

WP4: Data Forum

Øystein Godøy, Anna Irrgang

Assumptions

- INTERACT research stations generate data and metadata
 - Long term monitoring
 - Short term process studies
 - External data by individual scientists/ groups
- Research stations archive data and metadata (internal and external)
 - e.g. meteorological data
 - photos, maps, reports etc.
 - list of data acquired at the stations
 - information on data collection procedures (field diaries)

Objectives of the work

- Analyse & identify ✓
 - Current status and potential approaches to unified data management plan and system
 - Identify synergies with external activities
- Mitigate ✓
 - Step by step in a prioritised implementation
 - Basic principles outlined in a data management plan
 - Working with the community through the INTERACT Data Forum
- Establish a demonstrator catalogue ✓
 - Of available datasets
 - Go for the “easy wins” first
 - Through “INTERACT best practises”
- Link research stations ✓
 - with observation networks and data repositories
 - retaining station identity

Strategy

- Initially focus on discovery metadata ✓
 - Ensure interoperability and information flow/streams
 - To establish a unified view of the INTERACT data space
 - Expose this to external frameworks (metadata standards)
- Move to interoperability at the data level when data discovery is working ✓
 - Interoperability at the data level is required for interaction with larger frameworks
 - Bundling of similar data is required to be relevant for CalVal activities in larger programmes
 - While retaining the visibility of stations and scientists
- Develop guidance material ✓
 - For stations
 - For scientists
 - Based on existing efforts within disciplines, RDA, ICSU, WMO etc
- Improve visibility and relevance of Interact for e.g. WMO and SAON activities ✓
 - Through interaction with the Arctic Data Committee
 - Being pragmatic, working step by step, towards a long term vision

Deliverables and Milestones

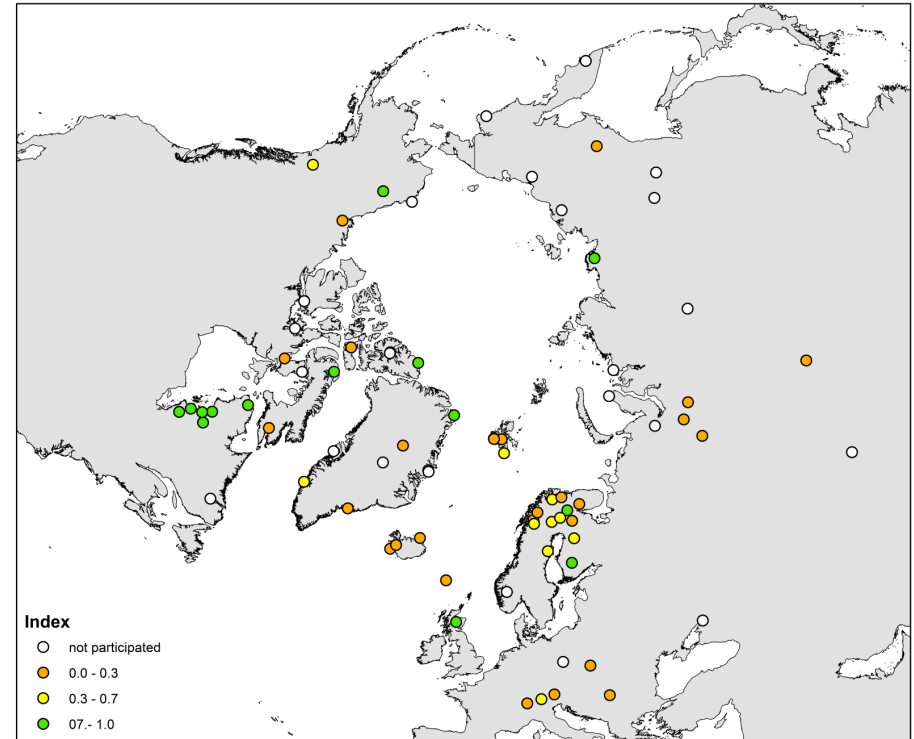
- D4.1: Data Management Plan (Month 6) ✓
 - Delivered
- D4.2: Report on current data flows (Month 12) ✓
 - Gap analysis and bottlenecks
 - Delivered
- D4.3: Field guide to data repositories (Month 24) ✓
 - Mentoring of potential providers of TA virtual access
 - Delivered
- D4.4: Data Policy (Month 24) ✓
 - Delivered

Data management plan

- The purpose of the Data Management Plan is to describe the data that will be created and how it will be shared and preserved.
 - Interact DMP is very high level
 - The goal is a unified view of INTERACT data that will improve the impact of INTERACT and individual stations.
 - The basic principles of INTERACT data management is that INTERACT is following a metadata driven approach.
 - This calls for INTERACT datasets described using standardised discovery and use metadata.
 - This shall ensure the data are archived and re-usable for future generations and relevant to technologically driven data analysis developments.
 - INTERACT relies on discipline specific efforts to establish interoperability at the data level.
 - But FAIRness decides
 - INTERACT promotes free and open access to data in line with the European Open Research Data Pilot (OpenAIRE).
- Selected recommendations
 - metadata and data products shall be free and open (Creative Commons attribution license),
 - although some data may have temporal restrictions
 - to be further detailed in the INTERACT Data Policy
 - shall use self explaining file formats/data encoding
 - Ensuring a lasting legacy
 - shall make data available in a timely fashion
 - data shall be archived in repositories with a long term mandate and in an interoperable form
 - promotes and encourages the implementation of globally resolvable Persistent Identifiers (e.g. Digital Object Identifiers) at each contributing data centre
 - To get credit to data providers

Current state of data management

- Survey circulated among station managers
 - used to guide development of Data Management Plan
- 16 multiple choice and free text questions addressing FAIR guiding principles
 - Wilkinson et. al (2016)
- 78% of INTERACT stations responded
- Considerable lack of knowledge on data management requirements
- For many stations responsibilities are unclear
- resulting in low or lacking data integrity/security



Field Guide to Data Repositories (1)

Fieldwork



Before...

Data management already starts before data collection in the field. A data management plan (DMP) is a good way to think through and document the data life cycle, including a sampling strategy, anticipated data formats, possible storages for long-term preservation, a publishing strategy and a data backup plan during data collection and processing. Prepare standardized metadata sheets before going to the field. This will make it easier for you to integrate metadata documentation into your workflow.



During...

Fieldwork time is limited - thorough preparation pays off. Fieldwork is the phase of data collection, however already thinking about the next steps in the data life cycle can reduce workload and save extra work back home. Making use of **best practices** for sampling may help you to make your results comparable to already published research data. The usage of standardized metadata sampling sheets will prevent you from forgetting to document important information. Since it is easy to lose data in the field, a regular backup of all data (incl. field notes) and metadata is crucial.



After...

Back home, data processing and analysis are the next steps. At this stage, sharing data with your colleagues from your research project and converting the data in **open data formats** will enhance their usability and usage. Do not forget to set up a backup system and use versioning while data processing. Long-term data preservation and data publication are the last two major steps. Think of the pros and cons of making your data openly accessible. The relevance and value of your data might rise enormously when used by your colleagues and further researchers. Embedded in a different scientific context, your research data may help answering scientific questions well beyond the scope of the initial research project.

Buzzwords and services

FAIR	Findable, Accessible, Interoperable, Reusable
DCC	Digital Curation Centre
DMP	Data Management Plan
DOI	Digital Object Identifier
OpenAIRE	European Open Research Data Pilot
RDA	Research data alliance
Re3data	Searchable online catalogue for data repositories



LOSS OF DATA

By Auke Herrema
CC BY www.mngerry.com/www.aukeherrema.nl



PUBLICATIONS AND DATA

By Auke Herrema
CC BY www.mngerry.com/www.aukeherrema.nl



Links and References

DCC: <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan>

European Commission (EU-C): (2016): Guidelines on FAIR Data Management in Horizon 2020:

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-dat-a-mgt_en.pdf

FAIR data: Wilkinson, M., et al., 2016: The FAIR guiding principles for scientific data management and stewardship. Nature Scientific Data, 3: 160018, DOI: 10.1038/sdata.2016.18,

<https://www.nature.com/articles/sdata201618>

INTERACT: www.eu-interact.org

OpenAIRE: <https://www.openaire.eu/>

RDA: <http://rd-alliance.github.io/metadata-directory/standards/>

Re3data: <https://www.re3data.org/>

DCC

EU-C

FAIR data

INTERACT

OpenAIRE

RDA


Re3data

Field Guide to Data Repositories (2)


Data management is needed in each research project in order to keep track of data handling, modification and storage. Data evolution in a project is commonly described as the **Data Life Cycle**.


Data Life Cycle




The **data life cycle** describes the different stages of data within a research project. Data practices should be guided by the **FAIR data principles** , meaning that the data is Findable, Accessible, Interoperable and Re-usable (Wilkinson et al., 2016).



The **FAIR data principles** are acknowledged by scientists all over the world as guidelines for data handling.


Already in the **Planning** phase of your research project, it is helpful to think through the life cycle of your data - starting from how to collect them and ending on how to preserve and publish them. A **Data Management Plan (DMP)** helps you to organize your data in advance and keep track of them during the course of the project. Guidance material for making a DMP is provided for example by the Digital Curation Centre .


During **Data Collection or Creation**, it is important that you do a thorough metadata  documentation. Metadata are data describing data. The diversity of data collected at INTERACT station is very high. The **best practices** for data collection and metadata standards differ with respect to the generated data.





 **Data Processing** implies its modification. Processing steps need to be documented in the metadata so that you keep track of data evolution. We recommend to use a self-explanatory versioning system, to make the alteration of the data transparent. A regular data backup during processing is crucial. This should include storing the data on two independent systems (e.g., PC and institutional server).

 The **Preservation** of your research data is a decisive step in the data life cycle. The best practice is to store your data and metadata in a trusted data repository .


 **Data Publication** is a necessary step to make your data accessible for the public. You can either publish your data and its corresponding metadata, or just the metadata. Use open data formats and follow metadata standards to ensure the findability, interoperability and reusability of your data. Embargos and licensing give you control over data access and usage. Publishing your data makes your research more trustworthy and may help increasing your scientific reputation. Data publication will not only help others to find your data but it also will help you to find your data after your research project ends!

 For the **Reuse** of data, you need to make them findable, accessible, interoperable and document them in a self-explanatory way. Having rich and standardised metadata increases the value of your data and the likelihood of its reuse. Providing access to your data keeps them relevant beyond the scope of your project and will help to exploit their full potential.


Data management within INTERACT

INTERACT is a Horizon2020 project and thus funded by the EU. Horizon2020 projects comply to the Open Data Pilot called OpenAIRE . Thus, by default all scientists using INTERACT stations are obliged to provide free and open access to data collected/created within the framework of the INTERACT network, or at least to research data which validate the results in scientific publications. If you can provide solid reasons for keeping these data closed, you can opt out of this obligation. The European Commission released some additional guidelines on FAIR data management in Horizon 2020 .

What are Metadata?

Metadata provide the context for data and contain descriptive information about the dataset (e.g., author, location, time, method of data collection, processing steps etc.). To ensure findability and reusability of data, it is important to use established data and metadata standards (e.g., ISO 19115 compliant). The used metadata standard depends on the data type and the scientific discipline. You will be requested to provide metadata during the data submission process to a trusted data repository. A list of general and discipline specific metadata standards are provided by the Research Data Alliance .

How to find a trusted repository

The search for a trusted data repository usually starts with exploring the repository of your institution or data repositories used by your research community. You can use the Re3data platform  to search for a suitable data repository for long-term storage and publishing of your data. Re3data is a searchable online catalogue with entries for more than 2000 data repositories. The data repository of your choice should fulfill the following criteria:


- Complies to the FAIR guiding principles
- Supports relevant metadata standards and data formats
- Provides persistent identifiers (e.g., DOI)
- Uses data licensing (e.g., creative commons)
- Ensures long-term data preservation.

READ MORE ABOUT INTERACT AND DATA MANAGEMENT ON:
www.eu-interact.org


Data Policy

- Aligned with OECD, GEOSS, IASC, and EU approaches
- The INTERACT data sharing principles are as follows:
 - I. There will be full and open² exchange of data, metadata and products³ shared within INTERACT, recognizing relevant international instruments and national policies and legislation;
 - II. All shared data and metadata will be made available⁴ through the INTERACT Data Management System with minimum time delay;
 - III. All shared data and metadata should be distributed free of charge or no more than the cost of reproduction;
 - IV. Data access may be restricted when data release could compromise the confidentiality of human subjects or cause harm to endangered species or other vulnerable subjects.
 - V. Users of data shall acknowledge in any publication or any other derived work, the contribution made by those who have created and worked up the data.
- This data policy applies to INTERACT members and third parties contributing to INTERACT or using data supplied through INTERACT. In the present form it is restricted to the Virtual Access component of INTERACT.

Project No. 730938
D4.4 – INTERACT Data Policy



Integrating Activities for Advanced Communities



D4.4- INTERACT Data Policy

Project No.730938– INTERACT

H2020-INFRAIA-2016-2017/H2020-INFRAIA-2016-1

Start date of project: 2016/10/01
Due date of deliverable: 2018/09/30 (Mx)

Duration: 48 months
Actual Submission date: 2019/02/08

Lead partner for deliverable: METNO
Author: Øystein Godey

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the Consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Document ID: D4.4_r6.doc
Date: 2018/10/??

Public

© INTERACT consortium
Page 1 of 9

FileEditViewHistoryBookmarksToolsHelp

Ontobee

Home

NBS 201

Flagg

Tutor

pand.

pand.

Hand

NetCDF

Year

GCM

Index of

NETGEAR

WISV

IAGC

Se x

+

+

+

←

→

↶

🏠

🔒

https://interact.met.no/datasearch

...

🔍

Search

⬇

📄

🖨

🔍

☰

⚙

Most Visited

🔗

Getting Started

🌐

NorDataNet

🌐

Arctic ...

📄

XPath and XSLT with I...

🌐

Parsing XML and HTM...

🌐

xslt.md.html

🏠

Dashboard

Content

Structure

Appearance

People

Modules

Configuration

Reports

Help

Hello steingod

Log out

⬆

INTERACT

Search For Data

ViewEdit

This is a preliminary test interface. The data found are not INTERACT data, but a sample collection for demonstration purposes.

This is an inclusive search meaning that the search criteria selected will be combined to retrieve results. All menus are collapsed initially, the are opened by "clicking" them. Geographical bounding boxes may be selected using the map and fine tuned by opening the bounding box menu. Institutions contains a list of all Institutions mentioned as PI Institutions in the datasets.

Always remember to cite data!

Full text search

Combine search keywords: e.g "radiosonde AND humidity"

synop AND eureka

Data Collection Period

Start Date

yyyy-mm-dd

End Date

yyyy-mm-dd

Bounding Box

Institutions


Investigator

Topics and Variables

Search


Reset

Geographical Extent



File Edit View History Bookmarks Tools Help
< Ontobee Home NBS 201 Flag Tutor pand. pand. Hand NetCDF Year GCM Index of NETGEAR WIS IAGC Se x > + v
https://interact.met.no/results?page=1
Search
Most Visited Getting Started NorDataNet Arctic ... XPath and XSLT with I... Parsing XML and HTM... xslt.md.html
Dashboard Content Structure Appearance People Modules Configuration Reports Help Hello steingod Log out

INTERACT Home Home Search for data Support News



Home » Search results

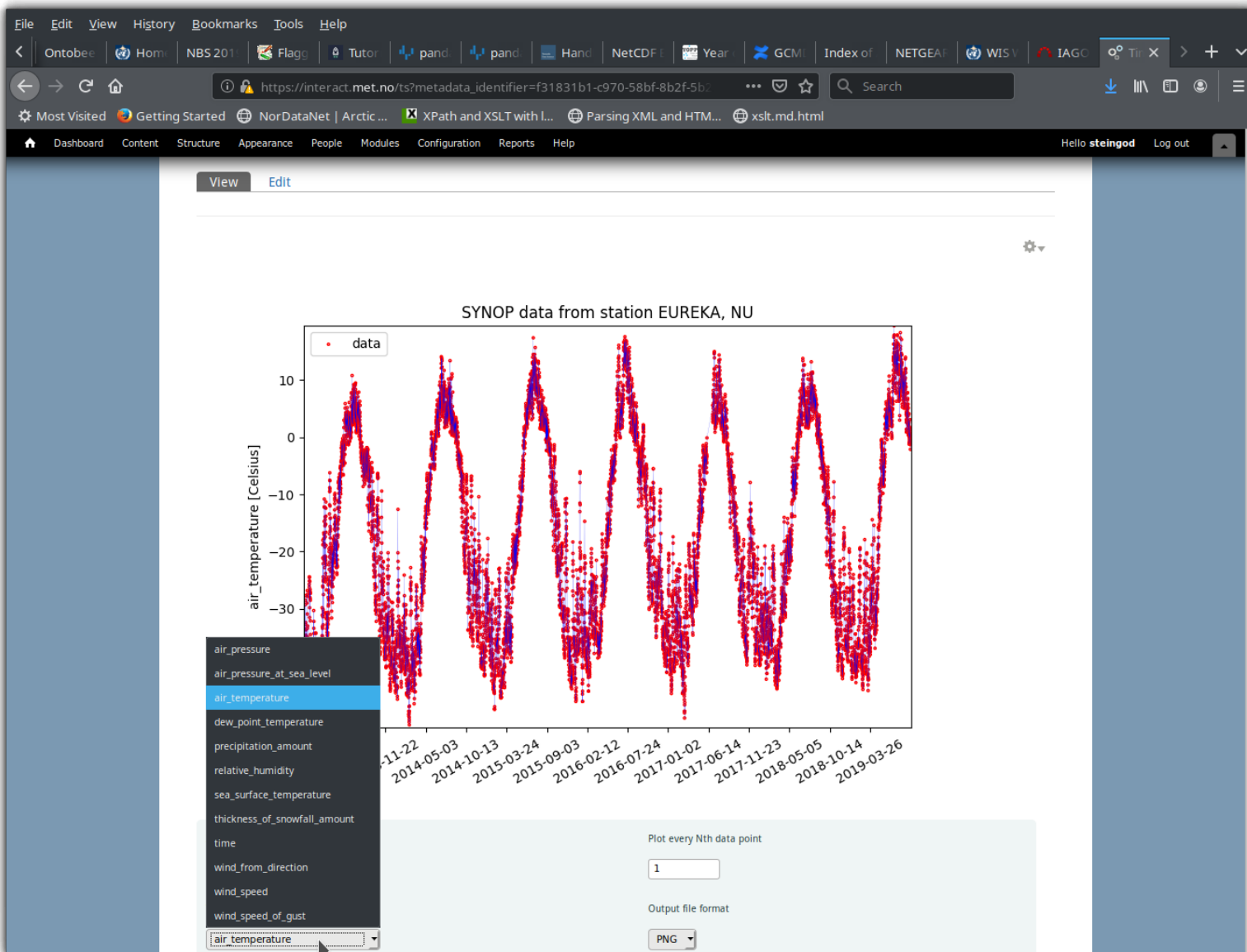
✔ Number of datasets found: 1

Search Results

<input type="checkbox"/> Dataset name	Institutions	Abstract	Collection period
<input type="checkbox"/> f31831b1-c970-58bf-8b2f-5b220213b545 View metadata Visualize timeseries Download as ASCII	Norwegian Meteorological Institute	Synoptic meteorological measurements from EUREKA, No extracted from the WMO Global Telecommunication System (GTS). Data are not quality controlled after extraction from GTS.	2013-01-01T12:00:00Z to

Prev 1 Next

Add to basket Back to search Share results



ASCII Data Download

View Edit

<input type="checkbox"/>	Standard name	Units
<input type="checkbox"/>	air_pressure	hPa
<input type="checkbox"/>	air_pressure_at_sea_level	hPa
<input type="checkbox"/>	air_temperature	Celsius
<input type="checkbox"/>	dew_point_temperature	Celsius
<input type="checkbox"/>	precipitation_amount	kg m-2
<input type="checkbox"/>	relative_humidity	percent
<input type="checkbox"/>	sea_surface_temperature	Celsius
<input type="checkbox"/>	thickness_of_snowfall_amount	cm
<input type="checkbox"/>	time	days since 1970-01-01 00:00:00
<input type="checkbox"/>	wind_from_direction	degree
<input type="checkbox"/>	wind_speed	m s-1
<input type="checkbox"/>	wind_speed_of_gust	m s-1

Output format

CSV

Submit

Back to results

Relevant activities

- SAON/IASC Arctic Data Committee
 - Interoperability Workshop and Assessment Process
 - Frascati November 2016
 - Polar Data Planning Summit
 - Boulder May 2018
 - Polar Data Architecture Workshop
 - Geneva November 2018
 - Third Polar Data Forum
 - Helsinki November 2019
- EU-POLARNET and EU-Arctic-Cluster (EU-Polar_Cluster) activities
- ENVRI-FAIR connecting ESFRI infrastructures to EOSC
 - Focus on gap analysis and mitigation for various domains
 - Strong focus on data documentation, semantic annotation and provenance

Ongoing work

- Updating INTERACT DMP
- Updating INTERACT DP
- Disseminating INTERACT Field Guide to repositories