



Data Sharing, Access, and Archiving in IPY

Mark A. Parsons

Co-Chair, IPY Data Policy and Management Subcommittee
Manager, IPY Data and Information Service

Electronic Geophysical Year Meeting, 5 March 2008

I would like to talk to you about data sharing--a core tenet of IPY.
Title serves as an outline for the talk and also the (very) basic strategy of IPY



Simple strategy

1. *Identify* the data
2. *Serve* the data
3. *Preserve* the data

Knowledge is power.

– Francis Bacon

Some pretty smart people make arguments that might suggest we should hoard data. But now, esp. with IPY, we are seeking societal knowledge to address global problems (diff. power structure)

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

– Albert Einstein

But some other, perhaps smarter, people show how that approach damages society.

...

When we consider what IPY is trying to accomplish, it is just the opposite. It is the enrichment of society we seek.

With this I would like to review the IPY data policy. Many of you may already be familiar with the policy, and a policy review may be a bit dry, but the policy drives our whole approach so it's worth a few minutes to understanding the policy is

but I think a quick review can guide us (or something)

IPY objectives challenge us to do things differently



- IPY has an interdisciplinary emphasis, with *active inclusion of the social sciences*.
- IPY will *link researchers across different fields* to address questions and issues lying beyond the scope of individual disciplines.
- IPY will strengthen *international coordination* of research and enhance international collaboration and cooperation
- IPY will *leave a legacy* of observing sites, facilities and networks, as well as individual data and data systems to support ongoing polar research and monitoring.

There are dozens of international data policies, so why does IPY need it's own?

The purpose of the IPY policy is to support the objectives of IPY, especially these objectives from the IPY Framework document (p10) which guided the discussion of the IPY Data Policy.

interdisciplinary and international mean broad data discovery and sharing and documentation for non expert. Plus timeframe creates need for promptness.

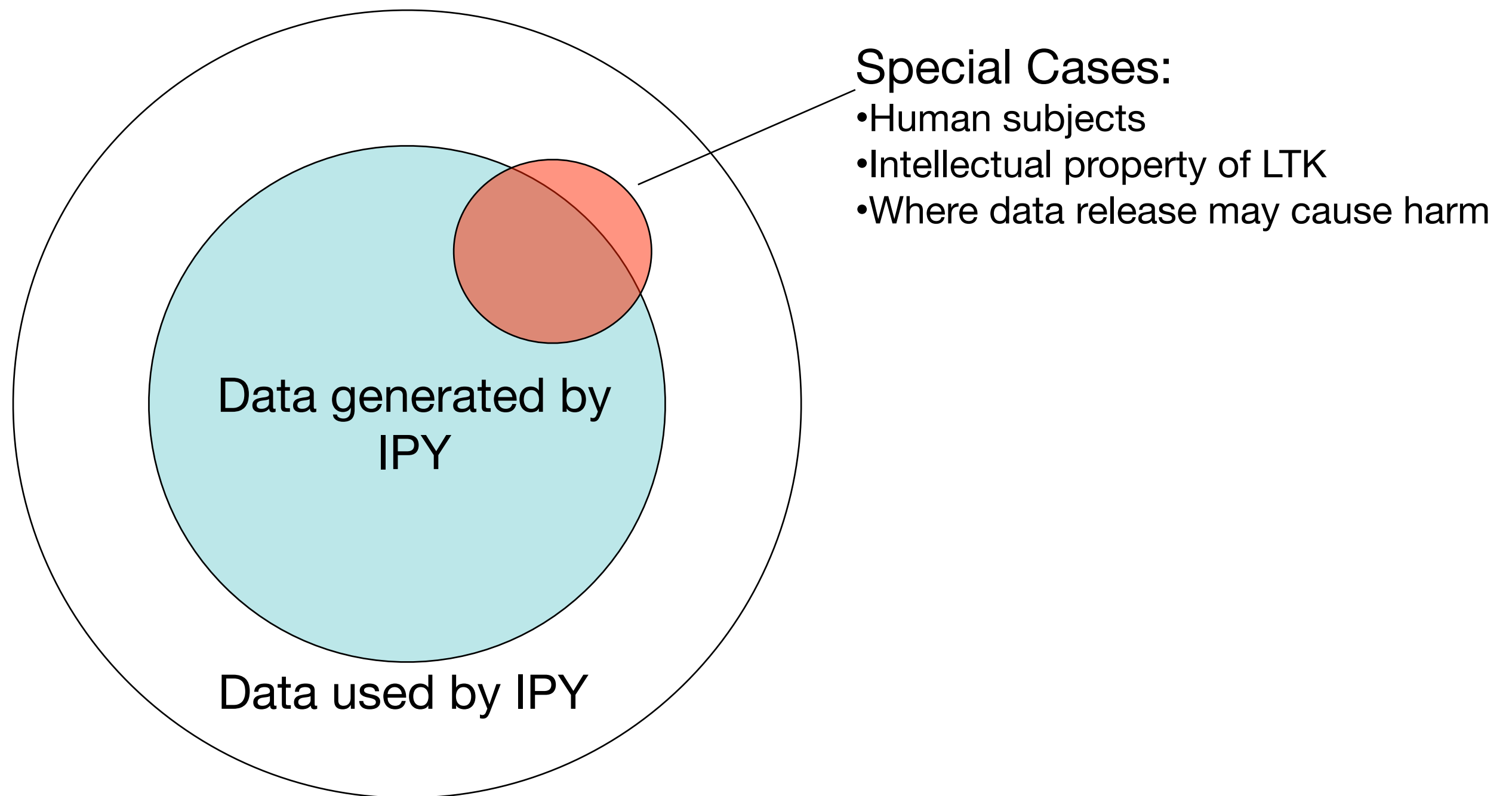
It is a mutual benefit to both data provider and analyst to share information as much as possible. This is the key element for furthering basic research as well as contributing significantly to longer term IPY accomplishments (e.g., new science, helping young scientists, education and outreach).

legacy gets into issues of preservation, system dev., and attribution.

Data can also support the broad education and outreach efforts of IPY by providing the raw material for applied educational approaches

IPY Data Policy—What are IPY Data?

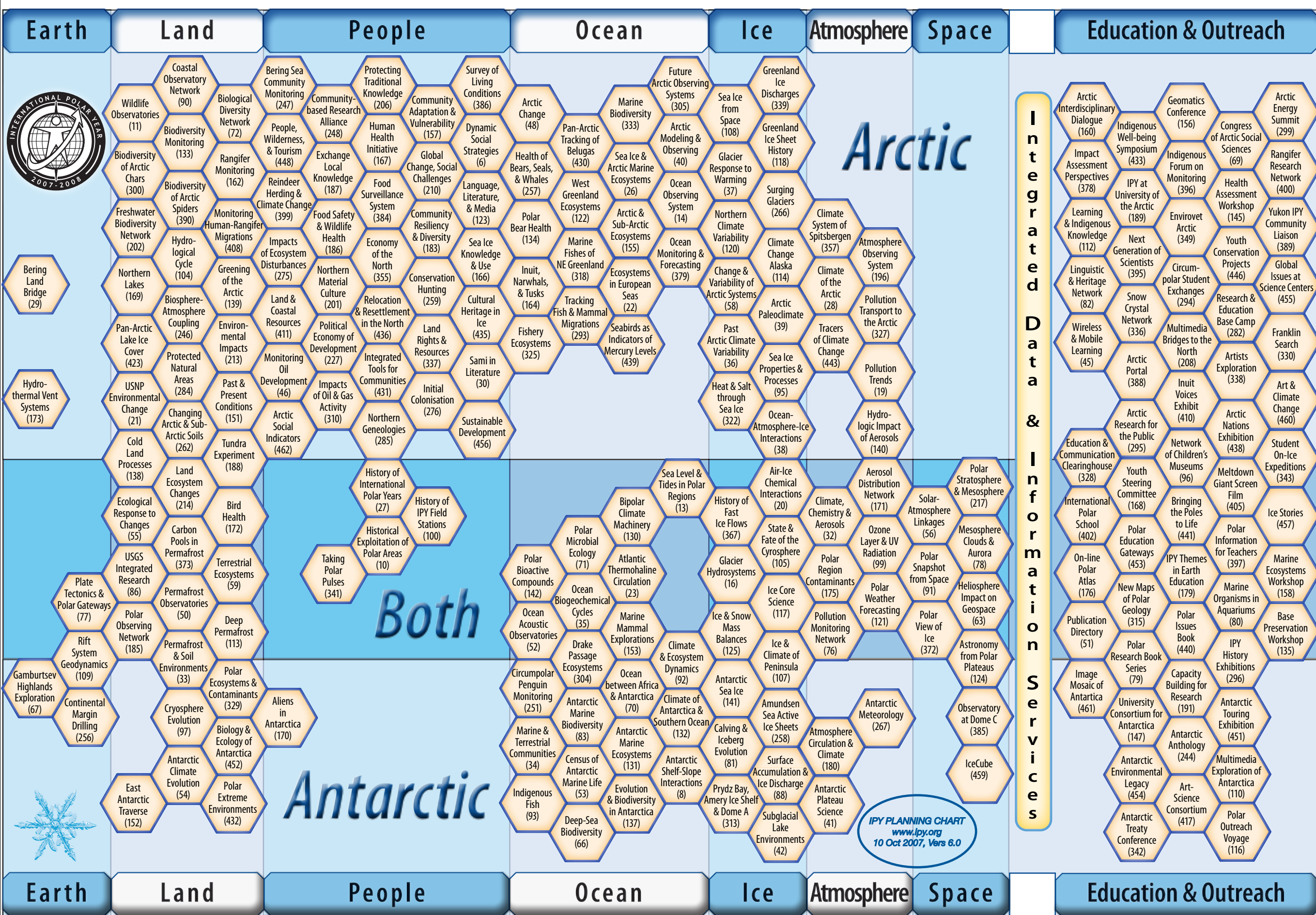
http://www.ipy.org/Subcommittees/final_ipy_data_policy.pdf



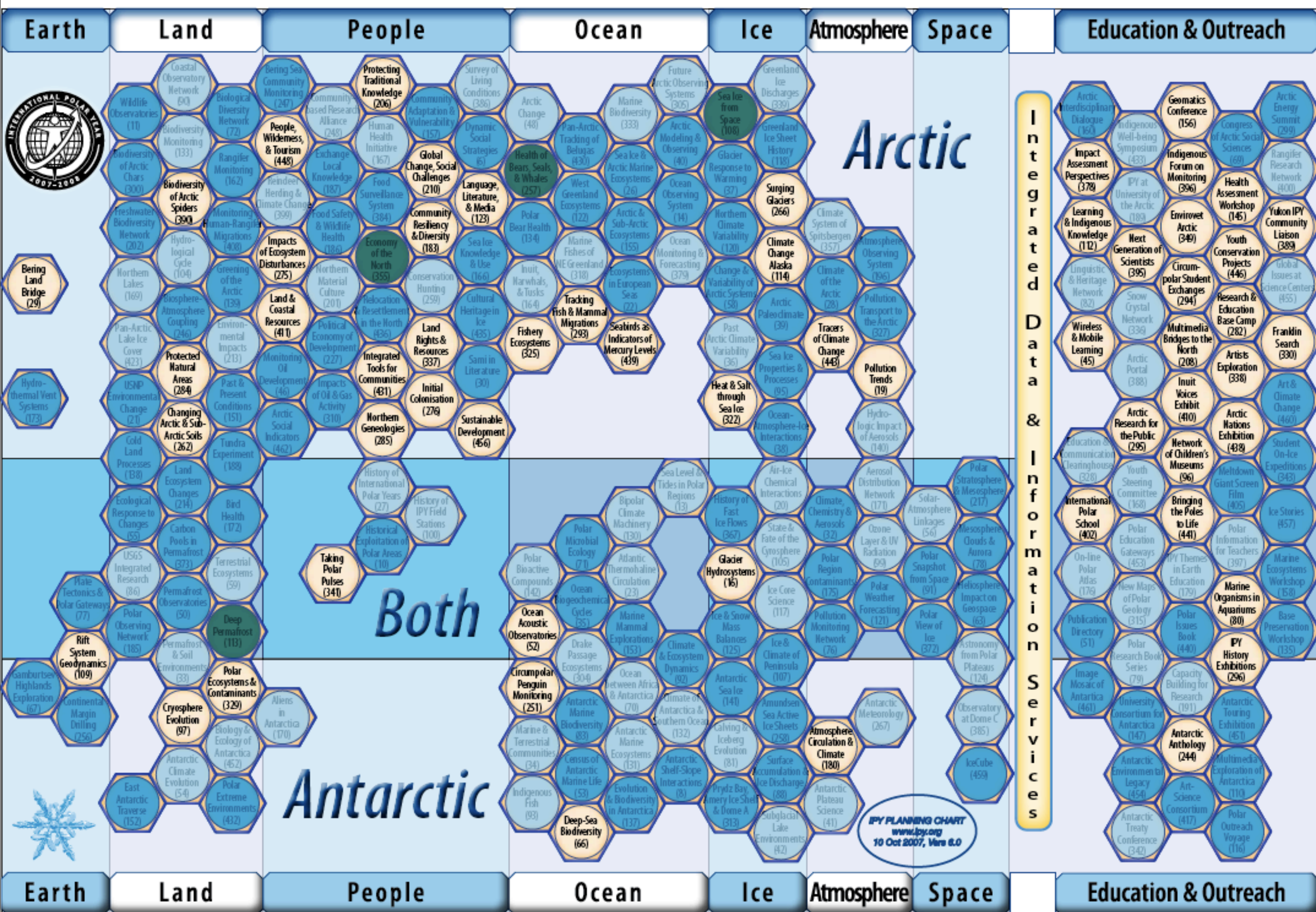
The first challenge was defining what the policy referred to. This also distinguishes the IPY policy from others (what it applies to)

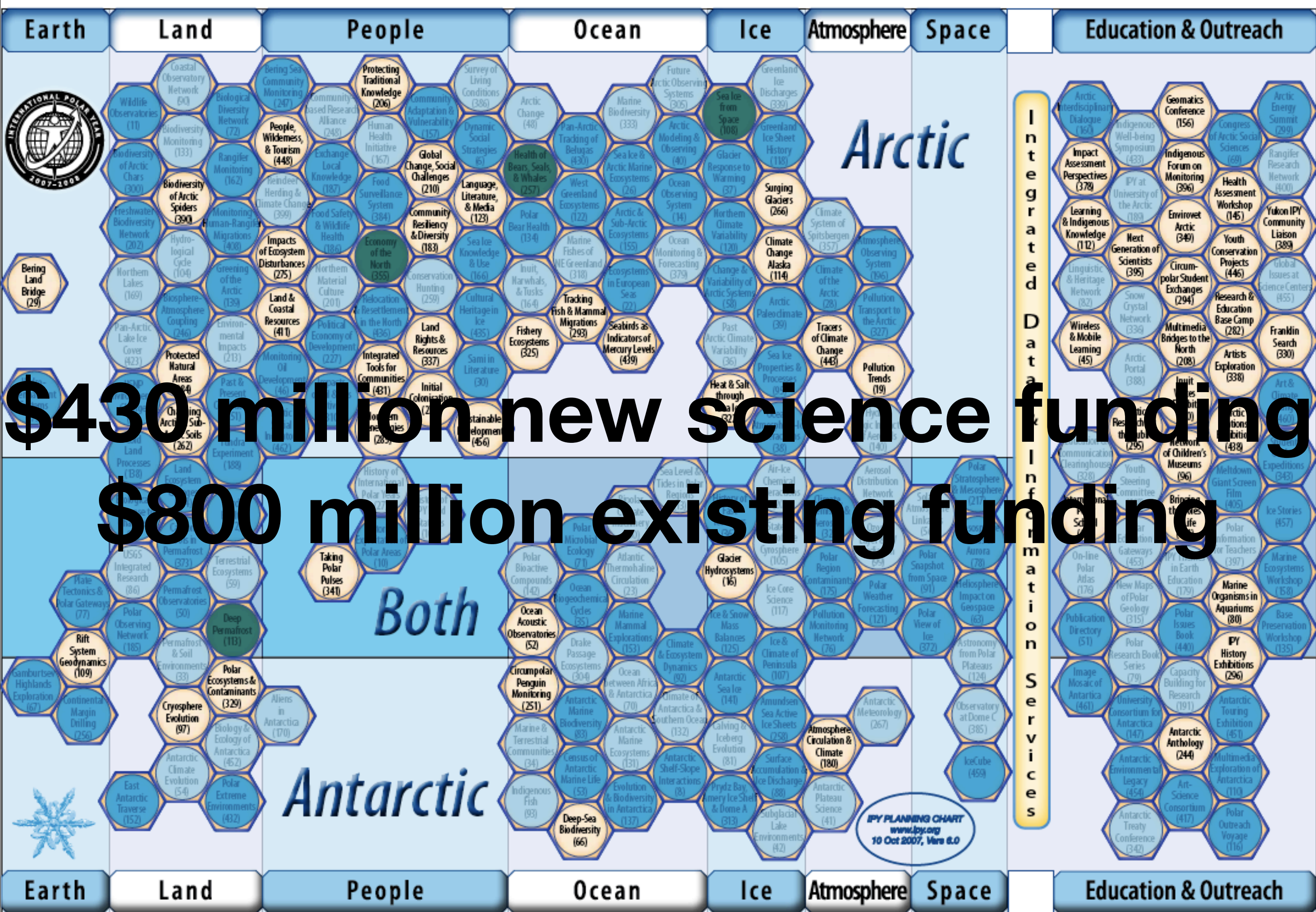
This also shows some unique considerations for IPY, driven by its interdisciplinary nature.

Hence the first part of the strategy--identify.



Have shown a few examples of the use of satellite data during IPY. Fair to say that perhaps most IPY projects benefit in some way from the navigation, communication, and measurement capabilities of satellites.







Data

The IPY Joint Committee requires that IPY data, including operational data delivered in real time, are made available [fully, freely, openly, and on the shortest feasible timescale](#).

The only exceptions to this policy of full, free, and open access are:

- where human subjects are involved, confidentiality must be protected
- where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised
- where data release may cause harm, specific aspects of the data may need to be kept protected

—IPY Data Policy

IPY will [set a new standard](#) in scientific cooperation as rapid and unrestricted data exchange becomes an accepted and enabling factor in daily research.

—IPY Science Plan

So that's the why, here's the what

rapid change creates an impetus for rapid data sharing

“shortest feasible timescale”: allow time for basic Validation and QC, of order of months, not years. May even want to consider broad “prerelease” of initial data with caveats in the spirit of the open source concept that “given enough eyeballs, all bugs are shallow.” (Raymond, 1999 p. 30). Data peer review.

Policy does not accommodate an embargo period.

“New standard” means a step change in increased cooperation

Raymond, ES. 1999. *The Cathedral and the Bazaar: Musings on Linux and Open Source by An Accidental Revolutionary*. Cambridge, MA, USA: O'Reilly.



Metadata

All IPY data must be accompanied by a full set of metadata that completely document and describe the data. In accordance with the ISO standard Reference Model for an Open Archival Information System (OAIS) (CCSDS 2002), [complete metadata may be defined as all the information necessary for data to be independently understood by users and to ensure proper stewardship of the data.](#)

Regardless of any data access restrictions or delays in delivery of the data itself, [all IPY projects must promptly provide basic descriptive metadata of collected data in an internationally recognized, standard format to an appropriate catalog or registry.](#)

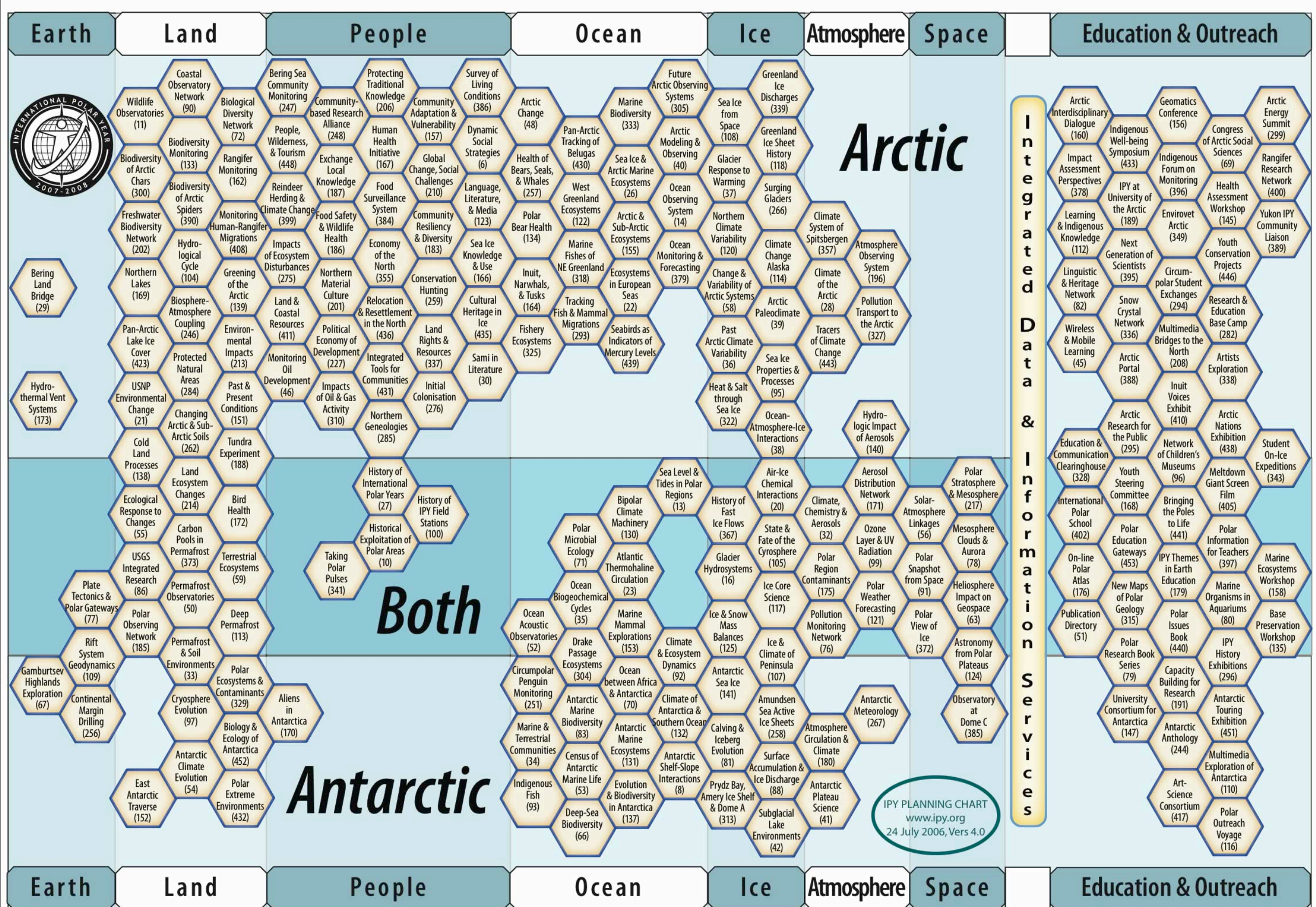
—IPY Data Policy



IPY Metadata Profile (and crosswalk)

- Basic who, what, where, when in either FGDC, DIF, THREDDS (ISO coming, but could use some help), plus some information on metadata provenance.
- The “bare minimum of information necessary to allow simple discovery across disciplines and to ensure we can track the heritage of the metadata in a broadly distributed data management environment.”
- Several portals working to establish a “union catalog”
- Details available at ipydis.org

With so many disciplines it is unlikely we will have a common standard for data formats and metadata standards, but the Data committee want to at least identify what data were collected and where they ended up so we created a minimal profile.

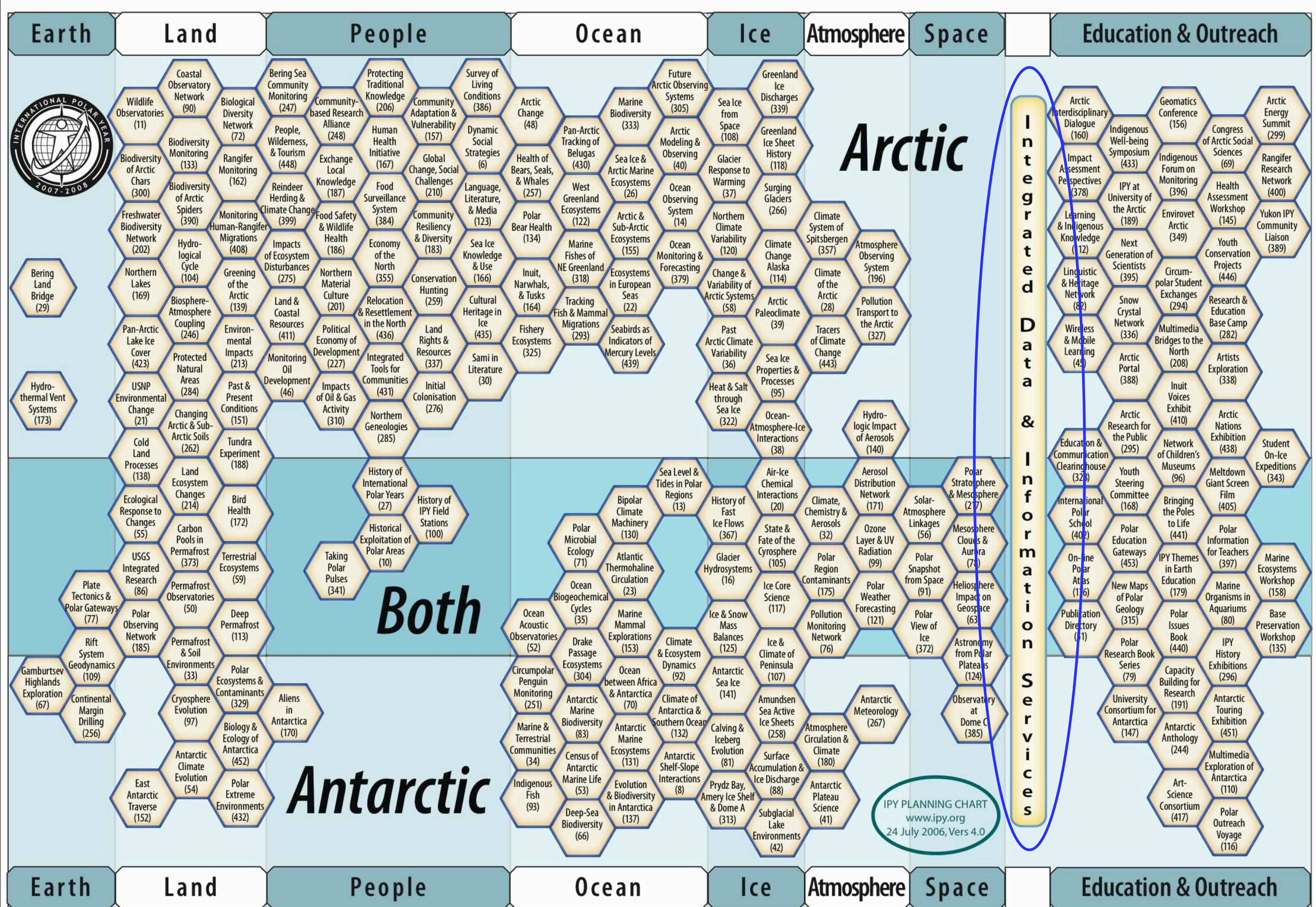


The IPYDIS like any other IPY project is a cluster of related projects, i.e. a loose federation of data centers and networks united in the cause of IPY data management and agreement on standards and approaches.

- The IPYDIS is
- An international federation of data centers, archives, and networks working to ensure proper stewardship of IPY and related data.
 - Promotes the infrastructure, practices, and international relationships necessary to ensure the capture, accessibility, sharing, and long-term preservation of data produced by IPY projects

In a simple sense you can think of data flowing from IPY projects into the DIS and then back out to the projects and the edu and outreach projects. But this is a overly simplistic view.

I'll discuss in a talk on friday how simplistic that is, but for now let me give you a few examples of what's out there.

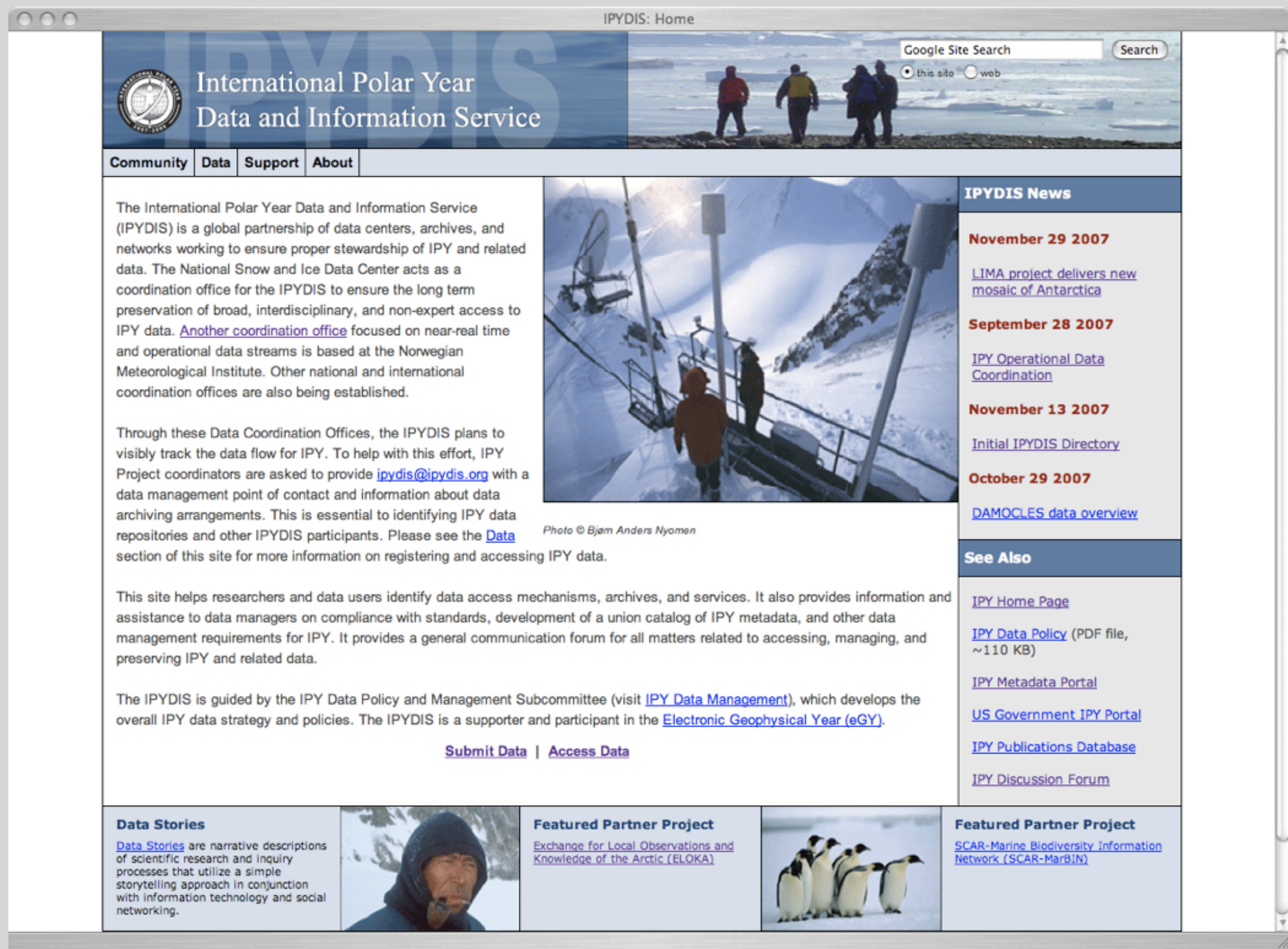


The IPYDIS like any other IPY project is a cluster of related projects, i.e. a loose federation of data centers and networks united in the cause of IPY data management and agreement on standards and approaches.

- The IPYDIS is
- An international federation of data centers, archives, and networks working to ensure proper stewardship of IPY and related data.
 - Promotes the infrastructure, practices, and international relationships necessary to ensure the capture, accessibility, sharing, and long-term preservation of data produced by IPY projects

In a simple sense you can think of data flowing from IPY projects into the DIS and then back out to the projects and the edu and outreach projects. But this is a overly simplistic view.

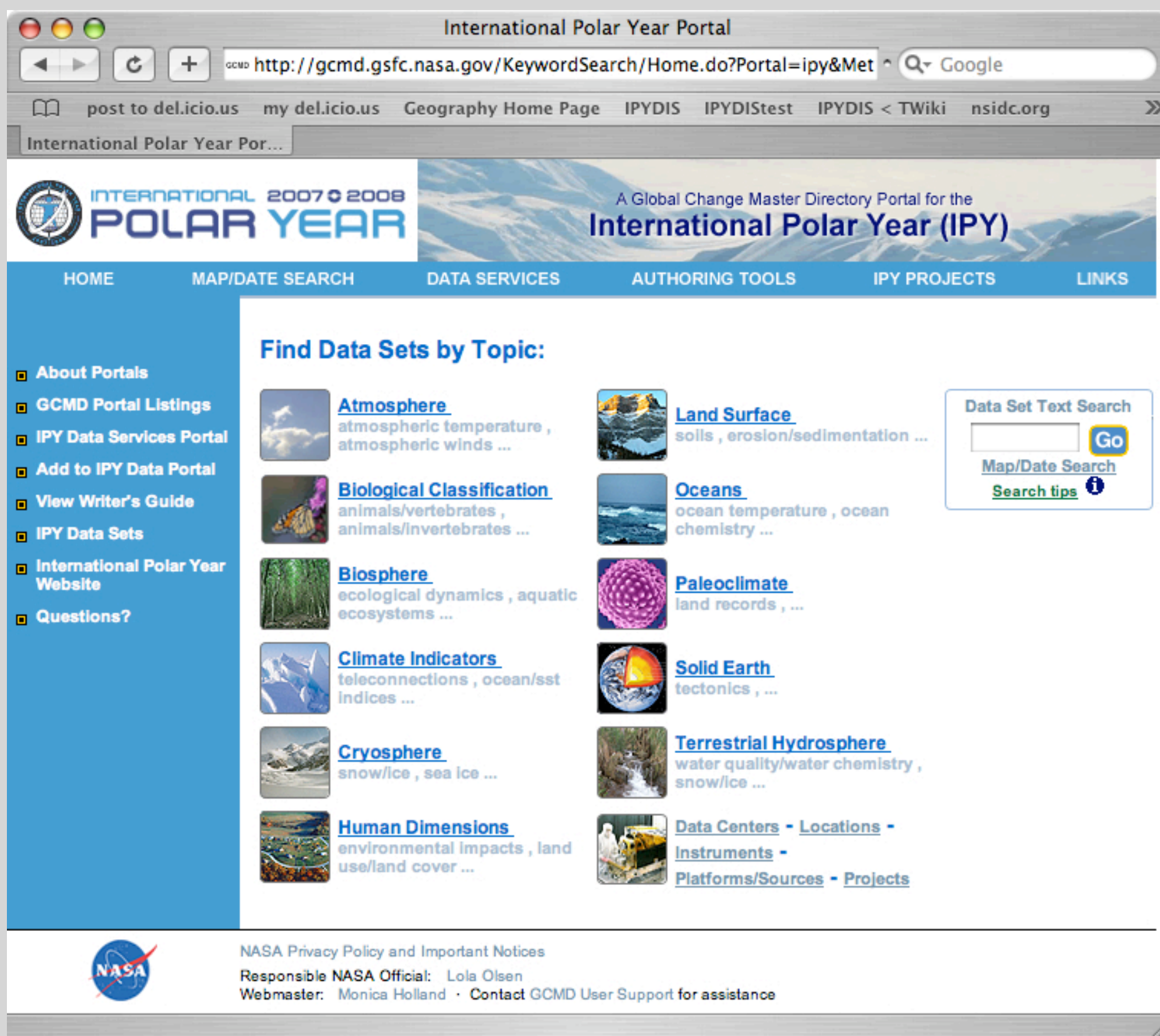
I'll discuss in a talk on friday how simplistic that is, but for now let me give you a few examples of what's out there.



IPY Data and Information Service—<http://ipydis.org>



I'm trying to be a hub, so this should be your first resource. All the examples I show quickly here are available through this site.




IPY Metadata Portal — <http://gcmd.nasa.gov/portals/ipy>
(others in progress)



Other systems be developed at national levels

IPYDIS: Directory



International Polar Year
Data and Information Service

Google Site Search

Search

☒ this site ☐ web

CommunityDataSupportAbout

Home > Community > IPYDIS Directory

Participate

Presentation Archive

Events

IPYDIS Directory

Community

IPYDIS Directory

This table shows all the substantially funded IPY projects and their data archives. This will be replaced by an interactive, searchable directory soon.

Please send additions and corrections to ipydis@ipydis.org

Project number	Title	Project Point of Contact	Archive Organization	Archive contact	Archive email address
6	Dynamic Social Strategies in Arctic Environments: Long-term Perspectives on Movement and Communication	Hans Christian Gulløv	Danish Polar Center	Danish Polar Center	
10	Large Scale Historical Industrial Exploitation of Polar Areas	Prof. Dr. Louwrens Hacquebord			
11	Arctic Wildlife Observatories Linking Vulnerable EcoSystems	Gilles Gauthier		Luc Courmoyer	luc.courmoyer@cen.ulaval.ca
14	Integrated Arctic Ocean Observing System	Prof Jean Claude Gascard			
21	Understanding environmental change and its biological, physical, social, subsistence and cultural effects in national parks and protected areas of Alaska, Chukotka, and the Yukon	Robert A. Winfree	Central Alaska Network Arctic Network Southwest Alaska Network	Scott Miller, Edward Debevec, Dorothy Mortenson	Scott_Miller@nps.gov Edward_Debevec@nps.gov Dorothy_Mortenson@nps.gov
22	POLARSTERN expedition "HERMES - the Nordic margin" in the framework of the EU funded Integrated Project HERMES (Hotspot Ecosystem Research on the Margins of European Seas)	Michael Klages			


IPYDIS Directory

Do you like this idea?
Please register your archive with the IPYDIS.

DADDI Mercury Brief Display Page


http://mercdev3.ornl.gov/daddi3/send/query?term1=sea+ice&term1attribute=fullText&op1

Google



DADDI

a working prototype focused on arctic coastal data



Metadata Summary

EmailHelp

Your search found: 173 documents.
Query: fullText:sea ice AND datasource:(ccin ciesin nsidcm ipy daac)

Filter by resource type

- digital (16)
- media (16)
- table (16)
- images (5)
- cd-rom (3)
- figures (1)
- image (1)

Filter by keywords

- Water Temperature (24)
- Ice Extent (23)
- Sea Ice Concentration (21)
- Ice Depth/Thickness (20)
- Ice Sheets (20)
- Salinity (19)
- Air Temperature (14)

Filter by decade

- 2000 (68)
- 1990 (49)
- 1980 (9)
- 1970 (3)
- 2010 (2)
- 1890 (1)

Viewing Documents 1 - 10 out of 173

Prev 1 2 3 4 5 6 7 8 9 10 Next

Return to Search

Sort By: Relevance Period of record Data source

Filter by data providers

- National Snow and Ice Data Center (149)
- International Polar Year (17)
- Canadian Cryospheric Information Network (7)

SEASONAL SEA-ICE MONITORING AND MODELLING SITE

01/01/1990 - 12/31/1995

SIMMS is a multi-year, multi-disciplinary field experiment that has many individual projects, each with autonomous goals and objectives. The science conducted has either directly or indirectly evolved from research relating to one of four general themes: i) sea ice energy balance; ii) numerical modeling of atmospheric processes; iii> remote sensing of snow covered sea ice; and iv) ecosystem studies.

Data provider: CANADIAN CRYOSPHERIC INFORMATION NETWORK

★★★★★★★★★★

Get data View full metadata

COLLABORATIVE INTERDISCIPLINARY CRYOSPHERIC EXPERIMENT (C-ICE)

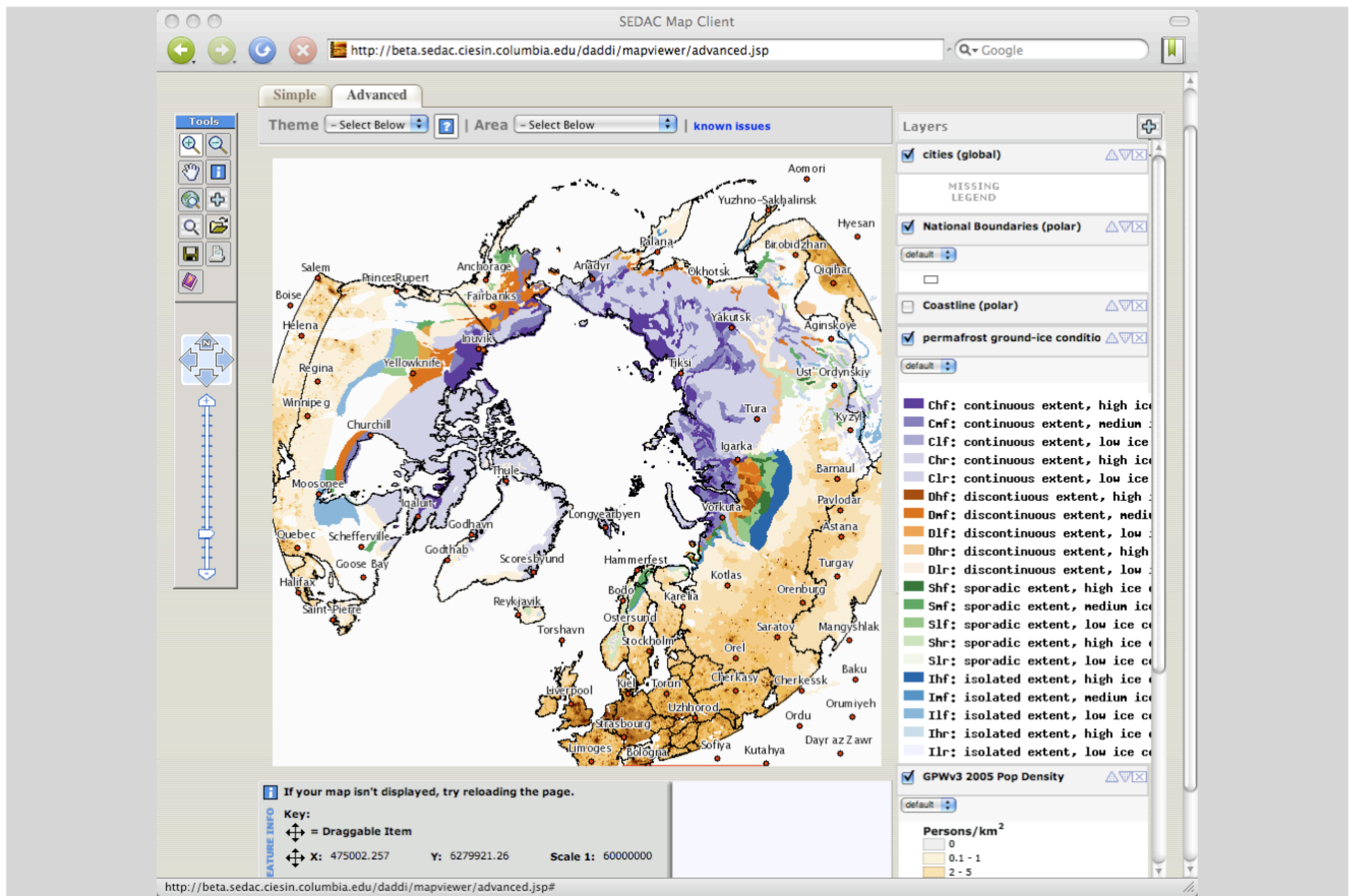
01/01/2003 - Present

The primary objective of C-ICE research is to assess the use of ice surface properties, derived from microwavelength remote sensing data, for modeling geophysical and biophysical processes within an

Discovery Access and Delivery of Data for IPY—<http://mercury.ornl.gov/daddi>
Facetted Data Search and Visualization for the Arctic



I mentioned the GCMD portal, this is an alternative interface to IPY data in conjunction with four data centers with Arctic coastal data.



DADDI Visualization and Integration

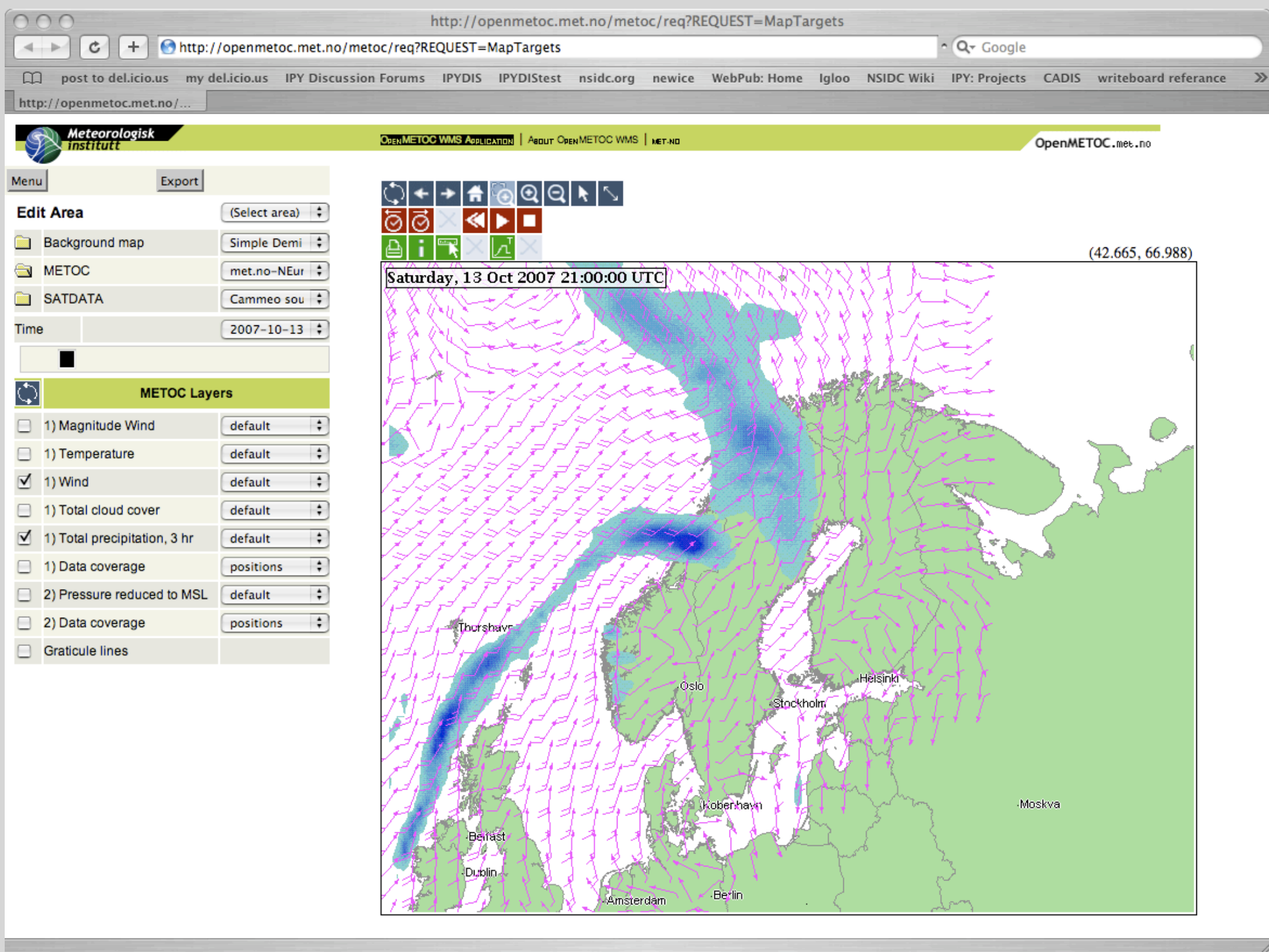




Operational Data Coordination—<http://ipycoord.met.no>



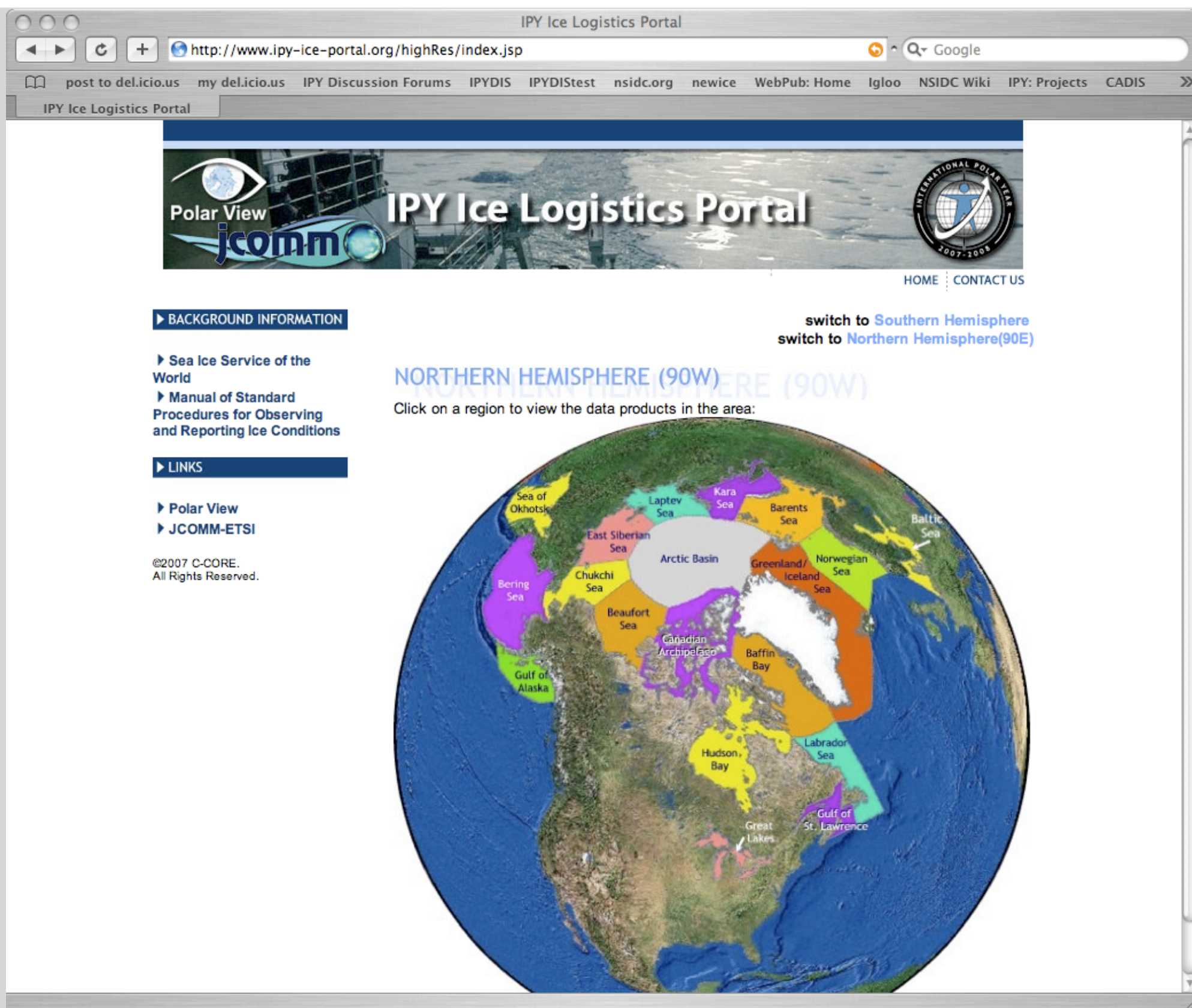
Operational data coordination by Øystein Godøy at the Norwegian Meteorological Institute.



ECMWF Surface Fields—<http://openmetoc.met.no/>



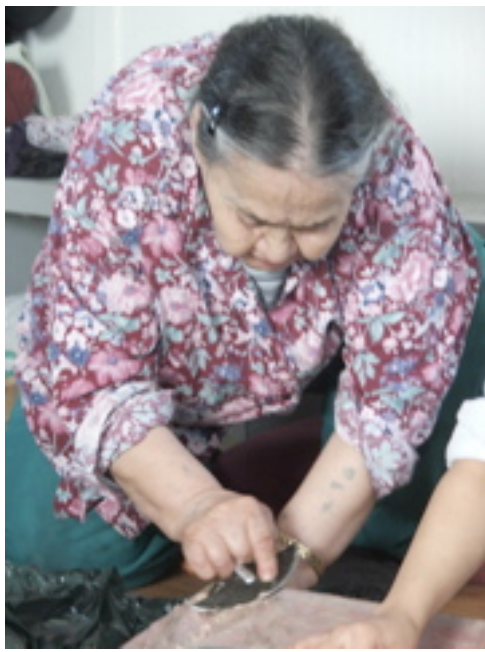
ECMWF will make data available to IPY scientists. METNO is currently working on technical solutions to provide IPY-users with access to ECMWF data. The technical solution will be WMO WIS compliant in time, although the first version probably will lack some WIS functionality. In the mean time, check out OGC WMS access to ECMWF-data through <http://openmetoc.met.no/>



IPY Ice Logistics Portal—<http://www.ipy-ice-portal.org>



Created by Polar View and
JCOMM



ELOKA

The Exchange for Local Observations and Knowledge of the Arctic

works to provide data management and user support to facilitate the collection, preservation, exchange, and use of local observations and knowledge of the Arctic.

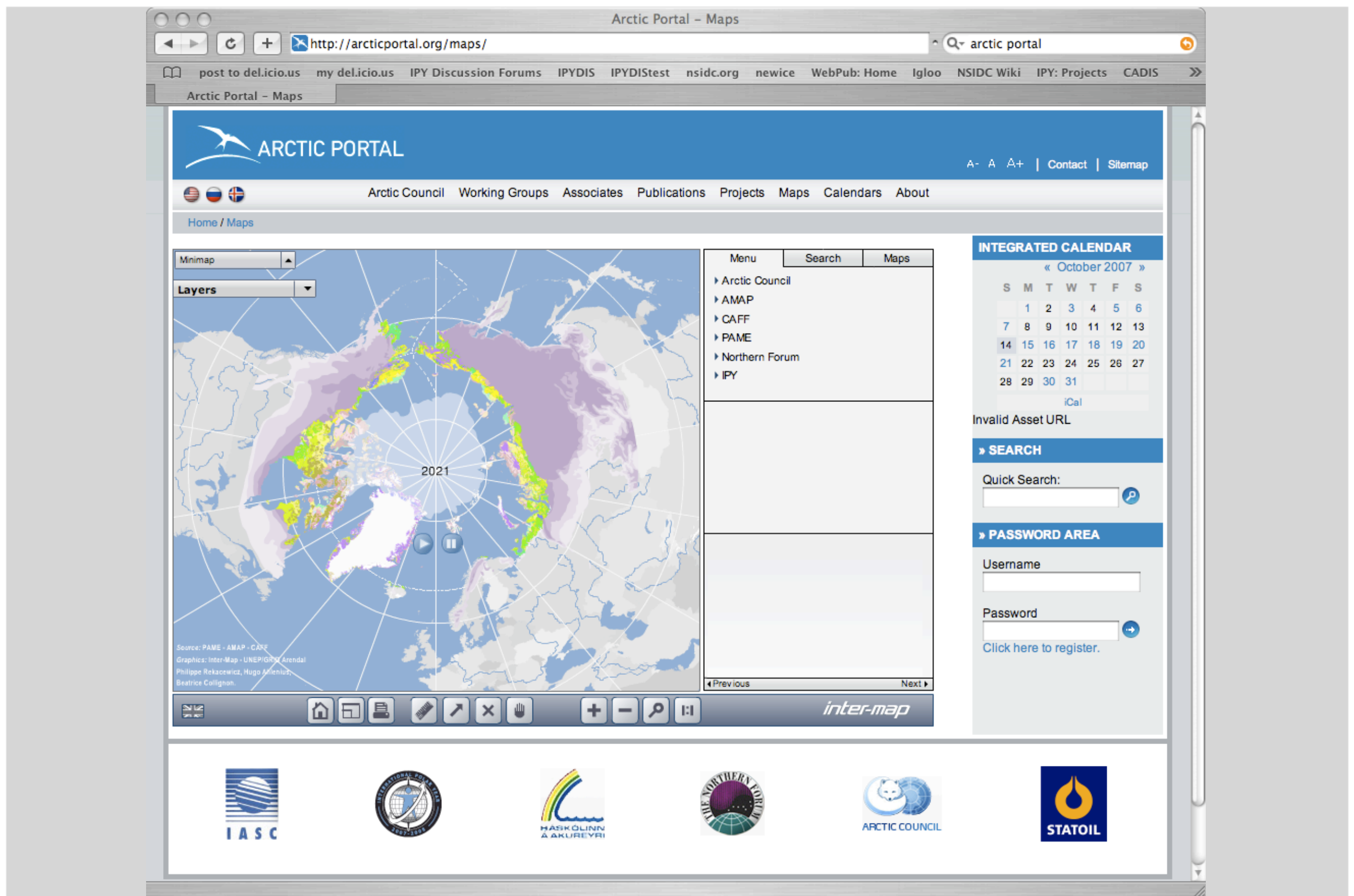
<http://nsidc.org/eloka>

PI: Shari Gearheard



National Snow and Ice Data Center
Supporting Cryospheric Research Since 1976





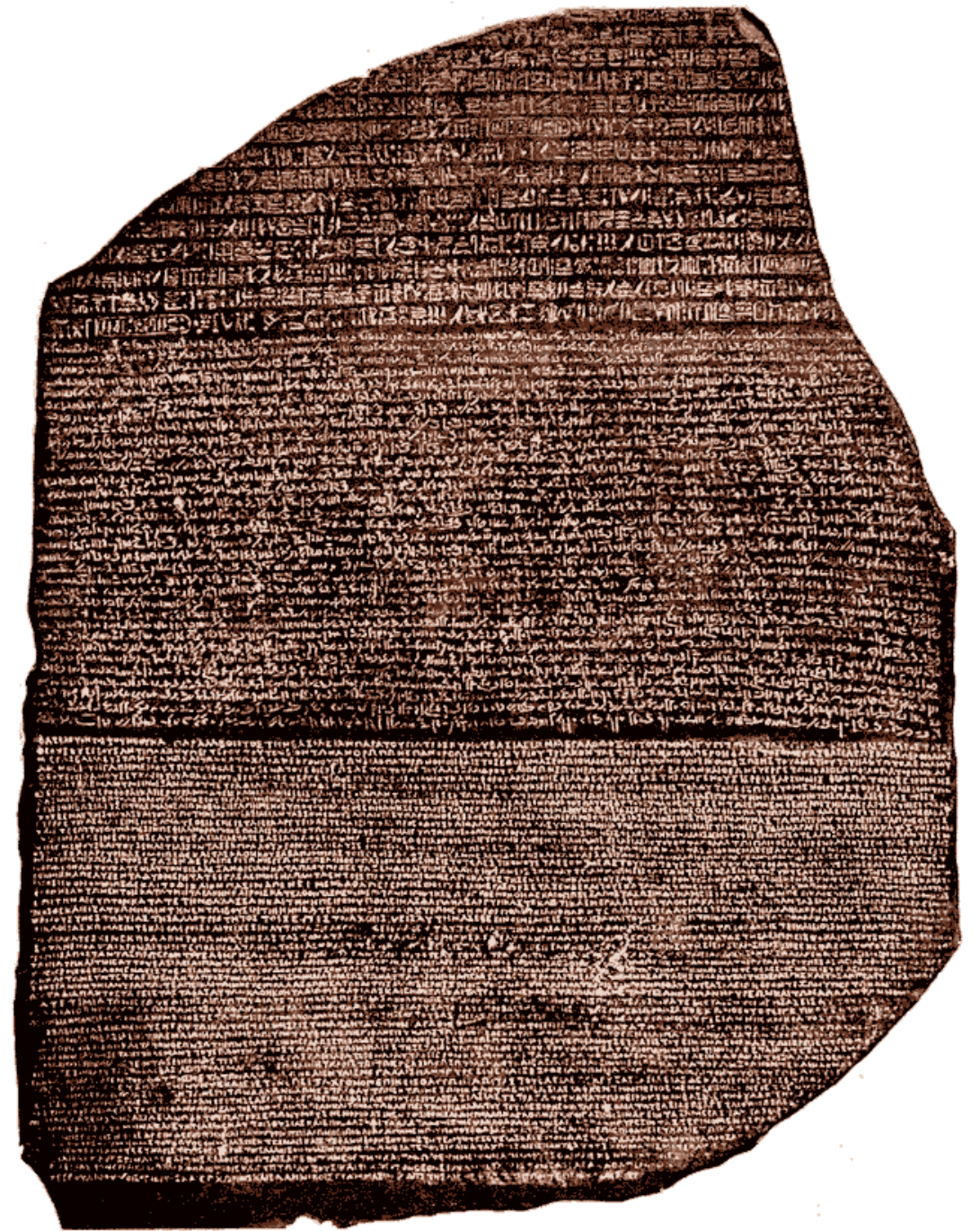
The Arctic Portal—<http://arcticportal.org>



The Arctic Portal team can help with project web site creation and maybe Google Earth.

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.



Projects have a responsibility to prepare data for preservation and plan transition

That's the what. So what about "How"



Preservation

An *immense* challenge, but...

More than 20 World Data Centers have expressed interest in archiving IPY data.

A first step to a new generation of science archives?



IPY Data Tasks

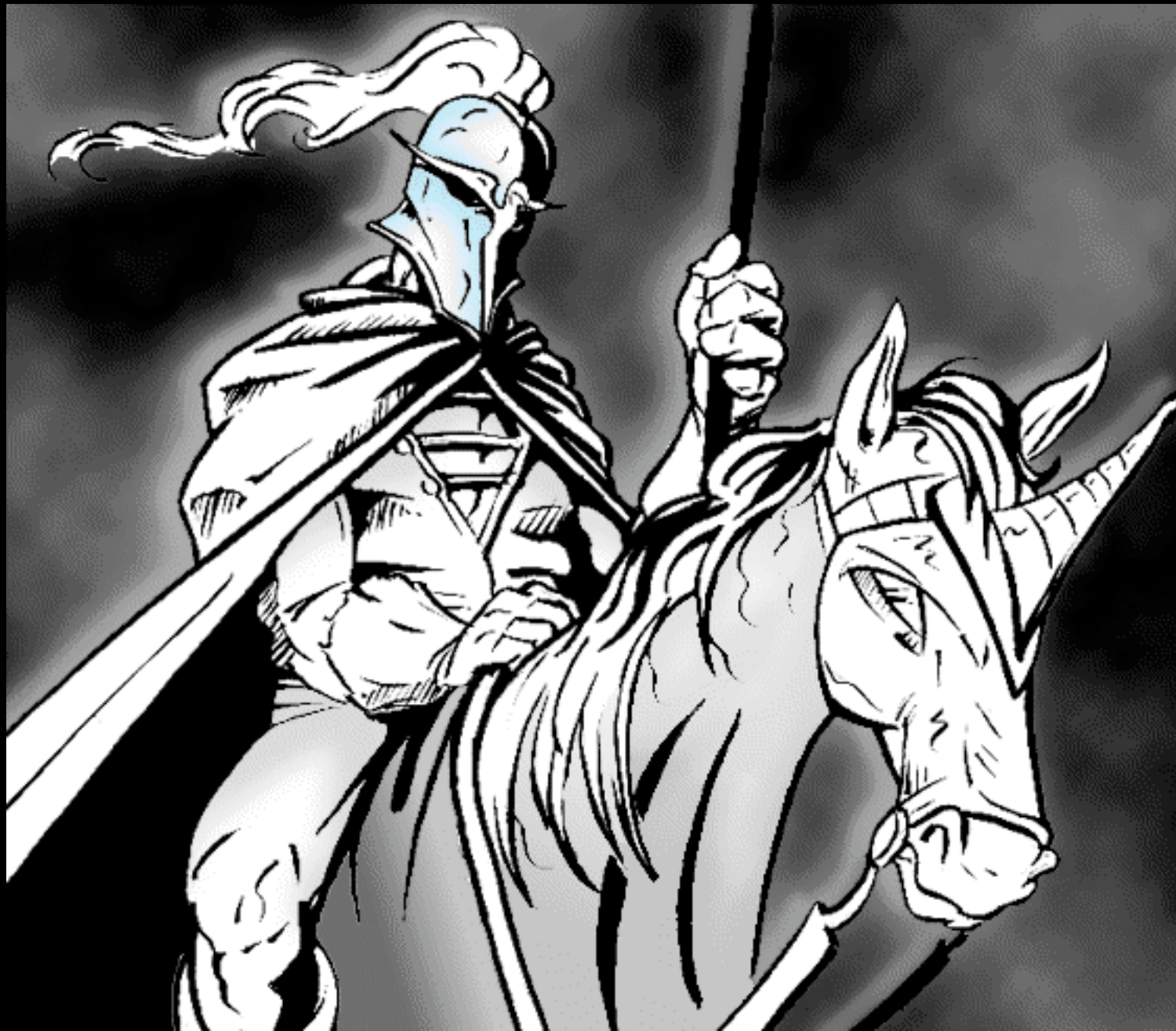
2007	2008	2009	2010	2011	2012
IDENTIFICATION					
		AVAILABILITY			
			PRESERVATION		
COORDINATION					

Identify and share the data—in progress but slow (not a tech. issue)
Goal: *all* metadata by Jun 2009

Serve the data—extremely variable, technical and social issues
Goal: ongoing demos of integration
all data available Mar 2010

Preserve the data—a major issue.
Goal: *all* data in secure archives by Mar 2012

Have shown a few examples of the use of satellite data during IPY. Fair to say that perhaps most IPY projects benefit in some way from the navigation, communication, and measurement capabilities of satellites.



William Craig's white knights of data sharing--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.

—

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