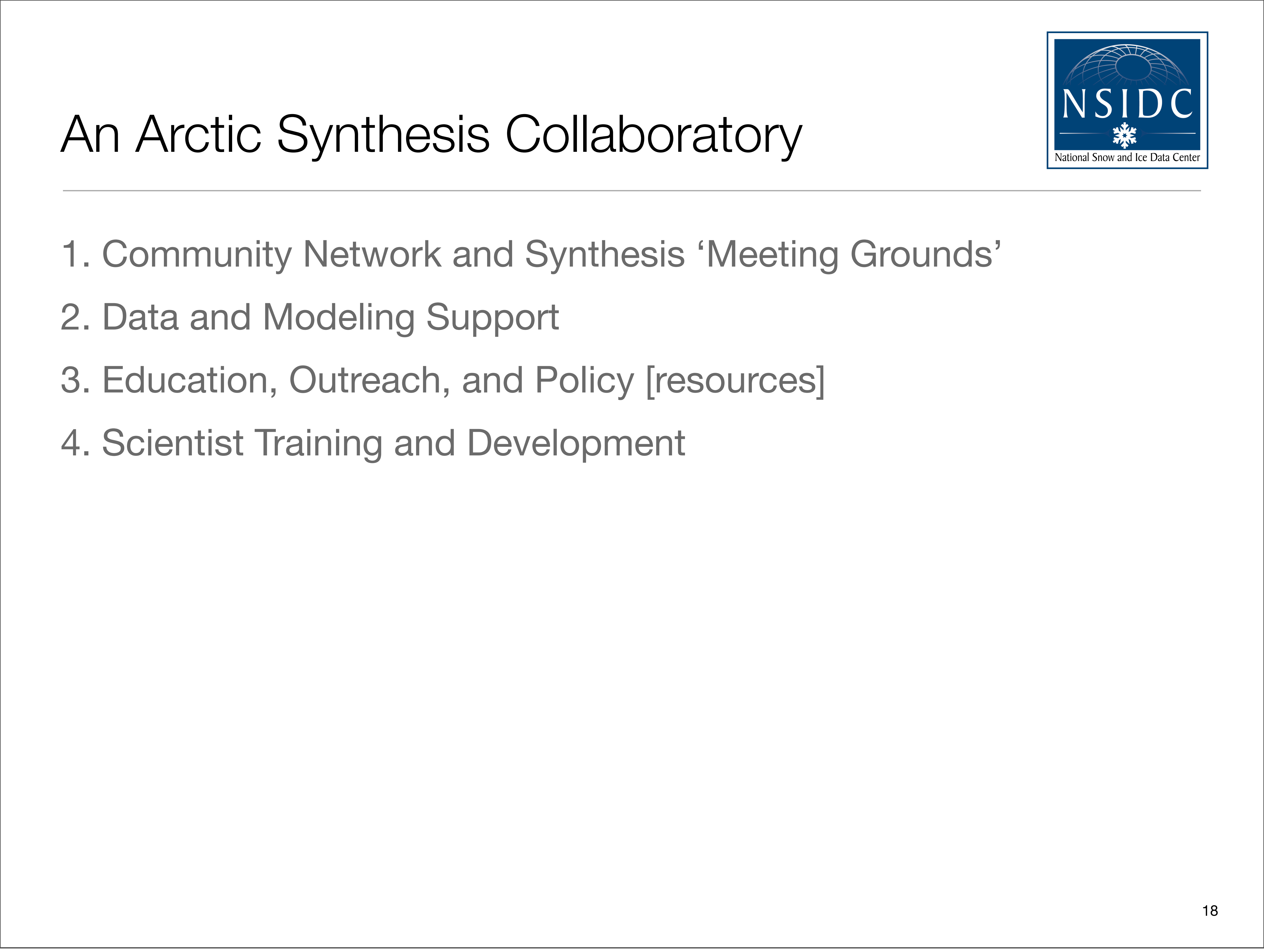
The restriction of knowledge to an elite group destroys the spirit of society and leads to its intellectual impoverishment.
– Albert Einstein

Another workshop of Canadian investigators working on International Polar Year projects revealed the tensions created by the IPY Data Policy's requirement for timely data release in that some investigators do not trust "outsiders" to use their data fairly or appropriately.

"Thieves and outsiders" Need incentive mechanisms including formal citation

Vannevar Bush, Director of the US Office of Scientific Research and Development, "As we may think"
Then IGY, Antarctic treaty and the WDCs fighting the secrecy of the cold war.
In the 90's things took off--FGDC and plans for specific infrastructure

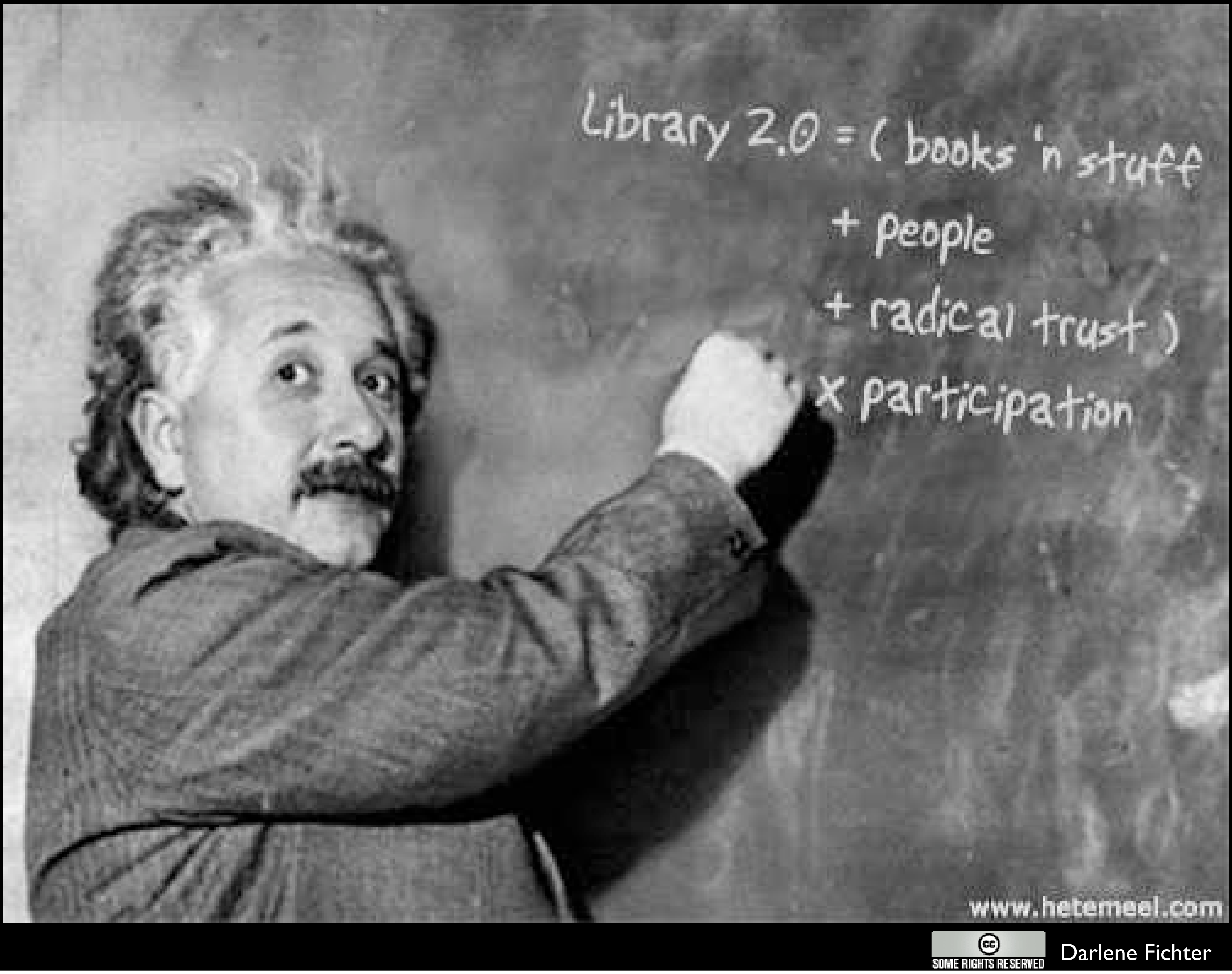
now the power is more diffuse and there is a need to empower scientists, developing countries, public hence more sharing



An Arctic Synthesis Collaboratory

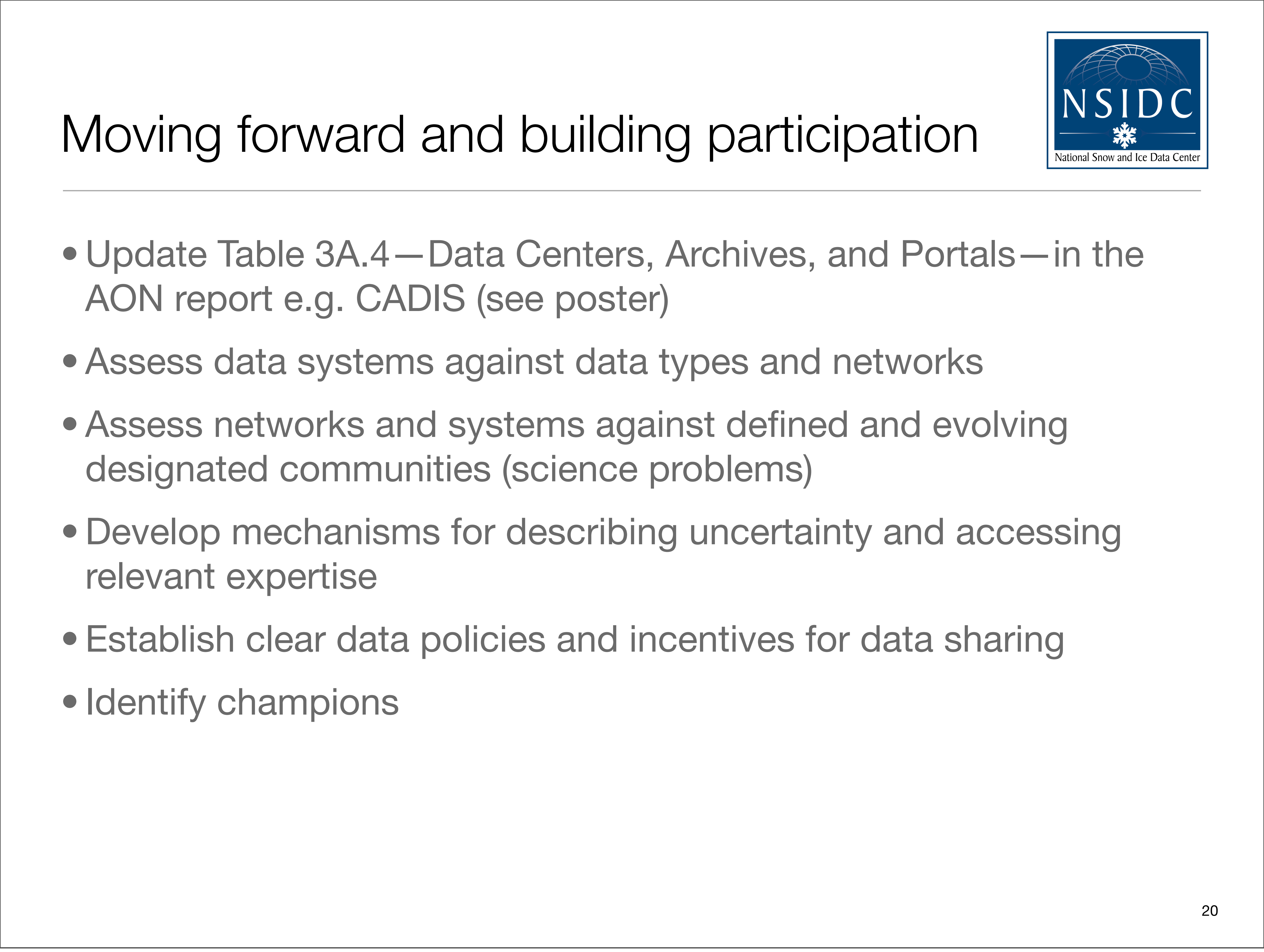
1. Community Network and Synthesis ‘Meeting Grounds’
2. Data and Modeling Support
3. Education, Outreach, and Policy [resources]
4. Scientist Training and Development

(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).



Darlene Fichter

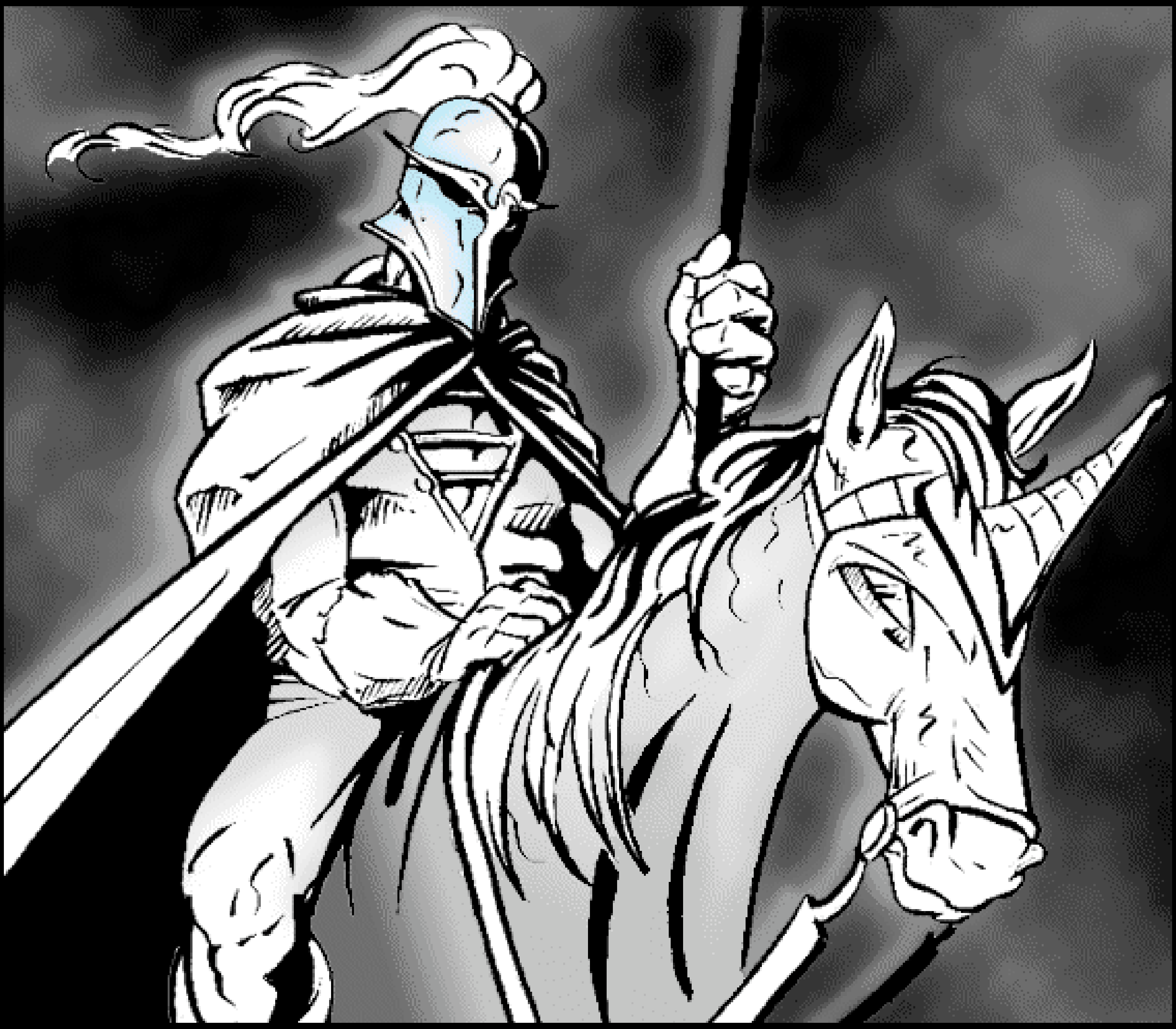
I return to this picture to emphasize the participation part of the equation. Participation is essential for the entire network--providers, users, *and data systems.* There needs to be an incentive and motivation factor for all to participate.



Moving forward and building participation

- Update Table 3A.4—Data Centers, Archives, and Portals—in the AON report e.g. CADIS (see poster)
- Assess data systems against data types and networks
- Assess networks and systems against defined and evolving designated communities (science problems)
- Develop mechanisms for describing uncertainty and accessing relevant expertise
- Establish clear data policies and incentives for data sharing
- Identify champions

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 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.



Willam Craig’s (2005) white knights--Altruism:

- Idealism--data sharing is good
- Enlightened self-interest--documentation is good, you gotta give to get, drive out bad data with their good data
- Involvement in a professional culture--members of larger committees, consortia, etc. Participate in annual conferences, etc.

SAON should identify champions--national or thematic coordinators.

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.

Craig, W. 2005 "White knights of Spatial Data Infrastructure: The role and motivation of key individuals." URISA Journal 16(2), 5-13.

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

Fetterer, F. & Knowles, K. (2002, updated 2004) Sea Ice Index. Boulder, CO: National Snow and Ice Data Center. Digital media. Retrieved 10 Nov. 2007 http://nsidc.org/data/seaice_index/.

GRANOVETTER, M. 1985. ECONOMIC-ACTION AND SOCIAL-STRUCTURE - THE PROBLEM OF EMBEDDEDNESS. AM J SOCIOL. 91:481-510.

Manley, W. F., L. R. Lestak, C. E. Tweedie, and J. A. Maslanik. 2005. Reduced-resolution radar imagery, digital elevation models, and related GIS layers for Barrow, Alaska, USA. Boulder, CO: National Snow and Ice Data Center. Digital media and CD-ROM.

Meier, W.N., VanWoert, M.L. & Bertoia, C. (2001) Evaluation of operational SSM/I algorithms. Annals of Glaciology 33, 102-108.

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.*

Zhang, T., O.W. Frauenfeld, J. McCreight, R.G. Barry. 2005. Northern Hemisphere EASE-Grid annual freezing and thawing indices, 1901 - 2002. Boulder, CO: National Snow and Ice Data Center/World Data Center for Glaciology. Digital media