Offener Zugang zu Forschungsdaten - eine Herausforderung

Jens Klump
Geoforschungszentrum Potsdam / Helmholtz Gemeinschaft
Hans Pfeiffenberger
Alfred Wegener Institut / Helmholtz Gemeinschaft

HELMHOLTZ GEMEINSCHAFT

Agenda

- *Intro*
  - **Das Ideal / die Haltung der Wissenschaftsförderer (die Mandate), ...**
  - **praktische Beispiele (aus der Erdsystemforschung)**
    - *Inkl. Beispiele für Grenzen des offenen Zugangs*
- *Diskussion*
  - **Definitionen OA zu Daten**
  - **Herausforderungen (=Probleme)**
  - **Lösungen (Ansätze)**

Worum geht es - finanziell?

- *Bruttowertschöpfung 2007* ***2171,2 Mia €***
  (Stat. Bundesamt)
- *2%* *davon für Forschung* *43,4 Mia €*
- *50%* *davon für Daten* *21,7*
- *33%* *davon nachnutzbar* *7,2*
- *33%* *davon als Beleg* *7,2*
- *34%* *davon Schrott* *---*

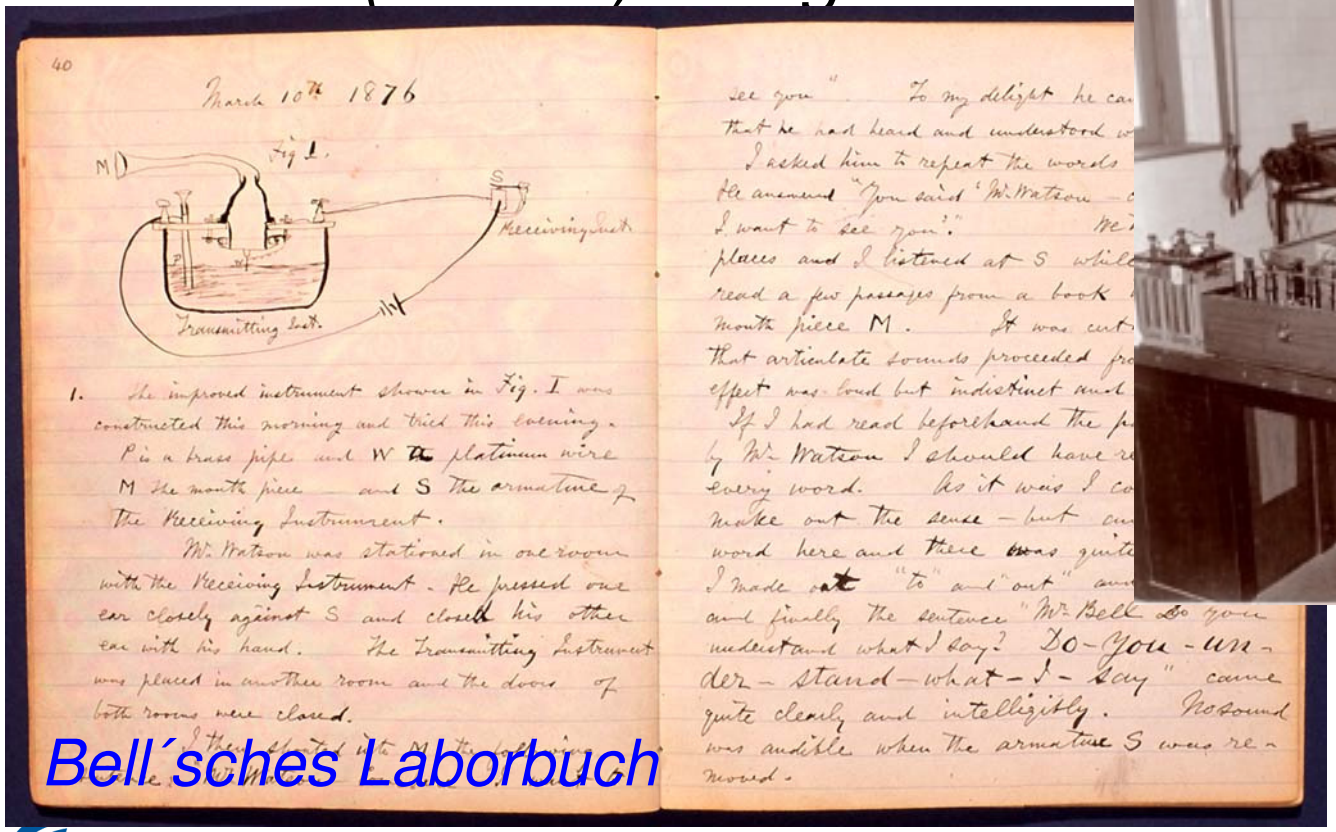
GwP: Wozu?

- *Verantwortung gg.über dem Steuerzahler: Maximaler Nutzen*
  - **Gute Wissenschaft : Nachvollziehbarkeit**
  - **Nicht durch (Daten-)Bürokratie behindert**
  - **... Optimale Nachnutzung von Daten, Data Sharing**

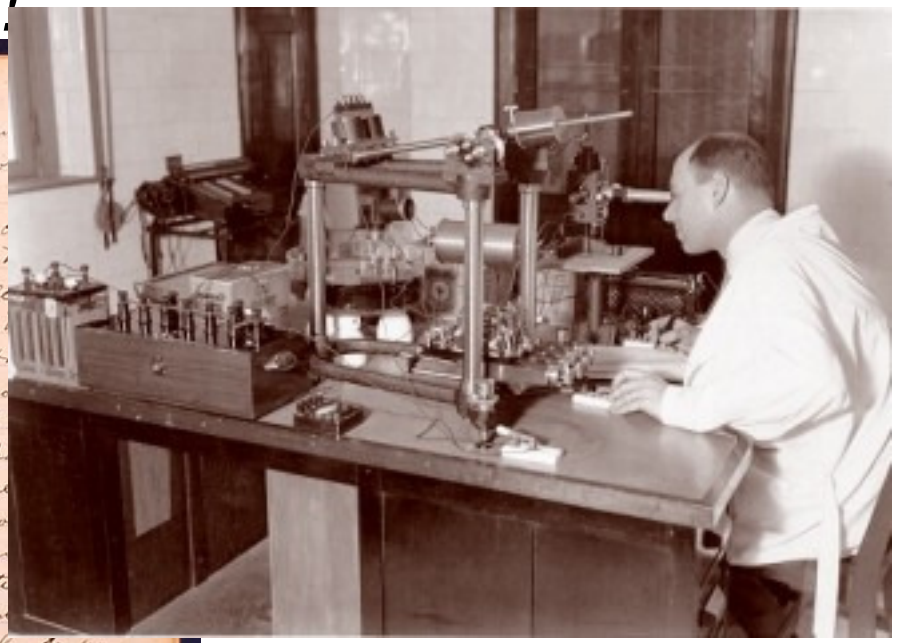
- *Fortschritt der Wissenschaft*
  - **Negativ-Erscheinungen vermeiden (z.B. Falschinformationen, Rauschen)**
  - **Nachnutzung von Daten, Data Sharing**

GwP: Was hat sich geändert?

- *Soziale „Kontrolle“ in größerer Community ?*
- *Anderer finanzieller Maßstab !*
- *Andere (Daten-) Mengenskala !*



Bell´sches Laborbuch



Am Helmholtz-Pendel

Nachnutzung, Data Sharing

HELMHOLTZ GEMEINSCHAFT

EUROHORCs Vision

A globally competitive ERA requires:

1. An effective European research policy, capitalising on cultural, geographic and scientific diversity
2. A stimulating education system
3. A single European labour market for researchers
4. Adequate funding for top quality curiosity-driven research
5. Cross-national funding, benchmarking of quality and shared scientific priorities for strategic research and bottom up researcher-driven programmes
6. Excellent research institutions
7. World-class research infrastructures
8. Open access to the output of publicly funded research and permanent access to primary quality assured research data
9. Effective and trusted bridges between science, society and the private sector
10. Openness to the world

EUROHORCs Roadmap

9. Common policy on Open Access and Permanent Access to research data

Addresses Vision point 8.

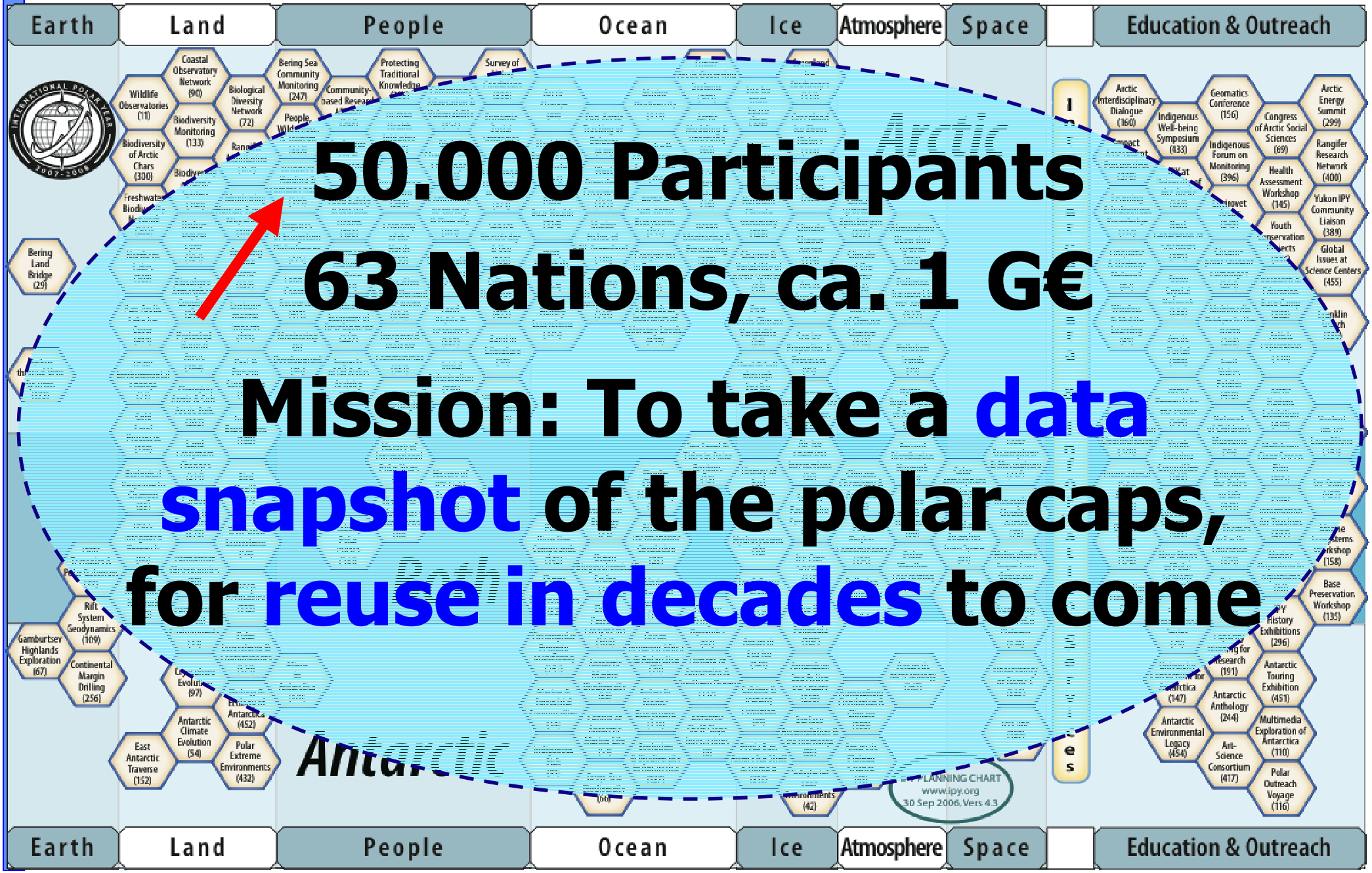
Whilst the crucial role of peer reviewed publications in both academia and research is recognised, there is also pressure to ensure that the results of publicly funded research are available quickly and publicly. Influential national and international funding and strategy bodies are formulating their own statements and policies. EUROHORCs Member Organisations, which account among them for over 18 billion Euros research funding in Europe, will develop a joint statement on Open Access. The formulation and adoption of such a common policy would have an immediate, beneficial and unifying impact.

The collection of research data is a huge investment. Permanent access to such data, if quality controlled and in interoperable formats, allows other researchers to use them, allows re-analysis of, for example, long time series and could play a role in ensuring research integrity. EUROHORCs and ESF will address how to best promote and ensure such permanent access to data generated with their funding.

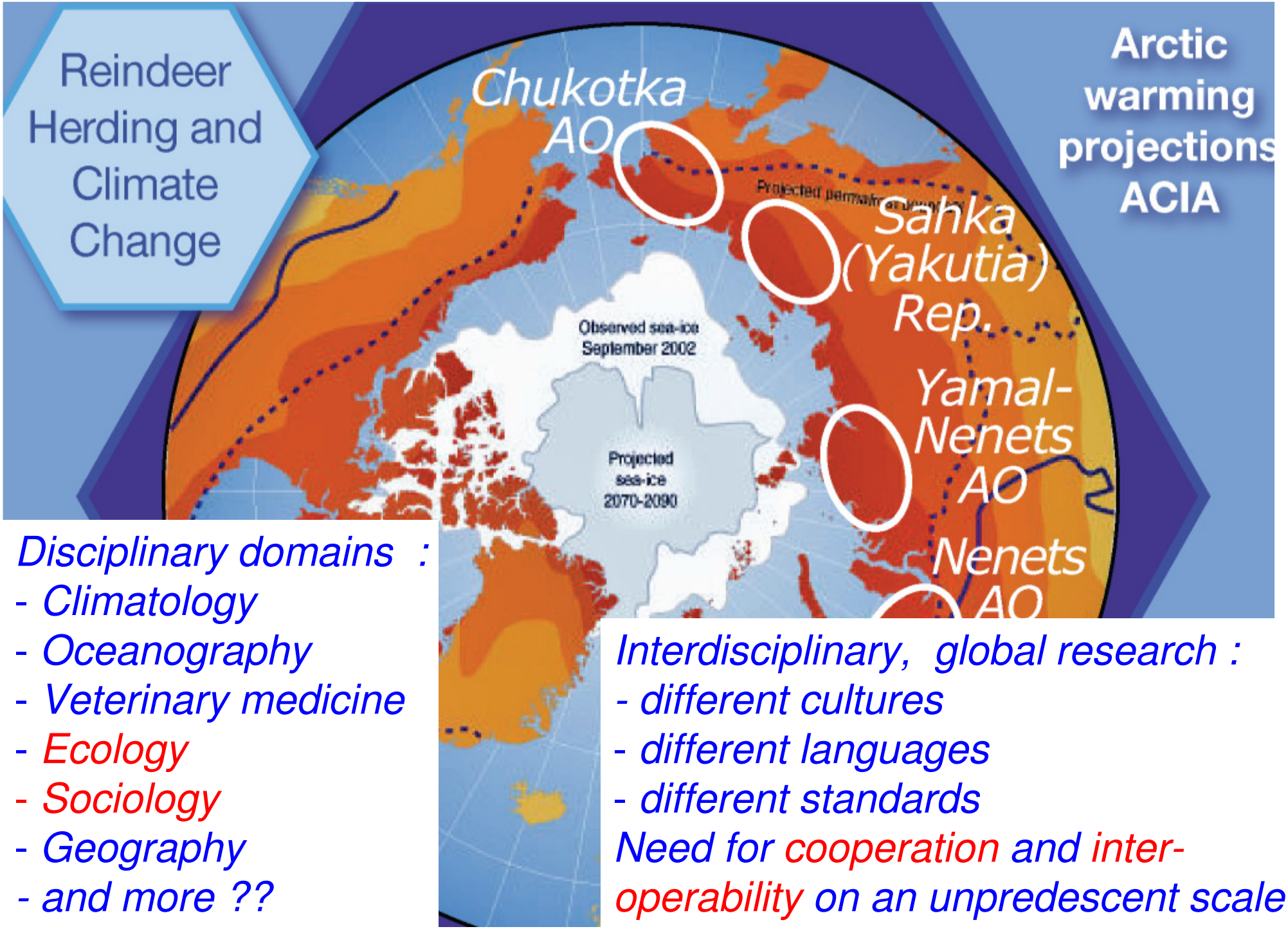
Allianz der Wissenschaftsorganisationen

- *11.Juni 2008: Initiative Digitale Information*
  - **... Es ist unbestreitbar, dass viele dieser Daten nach einer relativ kurzen Phase der Auswertung durch Einzelne oder kleine Gruppen dem Vergessen oder gar dem Verfall ausgesetzt sind. Hier sehen alle Wissenschaftseinrichtungen einen dringenden Handlungsbedarf hinsichtlich der systematischen Sicherung, Archivierung und Bereitstellung dieser Daten für die Nachnutzung durch Dritte.**

International Polar Year 2007-2008



Reindeer Herding and Climate Change



Disciplinary domains :
- *Climatology*
- *Oceanography*
- *Veterinary medicine*
- *Ecology*
- *Sociology*
- *Geography*
- *and more ??*

Interdisciplinary, global research :
- *different cultures*
- *different languages*
- *different standards*

Need for cooperation and inter-operability on an unpredescent scale

Data policy of the IPY 2007/2008

- *„... 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*

- *„... to ensure the lasting legacy of IPY, it is essential to ensure long-term preservation and sustained access to IPY data. All IPY data must be archived in their simplest, useful form and be accompanied by a complete metadata description.“*

- *„... it is the responsibility of individual IPY projects to make arrangements with long-term archives ...“*

PANGAEA®

Publishing Network for Geoscientific & Environmental Data

Data Description

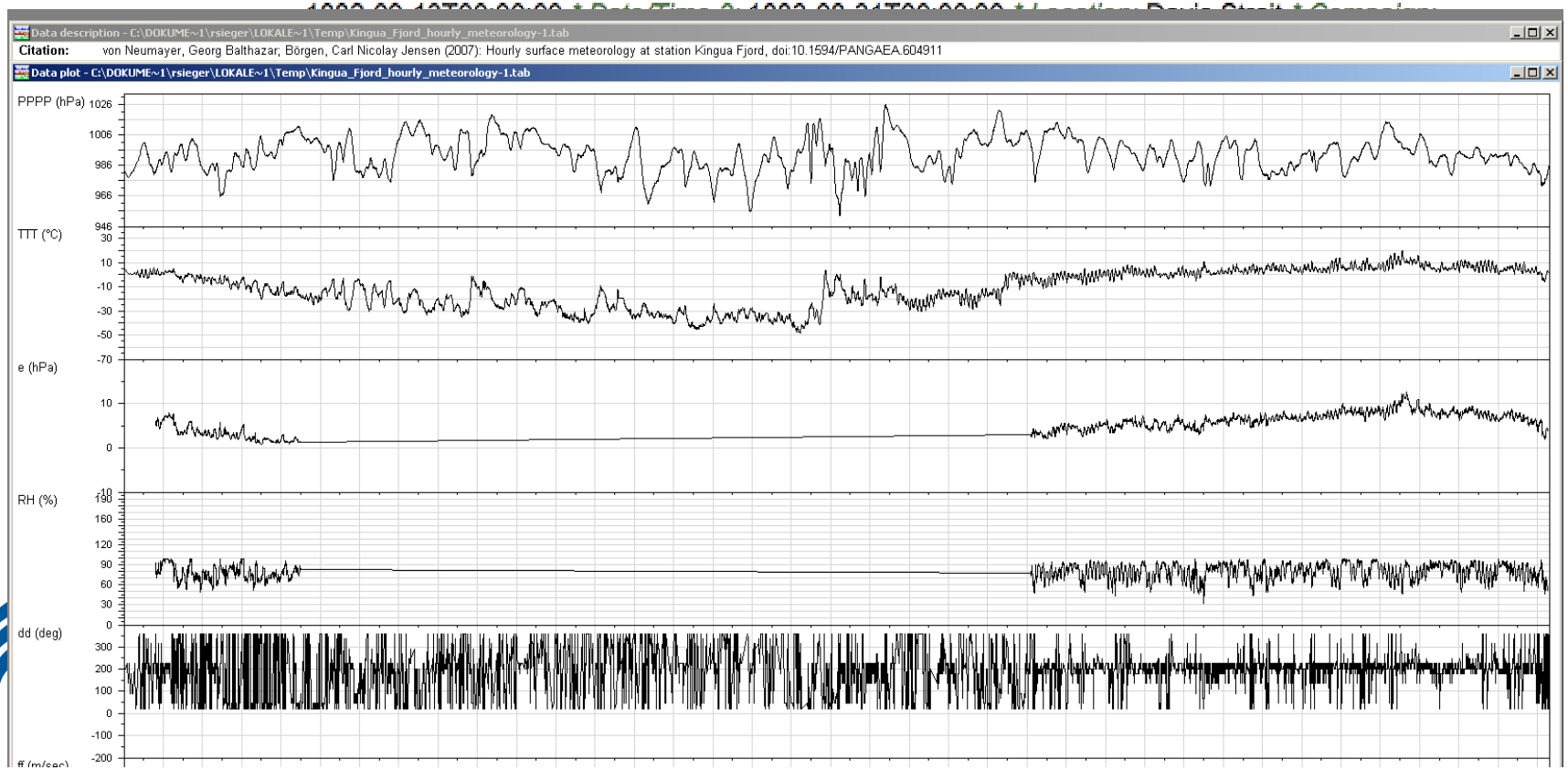
Citation: **von Neumayer, Georg Balthazar; Börgen, Carl Nicolay Jensen (2007):** Hourly surface meteorology at station Kingua Fjord, doi:10.1594/PANGAEA.604911

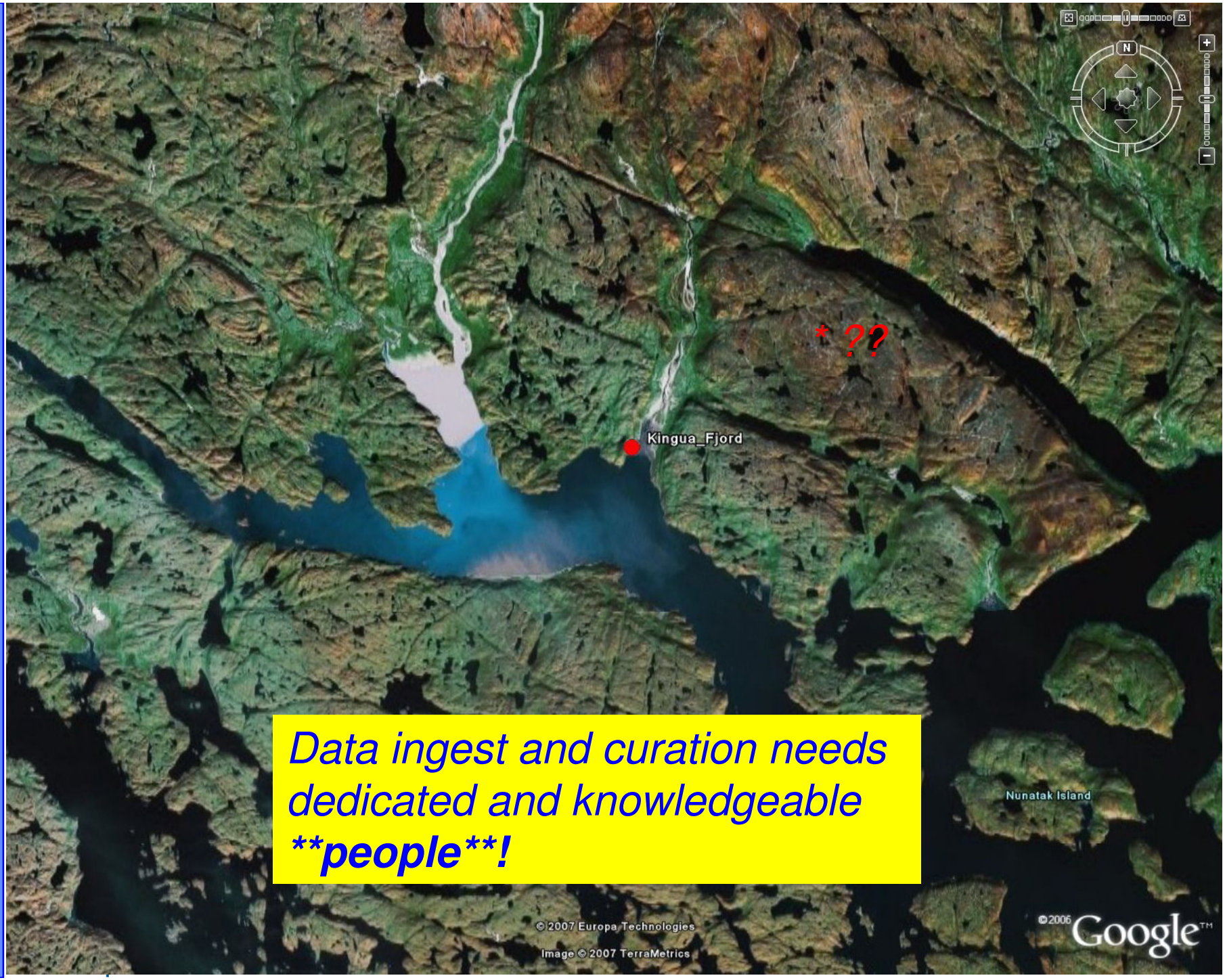
Reference(s): **von Neumayer, Georg Balthazar; Börgen, Carl Nicolay Jensen (1886):** Die Beobachtungs-Ergebnisse der Deutschen Stationen, Kingua-Fjord, *Berlin: Verlag von A. Asher & Co*, **1**, 736 pp

Coverage: *West:* -67.3450 * *East:* -67.3450 * *South:* 66.5948 * *North:* 66.5948

Date/Time Start: 1882-09-12T08:00:00 * *Date/Time End:* 1883-09-10T04:00:00

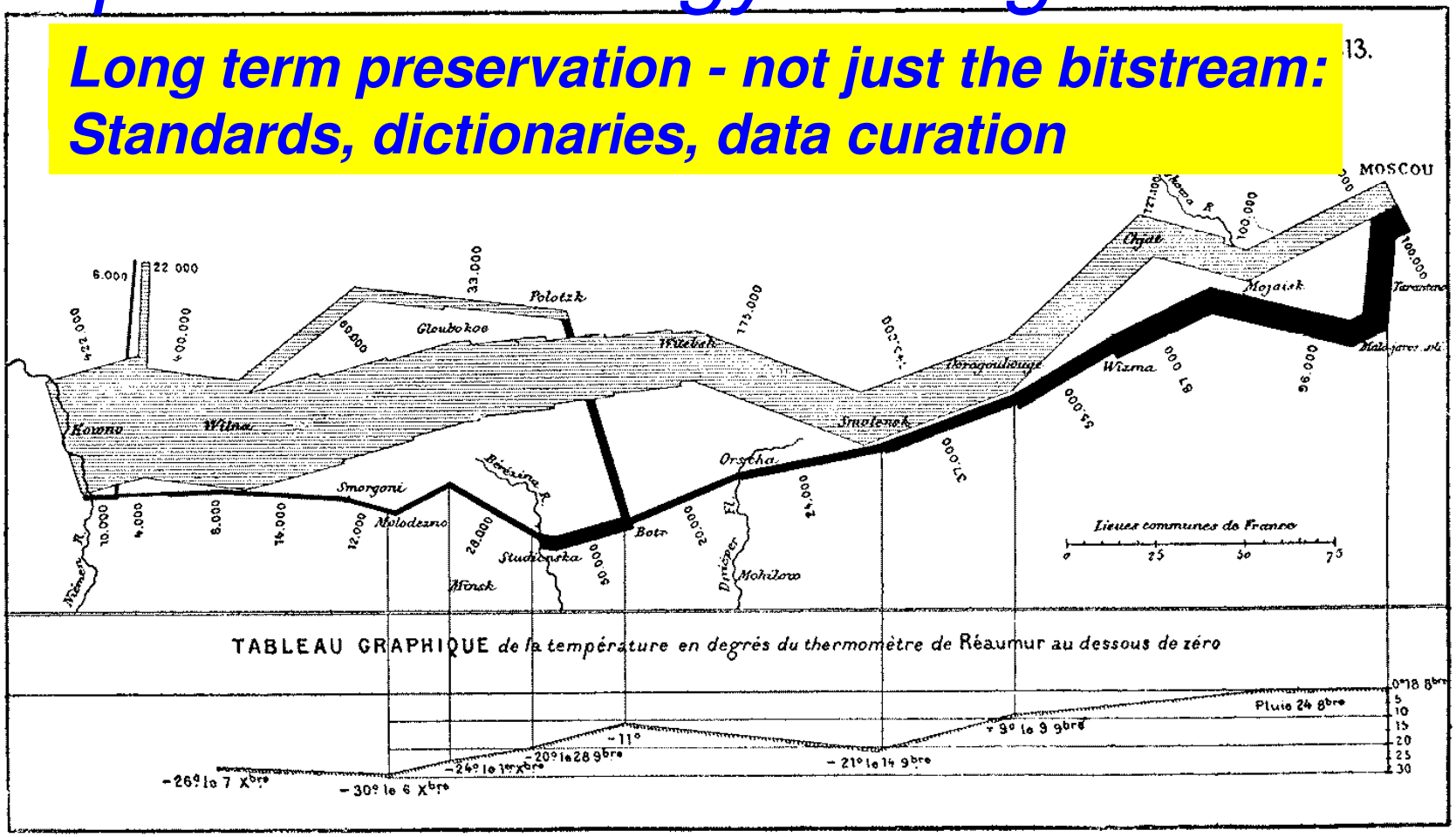
Event(s): Kingua_Fjord * *Latitude:* 66.5948 * *Longitude:* -67.3450 * *Elevation:* 10.6 m * *Date/Time:*




* ??
Kingua_Fjord
Nunatak Island
*Data ingest and curation needs dedicated and knowledgeable **people**!*
© 2007 Europa Technologies
Image © 2007 TerraMetrics
©2006 Google™

Expect terminology change over time

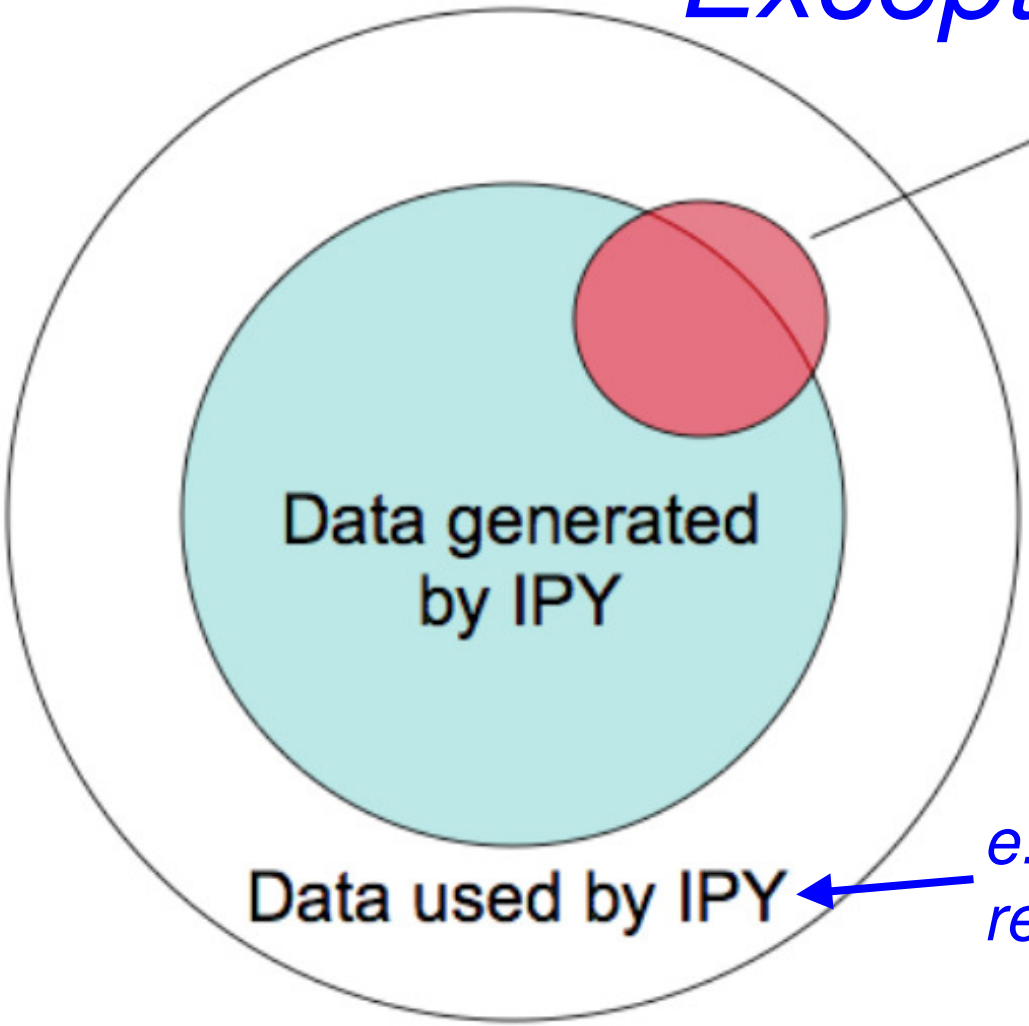
Long term preservation - not just the bitstream: Standards, dictionaries, data curation



Xbre = December 9bre = November 8bre = October

- *degree Reaumur = 0,8 x degree Celsius;*
- *1 Lieue commune de France = 4.452,2 m*
- *Wilna = Vilnius; Kowno = Kaunas*
- *9bre = Novembre !!*

Exceptions+Extensions



Special Cases:

- Human subjects
- Intellectual property of local and traditional knowledge
- Where data release may cause harm

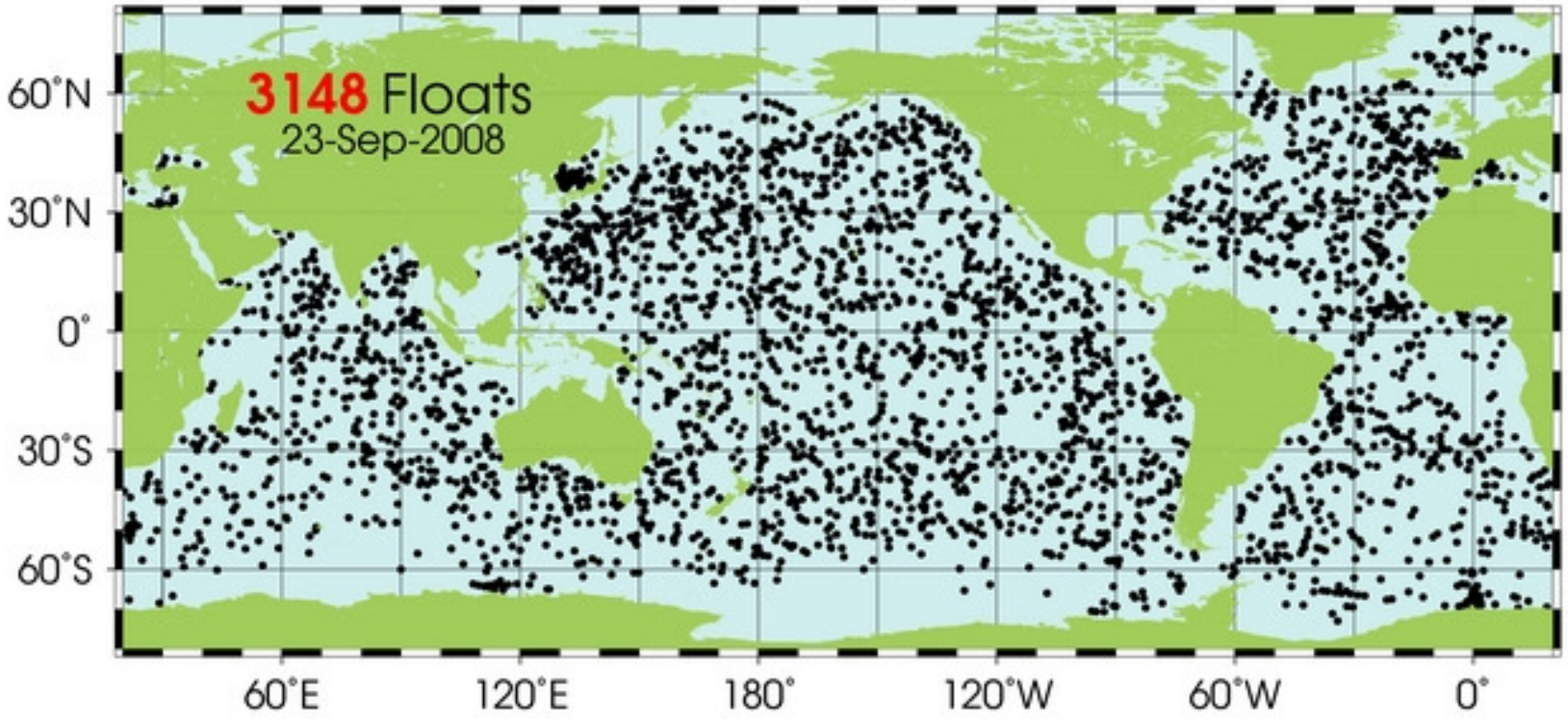
Some special cases may be present in the reindeer project cluster.

e.g.: Global atmospheric re-analysis data

Figure 1. Graphical definition of "IPY data" (inner blue circle), "IPY-related data" (outer circle), and special cases.

Existing repositories - meet the largest experiment on Earth

- *Hint: It is not CERN's LHC!*



6900499

NORWAY (Argo NORWAY)

880 Days

95 profiles at GDACs (origin Coriolis) including 0 DM profiles

Deployment Date: 13/04/2006 Lat : 64.6500 Lon: -.0216

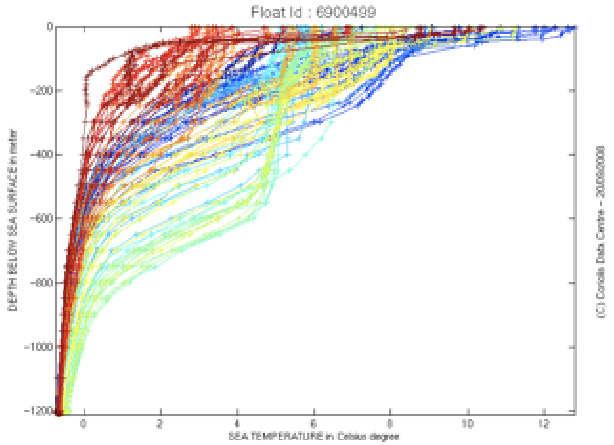
Latest Location Date: 09/09/2008 Lat: 67.0903 Lon: -9.0152

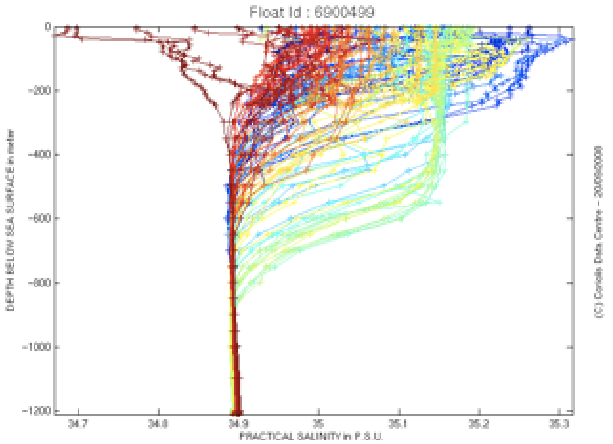
Web Products AIC Coriolis JMA CSIRO MEDS

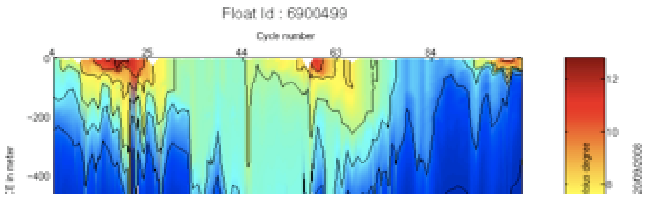
Data (netCDF) Profiles Metadata Trajectory Technical

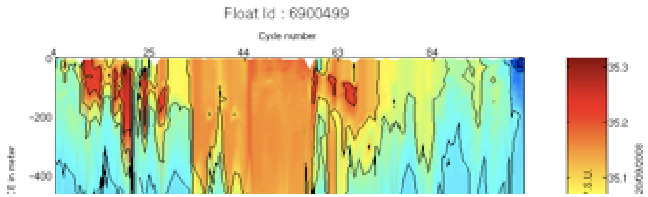
QC Altimetry QC

Subsurface Temperature - Subsurface Salinity (source IFREMER/Coriolis)









IPY poses technical challenges :

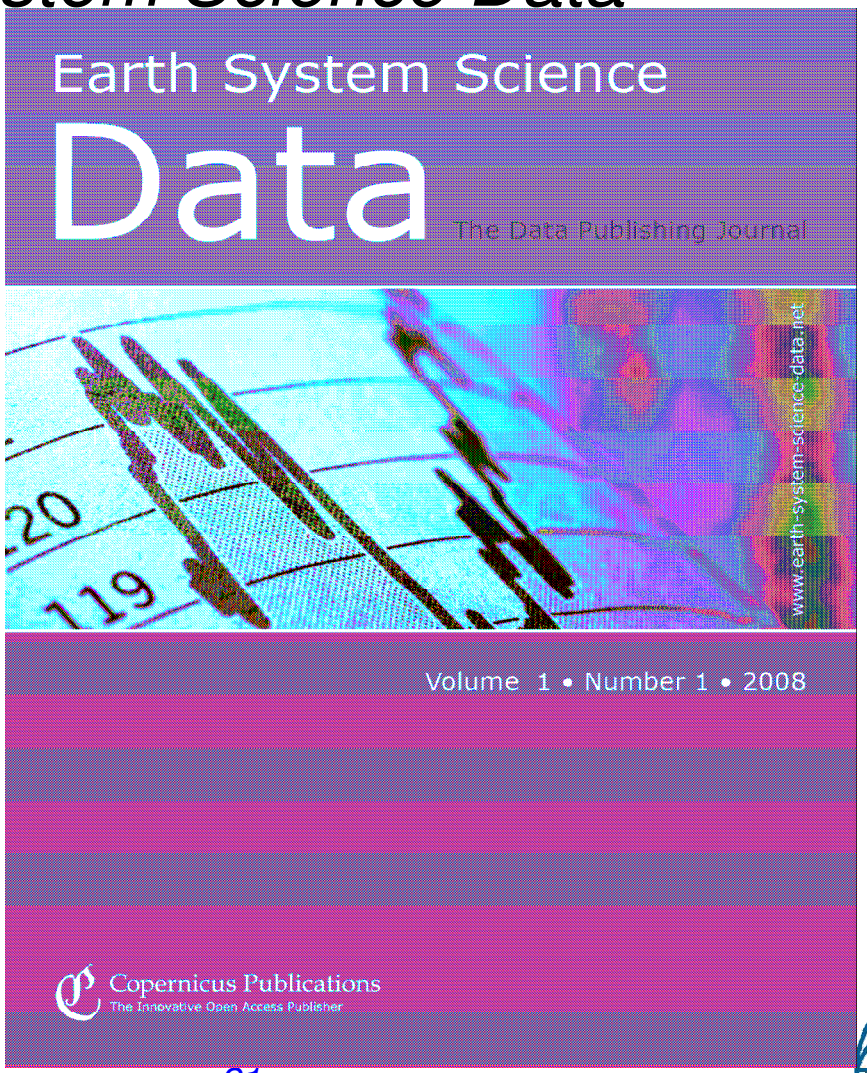
- *IPY data need to be identified and ingest into proper archives within a few years (capacities!)*
- *For many disciplines and in many countries, no commonly accepted practises and no certifyable repositories do exist* ***today***
- *There are* ***some*** *standards for discipline-spanning interoperability at the technical level, e.g.:*
  - **ISO 19115 (=> INSPIRE) metadata,**
  - **OAI harvesting, Open Geospatial Consortium access**
- *No trusted, scalable, long term system yet to deal with restrictions / rights on a global level (50.000 people, 63 nations, n*100 projects/VOs?)*

Need to use what we have (People!)

- *Weiter mit offene Fragen / Diskussion...*

Quality / Incentive

- *Data Publication: Earth System Science Data*
- *Addresses key problems:*
  - **missing reward**
  - **missing quality assurance**
- *Peer review*
- *Citable data*
- *to be on ISI list*



HELMHOLTZ GEMEINSCHAFT