The State of Arctic Data—the IPY experience

Mark A. Parsons, Taco de Bruin, Scott Tomlinson, Øystein Godøy, Helen Campbell, Julie Leclert, Ellsworth LeDrew, David Carlson,
and the IPY data community.

22nd International CODATA Conference
Stellenbosch, Cape Town, South Africa
25 October 2010

In fifty years time the data resulting from IPY2007-2008 may be seen as the most important single outcome of the programme.
—A Framework for the IPY (ICSU 2004)

Our vision:
Data are open, linked, useful, and safe.

A pragmatic assessment

- Data sharing and publication—open and linked

- Interoperability across systems, data, and standards—linked and useful

- Sustainable preservation and stewardship of diverse data—safe

- Governance and conduct of the virtual organization that coordinates data access and stewardship around the globe.—practicality
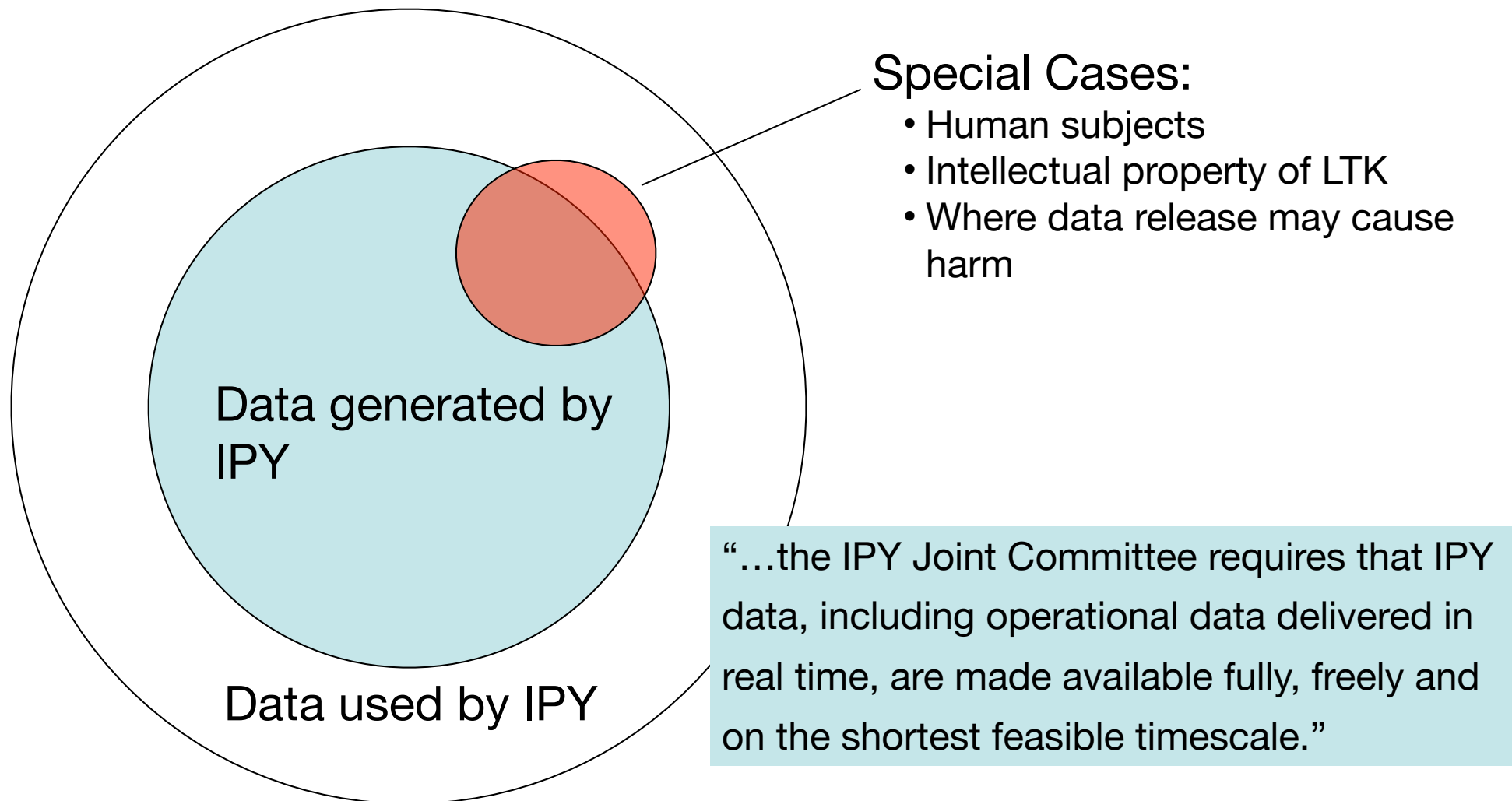
Open

Open Data is a philosophy and practice requiring that certain data are freely available to everyone, without restrictions from copyright, patents or other mechanisms of control.—Wikipedia
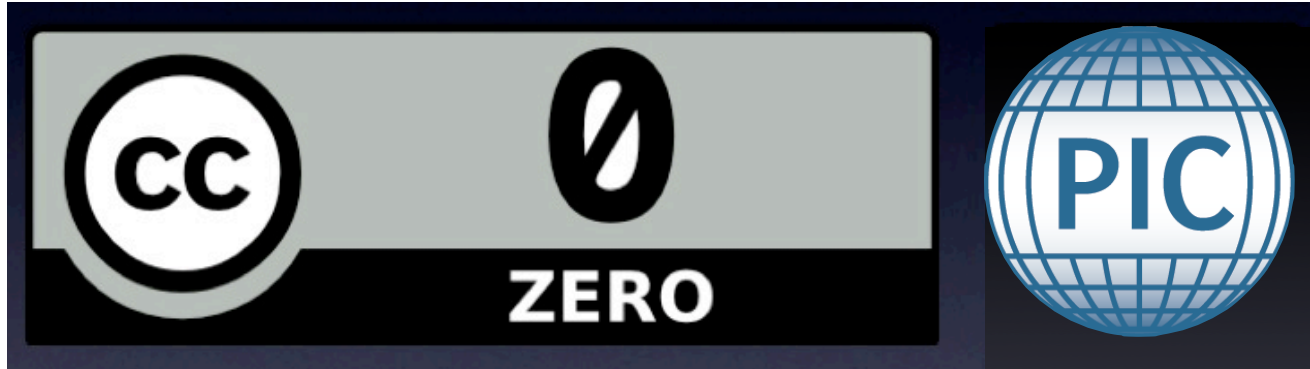
IPY Data Policy

http://www.ipy.org/Subcommittees/final_ipy_data_policy.pdf

Special Cases:
- Human subjects
- Intellectual property of LTK
- Where data release may cause harm

Data generated by IPY

Data used by IPY

"…the IPY Joint Committee requires that IPY data, including operational data delivered in real time, are made available fully, freely and on the shortest feasible timescale."

CC Zero waiver + norms

waive rights → public domain
+
attribution /citation through community
norms, not a contract
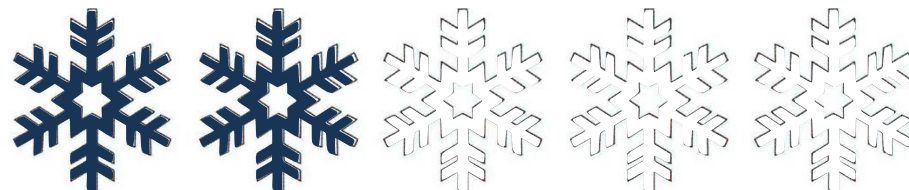
http://polarcommons.org/ethics-and-norms-of-data-sharing.php

Data sharing and publication

Data should be accessible soon after collection (online wherever possible) in a discovery portal such as the GCMD.

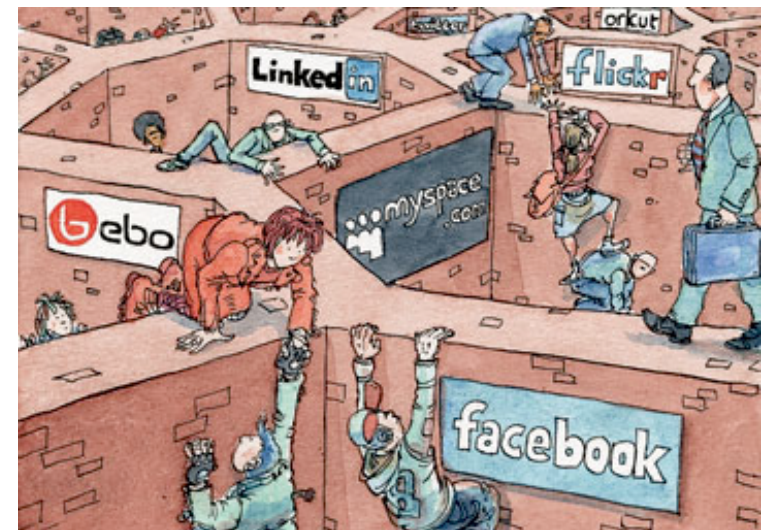Data users should provide fair and formal credit to providers.

Linked

The term Linked Data is used to describe a method of exposing, sharing, and connecting data via dereferenceable URIs on the Web.—Wikipedia

Linked Data

- Tim Berners-Lee's 4 steps emphasize that it's about relationships.

  1. Use URIs as names for things

  2. Use HTTP URIs so that people can look up those names.

  3. When someone looks up a URI, provide useful information, using standards (RDF, SPARQL).

  4. Include links to other URIs. so that they can discover more things.

- Current practice—registries and catalogs (e.g. GCMD)

© The Economist

Swedish IPY Data Centre

Australian Antarctic Data Centre

POLAR DATA CATALOGUE

NSIDC
National Snow and Ice Data Center
University of Colorado at Boulder

NIPR 国立極地研究所
National Institute of Polar Research

Regional OBIS Node for the Antarctic
SCAR–MarBIN
OBIS
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

Cooperative Arctic Data and Information Service
of the ARCTIC OBSERVING NETWORK
CADIS
AON
SEARCH

中国南北极数据中心
Chinese National Arctic and Antarctic Data Center
地球系统科学数据共享平台-极地区域数据共享运行服务中心
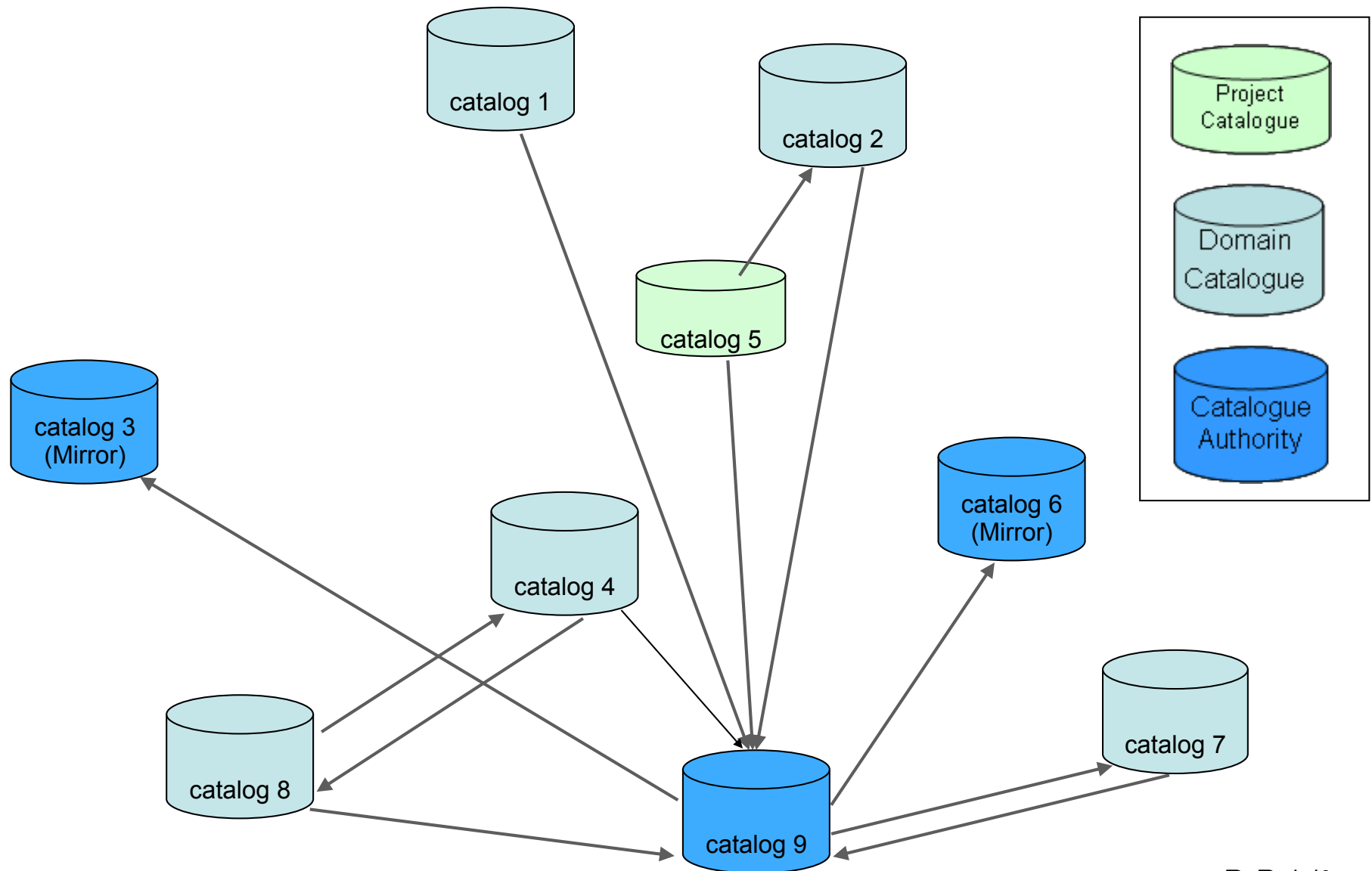Chinese National Committee for Antarctic Research

ELOKA Exchange for Local Observations and Knowledge of the Arctic

Международный полярный год
2007-2008

Вы вошли как Гость.
ВОЙТИ
Регистрация ?
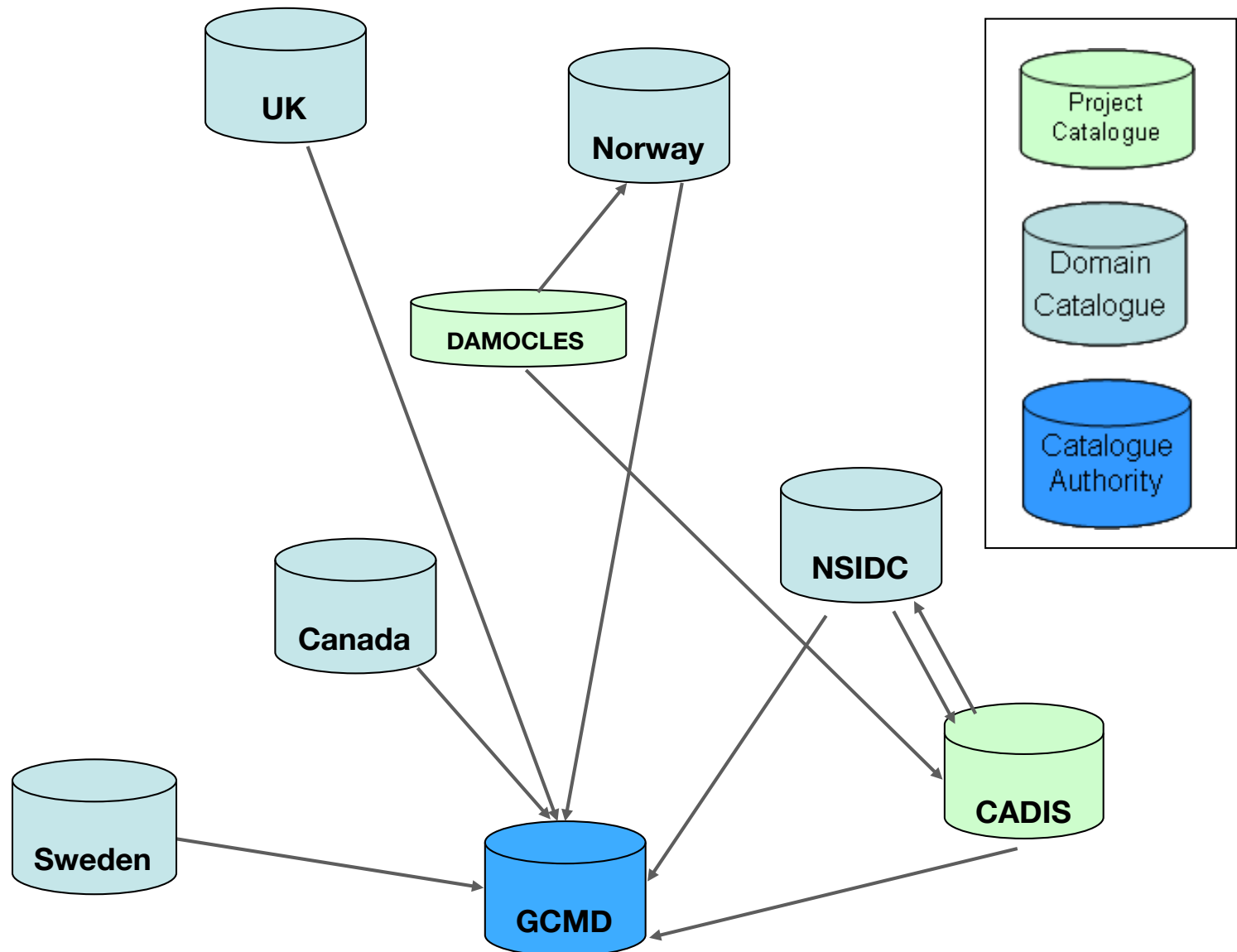
Global Change Master Directory
Discover Earth science data and services

A "Union" Catalog

An initial IPY Union Catalog (using OAI-PMH)

Interoperability—discovery

Metadata should be readily interchangeable between different polar data systems to enable data discovery across multiple portals.

Useful

Two (Over-Simplified) Worldviews
(borrowing from Ben Domenico & Stefano Nativi)

➤ **To the GIS community, the world is:**

  ✓ A collection of <u>**features**</u> (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).

  ✓ The <u>**features**</u> are <u>**discrete objects**</u> 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 <u>**parameters**</u> (e.g., temperature, velocity) that vary as <u>**continuous functions**</u> in (4-D) space-time

  ✓ Parameter behaviors are governed by a set of equations.

▸ **To the social scientist, the world is:**

  ✓ A complex, involved <u>**narrative**</u> with many players

  ✓ The narrative describes a **network of interactions** between human and non-human elements (including data)

Interoperable (Geospatial and Semantic)

An Arctic Spatial Data Infrastructure

- The result of two IPY "GeoNorth" conferences.

- Approved by all Senior Arctic Officials of the Arctic Council last fall

- Planning a task force, consisting of representatives from national mapping agencies and the Working Groups of the Arctic Council

- GeoNorth3 this summer in Tromsø

- Similar work already underway in the Antarctic under the auspices of the Standing Committee on Antarctic Geospatial Information

Semantic Sea Ice Interoperability Initiative

- A new collaborative project funded by the US National Science Foundation working to improve semantic interoperability of Arctic data.

- Builds from the International Polar Year and the IPY Data and Information Service

- Creating a sea ice ontology incorporating scientific knowledge and some elements of traditional knowledge and linked

- Working to create a community of semantic practitioners in the Arctic.

- Seeking to establish Arctic interoperability in regional and global data systems (e.g. GEOSS, WIS, SAON)

- PIs: Mark Parsons, Ruth Duerr, Siri Jodha Singh Khalsa—NSIDC
          Peter Fox, Deborah McGuinness—RPI

- Please talk to me to learn more

Fresh fallen snow accumulates in bands on the open water, while the blue and grey melt ponds and puddles cover parts of an ice floe in the background. Image courtesy of The Hidden Ocean, Arctic 2005 Exploration.
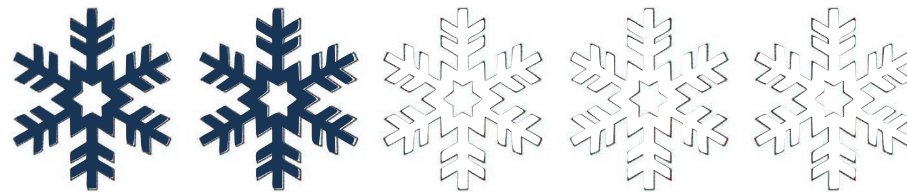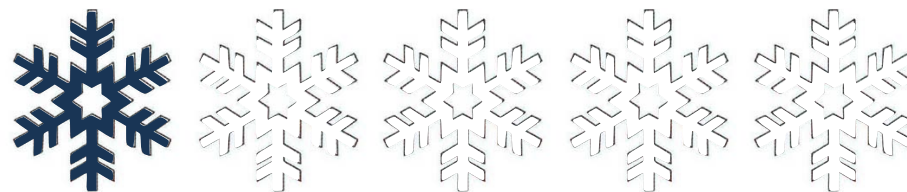
Interoperability—usability

Data from different projects, disciplines, and data centers should be easily understood and used in conjunction with each other in standard tools and analysis frameworks

Data should be well described so to be useful for a broad audience.

Safe

Safe from hackers, from obsolescence, from undocumented change, from loss, and from the ravages of time.

Preservation

All IPY data must be archived in their simplest, useful form and be accompanied by a complete metadata description. An IPY Data and Information Service (IPYDIS) should help projects identify appropriate long-term archives and data centers, but it is the responsibility of individual IPY projects to make arrangements with long-term archives to ensure the preservation of their data. It must be recognized that data preservation and access should not be afterthoughts and need to be considered while data collection plans are developed.
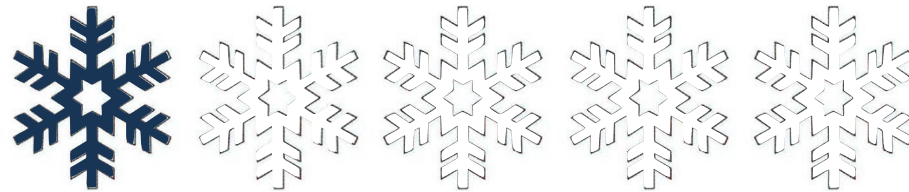
—IPY Data Policy
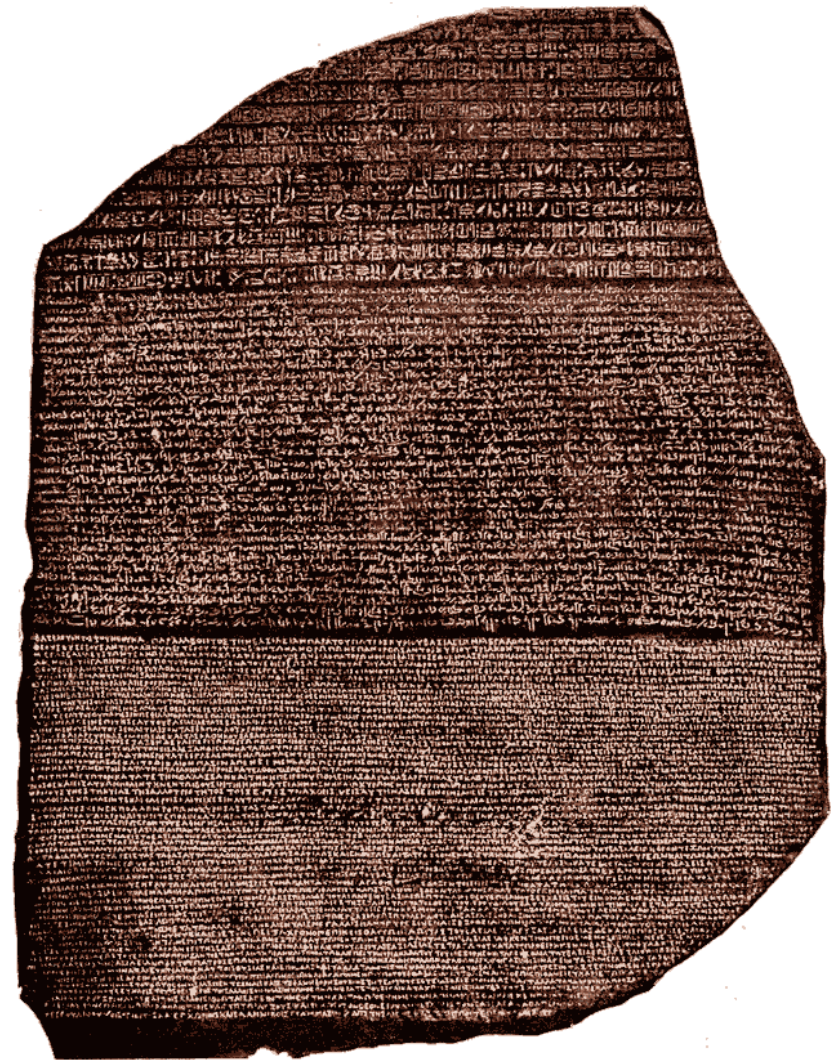
Preservation

All raw IPY data should be preserved and well stewarded in long-term archives following the OAIS Reference Model.

Data should be accompanied by complete documentation to enable preservation and stewardship.
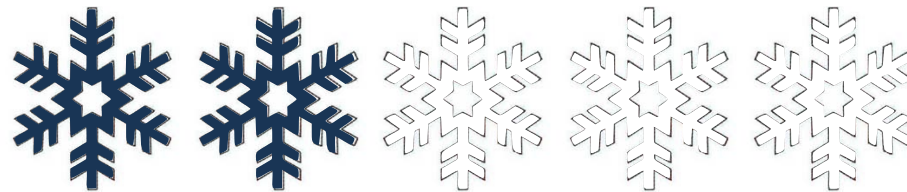
Coordination and Governance

Identify, evolve, or develop a sustained virtual organization to enable effective international collaboration on data sharing, interoperability, and preservation.

Some initial recommendations for scientists and data centers

- Investigators must publish their IPY data immediately.

- The scientific community needs to recognize the value of good data through citation, consideration of data publication in promotion and tenure review, and by training young scientists in data management.

- Data centers must develop partnerships with other data centers in other countries and other disciplines to enhance data accessibility and interoperability.

- Data centers should partner with their scientific community to explicitly meet their needs, provide easy submission tools, and make the data more useful and integrated with other data.

Some initial summary recommendations for national and international sponsors

- International sponsors must lead an aggressive initiative to ensure all IPY data are in secure archives by June 2012.

- IASC must develop an effective and pragmatic data strategy to ensure active pan-Arctic data sharing and collaboration.

- ICSU and WMO must continue to lead the global discussion to harmonize data policies around as much rapid openness as possible, while recognizing legitimate, moral restrictions.

- Funding agencies also need consistent and enforced policy.

- Funding agencies must support data archiving and insist that data they fund be archived and accessible. Agencies must also create new archives where appropriate ones do not exist.

- Agencies should take advantage of the interdisciplinary use cases generated by IPY science questions to support basic and applied research on improving interdisciplinary data management and interoperability.

Find the report at http://ipydis.org/documents/ or
Google "state of polar data"

Thank you
parsonsm@nsidc.org