

«Open data» and FAIR-principles applied to EMEP observation data

Kjetil Tørseth, NILU/EMEP-CCC



Norsk institutt for luftforskning
Norwegian Institute for Air Research

EMEP – «Open data» since the start

EMEP/CCC – REPORT 1/78
DATE: AUGUST 1978

CCC reports 1978

Preliminary data report - October 1977 - March 1978

Harald Dovland

EMEP/CCC-Report 1/78: [pdf](#)

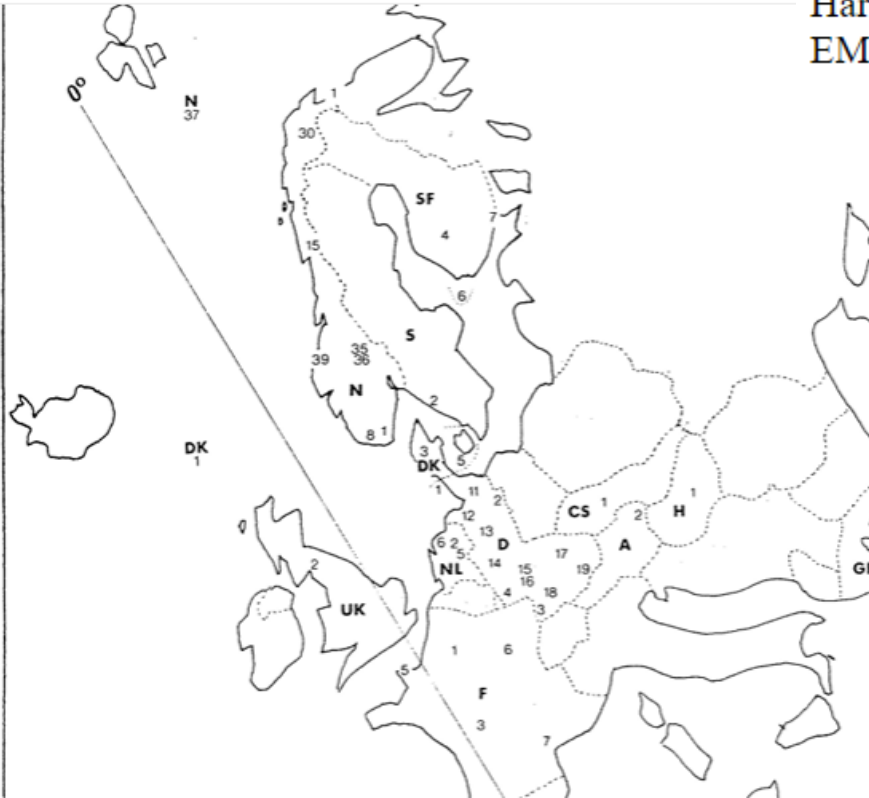
ECE Co-operative programme for
monitoring and evaluation of the long
range transmission of air pollutants
in Europe

PRELIMINARY DATA REPORT
OCTOBER 1977 - MARCH 1978

CONTENTS

	Page
1 INTRODUCTION	3
2 DATA FROM THE MONITORING STATIONS	3
3 GENERAL COMMENTS	10

ANNEX 1: Complete listing of data: 1 October 1977 -
31 March 1978



LANGENBRUGGE

GERMANY, F. R.

D 02 NOVEMBER 1977

DAY	AMOUNT MM	H+ UE/L	PH	PRECIPITATION				NA MG/L	MG MG/L	CL MG/L	AEROSOLS			TSP UG/M3	GASES	
				SO4-S MG/L	NO3-N MG/L	NH4-N MG/L	CA MG/L				SO4-S UG/M3	NH4-N UG/M3	H+ NE/M3		SO2-S UG/M3	NO2-N UG/M3
1	-	-	-	-	-	-	-	-	-	-	1.00	-	-	30	4.8	3.7
2	8.9	-	4.70	1.17	-	-	.7	-	-	.6	1.10	-	-	43	12.9	5.4
3	1.3	-	4.50	3.21	-	-	1.7	-	-	4.0	.70	-	-	20	8.6	4.9
4	.4	-	5.20	1.97	-	-	.8	-	-	2.7	.60	-	-	22	6.1	4.0
5	-	-	-	-	-	-	-	-	-	-	.80	-	-	32	10.4	6.1
6	.5	-	4.90	3.85	-	-	1.8	-	-	1.0	.80	-	-	22	3.6	3.6
7	-	-	-	-	-	-	-	-	-	-	1.30	-	-	36	11.1	7.0
8	4.4	-	4.80	1.66	-	-	.5	-	-	2.9	.70	-	-	22	4.6	4.4
9	1.7	-	5.10	3.45	-	-	1.4	-	-	1.0	.90	-	-	25	4.9	4.8
10	.1	-	-	5.86	-	-	2.8	-	-	5.1	1.20	-	-	34	5.3	4.9
11	1.7	-	5.40	1.68	-	-	1.0	-	-	2.5	1.40	-	-	38	10.2	5.9
12	.8	-	5.40	1.11	-	-	.6	-	-	10.6	.50	-	-	20	2.9	5.4
13	2.8	-	4.60	1.24	-	-	.6	-	-	3.5	.60	-	-	15	3.4	2.2
14	23.0	-	4.70	.99	-	-	.1	-	-	2.3	.50	-	-	14	3.4	3.3
15	1.8	-	6.00	1.06	-	-	.5	-	-	5.2	0.00	-	-	16	4.8	2.7
16	.2	-	6.50	1.72	-	-	.9	-	-	14.5	.30	-	-	14	3.4	2.6
17	.3	-	4.90	2.47	-	-	.7	-	-	.7	1.10	-	-	29	12.1	7.2
18	.3	-	4.70	4.20	-	-	.2	-	-	.2	1.10	-	-	22	3.7	4.5

ECONOMIC COMMISSION FOR EUROPE

CONVENTION ON LONG-RANGE TRANSBOUNDARY AIR POLLUTION

emep

Cooperative Programme for Monitoring
and Evaluation of the Long-range
Transmission of Air Pollutants in Europe

STRATEGY FOR EMEP 2000-2009



UNITED NATIONS

Vision

The Cooperative Programme for Monitoring and Evaluation of the Long-range Transmission of Air Pollutants in Europe (EMEP) will continue to be the main science-based and policy-driven instrument for international cooperation in atmospheric monitoring and modelling, emission inventories and projections, and integrated assessment to help solve transboundary air pollution problems. To this end it seeks to develop:

- **SCIENCE** - EMEP establishes sound scientific evidence and provides guidance to underpin, develop and evaluate environmental policies;
- **PARTNERSHIP** - EMEP fosters international partnership to find solutions to environmental problems;
- **OPENNESS** - EMEP encourages the open use of intellectual resources and products;
- **SHARING** - EMEP is transparent and shares information and expertise with research programmes, expert institutions, national and international organizations, and environmental agreements;
- **ORGANIZATION** - EMEP is organized to integrate information on emissions, environmental quality, effects and abatement options, and to provide the basis for solutions.

Framework [54] ⓘ
CAMPAIGN
CAPMoN
CARIBIC-GEOmon
COLOSSAL
CREATE
EANET
EMEP

Country [45]
>>All
Armenia
Austria
Belarus
Belgium
Bosnia and Hercegovina
Bulgaria
Cape Verde

Station [366] ⓘ
>>All
Abastumani
Abisko
Achenkirch
AGH University Krakow
Agia Marina Xyliatou / Cyprus Atmospheric Observatory
Aliartos
Amherd

Instrument type [84] ⓘ
>>All
abs_solution
abs_tube
ads_tube
aerosol_mass_spectrometer
air_UK
amalg_tube
AMS

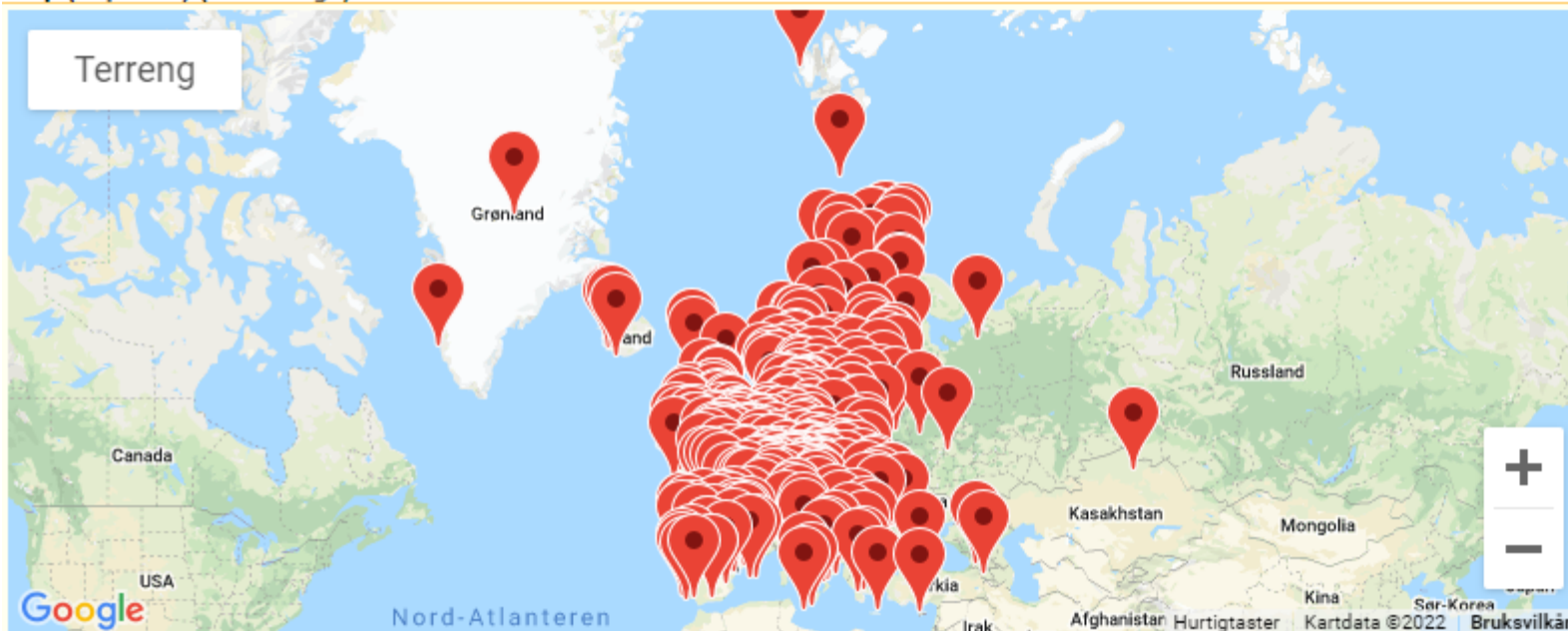
Component [496] ⓘ
>>All
1-2-3-trimethylbenzene
1-2-4-trimethylbenzene
1-3-5-trimethylbenzene
1-3-butadiene
1-butene
1-butyne
1-hexene

Matrix [29] ⓘ
>>All
aerosol
aerosol_humidified
air
air+aerosol
air+pm10
instrument
met

From >>All To >>All

Available datasets: 55667
Reset List datasets

Map (Populate) (Show large)



Additional resources

- Near-Real-Time data
- European Monitoring and Evaluation Programme (EMEP-CCC)
- Site descriptions - EMEP
- WMO Global Atmosphere Watch (GAW)
- Site descriptions - GAW
- Air mass trajectories
- Data submission
- About EBAS
- EBAS User Feedback Tracker



<http://ebas.nilu.no>

Aarhus Convention

From Wikipedia, the free encyclopedia

The **UNECE Convention on Access to Information, Public Participation in Decision-making and Access to Justice in Environmental Matters**, usually known as the **Aarhus Convention**, was signed on 25 June 1998 in the Danish city of [Aarhus](#). It entered into force on 30 October 2001. As of March 2014, it had 47 parties—46 states and the [European Union](#).^[1] All of the ratifying states are in Europe and Central Asia. The EU has begun applying Aarhus-type principles in its legislation, notably the [Water Framework Directive](#) (Directive 2000/60/EC). Liechtenstein and Monaco have signed the convention but have not ratified it.

The Aarhus Convention grants the public rights regarding access to information, [public participation](#) and access to justice, in governmental decision-making processes on matters concerning the local, national and transboundary environment. It focuses on interactions between the public and public authorities.

Directive on the re-use of public sector information

From Wikipedia, the free encyclopedia

Directive 2003/98/EC on the re-use of public sector information, known as the **PSI Directive**,^{[2][3]} now called **Open Data Directive**,^[4] is an [EU directive](#) that stipulates minimum requirements for [EU member states](#) regarding making [public sector](#) information available for re-use. This directive provides a common [legislative framework](#) for this area. The Directive is an attempt to remove barriers that hinder the re-use of public sector information through the [Union](#).

Directive 2003/98/EC

[European Union directive](#)



The digital economy is the worldwide network of economic activities, commercial transactions and professional interactions that are enabled by information and communications technologies (ICT).

No L 158/56

Official Journal of the European Communities

23. 6. 90

II

(Acts whose publication is not obligatory)

COUNCIL

COUNCIL DIRECTIVE

of 7 June 1990

on the freedom of access to information on the environment

(90/313/EEC)

What characterize EMEP data:

- Partly overlapping with data reported to other Frameworks (WMO-GAW, ACTRIS, AMAP, HELCOM ++)
- Largely produced by scientific institutions
- Mainly used by scientists for research or assessments

-> «Fair use data policy»

Open access

From Wikipedia, the free encyclopedia

Not to be confused with [Open source](#). For other uses, see [Open access \(disambiguation\)](#).

Open access (OA) is a set of principles and a range of practices through which [research](#) outputs are distributed online, free of access charges or other barriers.^[1] With open access strictly defined (according to the 2001 definition), or [libre](#) open access, barriers to copying or reuse are also reduced or removed by applying an [open license](#) for copyright.^[1]

The main focus of the open access movement is "[peer reviewed](#) research literature".^[2] Historically, this has centered mainly on print-based [academic journals](#). Whereas non-open access journals cover publishing costs through [access tolls](#) such as subscriptions, site licenses or [pay-per-view](#) charges, open-access journals are



Open source

From Wikipedia, the free encyclopedia

For a common use, see [Open-source software](#), and for other uses, see [Open source \(disambiguation\)](#).

Not to be confused with [Open access](#).

Open source is [source code](#) that is made freely available for possible modification and redistribution. Products include permission to use the source code,^[1] design documents,^[2] or content of the product. The **open-source model** is a decentralized [software development](#) model that encourages [open collaboration](#).^{[3][4]} A main principle of [open-source software development](#) is [peer production](#), with products such as source code, [blueprints](#), and documentation freely available to the public. The [open-source movement](#) in software began as a response to the limitations of [proprietary code](#). The model is used for projects such as in [open-source appropriate technology](#),^[5] and open-source drug discovery.^{[6][7]}

European Open Science Cloud

From Wikipedia, the free encyclopedia

The **European Open Science Cloud (EOSC)** is a [European Commission](#) initiative aiming at developing an infrastructure providing its users with services promoting [open science](#) practices. Besides being open science oriented, the envisaged infrastructure is built by aggregating services provided by several providers following a [System of systems](#) approach.

European Open Science Cloud

Type of project	research infrastructure
Owner	European Commission
Established	2015
Website	http://eosc-hub.eu/ 

Open Access | [Published: 15 March 2016](#)

The FAIR Guiding Principles for scientific data management and stewardship

[Mark D. Wilkinson](#), [Michel Dumontier](#), ... [Barend Mons](#) 

[+ Show authors](#)

[Scientific Data](#) **3**, Article number: 160018 (2016) | [Cite this article](#)

445k Accesses | **3925** Citations | **1948** Altmetric | [Metrics](#)

Findable – Accessible – Interoperable – Resuable

Findable

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource

Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

I3. (Meta)data include qualified references to other (meta)data

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

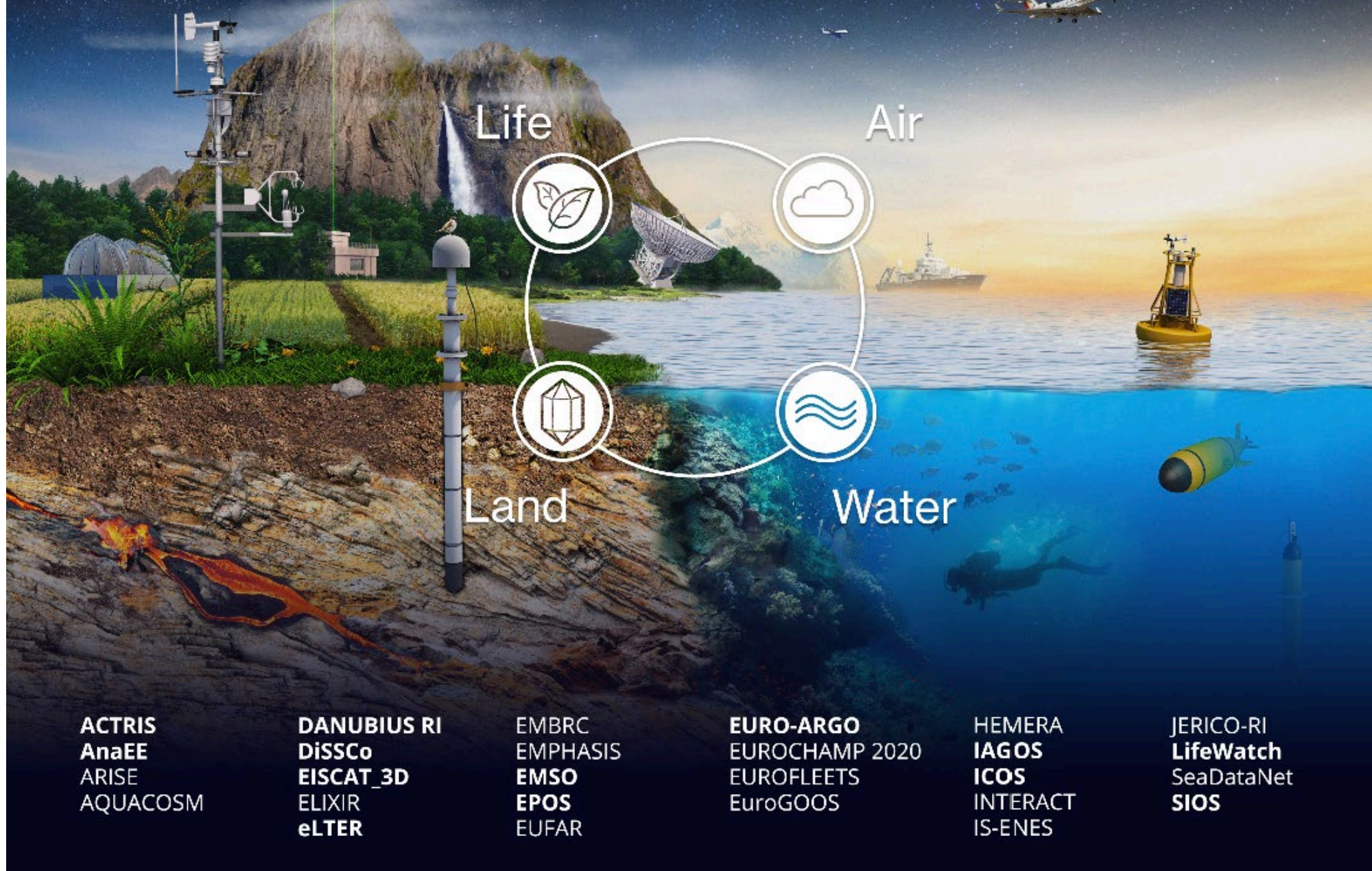
In the [Guidelines on FAIR Data Management in Horizon 2020](#), the European Commission states: "Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible."

Researchers, information managers and other stakeholders can rely on a framework of various international certification standards for digital repositories in order to assess and improve the quality of their work processes and management systems. "Trustworthy Digital Repository" (TDR) is a term often used in this respect.

In the European Framework for audit and certification of digital repositories three certification instruments, with increasing degrees of complexity and depth, are available:

- [CoreTrustSeal \(CTS\)](#): this is based on [Data Seal of Approval \(DSA\)](#) and [World Data System \(WDS\) van ICSU](#). All digital repositories that have one or more of these certifications are listed at <https://www.coretrustseal.org/why-certification/certified-repositories/>
- [Nestor Seal](#): verification according to DIN 31644
- [ISO 16363 certification](#)

ENVRI-FAIR (an EU EOSC-project)



FAIR Enabling Resources (FER)

FAIR Principles

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

FAIR Enabling Resource (FER)



Digital objects to achieve FAIRness

FAIR Implementation Profile (FIP)



Technology choices used for addressing each of the FAIR Principles

Results from the ENVRI-FAIRness Assessment

		2019	2020	2021
AIR	ACTRIS_DVAS	✓	✓	✓
	ACTRIS-Gres FIP	✓	✓	✓
	ACRIS-inSitu	✓	✓	✓
	ACTRIS_ARES	☐	✓	✓
	ACTRIS_CLU_FIP	✓	✓	✓
	ACTRIS-ASC	✓	✓	✓
	IAGOS	✓	✓	✓
	EISCAT_FIP	✓	✓	✓
WATER	ARGO	✓	✓	✓
	EMSO ERIC FIP	☐	☐	✓
	LW marine	✓	✓	✓
	SeaDataNet-CDI	✓	✓	✓
	SeaDataNet-Sextant	✓	✓	✓
LAND	EPOS	☐	☐	✓
LIFE	AnaEE	✓	✓	✓
	AnaEE-Crea	✓	✓	✓
	Danubius	☐	☐	✓
	DISSCo_FIP	☐	✓	✓
	eLTER-RI	✓	✓	✓
	LWERIC Ecosystem	✓	✓	✓
multi-domain	ICOS FIP	✓	✓	✓
	SIOS FIP	✓	✓	✓
Total count: 57		17	18	22

- FIP statistics
 - 22 communities
 - 57 FAIR Implementation Profiles
 - 178 FAIR enabling resources have been listed and declared

**Results from ENVRI FAIRness assessment, work led by Barbara Magagna and conducted by all partners all partners (Project Review|1 March 2022)*

FAIR enabling resource overlap

2021	ACTRIS_DVAS	ACTRIS_GRES	ACTRIS_InSitu	ACTRIS_CLU	ACTRIS-ARES	ACTRIS_ASC	IAGOS	EISCAT	ArgoGdac2	EMSO	lw-marine	SeaDataNet-CDI	SeaDataNet-Sextant	EPOS-ERIC	Anaee	AnaEE_CREA	DANUBIUS	DISSCo	eLTER-RI	LWERIC_Ecosystem	ICOS	SIOS	
ACTRIS_DVAS		11	11	8	8	11	9	2	6	4	4	2	5	3	7	4	4	3	4	5	7	7	125
ACTRIS_GRES	11		13	13	11	18	19	4	10	5	9	3	9	7	9	5	7	5	9	7	16	13	203
ACTRIS_InSitu	11	13		10	8	12	13	3	8	5	5	5	8	4	7	4	5	3	6	7	11	13	161
ACTRIS-ARES	8	11	8	8		9	12	3	7	4	5	1	4	5	4	4	5	2	5	5	13	7	130
ACTRIS_CLU	8	13	10		8	12	10	3	8	4	6	2	7	6	5	4	5	5	7	6	10	7	146
ACTRIS_ASC	11	18	12	12	9		14	3	9	5	8	2	8	5	8	4	6	5	7	6	10	11	173
IAGOS	9	19	13	10	12	14		4	11	6	10	7	12	7	9	5	7	6	12	10	20	13	216
EISCAT	2	4	3	3	3	3	4		5	2	3	3	3	2	2	2	1	1	5	1	7	2	61
ArgoGdac	6	10	8	8	7	9	11	5		8	7	6	10	3	6	5	6	4	5	7	14	6	151
EMSO	4	5	5	4	4	5	6	2	8		4	4	7	3	4	4	4	1	2	6	7	4	93
lw-marine	4	9	5	6	5	8	10	3	7	4		3	7	4	6	5	5	9	10	9	13	5	137
SeaDataNet-CDI	2	3	5	2	1	2	7	3	6	4	3		10	2	3	2	4	1	4	7	8	6	85
SeaDataNet-Sextant	5	9	8	7	4	8	12	3	10	7	7	10		4	8	5	6	3	8	9	12	9	154
EPOS-ERIC	3	7	4	6	5	5	7	2	3	3	4	2	4		3	3	1	4	7	4	8	5	90
Anaee	7	9	7	5	4	8	9	2	6	4	6	3	8	3		7	6	2	6	7	8	8	125
AnaEE_CREA	4	5	4	4	4	4	5	2	5	4	5	2	5	3	7		4	2	3	5	7	3	87
DANUBIUS	4	7	5	5	5	6	7	1	6	4	5	4	6	1	6	4		1	3	5	8	6	99
DISSCo	3	5	3	5	2	5	6	1	4	1	9	1	3	4	2	2	1		5	3	7	2	74
eLTER-RI	4	9	6	7	5	7	12	5	5	2	10	4	8	7	6	3	3	5		6	12	8	134
LWERIC_Ecosystem	5	7	7	6	5	6	10	1	7	6	9	7	9	4	7	5	5	3	6		9	6	130
ICOS	7	16	11	10	13	10	20	7	14	7	13	8	12	8	8	7	8	7	12	9		11	218
SIOS	7	13	13	7	7	11	13	2	6	4	5	6	9	5	8	3	6	2	8	6	11		152
	125	203	161	146	130	173	216	61	151	93	137	85	154	90	125	87	99	74	134	130	218	152	2944

Most frequent FERs	count
DOI Digital Object Identifier	33
CC BY 4.0 Attribution 4.0 International	29
HTTPS Hypertext Transfer Protocol Secure	27
PROV-O The PROV Ontology	25
NetCDF CF-1.7	23
DataCite	20
Open Data	19
ISO 19115 Geographic information - Metadata	16
NetCDF Network Common Data Form	14
Handle System	13
NVS NERC Vocabulary Service	13
REST Representational state transfer	13
DataCite Metadata Scheme	12
ORCID Open Researcher and Contributor ID	12
RDFS Resource Description Framework Schema	12
XMLS eXtensible Markup Language Schema	12
OPeNDAP Open-source Project for a Network Data Access Protoc	11

*Results from ENVRI FAIRness assessment, work led by Barbara Magagna and conducted by all partners all partners (Project Review|1 March 2022)

PIDs Recommended by ENVRI

Identified entity	PID type	metadata schema
humans	ORCID	ORCID record schema
organisations	Research Organisation Registry (ROR)	
instruments	Persistent Identifiers for eResearch (ePIC)	Persistent Identification of Instruments (PIDINST)
data products	Persistent Identifiers for eResearch (ePIC)	RI schema, complete metadata
	Digital Object Identifier (DOI) (mandatory)	DataCite
data pre-products	Persistent Identifiers for eResearch (ePIC)	RI schema
physical samples	Persistent Identifiers for eResearch (ePIC)	RI schema
	International Geo Sample Number (IGSN)	IGSN schema
software	Persistent Identifiers for eResearch (ePIC)	RI schema, complete metadata
	Digital Object Identifier (DOI)	DataCite
Internal documents	Persistent Identifiers for eResearch (ePIC)	

DOI-service, operational from 2022

Data Citation Index

Connecting data to the research it informs

[Go to product](#)[Contact us](#)

The Data Citation Index™ provides a single point of access to quality research data from global repositories across disciplines. Descriptive records are created for data objects and linked to literature articles in the Web of Science.™

As data citation practices increase, the Data Citation Index aims to provide a clearer picture of the full impact of research output, as well as to act as a significant tool for data attribution and discovery.

Vocabulary Landscape for ACTRIS/EBAS (EMEP)

ACTRIS internal:

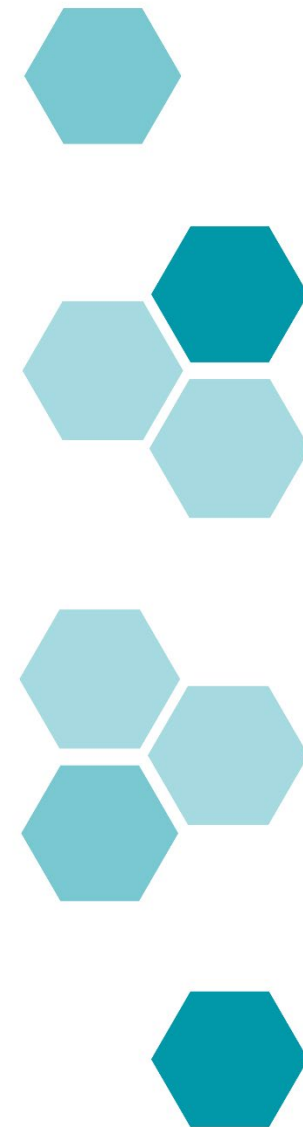
- Numerous expert groups growing together (aerosol, trace gases, in situ, remote sensing, chambers)
- Data centre nodes, TCs, HO.

ENVRI:

- Collaboration with other atmosphere RIs and other environmental domains.
- Capacity building by information scientists.

Relevant external vocabularies:

- **CF convention:**
represents ACTRIS user group, long experience, combines domain expertise with expertise on vocabulary.
- **GCMD Keywords:**
bias towards satellite remote sensing, but relevant user group, to be mapped to.
- **WIGOS, WIS:**
targeted towards experts, but relevant user group, to be mapped to.



Data management and Provenance

“Data provenance includes the data origin, what happens to it and where it moves over time. Data provenance gives visibility while greatly simplifying the ability to trace errors back to the root cause in a data analytics process.”

(Wikipedia contributors. (2022, July 6). Data lineage. In *Wikipedia, The Free Encyclopedia*. Retrieved 06:52, August 27, 2022, from https://en.wikipedia.org/w/index.php?title=Data_lineage&oldid=1096691134.)

How users get access to EMEP data

- 1) Self-service through web interface
- 2) Data on request to EBAS@nilu.no (NASA-Ames, NetCDF)
- 3) Machine to machine; data or metadata services
 - Services up and running
 - Copernicus/ECMWF
 - GEO Portal and NextGEOSS
 - SIOS
 - ACTRIS data portal
 - WMO WIS
 - MSC-W/AEROCOM
 - Services in development/implementation:
 - EOSC through ENVRI-FAIR
 - WMO – WIGOS4)

4) Data are duplicated and made available through secondary repositories



Licencing of data

- Benefits of licencing for data providers, data centers, organisations and users
 - define regulations on data use
 - reserve against any form of liability
 - a FAIR requirement
- Creativity and IPR rights
 - generally no IPR rights on «objective data»
 - funding support does not generate IPR
 - «Sui generis database rights»
- Creative Commons is a commonly used licence for Open Data
 - CC0 (no attribution) or CC BY4 (attribution required)
 - «Attribution» is not seen as a limiting condition towards «open data»
 - CC BY4 is strongly preferred for research derived data

Database Directive

From Wikipedia, the free encyclopedia

The **Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases** is a [directive](#) of the [European Union](#) in the field of [copyright](#) law, made under the [internal market](#) provisions of the [Treaty of Rome](#). It harmonises the treatment of [databases](#) under copyright law and the *sui generis* right for the creators of databases which do not qualify for copyright.

As of 2022 the directive is being reviewed as part of a proposed [Data Act](#). Public submissions closed on 25 June 2021,^[1] and a proposal for new harmonised rules on data was published on 23 February 2022.^[2]

Directive 96/9/EC

European Union directive



Title	Directive on the legal protection of databases
Made by	European Parliament & Council
Made under	Arts. 47(2), 55 & 95
Journal reference	L77, 1996-03-27, pp. 20–28

Licencing of EMEP data

To clarify for users that that EMEP data are free to be used for all purposes, and make sure that attribution is given to EMEP,

the EMEP Steering Body is hereby invited to take note of NILUs intention to (based on its Sui Generis database rights) introduce the Create Commons Open data license CC BY 4.0.

The licensing will affect all EMEP data in the EBAS database (<http://ebas.nilu.no>).

