



Data bases available at ICPs

Till Kirchner

8th joint session of EMEP SB & WGE
Sept 2022



OUTLINE

1. The ICP Forests data portal and data management
2. Data bases of other ICPs
3. Discussion



The ICP Forests data portal and data management

The ICP Forests data portal

<https://icp-forests.org/data>

Functionalities

- **Upload and test** monitoring data from participating countries
- **Download** monitoring data and reports on availability / quality
- **User management**
- **Overviews** on ongoing data submissions

The ICP Forests data portal

<https://icp-forests.org/data>

Technical components

- Portal software (self-developed)
- Database online documentation (self-developed)
- Database-Management-System (PostgreSQL)
- Data format (submissions, download): CSV

The ICP Forests data portal

Database (data model)

The complexity

21 surveys (deposition, meteo, ...) → 5 x Level I / 16 x Level II

└─→ **85 data tables**

└─→ About **2000 attributes** (columns)

└─→ About **65.5 millionen rows**

The ICP Forests data portal

Database (data model)

Online documentation*

is automatically generated from metadata in the database

- Definition / units of attributes
- Codelists for categorized variables
- Tests applied during the submission
- Changes of the data model over time (versioning)

* <https://icp-forests.org/documentation>

The ICP Forests data portal

Data submissions

A submission contains all data measured during one year by a partner for a specific survey (e.g. deposition 2019 from France)

Complex testing (about 2800 tests)

1. Minimum set of tables expected by the system
2. Structure of tables (columns, names, data types)
3. Tests (ranges – temporal - complex comparisons / calculations)

The ICP Forests data portal

Data access

Can be requested

- Standardized procedure
- Official data policy
- Differentiation between “internal” and “external” access

No request has ever been denied!

The ICP Forests data portal

Data access

Benefits of access after request

- participating countries stay informed about the use of their data
- cooperations of ICP Forests experts with external scientists can be coordinated
- PCC can report about ongoing activities / projects

The ICP Forests data portal

Data access

PCC documents ...

- projects using our data:
<http://icp-forests.net/page/project-list>
- scientific publications based on our data:
<http://icp-forests.net/page/scientific-publications>



**60 peer reviewed publications per year using
ICP Forests data or infrastructure!**

The ICP Forests data portal

Data access

Open dataset on Level II available:

- Plot description / Metadata on data availability
https://icp-forests.org/open_data/
- Can be used ...
 - to identify in detail what we can offer before requesting our data
 - as a set of describing attributes to be published with own results without further permission

The ICP Forests data portal

Ongoing continuous developments

- Adapt data model to changes in the ICP Forests Manual (new methods / attributes)
- Corrections of logical gaps / inconsistencies
- improvements of test routines

Recent success (milestone):

Meanwhile we are able to test all existing data against the newest check routines automatically

The ICP Forests data portal

Future plans

So far we are (just) providing the original monitoring data submitted (raw data).

We are working on gap filling and aggregation algorithms to provide harmonized and aggregated datasets in future.

The ICP Forests data portal

FAIR principles

Findability | **A**ccessibility | **I**nteroperability | **R**eusability

- + website
- + online docu
- + standardized citation (data policy)
- No DOI for datasets

The ICP Forests data portal

FAIR principles

Findability | **A**ccessibility | **I**nteroperability | **R**eusability

- + data portal (newest version can be accessed)
- standardized data request takes some weeks

The ICP Forests data portal

FAIR principles

Findability | **A**ccessibility | **I**nteroperability | **R**eusability

NOTHING SO FAR

The ICP Forests data portal

FAIR principles

Findability | **A**ccessibility | **I**nteroperability | **R**eusability

description of methods (manuals) +
documentation of the database +
gap filling and aggregation methods will be published in future +

A photograph of a forest scene with tall trees and green undergrowth. A large white circle is overlaid on the left side of the image, containing the text 'ICP IM'.

ICP IM



The ICP IM data base

- Data from 48 active sites from 15 active countries, also data from further previously active sites
- Data submission for at least part of the last five years: Austria, Czech Republic, Estonia, Finland, Germany, Ireland, Italy, Lithuania, Norway, Poland, the Russian Federation, Spain, Sweden and Switzerland
- 24 sub-programmes (e.g. soil chemistry, foliage chemistry...)
- Mostly started in mid 1990's but older data available from some sites and countries (oldest data from 1980s)



Structure

- Database (2022)
 - SQLite in a single file
 - 200 000 samples
 - 1 300 000 data values
 - During data submission, python script autogenerates and supplies data provider with a validation report indicating problems
- Advantages of file database
 - No server needed to be maintained
 - No cost for software - open source
 - Easy to integrate into python programming language



Open data?

- Data currently available by request to Programme Centre
- Open data discussed, and no fundamental objection by NFPs that participated in the 2022 IM TF meeting
- However some details to be elaborated
 - Completely open online or via Programme Centre?
How best to track usage?
 - How to include soft supporting data, e.g. site specific knowledge of droughts, storms, insect attacks etc.?

A photograph of a forest scene with tall trees and green undergrowth. A large white circle is overlaid on the left side of the image, containing the text 'ICP WATERS'.

ICP WATERS

Data access ICP Waters

Q1 – obstacles to open access

- All code relating to data storage, manipulation and analysis for water chemical data ICP Waters (is already Open Source and available online via GitHub)
- Raw (water chemical) data are publicly funded, and sometimes already available through national portals. However, it is not certain if data providers are interested that ICP Waters gives full access to raw data to third parties

Data access ICP Waters

Q1 – obstacles to open access

- Databases are maintained at source institutes, no automated updates of changes in raw data to ICP Waters (problematic to have two sources for the same data; unless all institutes provides APIs (allows for automated data transfer)
 - Not common for biological data
(less standardized, need more quality assurance)

Data access ICP Waters

Q2 – FAIR principles

- NIVA (where program centre is located) works mostly according to FAIR principles (as are probably most institutes where NFCs are located), but full adherence is (partly) a funding issue
- ICP Waters is not functioning as data portal, ICP Waters role is to collect, quality assure and assess data and publish summaries in reports

Data access ICP Waters

Q3 - what type of data should be made available

- Summaries of data, code, extended datasets used in reports are open access (in repositories) (is our current practice; but no DOIs (yet))

Q4 - compatibility with open access principles

- Yes (if the question is correctly understood)

A photograph of a forest scene with tall trees and green undergrowth. A large white circle is overlaid on the left side of the image, containing the text 'ICP VEGETATION'.

ICP VEGETATION



ICP Vegetation data access

Q1 Barriers

- Participants provide data – data ‘ownership’ sensitivities
- Small ‘ad hoc’ datasets – if these become readily available QA and QC is needed on a dataset we might never use thoroughly. There is a trade-off for cost:benefit
- Identification of when a dataset is ‘finalised’ – in terms of completeness and QA, and in terms of science

Q2 FAIR principles

- Host institution of the ICP adheres (although note the process is the same for very small ad-hoc datasets as for larger datasets, so in practice the small ones may not be included in the formal repository)
- In practice the repository is large, and it can be difficult to know what to look for (in terms of what is available)



ICP Vegetation data access

Q3 Types of data

- The repository we use can ingest maps as well as files of data. All with supporting documentation and a DOI.
- The repository data is clean and validated datasets, and often summarised data (not raw data)
- Moss data is not in the institute repository – available on request (freely available for use within UNECE, otherwise after checking with the data provider). A small amount of summary data for the moss survey is on the ICP Veg website, with a note saying more is available from ICP Veg.

Q4 Compatibility with OA principles

- Yes



DISCUSSION

DISCUSSION

Potential obstacles or difficulties to propose open access to the data and source code?

FAIR principles

(findability, accessibility, interoperability, and reusability)

What data should be publicly available?

(raw data, only validated data, data products)