Data Sharing Policies
The International Polar Year Case

Hans Pfeiffenberger
Alfred Wegener Institut / Helmholtz Gemeinschaft
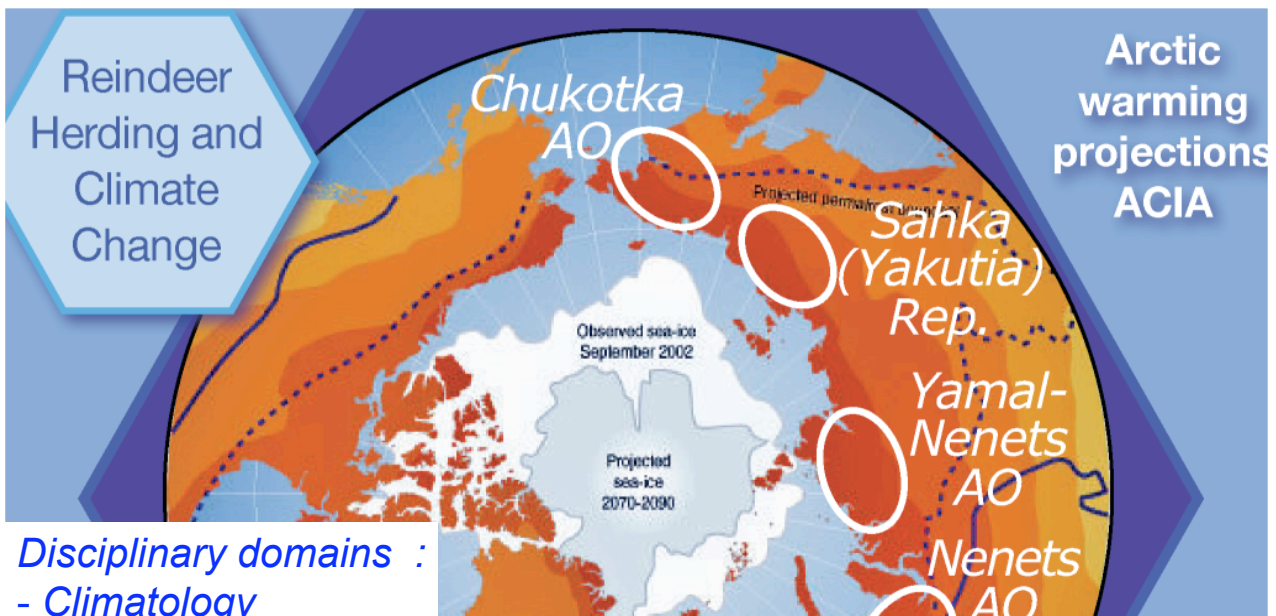
HELMHOLTZ
| GEMEINSCHAFT

Agenda

- *What is the IPY (and its mission)*
- *What are the challenges*

- *Official Data Policy of IPY*
- *Real World limitations and obstacles*

- *Implications and challenges for*
  - **repositories in general and**
  - **AAI matters in particular**

HELMHOLTZ
GEMEINSCHAFT

International Polar Year 2007-2008

Earth	Land	People	Ocean	Ice	Atmosphere	Space		Education & Outreach

50.000 Participants
63 Nations, ca. 1 G€
Mission: To take a data snapshot of the polar caps, for reuse in decades to come

Earth	Land	People	Ocean	Ice	Atmosphere	Space		Education & Outreach

---

Reindeer Herding and Climate Change

Chukotka AO

Sahka (Yakutia) Rep.

Yamal-Nenets AO

Nenets AO

Arctic warming projections ACIA

Observed sea-ice September 2002

Projected sea-ice 2070-2090

Projected permafrost ...
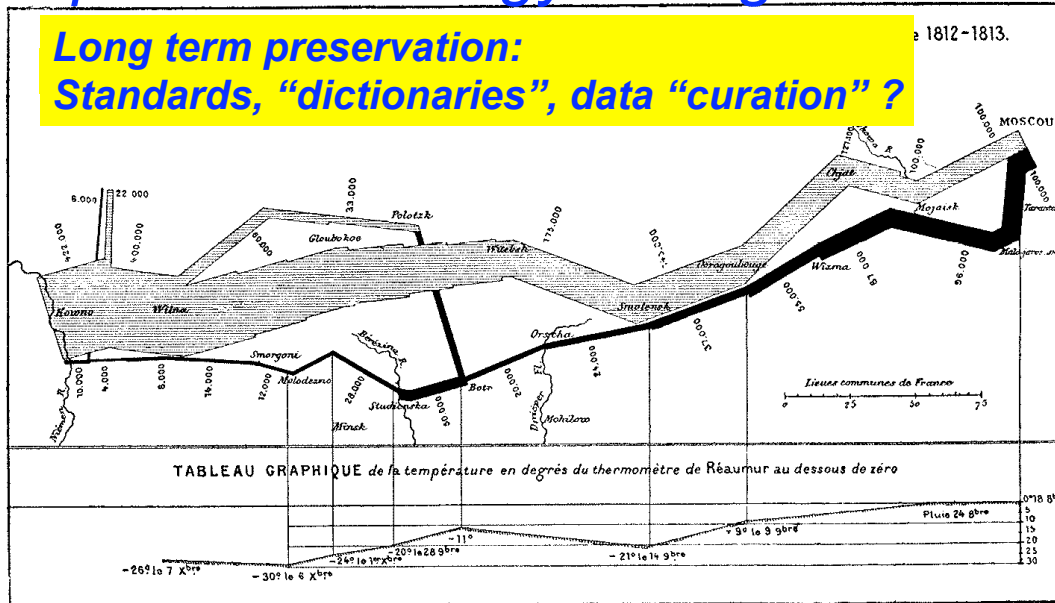
Disciplinary domains :
- Climatology
- Oceanography
- Veterinary medicine
- Ecology
- Sociology
- Geography
- and more ??

Interdisciplinary, global research :
- different cultures
- different languages
- different standards
Need for cooperation and inter-operability on an unpredescent scale

Expect terminology change over time

Long term preservation:
Standards, "dictionaries", data "curation" ?

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

Data policy of the IPY 2007/2008

- „In order *to be considered as officially part of IPY,*
each Project must follow the IPY 2007-2008 Data Policy"

- „... 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 ..."

Has all been said about policy ?

- Ideal (or lazy) World (with insignificant exceptions)
  IF

  > 80/20 rule is (naively) applied

  > **AND**

  > policy = practise is (naively) assumed

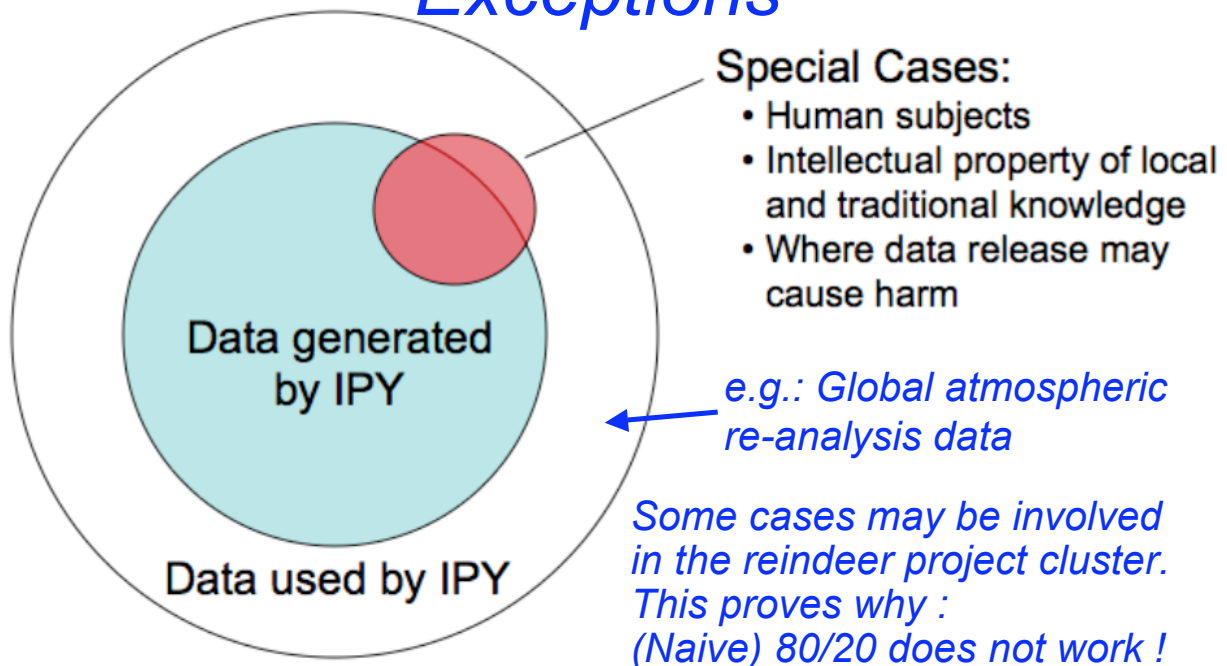 **THEN**

  > END of talk

- Real World
  ELSE

  > deal with exceptions (since they <u>are</u> significant)

  > **AND**

  > deal with objections, reservations, imperfections

Exceptions

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

e.g.: Global atmospheric re-analysis data

*Some cases may be involved in the reindeer project cluster. This proves why :
(Naive) 80/20 does not work !*

Data generated by IPY

Data used by IPY

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

Deal with exeptions

- *Data „owners" need trustworthy assertions, e.g.:*
  - **„I am doing non-commercial research" (ECMWF reanalysis)**
  - **„(S)he is a member of a trustworthy group" (nesting sites)**
  - **„(S)he needs to know" (sociological details)**
- *Repositories need to implement owners „policy"*
  - **Fine grained access rights („policy" decision/enforcement)**
  - **YES/NO or „selective availability" (taylored precision)**
- *Who are the autorities?*   *Mind the long term !!*
  - **The user-individual himself (license agreement)**
  - **The owner-individual/organisation herself/itself**
  - **A group (a project PI, a peer group, a learned society)**

Objections

- *„These are my data, I wish to exploit them exclusively for years..."*
- *„I would need to explain the data to a user anyway"*
- *„I am required to sell the data by my funding agency"*
- *....*

- *Except for the last argument, all other boil down to:*
  - **I have put much work into the data and**
  - **would need to put more work into it to make it re-usable.**
  - **Where is the compensation?**

More „Objections" to Open Access

- *Would like to have a contract ...*
  Granting offer of co-authorship if publishing about data

- *Would like each user to register,*
  be alerted about each download by registered users

- *Restrict access to*
  - **Personal use for one year**
  - **project (cluster) members for 3 years**
  - **public thereafter**

  - **This is the policy of the „Network for the Detection of Stratospheric Change" (15 reference observatories)**

HELMHOLTZ
| GEMEINSCHAFT

Deal with objections

- *The IPY Joint Committee has no means to enforce the data policy*

- *Build a culture of sharing (per discipline!)*
  - **„evangelism"**
  - **incentives : proper citation, data journals (CV!)**
- *Make repositories useful and reliable*
  - **high quality data !! (certification)**
  - **compatible with users' tools practises and pradigms (VO!)**
- *High level of scepticism of funders due to bad prior experience (junkyards or black holes)*

HELMHOLTZ
| GEMEINSCHAFT

Imperfections affecting IPY policy

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

Deal with imperfections

- *„An IPY Data and Information Service (IPYDIS http://ipydis.org) should help projects identify appropriate long-term archives and data centers ...“*
  - **IPYDIS is a framework for global cooperation to tackle IPY data challenge**
- *Unfortunately, this **IPY project (!!)** is un(der)funded*
  - **Realistic funding would be in low % range of research cost**
  - **Even given „realistic" money, too late to build capacity *during IPY* itself (but it is worth doing *for IPY*!)**
  - **However, some realistic national funding for national data (Canada), some for specific data types (remote sensing, meteorology) is available**
  - **Must work on a number of problems simultaneously, on a best effort basis, provide „glue", bridge gaps, ...**

Conclusion

- *The IPY data challenge is an opportunity*
- *The problem seems to be overwhelming :
  „deluge" of contexts, as well as data,*
- *but it __must__ be solved : Much is to be gained (or lost...)*
- *We must (and can) work on IPY data*
  - **pragmatically**
  - **addressing scientists / disciplines need & concerns**

- *IPY data as a whole **could be a proving ground** for all
  repository / data infrastructure concepts*
- ***IPY data will be a data treasure for future generations***

HELMHOLTZ
| GEMEINSCHAFT