Data Licences, Use-Tracking, Repositories

M. Fiebig
NILU – Norsk Institutt for Luftforskning
O. Godøy
Norwegian Meteorological Institute

NorDataNet
Norwegian Scientific Data Network

Intellectual Property (IP)

"Intellectual property refers to creations of the mind, such as inventions; literary and artistic works; designs; and symbols, names and images used in commerce."

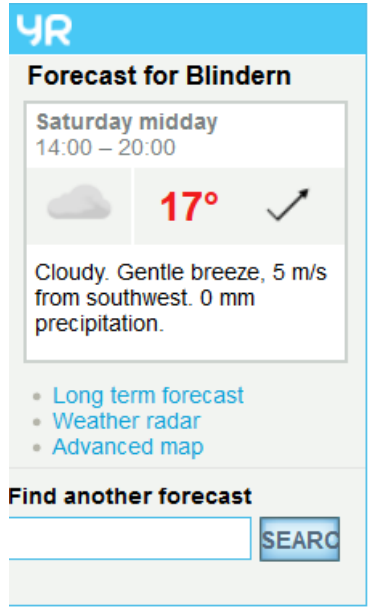
-World Intellectual Property Organization

Photo Attribution:
https://theforevergreenstateofmind.files.wordpress.com/2012/07/lightbulb3.png

Fact and Copyright

- Facts cannot be protected by copyright!
  - For example, temperature at Blindern is
- "Databases" / Data collection can be copyright protected
  - If it contains additional value / documentation / annotations

Photo Attribution:
http://www.romakerlaw.com/blog/wp-content/uploads/2015/07/doctelportal.jpg



A Few Terms in the Start

- **Data Licence:**
  A **licence** is an official permission or permit to do, use, or own something (as well as the document of that permission or permit). (Cambridge English Dictionary)
- **Data Policy:**
  - a specified course of action adopted for the sake of expediency, facility, etc.
  - an ethical statement on a dataset should be used.

- Data licences and policies live in different realms (legal, ethical)
- They can be combined.

NorDataNet
Norwegian Scientific Data Network

Why Use a Data Licence?

For data provider:

- Legally binding.
- Needs to be issued by data owner (institute, sponsor, …)
- Specifies exactly what is allowed to be done with the data.
- Can require attribution (of data provider).

❑ Data ownership needs to be confirmed!!!

For data user:

- Certain that data use is lawful.
- Knows the exact conditions for data use.

❑ Usually not the principal investigator

NorDataNet
Norwegian Scientific Data Network

Identifying Ownership and Control

- Funder agreement
- Institutional intellectual property policy
- Private company data use agreement
- Attached license
- For NRC and EU projects:
  data ownded by contract partner (RPO / institute).

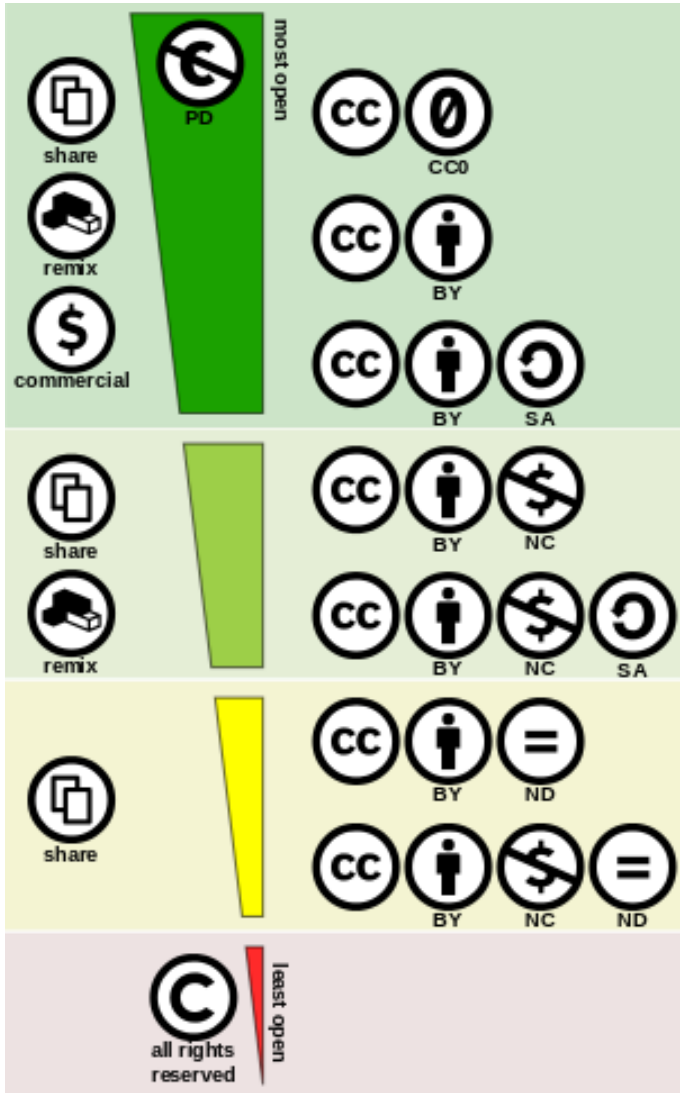
Photo Attribution: http://cdn.gobankingrates.com/wp-content/uploads/nashville-housing-market-overview-300x238.jpg

Creative Commons (CC) Licences

- Invented for sharing content with others.
- Giving both sides legal certainty (liability waived, use conditions specified.
- Original version of December 2002.
- Latest version 4.0 of November 2013
- Mature and widely used.

NorDataNet
Norwegian Scientific Data Network

CC licence spectrum



most open

least open

Commonly used are 7 CC licences (combination of conditions)

Most open: **CC0** (public domain waiver)

Least open: „all rights reserved“ (copyright owner hold all rights for own use)

- **CC-BY-NC-ND**: share only, credits required, non commercial use, no modification
- **CC BY-NC**: credits required, allows to share and remix, non-commercial only
- **CC BY-SA**: credits required, allows to share and remix, same legal conditions
- **CC BY 4.0** (Attribution 4.0 International): latest version of CC licences, applicable to most jurisdictions: allows to
  - to copy and redistribute the material in any medium or format,
  - Remix, transform, and build upon the material for any purpose, even commercially.

NorDataNet
Norwegian Scientific Data Network

Norwegian Licence for Open Government Data (NLOD)

- Developed in parallel with CC licences.
- Refers to Norwegian law.
- Recommended for Norwegian public data.
- Latest published version (landing page): https://data.norge.no/nlod/en
- Latest version compatible with, for practical purposes identical with, CC-BY 4.0.

NorDataNet
Norwegian Scientific Data Network

The full slide deck may be downloaded from:

http://www.dataone.org/education-modules

Suggested citation:

DataONE Education Module: Metadata. DataONE. Retrieved Nov12, 2016. From http://www.dataone.org/sites/all/documents/L07_Metadata.pptx

Copyright license information:

Data Policies, WMO Global Atmosphere Watch Example

"For scientific purposes, access to GAW data is unlimited and provided without charge. By their use you accept that an offer of co-authorship will be made through personal contact with the data providers or owners whenever substantial use is made of their data. In all cases, an acknowledgment must be made to the data providers or owners and to the data centre when these data are used within a publication."

- "Fair-use" data policy
- Grey-zone "substantial use": co-authorship if arcticle wouldn't be possible without that data
- Don't underestimate the value of and the work needed for good data.
- Data providers depend on attribution!!!
- In this case, it is always a good idea to contact data providers to reach agreement. Data providers often have additional information despite metadata.

NorDataNet
Norwegian Scientific Data Network

How to Cite Data

Similar to citing a published article or book

- Provide information necessary to identify and locate the work cited

Broadly-applicable data citation Standards have not yet been established; use standards adopted by relevant academic journal, data repository, or professional organization

- ❑ Data needs to be identified by persistent identifier (PID)
- ❑ Many journals require data used in article to be accessible by PID / DOI.
- ❑ Use repository that issues DOI.

NorDataNet
Norwegian Scientific Data Network

What are examples of Persistent Identifiers?

A **persistent identifier** should be included in the citation:

- DOI (Digital Object Identifier)
  - Globally unique, alphanumeric string assigned by a registration agency to identify content and provide a persistent link to its location.
  - May be assigned to any item of intellectual property that is defined by structured metadata
  - Examples:10.1234/NP5678, 10.5678/ISBN-0-7645-4889-4; 10.2224/2004-10-ISO-DOI

- Researcher identifier: ORCID (Open Researcher & Contributor ID)
  - Central registry of unique identifiers for individual researchers to address author name ambiguity
  - Transparent linking mechanism between ORCID and other author ID schemes

NorDataNet
Norwegian Scientific Data Network

Examples of information needed in a citation

Author/Principal Investigator/Data Creator

Release Date/Year of Publication – year of release, for a completed dataset

Title of Data Source – formal title of the dataset

Version/Edition Number – the version of the dataset used in the study

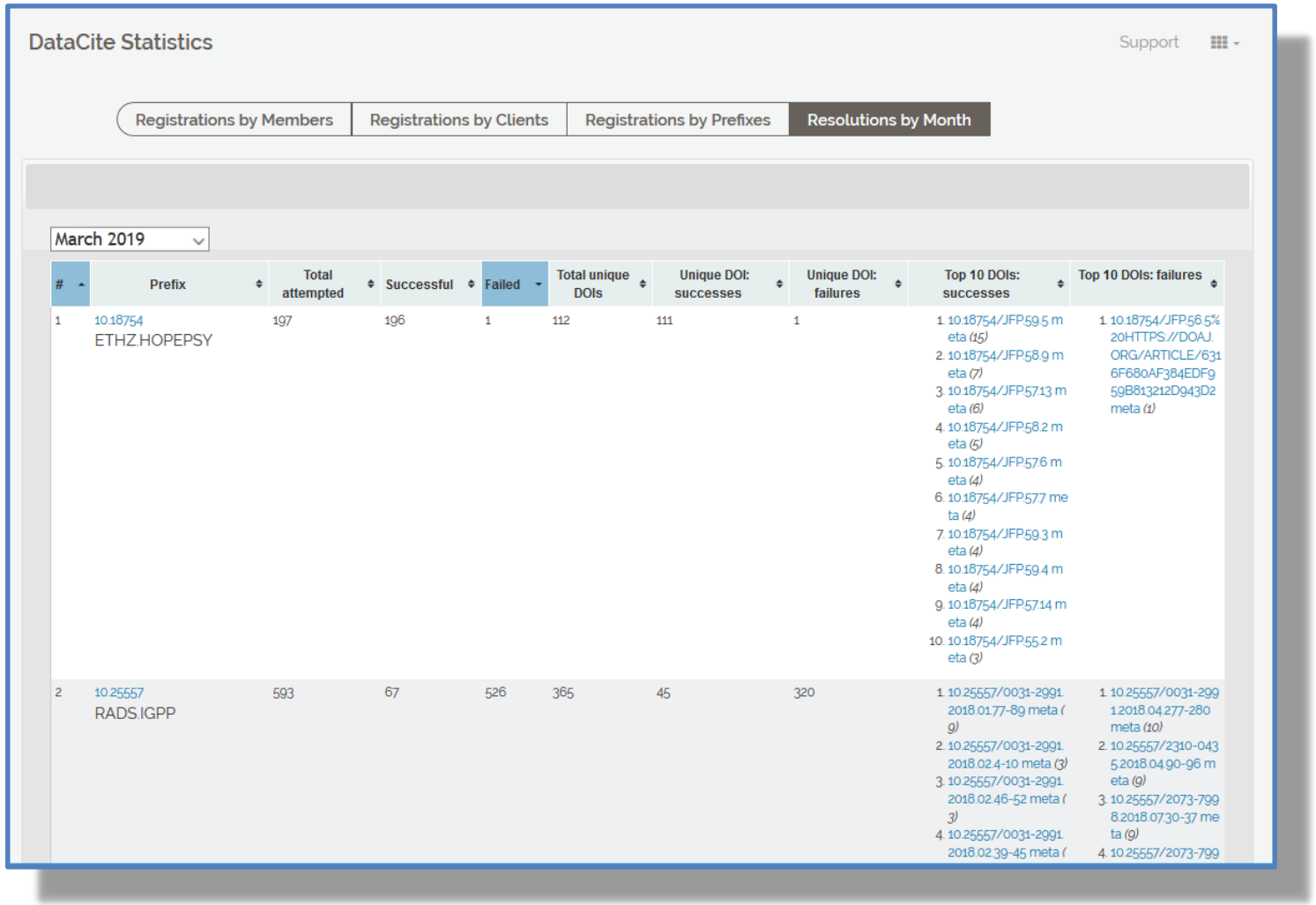
Format of the Data – physical format of the data

3rd Party Data Producer – refers to data accessed from a 3rd party repository

Archive and/or Distributor – the location that holds the dataset

- A well-chosen repository creates the citation metadata for you (from the metadata you provided)

DataCite: Metrics on Data Use


DataCite Statistics

Support

Registrations by Members | Registrations by Clients | Registrations by Prefixes | Resolutions by Month

March 2019

#	Prefix	Total attempted	Successful	Failed	Total unique DOIs	Unique DOI: successes	Unique DOI: failures	Top 10 DOIs: successes	Top 10 DOIs: failures
1	10.18754 ETHZ.HOPEPSY	197	196	1	112	111	1	1. 10.18754/JFP.59.5 meta (15) 2. 10.18754/JFP.58.9 meta (7) 3. 10.18754/JFP.57.13 meta (6) 4. 10.18754/JFP.58.2 meta (5) 5. 10.18754/JFP.57.6 meta (4) 6. 10.18754/JFP.57.7 meta (4) 7. 10.18754/JFP.59.3 meta (4) 8. 10.18754/JFP.59.4 meta (4) 9. 10.18754/JFP.57.14 meta (4) 10. 10.18754/JFP.55.2 meta (3)	1. 10.18754/JFP.56.5%20HTTPS://DOAJ.ORG/ARTICLE/6316F680AF384EDF959B813212D943D2 meta (1)
2	10.25557 RADS.IGPP	593	67	526	365	45	320	1. 10.25557/0031-2991.2018.0177-89 meta (9) 2. 10.25557/0031-2991.2018.024-10 meta (3) 3. 10.25557/0031-2991.2018.0246-52 meta (3) 4. 10.25557/0031-2991.2018.0239-45 meta (	1. 10.25557/0031-2991.2018.04277-280 meta (10) 2. 10.25557/2310-0435.2018.0490-96 meta (9) 3. 10.25557/2073-7998.2018.0730-37 meta (9) 4. 10.25557/2073-799

Currently:

- Statistics how often DOIs were resolved

Future:

- Statistics by author, funding agency, ...
- Accessible by API

- ❑ Only counts DOI resolutions.
- ❑ Statistics on data citations still pending

NorDataNet
Norwegian Scientific Data Network

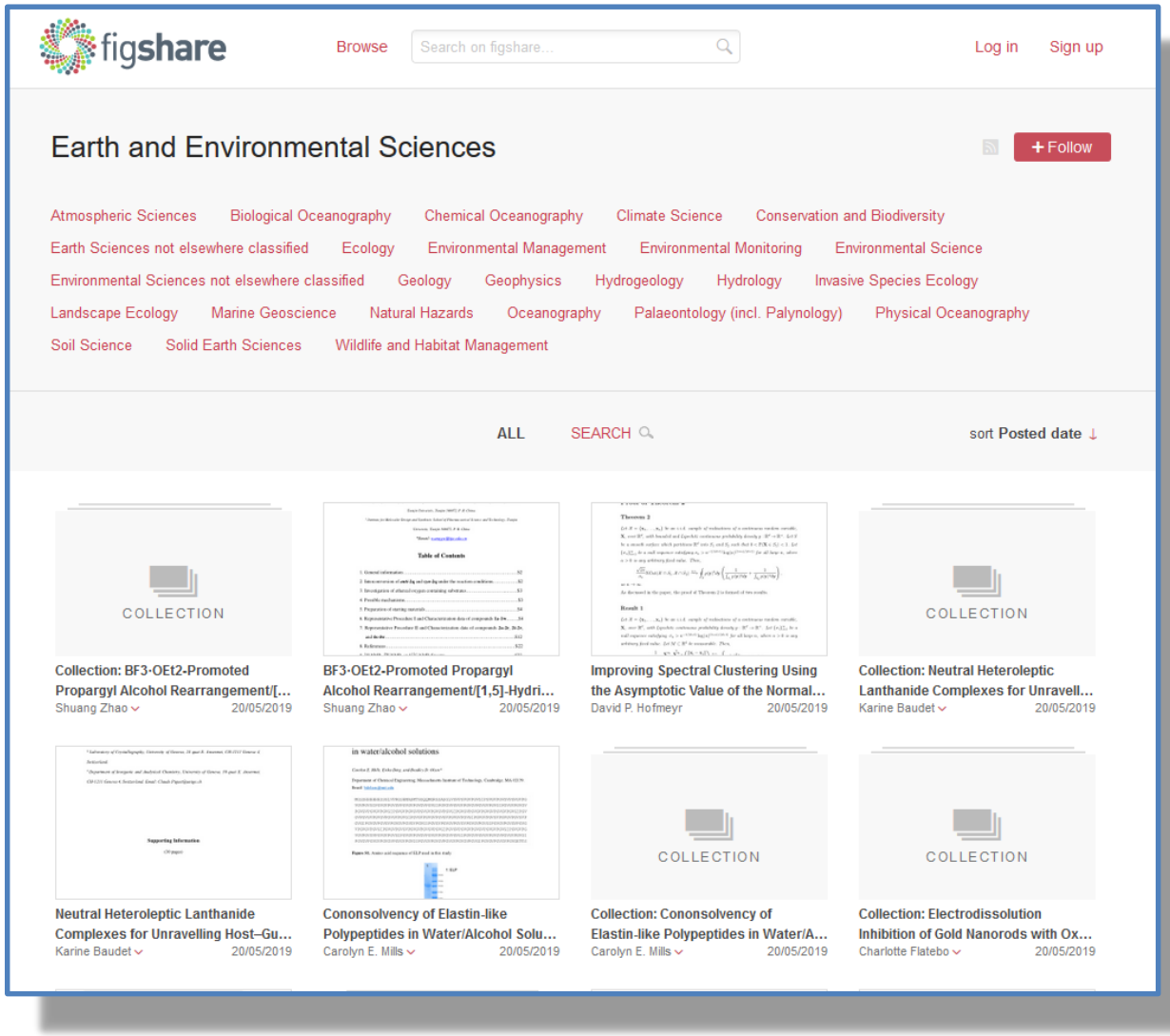
Which Repository to Choose

- **Data Type**
  - **Primary Data:** original dataset that hasn't been published before. Requires domain specific metadata. Often domain specific repository.
  - **Secondary Data**: dataset assembled / produced from other dataset(s). Requires provenance to be included. Can be generic repository.
- **Domain:**
  - Each domain has domain specific metadata (discovery, use)
  - Domain specific repositories have expertise in "their" vocabulary, metadata standards, observation methods, quality control

- ❑ Choosing the right repository solves issues for you.
- ❑ Domain specific repo has better FAIRness.

NorDataNet
Norwegian Scientific Data Network

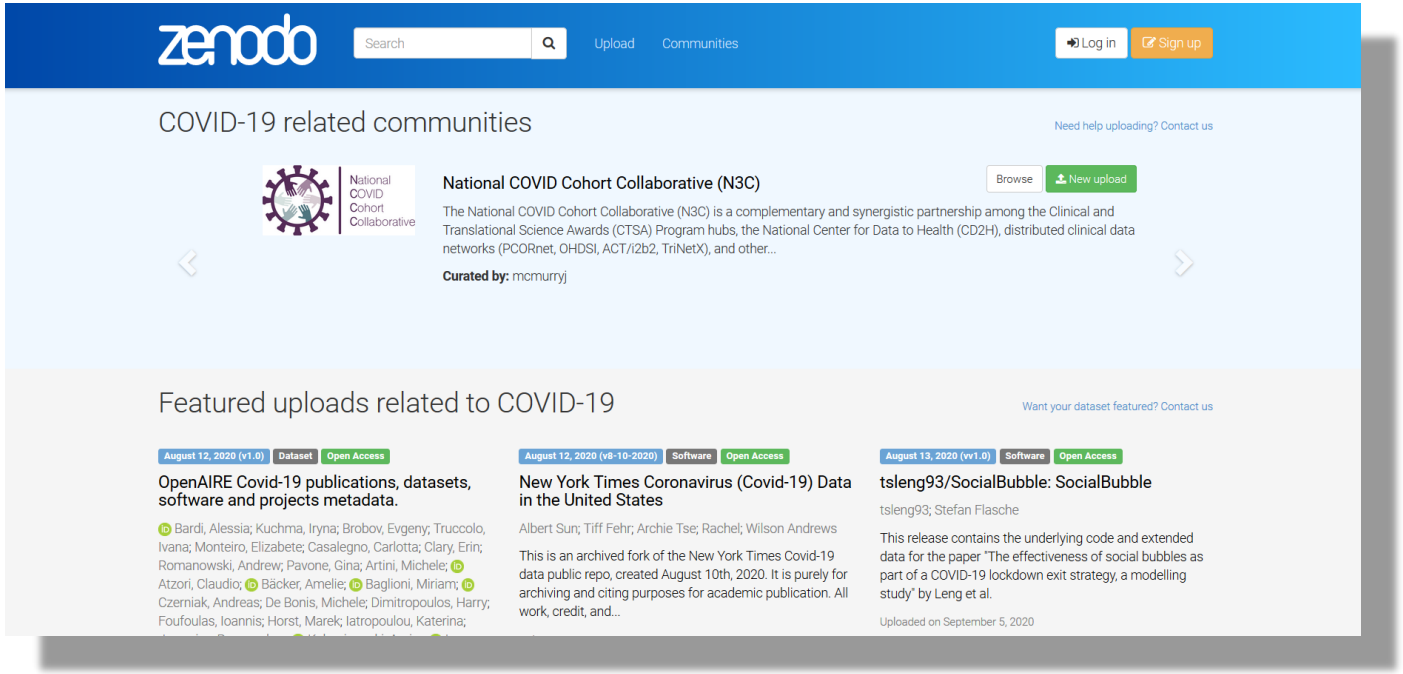
Generic Data Publishers: Figshare



- Commercial
- Generic
- Archiving
- Identification (DOI)
- Semantic search (limits re-usability)
- Suited for secondary datasets

NorDataNet
Norwegian Scientific Data Network

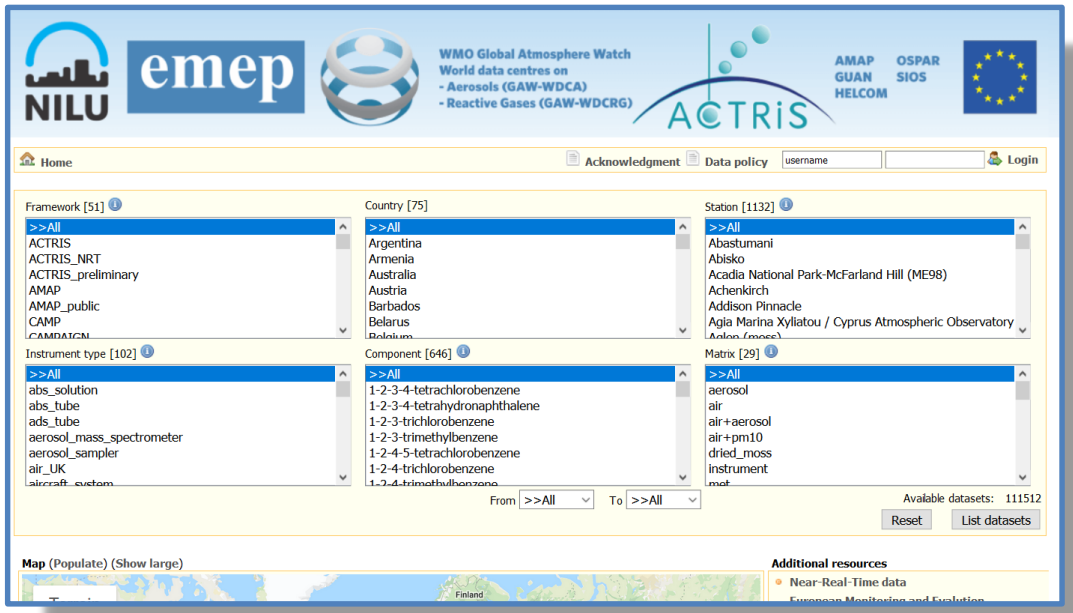
Generic Data Publishers: Zenodo



- https://zenodo.org/
- Operated by CERN.
- Recommended by EU commission.
- Generic
- Archiving
- Identification (DOI)
- Semantic search (limits re-usability).
- Suited for secondary datasets.

NorDataNet
Norwegian Scientific Data Network

Domain Archive: Atmospheric Observations



WMO Global Atmosphere Watch (GAW) topic data centres:

- **Precipitation chemistry**
- **Ozone & UV**
- **Greenhouse Gases**
- **Reactive gases**
- **Aerosol**
- **Radiation**
- **Remote Sensing**

- **Better Re-Usability**
- Archiving
- Identification (DOI)
- Structured search, high re-usability, specially for expert user
- Easy to narrow search exactly to targeted data.
- NorDataNet offers topical archives optimised for data types, their (use) metadata, and their user community

NorDataNet
Norwegian Scientific Data Network

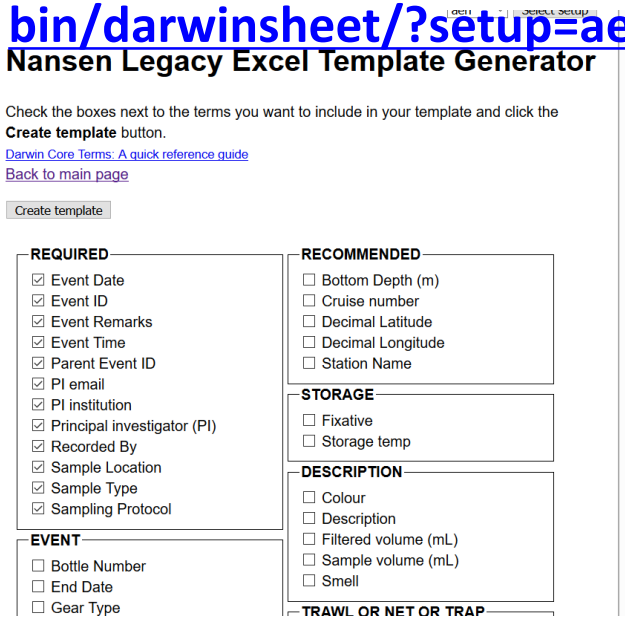
Domain Archive: NorDataNet Archives

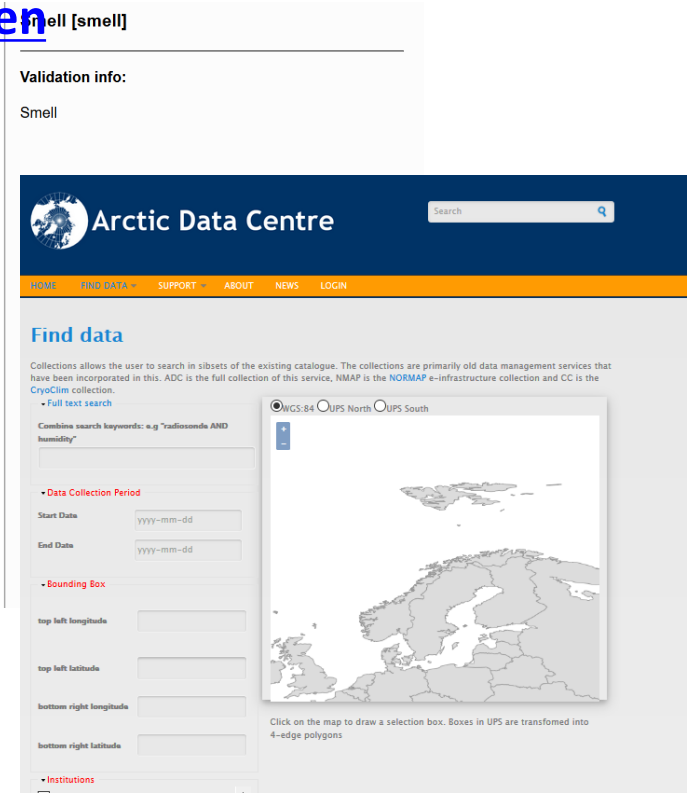
Norwegian Marine Data Centre

https://www.nmdc.no

NorDataNet Nansen Legacy Tools

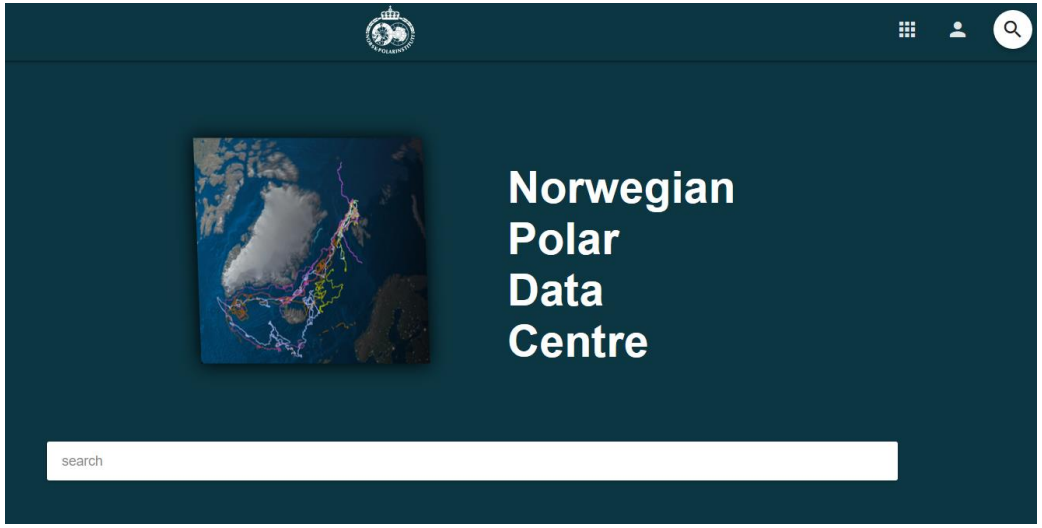
https://www.nordatanet.no/cgi-bin/darwinsheet/?setup=aen





Arctic Data Centre

https://adc.met.no/

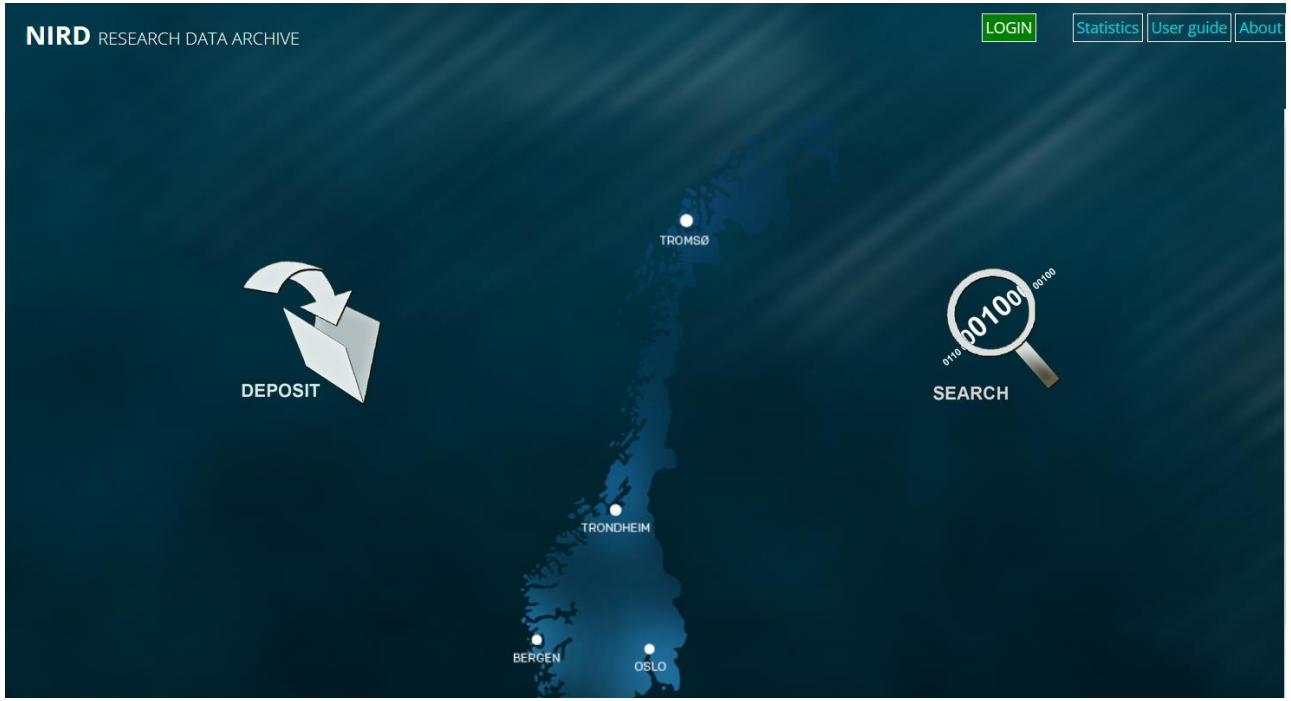


Norwegian Polar Data Centre

https://data.npolar.no/home/

NorDataNet
Norwegian Scientific Data Network

NIRD Research Data Arvive



- https://archive.sigma2.no/
- General purpose archive
- Identification (DOI)
- Suited for large datasets as generated by geoscientic models
- Well-suited for model data.
- Recommended by UiO.
- Metadata entered manually
- Agnostic to data file content.

NorDataNet
Norwegian Scientific Data Network