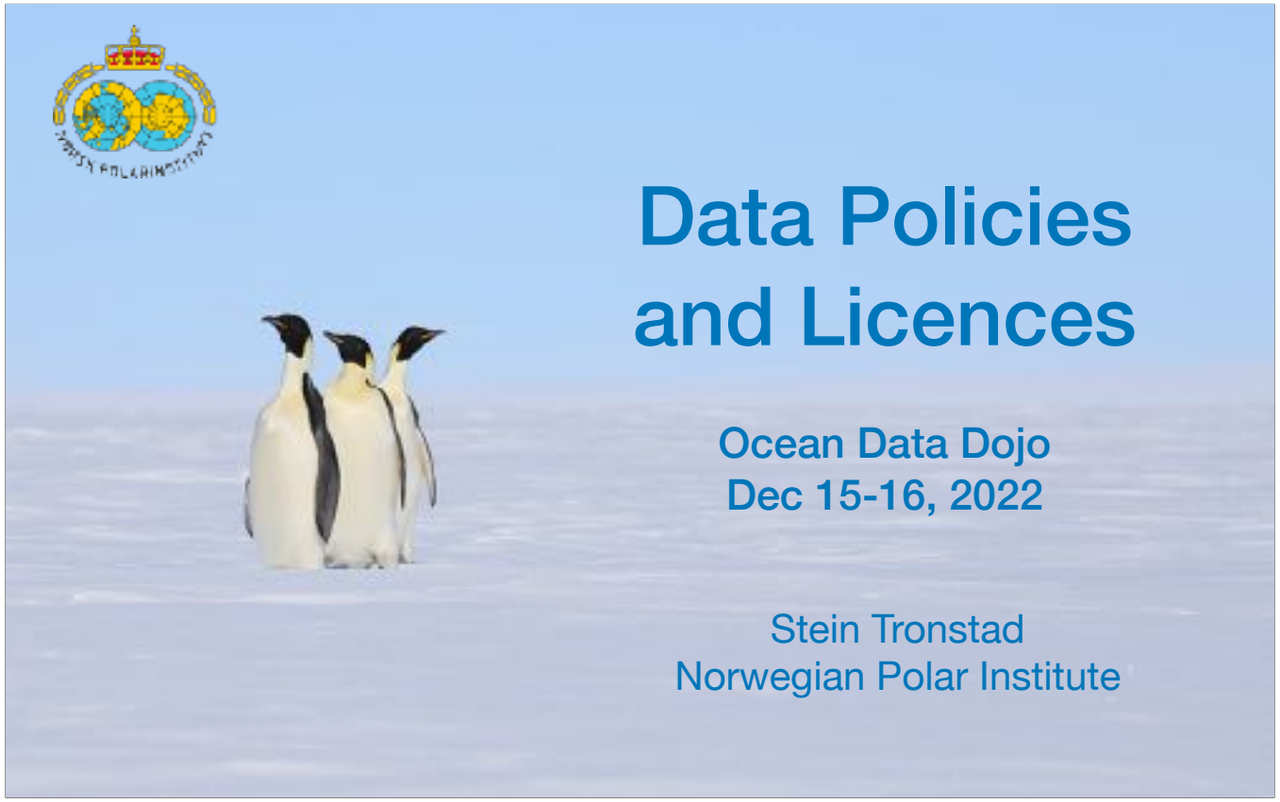
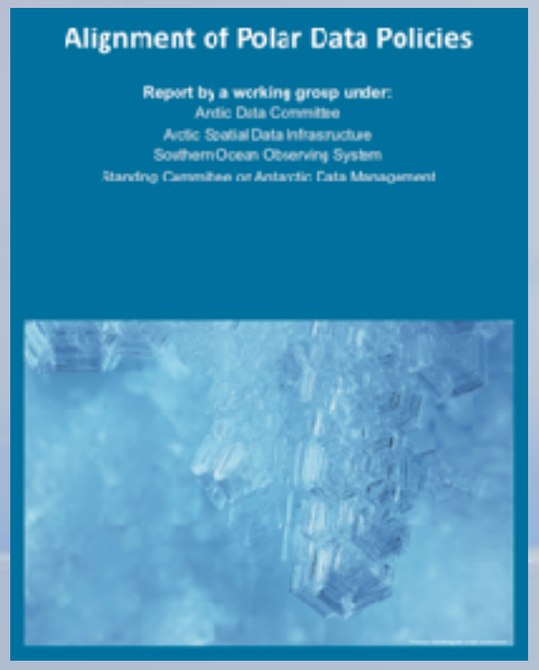


Data Policies and Licences

Ocean Data Dojo
Dec 15-16, 2022

Stein Tronstad
Norwegian Polar Institute



https://doi.org/10.5281/zenodo.5734900

*The process has been driven by a working group under IASC, SAON, SOOS, and Arctic SDI.
*A final report was published in November this year.
*The report recommends 10 fundamental principles for polar data policies.

Why Data Policies?

- Promote scientific cooperation
- Preserve scientific assets
- Transparency, reproducibility
- Efficiency, quality of science
- Scientific productivity
- Innovation

Public data are public assets, maximise their benefit to society
Access, sharing, reuse, publication, archiving, documentation…

Setting expectations among scientists:
* how and what data to share
* how to treat data shared by others.

Instruments for science coordinating bodies, funding agencies and research institutions
in optimising the societal benefits and scientific productivity of their investments in scientific data collection
— and strengthen the understanding of global change and challenges

«provide a framework for these data to be handled in a consistent manner, and to strike a balance between the rights of investigators, the rights of indigenous peoples, and the need for widespread access» (IPY DP)

…and highlight gaps in knowledge, induce innovation and the proliferation of ideas, and stimulate the search for new knowledge

The Legacy

- Antarctic Treaty (1961)
- Europe: PSI Directive (2003)
- IOC Oceanographic Data Exchange Policy (1999, 2003, 2019)
- ICSU (2004), CODATA, WDS
- GEO (2005)
- IPY Data Policy (1882, 2007)

Antarctic Treaty

”1. In order to promote international cooperation in scientific investigation in Antarctica, as provided for in Article II of the present Treaty, the Contracting Parties agree that, to the greatest extent feasible and practicable:

...

(c) scientific observations and results from Antarctica shall be exchanged and made freely available.”

Advent of the internet

PSI directive (2003); public sector information

- a framework of rules governing the reuse of existing documents held by the public sector

ICSU (ICS) (2004): Scientific Data and Information

Report of the CSPR Assessment Panel (Committee on Scientific Planning and Review)

- 56 recommendations (long-term strategic framework, DM, access, interoperability, IPR, …)
- continue to actively promote the principle of full and open access to scientific data
- WDS in 2008

GEO (2005), strategic set of data sharing principles:

• There will be full and open exchange of data, metadata and products shared within GEOSS, recognizing relevant international instruments, and national policies and legislation.

• All shared data, metadata, and products will be made available with minimum time delay and at minimum cost.
• All shared data, metadata, and products free of charge or no more than cost of reproduction will be encouraged for research and education.

OECD Principles and Guidelines for Access to Research Data from Public Funding (2007):

Scientific databases are rapidly becoming a crucial part of the infrastructure of the global science system.

Effective access to research data, in a responsible and efficient manner, is required to take full advantage of the new opportunities and benefits offered by ICTs

* Open access to research data from public funding should be easy, timely, user-friendly and preferably Internet-based.
* Openness, flexibility, transparency, legal conformity, protection of IPR…

IPY Data Policy:

* Made available fully, freely, openly, and on the shortest feasible timescale.
* Full set of metadata
* Long-term preservation
* Data acknowledgment, formal citations

Policies Examined

- International Science Council (ICSU, CODATA, WDS 2004 - 15)
- UNESCO Rec. on Open Science (2021)
- WMO, World Meteorological Organisation (2021)
- IOC, Intergovernmental Oceanographic Commission (2019)
- OECD (2006)
- Group on Earth Observations, GEO (2015)
- Antarctic Treaty (1961)
- Arctic Council (2017)
- EU (-2019, 2020)

The report starts out by examining overarching data policies and recommendations of relevant global and regional organisations, some of them quite recent.

(Arctic Council = Agreement on Enhancing International Arctic Scientific Cooperation (2017))

Other Policy Drivers

- The drive towards open data
- Limits to openness and timeliness
- The ‘data deluge’, ‘big data’ and cloud computing
- New cost models and big data infrastructure
- The FAIR, CARE, and TRUST principles
- Ethical considerations, responsible reuse
- Demand for transparency and trust in science
- Legal instruments for data sharing
- Metrics (scientific impact)

As a basis for recommendation, the report also examines a number of external data policy drivers.

TRUST = Transparency, Responsibility, User focus, Sustainability, and Technology
CARE = Collective benefit, Authority to control, Responsibility, Ethics
FAIR = Findable, Accessible, Interoperable, Reusable

Shared Elements

- «Open by default and design»
- «Ethically open» or «As open as possible, as closed as necessary»
- «Free», «timely», «unrestricted»
- From Findable, Accessible to FAIR
- Data management plans
  - Curation, preservation, sustained access
- IPR recognition
  - Acknowledge and be acknowledged
- New legal instruments, label as reusable

Machine readability or linked data?
Ethically open: privacy, safety, security, environment protection, ethics

Recommended Core Principles 1-5

1. Data must be ethically open
2. Data should be free
3. Data must be provided in a timely manner
4. FAIR principles should be applied to the greatest extent practicable
5. All data must be accompanied by a complete set of metadata

The report concludes by recommending 10 fundamental principles for aligned polar data policies.

Rec. 5 - 7 are elements of the FAIR principles, but stated explicitly because they are important even if data cannot be made fully FAIR.

Recommended Core Principles 6-10

6. Data should have persistent and globally unique identifiers
7. Data must be labelled as reusable
8. Data should be attributable and attributed
9. Data must be appropriately preserved for the long term
10. Data management and long-term curation must be planned and resourced

Processes have already started to form new or revised data policies for the Arctic Council and for SCAR based on the 10 recommended principles.

ADC, SAON, IASC should follow. How?

For the Arctic Council a milestone was achieved during the last Arctic Council Ministerial in 2021 (ext slide).

Aligned Data Policies

- Scientific Committee on Antarctic Research (2022)
- Southern Ocean Observing System (2022)
- Arctic??
- IOC?

Arctic Council Ministerial 2021

«We, the Ministers representing the eight Arctic States, …

52
Recognize that adequately responding to rapid changes in the Arctic environment requires access to sustained observational networks and reliable data to facilitate understanding and informed decision making,
welcome progress on implementing guiding principles on management of and access to data and
facilitate that data, generated by the Council, is findable, accessible, interoperable, reusable, and widely shared …»

Excerpt from the Declaration of the Foreign Ministers of the Arctic States at the 12th Ministerial meeting of the Arctic Council, held in Reykjavik, Iceland, 20 May, 2021.

For the Norwegians

- Meld. St. 22 (2020–2021): Data som ressurs
  - «National principles for data sharing and use»
- Retningslinjer ved tilgjengeliggjøring av offentlige data (2017)
  - Open standard licences, Free, No login walls, Visible, Updated…
- Nasjonal strategi for tilgjengeliggjøring og deling av forskningsdata (2017)
  - «Open when possible, protected when necessary»; Promote reuse
- Forskningsrådets policy for åpen forskning (2020)
  - FAIR, International standards, Citation, Budget DM, Provide source code…

+ Individual institutions' policies
+ Digitaliseringsstrategi for offentlig sektor, rammeverk for digital samhandling m.m.

2. Data skal åpnes når de kan, og skjermes når de må.
   Open when possible, protected when necessary
2. Data bør være tilgjengelige, gjenfinnbare, mulige å bruke, og kunne sammenstilles med andre data.
   FAIR
3. Data skal deles og brukes på en måte som gir verdi for næringslivet, offentlig sektor og samfunnet.
   Share and use for added value to industry, public sector and society
4. Data skal deles og brukes slik at grunnleggende rettigheter og friheter respekteres, og norske samfunnsverdier bevares.
   Respect fundamental rights and freedoms and protect societal values

Legal Interoperability

Report and recommendations on licensing and sharing of research data (RCN 2021)

Legal Interoperability

- Facilitate lawful access and reuse
- Determine rights and responsibilities
- Balance legal interests
- Clear and transparent rights declarations
- Harmonised data reuse rights
- Appropriate attribution and credits

RDA/CODATA Legal Interoperability WG,

1. Facilitate the lawful access to and reuse of research data
2. Determine the rights to and responsibilities for the data
3. Balance the legal interests
4. State the rights transparently and clearly
5. Promote the harmonization of rights in research data
6. Provide proper attribution and credit for research data

Legal Interoperability

- Legal constraints on interoperability
  - Privacy (GDPR), security, sensitivity
  - Patents, trade secrets,
  - IPR (IPR Act § 2), legal database protection (IPR act § 24)
- FAIR R 1.1: (Meta)data are released with a clear and accessible data usage license
  - «Label as reusable»

(Developments since Wilkinson & al. 2016; data licencing a recent tool)

FAIR:
- What usage rights do you attach to your data?
- Ambiguity could severely limit the reuse by organisations that struggle to comply with licensing restrictions. Clarity of licensing status will become more important with automated searches involving more licensing considerations.
- The conditions under which the data can be used should be clear to machines and humans.

Licences and Waivers

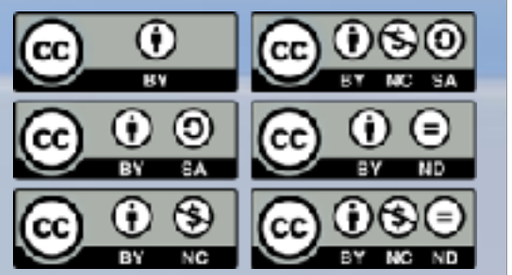
- Data licence (CC BY, NLOD etc): A reuse permit stipulating the conditions for reuse
  - Required if data have legal (rights) protection («label as reusable»)
  - Seen as a way to ensure reuse according to ethical norms?
- Rights waiver (CC0): Label as reusable even if IPR exist

CC0 is not equal to PDM

Purpose?

- Demand attribution?
- Renounce responsibility for reuse?
- Deny or require identication of changes?
- Require non-commercial reuse?
- Limit derivatives?
- Anything else?

«Open and freely accessible»?



Attribution:
Forskningsetikkloven §4 / Act on ethics and integrity in research:
* Forskere skal opptre med aktsomhet for å sikre at all forskning skjer i henhold til anerkjente forskningsetiske normer.
* Researchers must act with care to ensure that all research takes place in accordance with recognised ethical norms for scientific research.

Forskningsetiske retningslinjer/guidelines for ethical science
* Forskere skal følge god henvisningsskikk som sikrer krav til etterprøvbarhet og gir grunnlag for videre forskning.
* Proper referencing to ensure reproducibility and allow further research

Licensing Recommendations

- The licence terms should agree with the permitted and intended data reuse
- Research data should have international standard licences that:
  - have as few access and purpose constraints as possible
  - promote the principles of legal interoperability for research data
  - are both human and machine readable
- Metadata should always have a rights waiver (CC0, PDM)
- Licences protecting creative copyrights should only be used for data satsifying criteria for copyright or database protection

CC:
The licence terms and conditions apply to the database structure (its selecction and arrangement, to the extent copyrightable), its contents (if copyrightable), and in those instances where the database maker has SUI GENERIS DATABASE RIGHTS, to the rights that are granted those makers. It is possible for licensors to licence some rather than all of the rights they have in a database.
Creative Commons advises against this practice.

CC BY, CC0 (v. 4.0 universal, no need for porting)
NLOD
ODC-By

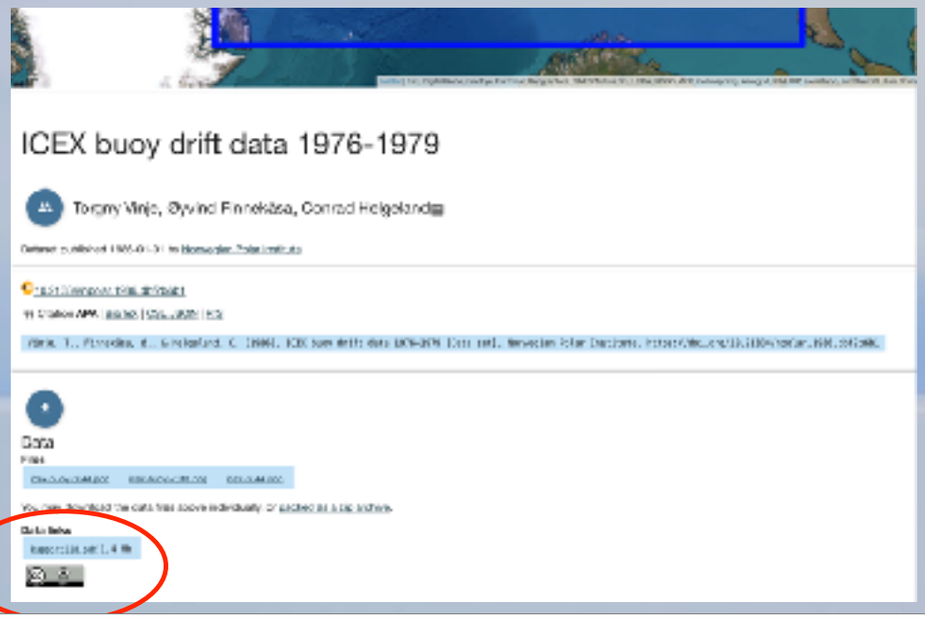
Questions

- Do you own or have the authority to licence the data?
- Will the licence unnecessarily constrain data access?
- Will the licence unnecessarily constrain the intended reuse?
  - E.g. by attribution stacking?
- Is the licence appropriate under the actual IPR?

«Bruk av CC-lisenser på data uten opphavsrettslig vern som et virkemiddel for å oppnå kildehenvisning og sitering, bidrar til forvirring rundt hva som er vernet av loven og hva lisensgiveren ønsker å oppnå med lisensen. Slik bruk bør derfor unngås.» (s. 31)

«Der forskningsdataene ikke er underlagt opphavsrett, men allikevel utstyrt med en CC-lisens, er dette problematisk og forvirrende for den som ønsker å ta i bruk forskningsdataene. Vilkår om navngivelse er også strengere enn det som er god henvisningsskikk.» (s. 37)

Applying a Licence or Waiver



Machine-readable, in metadata