D8.1
Atmosphere subdomain FAIRness Assessment

Work Package	WP8
Lead partner	UVSQ
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30/09/2019
Submission date	30/11/2019

Deliverable abstract

This document is the Deliverable 8.1 of the ENVRI-FAIR EU project conducted by the ESFRI Cluster of Environmental Research Infrastructures (ENVRI). Here, a first self-FAIRness assessment was done by the five Atmospheric RIs of the ENVRI-FAIR subdomain work package (WP) 8, comprising ACTRIS, IAGOS, ICOS-Atmosphere, EISCAT-3D and EISCAT. It used a questionnaire based on the 15 concepts of FAIRness described in Wilkinson et al. (2016). The assessment shows that all of these RIs find room for improvement in all four of the FAIR domains: Findability, Accessibility, Interoperability, Reusability.

After a thorough gap analysis, the WP8 as a whole could identify a common plan for a number of concrete developments where action could start immediately. These actions are further detailed in ENVRI-FAIR MS35 and an actual implementation plan is currently being drafted under task Task 8.3 of the project (Define atmospheric RI technological implementation plan and common procedures).

Furthermore, two topics that were found to be especially challenging are introduced in the last section of the document. These deal with: the use of semantic web, ontology and vocabularies and exhaustive provenance.

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Rivier L.	UVSQ	15/11/2019
Contributing Authors	C. Lund Myhre; D. Boulanger; M. Feibig; Lara Ferrighi; J. Tarniewicz; A. Tjulin	ACTRIS, IAGOS, ICOS, EISCAT-3D, SIOS	
Reviewer(s)	Alex Vermeulen, Paolo Laj		17-23/11/2019
Approver	Andreas Petzold		30/11/2019

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	11/11/2019	First draft	Cf above
V1.0	15/11/2019	Second draft	Same as above

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY

A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at http://doi.org/10.5281/zenodo.3465753.

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

1. Summary

This document is the Deliverable 8.1 of the ENVRI-FAIR EU project conducted by the ESFRI Cluster of Environmental Research Infrastructures (ENVRI). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Ecosystems and thus the Earth system in its full complexity. Here, a first self-FAIRness assessment was done by the five Atmospheric RIs of the ENVRI-FAIR subdomain work package (WP) 8 , comrising ACTRIS, IAGOS, ICOS-Atmosphere, EISCAT-3D and EISCAT. It used a questionnaire based on the 15 concepts of FAIRness described in Wilkinson et al. (2016)[1]. The assessment shows that all of these RIs find room for improvement in all four of the FAIR domains: Findability, Accessibility, Interoperability, Reusability.

After a thorough gap analysis, the WP8 as a whole could identify a common plan for the following concrete developments where action could start immediately:

- Consolidation of consistent use of PIDs throughout data production workflow
- Use of common standard interfaces for metadata and data
- Indexing of data resources in WIS, GEOSS
- Use of domain vocabulary / ontology
- Common use of authentication schemes
- Consistent documentation of provenance throughout data production workflow
- Development of semantic search for atmospheric ENVRI RI user interfaces

These actions are further detailed in ENVRI-FAIR MS35 and an actual implementation plan is currently being drafted under Task 8.3 of the project (Define atmospheric RI technological implementation plan and common procedures). An upcoming WP8 meeting is organised Dec 11-12, 2019 in Paris to further elaborate implementation plans for these actions and discuss tools to be used for developments. For this step of the work, the guidance of the ENVRI-FAIR transverse WP 5 and 7 will be pivotal to make sure that, if not common tools are used always, at least interoperability is ensured.

2. Introduction

This deliverable is part of the European project ENVRI-FAIR. It emanates from WP8 (Atmospheric subdomain) and is done under Task 8.2: Analyse the capabilities for FAIRness of the atmospheric RIs. It provides a first assessment of FAIRness from the five RIs that constitute the Atmospheric subdomain in ENVRI-FAIR. A gap analysis proposes then ways of improvements to become more FAIR.

WP8 Atmospheric subdomain

In ENVRI-FAIR, WP8 regroups five European RIs of the Atmospheric subdomain. These RIs target the composition of the atmosphere and its physical state, from the ground level to ionosphere, including space weather. The RIS involved in WP8 are ACTRIS, EISCAT, IAGOS, ICOS-Atmosphere, and SIOS. The overall aim of the WP is to improve the level of FAIRness of the involved RIs. To this aim, the work is organised in six tasks: in addition to the coordination task 8.1, task 8.2 provides a gap analysis in FAIRness, task 8.3 organises detailed implementation plans for each RI based on the analysis of 8.2. The actual implementation work related to increasing FAIRness at the RI level is regrouped in task 8.4 to promote synergies in the solutions that are chosen. Task 8.5 is set up to demonstrate the

[1] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al., 2016: The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, 3, 160018, doi:10.1038/sdata.2016.18.

new interoperability-based atmospheric services, and task 8.6 performs an assessment and provides recommendations for the future strategy.

WP8, Task 8.2: Analyse the capabilities for FAIRness of the atmospheric RIs

The objective of this task is to determine the level of FAIRness of the Atmospheric RIs. The aim of D8.1 is to provide a first self-assessment of FAIRness of the Atmospheric RIs, to make a gap analysis and derive from it a plan to improve FAIRness of the RIs during the ENVRI-FAIR project. This assessment will enable the identification of common developments and implementations to be made at the RI level. The latter are then further elaborated in collaboration with task 5.1 (Data service requirement tracking, technology landscape and FAIR data services gap analysis on the RI level) and 7.1 (Customised consultation and support to RIs for FAIR data service co-design), as well as with target end-users to ensure applicability. Emphasis will be made on machine actionable FAIRness developments.
Technological analysis of the capabilities for FAIRness includes:
- Identifying potential drivers for the implementation process to enhance interoperability within the subdomain and at domain level, (link to WMO standards (WIS/WIGOS), ICSU WDS, OGC, etc.)
- Set up the subdomain roadmap to FAIRness with specific emphasis on subdomain priorities through real common targets as defined in the objectives of this work package.

The task also includes the analysis of the policies and governance at atmospheric subdomain level and documents the existing systems, polices/licenses/condition of use among the involved RIs. On this last topic please refer to the ENVRI-FAIR milestone 34 already completed.

Link to ENVRI-FAIR WP5: Common requirements and testbed for (meta)data services, community standards and cataloguing

The main objectives of this ENVRI-FAIR WP5 are as follows:

1) Provide an up-to-date analysis on the gap(s) each individual RI needs to bridge in order to meet its interoperability and FAIR requirements
2) Select the common development targets for (metadata and data) services that will be implemented in the subdomains in WP8-WP11
3) Design, develop and implement the ENVRI-FAIR Catalogue of EOSC services
4) Design and provide guidelines for testing and validation ENVRI-FAIR services
5) Synthesise and demonstrate the readiness of ENVRI-FAIR services for EOSC and formulate a strategic roadmap for future development

Thus a gap analysis is also made in WP5, it is built from the data gathered by the subdomain work packages. Also, the analysis differs from the one performed at subdomain level, as it is done across all domains providing a common analysis tool to interpret the results. The automatised tool is built to provide as much objectivity as possible in assessing FAIRness.

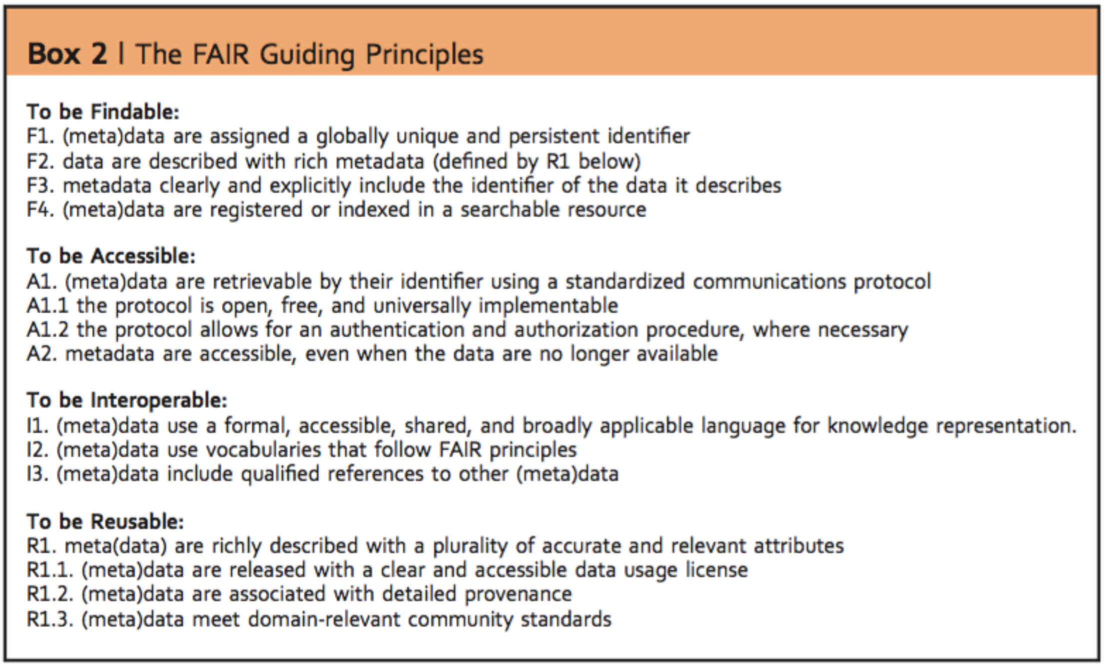
The FAIR principles

This section provides a short summary the FAIR guiding principles. In a world becoming increasingly digital, the challenge of correctly handling and using this enormous amount of data is increasing. The so called FAIR principles were first published in a short landmark paper by Wilkinson et al. in 2016 where a group of academic and private stakeholders proposed a set of guiding principle optimising data usage in a machine actionable way, i.e.;

without human intervention. FAIR is the acronym for Findable, Accessible Interoperable, Reusable. To make it over-simple, one can relate these four concepts to some basic questions:

Findable: Can I search the data and can I find it? i.e. is it properly identified?
Accessible: Once found, can I for example download the data?
Interoperable: What kind of format has the data? Are the associated metadata provided in a standard format? Can it be machine only actionable?
Reusable: Can I reuse the data knowing the proper license attached to the use of the data; how I should cite or give attribution when using the data? The notion of Provenance of the data is addressed here.

Each of these four principles were further divided in four, more technical, questions; as shown in the figure 1 below.



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

Figure 1: The FAIR Guiding Principles, from Wilkinson et al., 2016.

The first self-assessment of FAIRness performed in the ENVRI-FAIR project is done along these 15 principles. The rest of the document gives the result of this first FAIRness self-assessment for the RIs of the Atmosphere subdomain: ACTRIS, EISCAT, IAGOS, ICOS-Atm, SIOS. The next section gives a brief summary of the scope of these different RIs in terms of the atmospheric parameters they measure. It also gives an idea of the level or readiness of the infrastructure in terms of being completely or on the way to becoming fully operational.

3. RIs in Atmosphere subdomain

ACTRIS – Aerosol Cloud and Trace gases Research Infrastructure

The ACTRIS Research Infrastructure operates three topical databases that are linked via a metadata portal. In the upcoming years, more data centres will be added. All data are available through a common portal. First datasets date back from 2001. They are results from earlier (FP5-7)-projects such as CREATE, EUSAAR, EARLINET, EARLINET-ASOS, CLOUDNET, ACTRIS-FP7 and ACTRIS-2. The number of active data users per month is around 80 to 100, and around 20 000 measurement years of data sets are currently downloaded per year. ACTRIS has currently 96 sites (50 operational on a continuous basis) and around 110 different atmospheric variables are measured, among the most important:

- eight cloud profile variables, twelve aerosol profile variables, twelve aerosol in situ variables, about 75 different trace gases measured at the surface,
- 30 different methodologies, time resolution ranging from seconds to 1 week(s),
- NRT data available is available from 25 sites and for around ten variables,
- many in situ sites are collocated with ICOS,
- One of the sites, Zeppelin, is also contributing to SIOS.

EISCAT_3D – European Incoherent Scatter Scientific Association

EISCAT_3D is an ESFRI Landmark. It is in its implementation phase and the first set of data is expected to be collected in 2021. However, the EISCAT Scientific Association has operated incoherent scattering radar systems since 1981 and thus has scientific data covering a time span longer than 35 years. The data collected are incoherent scatter radar data. The bulk of the stored data consists of auto-correlations (as time lags, in a number of view directions, ranges, and sites). Data derived from these auto-correlations are the physical parameters (n_e, T_e, T_i, v_d, etc.) describing the ionosphere. The EISCAT systems also produce specialised data including for instance observations of meteors and space debris. EISCAT_3D will in its initial stage of construction include three sensor sites in northern Scandinavia observing a common volume in the uppermost parts of the atmosphere. The storage and access to data from the radar system will be handled through a centralised portal.

IAGOS – In-service Aircraft for a Global Observing System

IAGOS performs in situ observations on mobile platforms (airliners), measuring about ten variables (CO, O_3, H_2O, NO_x, NO_y, cloud particles, aerosols, CH_4, CO_2), and conducting aircraft measurements (air temperature, wind, etc.). The data are available in high time resolution, meaning observations every four seconds. Two packages are available: Package 1: including one instrument per variable (O_3, CO, H_20, clouds) and Package 2 optional: measuring one variable (NO_y, GHG, aerosols). The same level of quality is assured, whereas uncertainties are specific to each instrument. Data are available in Near-Real-time (NRT) within three days for Copernicus: data assimilation, model validation; and planed in Real-Real-time (RRT), within three hours delay but only for vertical profiles.

The IAGOS Data Centre is operational since July 2011. The IAGOS RI is based on former projects, namely MOZAIC (1994-2014) and CARIBIC (since 1997). The number of users per month evolves around 15. The type of users being mainly operational services (CAMS) and academic scientists (trends, model validation and assimilation, process studies, satellite validation). The infrastructure currently counts a global coverage of eight airliners.

ICOS-Atm – Integrated Carbon Observation System ICOS Atmosphere

ICOS Research Infrastructure is coordinated and integrated by the ICOS European Research Infrastructure Consortium (ERIC). ICOS ERIC was established in 2015 and is operational as ESFRI Landmark infrastructure since 2016. The first year of observations is 2016. Earlier data is available starting 1985 from earlier EU projects like CarboEurope, CHIOTTO, InGOS, ICOS Inwire and ICOS PP. ICOS is a multi-domain infrastructure that provides high quality long-term observations that support carbon cycle science from the atmosphere, ecosystem and marine domains. The Atmospheric Thematic Centre gathers all raw data from the atmosphere domain and processes this data to automated quality controlled NRT data (delay one day),

that is provided to Copernicus. They also produce yearly releases of final fully quality controlled data. All data and metadata is stored in the ICOS central repository at ICOS Carbon Portal and is available from there.
Variables for the Atmospheric part of ICOS are in situ time series of greenhouse gases at 21 existing atmospheric sites today, around 30 sites will be available by the end of 2020. Several sites are collocated with ACTRIS. Several sampling heights are often available along mast often exciding 100 m height. Continuous data is available for the following variables: CO_2, CH_4, CO, meteorological data. On a weekly basis, flasks are sampled for the following variables: CO_2, CH_4, N_2O, SF_6, CO, H_2, ^{13}C and $^{18}0$ of CO_2.
The instrumentation used includes infrared spectroscopy for the GHG, with stringent QAQC multi-level procedures. NRT data are available on a daily basis.

SIOS – Svalbard Integrated Arctic Earth Observing System

SIOS regional observing system entered its operational phase in January 2018. It includes distributed data management with contributing data centres hosting the data. The central node harvests discovery metadata and builds services on top of the data, these are depending on the documentation standard used for the data as well as the availability of standardised web services at the data centres. Currently harvesting from 4 data centres, 3 further data centres are in the process of being integrated. However, the RI is also relying on existing datasets and sites, for which some very long time series are available. The treated variables and types of data are in situ and remote sensing observations. Real time data are available. The number of core observation facilities is not yet decided. Data may be connected to other RIs such as ACTRIS, ICOS or operational programs such as WMO-GAW, EMEP. Currently the number of registered users of the portal evolves around 170.

Essential Climate Variables as common denominator for the Atmospheric subdomain

Essential Climate Variables (ECVs) are physical, chemical or biological variables identified to critically contribute to the characterisation of the Earth's climate. They are needed to understand and predict the evolution of climate, to guide mitigation and adaptation measures, to assess risks and enable attribution of climate events to underlying causes, and to underpin climate services. They are required to support the work of the UNFCCC and the IPCC.

In ENVRI-FAIR ECVs are used to frame the ensemble of variables targeted for FAIRness analyses. The atmospheric ECVs that are covered by the Atmospheric subdomain in WP8 are:

- **At the surface:** air temperature, wind speed and direction, water vapor, pressure, precipitation, surface radiation budget
- **In Upper-air:** temperature, wind speed and direction, water vapor, cloud properties, Earth radiation budget, lightning
- **Atmospheric composition:** carbon dioxide, methane, other long-lived greenhouse gases, ozone, aerosol, precursors for aerosol and ozone

For the atmospheric subdomain it is important to also take into consideration additional variables included in the WMO-GAW programs and variables characterising ionospheric conditions. In total it amounts to some 120 atmospheric variables covering composition, optical and physical properties and characteristics from ground level to ionosphere. These variables constitute data with different level of FAIRness in the various RIs of the Atmospheric subdomain.

4. Approach applied for gathering FAIRness information

This first self-assessment of FAIRness at the Atmospheric RI level is based on an extended version of the statements given in Figure 1. The raw data for the FAIRness assessment was collected via online questionnaires. To establish the questionnaires, it was decided at the ENVRI-FAIR Kick-Off meeting in January 2019 to cooperate with the GO FAIR initiative that has developed analytical methods for assessing FAIRness of data and services. WP5 and WP7 were also instrumental in developing these questionnaires.

The resulting answers from the five WP8 Atmosphere RIs to a first questionnaire are given in Appendix B of this document. Following up this first questionnaire, a second one was sent out to the RIs, aimed at collecting machine readable information. It was also more detailed and more technical. The exploitation of the questionnaire is being done in WP5 by B. Magagna et al. and will be reported in ENVRI-FAIR WP5 Deliverable 5.1 by end of 2019.

Furthermore, two internal ENVRI-FAIR meeting were organised in direct link to FAIRness assessment:

- in Amsterdam on June 11,12, organised by WP8
- in Lund on Oct 30, organised by WP5

Also, the WP8 coordination organised internal WP8 regular teleconference on a six-week basis. It also participated in the regular WP5-WP7 teleconferences.

5. FAIRNESS assessment: analysis of strengths and gaps with considered solutions at the individual RI level

Annex B of this document gives the answers of the five Atmospheric RIs to the first FAIRness assessment questionnaire. Based on this material, this section provides a synthesis in terms of strengths and gaps in FAIRness for the individual RIs together with proposed solutions to improve FAIRness.

ACTRIS

Strengths

- For two out of the five data curation units in ACTRIS, dedicated metadata catalogues exist using standardised protocols and standards for metadata and data exchange via machine to machine interfaces.
- All ACTRIS level 2 data, across the units, is findable via a common human user web interface.
- For three out of the five data curation units, machine to machine access is possible, either through standardised protocols or custom REST APIs.
- All data centre units follow or aim at following the Climate and Forecast (CF) standard for vocabulary/names, and three out of five data curation units are providing data in the same file format (NetCDF), in addition to their legacy file formats.
- For two out of the five data curation units, centralised processing assure the full traceability and reprocessing capability.
- Detailed descriptions of workflows are available for each data centre unit, making the task of linking provenance information to the metadata easier.

Gaps and planned activities

Findability

Findability includes the aspects of indexing data in relevant searchable resources, persistently identifying the data, and describing the data with rich metadata.

Currently there is no dedicated metadata catalogue for all ACTRIS data. Metadata catalogues exist, but on data centre unit level as a component of the individual data repository. The current way to access ACTRIS data is through the ACTRIS web portal, a user interface on top of a metadata database that collects metadata from the ACTRIS data centre (DC) units: In Situ, ARES, CLU, and GRES via custom APIs. In the future, ASC (Atmospheric Simulation Chamber) will be added as extra DC.

In the future, the aim is to collect all ACTRIS metadata into a single metadata catalogue, providing discovery metadata for all ACTRIS level 2 data using unique and persistent identifiers, and index the metadata resources in different data discovery portals and repositories, including data discovery portals like WIS, WIGOS, GEOSS, and SIOS.

In addition to this, the plan is to also include the Atmospheric Simulation Chamber unit of ACTRIS into the metadata catalogue and data portal of ACTRIS.

In ACTRIS, primary data identification will happen at the level of the data repositories. Data products will receive persistent identifiers (DOIs) at fixed granularity, which will be comparable across ACTRIS wherever possible (e.g. one primary DOI per one data submission per individual instrument). This function is presently being implemented across ACTRIS.

In addition, several ACTRIS DC units currently offer DOIs for pre-defined data collections. This function will be moved to the ACTRIS data portal, where users will be able to coin DOIs for self-defined data collections.

Accessibility

Accessibility includes the aspects of retrieving data by identifier, using an open and standardised protocol, as well as long-term availability.

Currently there is no standardised machine actionable communications protocol used for providing access to all of the ACTRIS data. Access to all ACTRIS data is only provided via the ACTRIS web portal and web interface. Still, some data centre units provide access to data via standardised communications protocols as a component of the individual data repository. Some of the units have implemented Thredds with OpenDAP protocol and/or Open Archive Initiative Protocol (OAI-PMH). There are also some units using REST APIs for machine to machine interaction.

The plan is to implement a standardised solution for accessing all ACTRIS level 2 data, allowing machine to machine access to all ACTRIS level 2 data through a standardised communications protocol.

Interoperability

Interoperability comprises the use of openly available and well-documented vocabulary and ontologies.

Since there is no centralised ACTRIS metadata catalogue nor there are specific access protocols supporting ACTRIS data, interoperability is not addressed in an harmonised way in the ACCESS unit of the data centre. There are solutions supporting interoperability of ACTRIS data at the data centre unit level, for some of the units.

The goal is to harmonise the efforts of the individual data curation units, providing standardised metadata, using broadly accepted vocabularies for all ACTRIS level 2 data, implementing iso19115/iso19139 and CF standard vocabulary, in addition by making sure that the ACTRIS data centre uses schemas that are available in common registries. This will be done in the ACCESS unit, and on the data curation units.

The topic of domain-specific vocabularies and ontologies will be worked on in the ENVRI-FAIR project.

Reusability

As of today, no license has been granted on ACTRIS data. Provenance information exists to varying degrees, but is not standardised.

The new and approved ACTRIS data policy recommends using a CC-BY 4.0 Attribution license. It remains to be decided what ACTRIS data it will cover. For provenance information, the aim is to follow domain relevant community standards and best practices.

EISCAT_3D

The EISCAT_3D system is under construction and the first data is expected to be obtained at the end of 2021, which means that there are no EISCAT_3D data available at the present. However, EISCAT Scientific Association has been operating incoherent radar systems and collected data since 1981. The FAIRness assessment here is based on the status for the present data sets.

There are two different fundamental types of archived EISCAT data: Low-level data which consists of receiver voltage levels and spectral data, and high-level data consisting of the ionospheric physical parameters. These data types are handled in slightly different manners

Strengths

- EISCAT Scientific Association owns all data produced by its facilities.
- The existing EISCAT data policy is an agreement signed by all EISCAT member countries, and its general aim is to have a large degree of FAIRness for the data produced by the EISCAT facilities.
- With data policies already in place and full control of the data management, there are no fundamental formal barriers to implement a FAIR data management system to handle the EISCAT_3D data volumes when the new radar system is ready and operational.

Gaps and planned activities

Findability

- No persistent identifiers are used for EISCAT data
- The EISCAT data has a weak metadata registry, which is not following any of the established standards

Accessibility

- There are no standardised solutions in general for the access of the present EISCAT low level data
- The system to access high-level data (Madrigal) is not known outside the international geospace community
- Work is needed on the authentication and authorisation to cover the different foreseeable cases

Interoperability

- The only standardised metadata at present is the time of the experiment
- Work has started to ensure that metadata will be included, enabling interoperability standards

Reusability

- Gaps exist in the standardisation in areas regarding how the data is produced, processed and validated
- Work has started to ensure that metadata will be included, enabling provenance attributes

The main prioritisation for all different aspects of FAIRness is to work towards an implementation following standards adopted by the atmospheric user community.

IAGOS

Strengths

- The IAGOS Data Portal proposes a rich and efficient search service for data, with possibilities of download and visualisation
- Metadata and data products are well findable and accessible through the IAGOS Data Portal
- A IAGOS metadata catalogue is implemented. Metadata is available in JSON and XML (19115 / INSPIRE compliant) and through a CSW endpoint.
- IAGOS has a simple workflow compared to other RIs
- Data and metadata are homogeneous for the whole RI
- IAGOS supports many protocols to access data and metadata
- Workflows are in place for metadata and data control before publishing (using Apache Camel)
- Used and standard data format: NetCDF
- User authentication system for monitoring use

Gaps and planned activities

Findability

- IAGOS data are not indexed yet in external data portals. Indexation in portals such as GEOSS or WIS is planned.
- Semantic search is currently not available on the IAGOS data portal. It’s planned for the coming years, using opensearch standard (low priority)

Accessibility

- The IAGOS Data Centre has no DMP yet. It is in progress and planned for the end of 2020.
- The IAGOS Data User license is not clearly defined. Discussions are in progress and the license should be defined by the end of 2020. It is connected to the DMP action.
- The IAGOS repository is not certified yet. It is planned for 2021. We will apply for a CoreTrustSeal certification. In that context IAGOS is supported by a FAIRsFAIR initiative.
- No explicit persistency policy is available in metadata. It should be resolved by the repository certification.
- IAGOS needs to improve machine-to-machine access. It will be done by the implementation of RESTful services for data and metadata access (in progress) and the implementation of standard services such as OGC, OpenDAP (started)
- Authentication and authorisations are managed locally by the Data Centre. Activities are ongoing to implement a system allowing connections with ORCID, etc. It will allow the authentication for the UI interface and the web services.
- Data access is open but users need to register and connect to access the data. It is necessary to analyse use of data, user behaviour, etc.

Interoperability

- The metadata schemas are not registered in common registries.
- Not all categories in metadata are marked up with vocabularies.
- No standard vocabularies are used. IAGOS needs the publication of vocabularies developed by the French Atmospheric community for parameters, instruments and platforms names

- No controlled vocabularies are used yet. It needs to be done within the subdomain.
- Implementation of standard services is in progress: OGC, OpenDAP using the server THREDDS

Reusability

- No PID used for dataflow management, except at product level (DOI): it is planned to use the ePIC system. Datasets, sensors, etc. will have PID assigned.
- No provenance information is included into the metadata. It is planned to use PROV-template
- The metadata is not yet machine interpretable
- A compliance validation service is missing: it is planned to propose data format checkers

ICOS-Atmosphere

Strengths

- Harmonised data formats
- Strong data identification: raw data and aggregated data (hourly means) are minted with Persistent Identifiers for eResearch (ePIC) PIDs: vital for transparency, reproducibility and reusability
- Focus on attribution of data to data providers coupled to data usage statistics as part of the metadata
- Data and metadata are findable and accessible
- Linked data implementation, RDF and SparQL endpoints
- Open data access and user registration
- Portal is multi-domain integrating atmosphere, marine and ecosystem data in one system using a common ontology

Gaps and planned activities

Findability

- develop search capability with elaborated filters (bounding box/ conditions/ keywords)
- on demand level data merging (lev2 + nrt)
- enhance metadata on landing pages with downloadable sheets in XML/HTML format in ISO norm
- extend search area by including related publication/ dataset/ image/ web resource/ software in search options

Accessibility

- Discovery of metadata for the data resources (e.g. type of instruments)
- subsetting of the data sets (i.e. by period of time)
- NetCDF files

Interoperability

- Provide the ICOS Atmosphere vocabulary
- Ontology for the ENVRI atmospheric subdomain
- Connect to Global Earth Observation System of Systems (GEOSS)

Reusability

- Provide the big steps of the provenance information. The information is available and structured at the provider level (ATC) but not at the distribution level (Carbon Portal).

SIOS (atm)

SIOS is a regional observing system for long-term measurements in and around Svalbard, with the scope of integrating existing data centres through a distributed data management system (SDMS) which harvests, indexes and makes available data from different contributing centers. Each data centre has its own procedures and technical solutions tailored to the needs and the use of that data centre. SIOS is multidisciplinary and promotes integration of data through a dedicated working group involving all partners.

Strengths

- SIOS web portal has a searching interface where datasets can be searched by means of several filters, i.e. Full Text Search, Time Interval, Geographical location, Institutions, Principal Investigator, Science Keywords.
- Several access types can be provided if made available by the data centres.
- Well established standards are used for accessing metadata (OAI-PHM) and data (http/OPeNDAP/WMS).
- Services (visualisation, shopping-cart, subsetting, variable extraction, reprojection) are offered.
- Interoperability at the metadata level is at a relatively mature stage, as datacenters support standards with controlled vocabulary for discovery metadata.
- SIOS is promoting free and open access.

Gaps and planned activities

Findability

- UIDs (PIDs better) are required, DOIs recommended. SIOS relies on repositories as stated in the SIOS Data policy. Not all centers have DOIs. DOIs should be in place at least for core data, i.e. for variables that are critical to answer the key research questions as defined within SIOS.
- OAI-PMH is still not fully implemented in some data centres. Planning to complete the implementation.
- Improve searching interface. Planning to provide faceted searches.
- Some work still has to be done to have correct indexing of metadata. Indexing tool (Java) is not optimal. Manual contribution is still very high. Planning to replace the Java-based indexing tool with a lighter and more sustainable python-based client (pysolr) for SolR.
- Planning to look into Semantic Search in line with the Polar Semantic Working Group, ENVO, Sweet ontology and RDA efforts.

Accessibility

- Not all data centres use standardised protocols (e.g. own REST interface to access data, but has discovery metadata in a standard protocol).
- OPeNDAP is not fully implemented in some data centres. Planning to complete the implementation and provide OPenDAP access.
- Implementation of authentication schemas is not a priority. Planning to follow the developments within WP8, in particular with respect to eduGAIN and ORCiD.

Interoperability

- Better use of controlled vocabularies: improve on semantic translations, mapping of keywords and vocabulary. Planning to actively participate in the vocabulary/ontology working group as organised within WP8.
- At the data level the CF standard names and CF conventions are used, but not at all contributing centers. Planning to work on harmonisation of this.
- Planning to work on WIGOS (WMO Integrated Global Observing System) metadata standards for station catalogue, for better integration.

- Interoperability at the data level is very low, if not absent, due to lack of standardisation of metadata. Planning to actively participate in the vocabulary/ontology working group as organised within WP8.
- Planning of implementing OGC CSW (providing ISO19139) and possibly OpenSearch.

Reusability

- Provenance not widely explored. According to CF convention the element history is used. Planning to actively participate into the provenance working group to explore possibilities for provenance attributions, particularly for embedding this in the data.
- Data Licence is not always referenced in a clear manner. Planning to harmonise this.
- SIOS provides a dataset validation service for netCDF, but this could be improved.

6. Proposed action plan

This first self-assessment of FAIRness enabled WP8 atmosphere subdomain as a whole to agree on a list of actions to improve FAIRness, summarised here in two groups corresponding to two different times for implementation (short term, and medium term). More details are given in ENVRI-FAIR MS35: "Implementation plan, defining the starting point". A detailed implementation plan of these actions will be developed in Task 8.3 of WP8 in collaboration with WP5 and WP7.

Actions for short term implementation

- Consolidation of consistent use of PIDs throughout data production workflow
- Common standard interfaces for metadata and data
- Indexing of data resources in WIS, GEOSS
- Domain vocabulary / ontology for observed parameters, discovery and use metadata
- Common use of authentication schemes
- Consistent documentation of provenance throughout data production workflow
- Recommendations for licenses on metadata and data
- Semantic search for atmospheric ENVRI RI user interfaces

Actions for middle term implementation

- Common metadata standards and interfaces for use of metadata
- Machine-readable license and attribution metadata.
- Common strategy for structured search interfaces, including common base set of searchable items
- Traceable post-production user feedback services.
- Data indexing in further data portals
- Standards for RESTful APIs for metadata and data.
- Common interfaces for data, facilitating machine readability of data, e.g. in Virtual Research Environment (VRE)s

7. Two main challenging topics: Semantic web and Provenance

Further to the proposed action plan described above, this section provides introductory general considerations on two topics identified in the Atmospheric subdomain, to be especially challenging on the road to increasing FAIRness. These are:

1. The use of semantic web, ontology and vocabularies in becoming more FAIR
2. Provenance in the framework of ENVRI

Further references to deepen the subjects are found in the text.

The use of semantic web, ontology and vocabularies in becoming more FAIR

Using the semantic web means using the web as a database. It implies migrating from a web where data are more or less static, community and/or portal-dependent and strongly linked to the entities which produce them, to a web where data can be directly interpreted by machine, easily (re)used in a trans-community horizon, all with a reinforced authorship. Adopting the FAIR principles is a complex task that involves not only knowledge of data, but also awareness of metadata, protocols, policies, and community agreements. The FAIR principles have established the importance of using standards vocabularies or ontologies to describe FAIR data and to facilitate interoperability and reuse. And the Semantic Web offers the technologies to apply FAIR principles.

The FAIRification process consists of the following steps (See https://www.go-fair.org/fair-principles/fairification-process/ for a workflow overview):

- First step is to analyse data to model it. Most infrastructures use relational database to store their data and the relational schema provides information about the dataset structure, the types involved (the field names), cardinality, etc. This analysis allows definition of the structure of the data, the relations between the data elements.
- Then a semantic model has to be defined. It describes the meaning of entities and relations in the dataset accurately, unambiguously, and in a computer-actionable way. A good semantic model should represent a consensus view in a particular domain, for a particular purpose. Semantic models often contain multiple terms from existing ontologies and vocabularies. A vocabulary is a computer-readable file that captures terms, their URIs, and descriptions. An ontology is a formal representation using a set of concepts within a domain and the relationships between those concepts. It usually uses standard vocabulary with hierarchies, meaningful relations among concepts, and their constraints. These conceptual models allow us to classify the data models and data items using the provided terms, concepts, and conceptual structures. One shall strive to find existing ontologies and reuse existing semantic resources (thesauri, ontologies, formal ontologies, …) and in the worst case scenario, to create a new one. There is a rapid increase of the number of ontologies and semantic repositories, FAIR principles are now also applied to ontologies themselves and associated semantic artefacts (controlled vocabularies, thesauri, ontologies, codelists, …) so that they can be easily findable.
- Next step is to make data linkable. The non-FAIR data can be transformed into linkable data by applying the semantic model explained in the previous step. Currently, this is done using Semantic Web and Linked Data technologies. This step promotes interoperability and reuse, facilitating the integration of the data with other types of data and systems. Linked data is a solution for integrating heterogeneous and multi-disciplinary data, as in the atmosphere subdomain of ENVRI-FAIR.
- An important step is to define proper and rich metadata for the dataset. Rich includes the possibility that metadata can be added after data acquisition, a posteriori, to better qualify dataset.
- Finally, one needs to deploy FAIR data resource, together with relevant metadata and license, so that the metadata can be indexed by search engines and the data be accessed, eventually using authentication and authorisation if required.

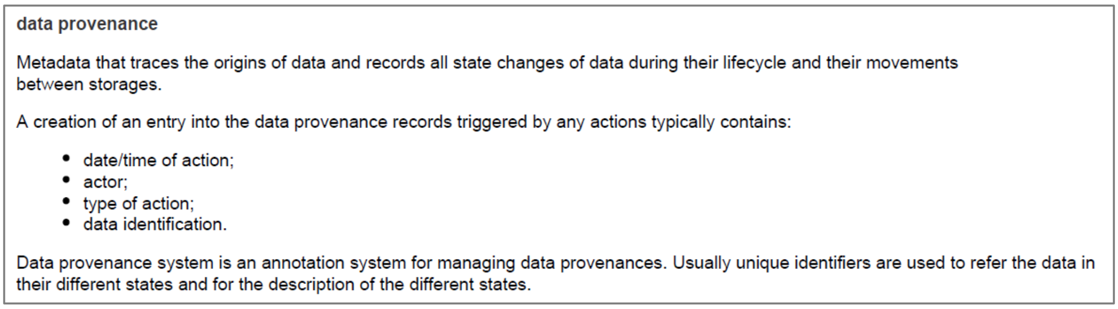
Provenance in the framework of ENVRI

In this sub-section we give general consideration about Provenance in the framework of ENVRI by providing a short and subjective summary, in quoting the recommended deliverable 8.5 produced during the ENVRIplus project. This is made for readers wanting to get acquainted with basic aspect of Provenance and get a review of available tools related to Provenance and best practices that can be found in the ENVRI community. The idea here is to promote the further consultation of the deliverable document available on the web at http://www.envriplus.eu/wp-content/uploads/2015/08/D8.5-Data-provenance-and-tracing-for-environmental-sciences-system-design.pdf.

The ENVRIplus deliverable 8.5 is entitled “Data provenance and tracing of services for environmental sciences: system design” it was produced as part of ENVRIplus WP8 on “Data Curation and Cataloguing”. The deliverable production was led by Barbara Magagna from Umweltbundesamt GmbH (Environment Agency Austria).

The W3C gives the following definition for data provenance: information about entities, activities, and people involved in producing a piece of data.

In the Information Viewpoint of the ENVRI Reference Model (RM), data provenance is defined as:



data provenance

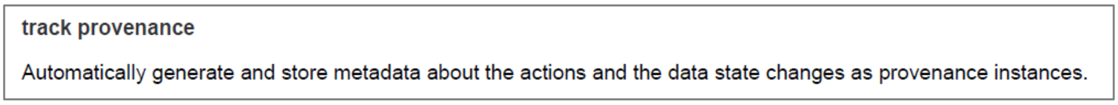
Metadata that traces the origins of data and records all state changes of data during their lifecycle and their movements between storages.

A creation of an entry into the data provenance records triggered by any actions typically contains:

- date/time of action;
- actor;
- type of action;
- data identification.

Data provenance system is an annotation system for managing data provenances. Usually unique identifiers are used to refer the data in their different states and for the description of the different states.

The Information Viewpoint also defines one action related to provenance management that is “track provenance”. This action is for the entire data lifecycle as an activity that must be performed whenever there is a change in the state of a data or metadata object. The purpose is to make evident the need to implement provenance tracking as a continuous, parallel activity within the data lifecycle.



track provenance

Automatically generate and store metadata about the actions and the data state changes as provenance instances.

The main motivation for provenance is that it is central to the requirement that scientific research should be reproducible improving upon credibility and trustworthiness. A prerequisite to reproducibility of scientific conclusions is traceability.

Data (but not only, also software versions, workflows, etc.) are useful if accompanied by context on how they are captured, processed, analysed, and validated. With also other relevant information that enables interpretation and use. This is what provenance is about. Originally, Provenance was used to keep track of the chain of ownership of cultural artifacts e.g. paintings.

Provenance is not metadata. It is only a kind of metadata if it gives information on how/from where the resource was derived. Ex: file size can be a metadata but it is not provenance-metadata.

Provenance for processing is highly facilitated by the presence of workflows that automates the process. Nevertheless, for activities that make use of notebooks for processes that rely on manual intervention (as opposed to automated processes), ENVRIplus D8.5 mentions, for example, the existence of add-ons to spreadsheets enabling users to record their actions into a provenance log. Still the provenance record has to be manually constructed.

Many scientists do not operate within the confines of a particular workflow system or data processing platform, preferring to run their own scripts, typically in their own environment (e.g. their office laptop). In this case there are still ways to (partially) automate the generation of provenance data. One way is to use tools that extract provenance data from

specially annotated scripts, e.g. the NoWorkflow system by Murta et al. (2015)[2] for retrospective provenance and the accompanying YesWorkflow system by McPhillips et al. (2015)[3] for prospective provenance. Many scientific workflow management systems supporting provenance exist: e.g. Kepler, Pegasus, Taverna and dispel4py from the seismology community, see references in ENVRIplus D8.5 document. Provenance analysis is more difficult when parallel processing is involved but here again some solutions exist.

ENVRIplus established that "The use of workflow systems is already established within the environmental sciences, supported by many research infrastructures. LifeWatch makes use of Taverna, for example, while the VERCE project (operating as a contributor to EPOS) implemented provenance facilities based on W3C PROV linked to the dispel4py workflow description framework (Atkinson et al., 2015[4]), which can be queried via a custom provenance explorer GUI as described below." In the ENVRI community the two advanced RIs on the overall topic of Provenance are EPOS and IS-ENES.

Before providing a general technology review ENVRIplus D8.5 provides a list of "Challenges" in Provenance, summarised here:

- It should be accessible at runtime.
- One should pay attention to granularity tradeoff (provenance data may exceed the size object itself …)
- One should use standard for domain semantics and annotations.
- The provenance information should be interoperable and trustworthy (mention of blockchains).
- Incompleteness and fragmentation puts provenance at risk
- Finally, provenance information should be easy to use including well-arranged presentation and visualisation.

Fortunately, ENVRIplus D8.5 mentions the existence of a number of core standards for provenance, though the two highest profile standards are the Open Provenance Model (OPM) and W3C's PROV recommendation. A number of useful tools have been made available online for use by researchers and other users of provenance data in order work to with these standards. There is already a widely used and acknowledged standard for provenance (W3C –PROV documents) one can rely on. The PROV ontology provides a generic model for implementing provenance applications that can represent, exchange and integrate provenance information generated in different systems and under different contexts.

Discovery and retrieval of Provenance data

In order to be readily available for analysis, provenance data must be provided in a findable and accessible way. Dedicated services to store provenance documents, such as ProvStore, often include individual means to discover and view hosted resources, such as via a dedicated REST API. Due to the potential complexity of large provenance graphs, their direct visualisation can easily exceed the capacity of the medium (paper, screen) and/or the viewer. A number of different techniques have therefore been applied in order to tackle this visual overload. They can be regrouped into two main categories: graph summarisation and semantic zoom. Graph summarisation uses tools that first transforms the provenance information and then output synthetic information e.g. information on chronological vicinity of

[2] Murta, L., Braganholo, V., Chirigati, F., Koop, D. and Freire, J., 2015: noWorkflow: Capturing and analyzing provenance of scripts, In: Provenance and Annota on of Data and Processes, Ludäscher, B. and Plale, B. (Eds.), Springer International Publishing, pp. 71–83.

[3] McPhillips, T., Song, T., Kolisnik, T., Aulenbach, S., Belhajjame, K., Bocinsky, K., Cao, Y., Chirigati, F., Dey, S., Freire, J., et al., YesWorkflow: A user-oriented, language-independent tool for recovering workflow information from scripts, arXiv preprint arXiv:1502.02403.

[4] Atkinson, M., Carpené, M., Casarotti, E. et al., 2015: VERCE delivers a productive e-science environment for seismology research, Proceedings of the 2015 IEEE 11th International Conference on e-Science, 224–236.

the recorded events that allows for example identification of data re-use between different users and interactions between different workflows.

The recommendation section 6 of the ENVRIplus D8.5 lists further important aspects to take into account by ENVRI RIs when implementing a provenance framework. Quoting:

Identification and citation. The persistent identifiers assigned to data collections and other resources provide the preferred way to refer to entities involved in various forms of activity, and thus in provenance traces. It is important that the relationships between the digital objects (assets) are recorded such as: a new version generated by a particular piece of software executed by a particular person.

Curation. Curation activities should include provenance management; provenance traces can be used to locate resources and judge their condition with regard to accessibility and preservation. Conversely, provenance should provide the graph of relationships between curated digital objects (assets).

Cataloguing. The generation of metadata for external (joint) catalogues should be based partly on provenance records, whether integrated in the source metadata or elsewhere in the source research infrastructure, especially where mapping between metadata standards is involved. The full provenance trace of a given resource should be accessible via any catalogue that contains that resource's metadata. Changes to catalogues should also result in provenance traces that can be used to assess the catalogue themselves. There are particular implications when metadata from a RI catalogue is harvested into a common catalogue – in itself a provenance action - but also requiring the provenance traces to be harvested using the ENVRI canonical format.

Processing. All activities on the part of a common processing platform should be recorded in the provenance trace of the processes themselves and that of any datasets modified or new data created. This is constrained by catalogue information concerning rights, licences and appropriate security and privacy constraints.

8. Conclusion

A first self-FAIRness assessment was done by the five atmospheric RIs of the ENVRI-FAIR WP8: ACTRIS, IAGOS, ICOS-Atmosphere, EISCAT-3D and EISCAT. It used a questionnaire based on the 15 concepts of FAIRness described in Wilkinson et al. (2016). The assessment shows that all of these RIs find room for improvement in all four of the FAIR domains: Findability, Accessibility, Interoperability, Reusability.

After a thorough gap analysis, the WP8 as a whole could identify a common plan for the following concrete developments where action could start immediately (see section 6):

- Consolidation of consistent use of PIDs throughout data production workflow
- Use of common standard interfaces for metadata and data
- Indexing of data resources in WIS, GEOSS
- Use of domain vocabulary / ontology
- Common use of authentication schemes
- Consistent documentation of provenance throughout data production workflow
- Development of semantic search for atmospheric ENVRI RI user interfaces

These actions are further detailed in ENVRI-FAIR MS35 and an actual implementation plan is currently being drafted under Task 8.3 of the project. An upcoming WP8 meeting is organised Dec 11-12, 2019 in Paris to further elaborate implementation plans for these actions and discuss tools to be used for developments. For this step of the work, the guidance of the

ENVRI-FAIR transverse WP 5 and 7 will be very important to make sure that, if not common tools are used always, at least interoperability is ensured.

Furthermore, two topics that were found to be especially challenging are introduced in the last section of the document. These deal with: the use of semantic web, ontology and vocabularies and exhaustive provenance.

Appendix A: Glossary

ACDD	Attribute Convention for Data Discovery (for NetCDF)
API	Application Programming Interface
B2HANDLE	EUDAT minting, storing, managing and accessing persistent identifiers
CAS	Central Authentication Service
CC-BY	Creative Commons Attribution License
CDI	Common Data Index (metadata format and data access system by SeaDataNet)
CF	Climate and Forecast (semantics for NetCDF)
CMEMS	Copernicus Marine Environment Monitoring Service
COPERNICUS	A major earth observation programme run by European Commission and European Space Agency
CSR	Cruise Summary Report
CSW	Catalogue Service for the Web
DMP	1) Data Management Plan 2) Data Management Platform (WP9)
DOI	Digital Object Identifier
DSA	Data Seal of Approval
ECV	Essentia Climate Variable
EDIOS	European Directory of ocean Observing Systems
EDMED	European Directory of Marine Environmental Datasets (SeaDataNet)
EDMO	European Directory of Marine Organisations
EDMERP	European Directory of Marine Environmental Research Projects
EML	Election Markup Language
EMODNET	European Marine Observation and Data Network
EMSO	European Multidisciplinary Seafloor and water column Observatory
ENVRI	1) A environmental RI cluster FP7 project 2) Environment research infrastructures (in ESFRI level or upcoming) as a community
ENVRIplus	A environmental RI cluster H2020 project
EOSC	European Open Science Cloud
EOV	Essential Ocean Variable(s)
ERDDAP	NOAA developed science data server technology
ERIC	European Research Infrastructure Consortium (legal entity type)
EUMETNET	Grouping of 31 European National Meteorological Services
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable Accessible Interoperable Reusable
FAQ	Frequently Asked Questions
FORCE11	a community to help facilitate the change toward improved knowledge creation and sharing
GBIF	Global Biodiversity Information Facility
GCMD	Global Change Master Directory
GDAC	Global Data Assembly Center
GEMET	GEneral Multilingual Environmental Thesaurus
GEO	Group on Earth Observation (System of Systems)
GEOSS	Global Earth Observation System of Systems
GOFAIR	An international programme on FAIR implementation
GOOS BGC	Global Ocean Observing System Biogeochemistry Panel
GUI	Graphical User Interface

ICOS	Integrated Carbon Observation System
ICT	Information and Communications Technology
IMIS	Integrated Marine Information System
INSPIRE	Infrastructure for Spatial Information in the European Community
iRODS	Open Source Data Management Software
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
LW	LifeWatch
Marine-ID	Registration and authentication services for marine data services
MDA	Marine Data Archive
NetAPP	Hybrid cloud service
NetCDF	Network Common Data Format
NVS	NERC Vocabulary Services
NOAA	US National Oceanic and Atmospheric Administration
OAUTH	Open Authorization (standard)
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OBIS	Ocean Biogeographic Information System
ODIP	Ocean Data Interoperability Platform
OGC	Open Geospatial Consortium
OpenDAP	Open-source Project for a Network Data Access Protocol
ORCID	Open Researcher and Contributor ID
OWL	Web Ontology Language
PID	Persistent Identifiers
PROV-O	Web Ontology Language encoding of the PROV Data Mode
QA/QC	Quality Assurance/Quality Control
RDF	Resource Description Framework
RI	Research Infrastructure
RSS	Really Simple Syndication
SAML	Security Assertion Markup Language
SEADATANET	SeaDataNet pan-European infrastructure for marine data management
SME	Small or medium Enterprise
SparQL	SparQL Protocol And RDF Query Language
SWOT	Analysis on Strengths, Weaknesses, Opportunities and Threats
VRE	Virtual Research Environment
W3C	World Wide Web Consortium
WMO	World Meteorological Organisation
WoRMS	World Registry of Marine Species
WPS	Web Processing Services
YAML	Yet Another Mockup Language

Appendix B: Questionnaire 1: responses from WP8 Atmosphere

Research Infrastructure Name	Please provide the URL of one of the datasets in scope for your answers	Please provide the URL to the discovery portal in which the dataset can be downloaded	Which repositories do you use for data?	Which repository software is being used?	Which repositories do you use for metadata?	Do your repositories use PIDs? If so which PID systems?
ICOS	https://data.icos-cp.eu/objects/-ffoiHjX5NDN0Vq_fKuVmas0	https://data.icos-cp.eu/portal	Local (ICOS servers in Lund), B2SAFE (EUDAT/CDI) at CSC, iRODS at SNIC (Swedish national storage system)	HTTP API for B2STAGE access to B2SAFE, iRODS Java client library Jargon 4.3.0.1	local (at ICOS Carbon Portal production server) versioned RDF triple store	Yes. Handle system generic PIDs and Datacite DOIs. In both cases we have ICOS-specific prefixes.
IAGOS	http://doi.org/10.25326/06	http://www.iagos-data.fr	Local repository at the Laboratoire d'Aérologie in Toulouse, France Local database servers with discs Managed by IAGOS	Homemade : NoSQL database (MongoDB) dupplicated files archive linux file system	Metadata stored in local MongoDB database and in data files headers Dupplicated in the metadata repository of the French Atmospheric Cluster AERIS for harvesting (Geonetwork)	Use of DOI for qualified datasets collection (L2 to L4) Small amount of datasets (collection level) Use of DOI fragments planned in order to manage subsets of the collections The IAGOS repository doesn't use PID yet for workflow management
EISCAT	https://www.eiscat.se/schedule/tape2.cgi?exp=manda_zenith_4.00v_SW&date=20140210	https://www.eiscat.se/schedule/	Local storage at EISCAT headquarters	Home-made (by EISCAT; apache, python, sql)	Internal	No, but SQL resource ID exists
EISCAT	https://www.eiscat.se/madrigal/cgi-bin/madExperiment.cgi?exp=2014/eis/10feb14&displayLevel=0	https://www.eiscat.se/madrigal/	Local storage at EISCAT headquarters (EISCAT data); Madrigal is a distributed world-wide data base	Home-made (by international team; apache, python, C, …)	Distributed between madrigal sites	No
SIOS	http://thredds.met.no/thredds/catalog/met.no/observations/stations/catalog.html?dataset=met.no/observations/stations/SN99938.nc	https://sios-svalbard.org/metadata_search	SIOS is a distributed system. This specific dataset is hosted by the Norwegian Meteorological Institute.	Inhouse for this dataset and most other datasets in SIOS hosted through partner repositories.	SIOS is centrally harvesting discovery metadata from partner repositories using OAI-PMH. The preferred structure is GCMD DIF, but ISO19115 is also used. Support for OGC CSW/ISO19115 with specific terminologies for parameter and URL descriptions are required.	SIOS relies on partner repositories. UIDs is a requirement, PIDs (DOI) is a recommendation currently.
ACTRIS	http://ebas.nilu.no/DataSets.aspx?stations=NO0002R&projects=ACTRIS&InstrumentTypes=dmps&components=particle_number_size_distribution&fromDate=2016-01-01&toDate=2017-12-31	http://actris.nilu.no	"Several solutions for various units within the DC. There are six units: ACTRIS data and services access unit (ACCESS) ACTRIS In situ data centre unit (In situ) ACTRIS Aerosol remote sensing data centre unit (ARES) ACTRIS Cloud remote sensing data centre unit (CLU) ACTRIS trace gases remote sensing data centre unit (GRES) ACTRIS Atmospheric simulation chamber data centre unit (ASC) ACCESS: This unit is a centralised metadata catalogue on local server In Situ data: http://ebas.nilu.no, centralised repository running on local server at NILU ARES: http://access.earlinet.org/EARLINET/ Centralised local repositories at DC CLU; http://cloudnet.fmi.fi, centralised repository running on local server at FMI GRES: in under the responsibility of the AERIS ASC: in Toulouse under the responsibility of the AERIS: in Paris under the responsability of the AERIS "	"Several solutions for various units within the DC: ACCESS: home-made, technology stack: oracle relational database, Perl, .NET In situ: home-made, technology stack: SYBASE relational database, Python, .NET, ARES: Database : PostgreSQL, MariaDB Interfaces home-made (PHP, Java, C++) & THREDDS, Processing: C, Python, FPC, Django CLU: home-made, technology stack: mariaDB relational database, Python, PHP GRES: Home-made, technology stack : mongoDB database, Vue.js ASC: Home-made, technology stack : mongoDB database, Vue.js"	"Several solutions for various units within the DC: ACCESS: Oracle relational database In situ: SYBASE relational database ARES: PostGre SQL & MariaDB Additionally metadata information are stored within the file CLU: MariaDB relational database GRES: mongoDB database ASC: mongoDB database"	"Several solutions for various units within the DC: ACCESS: partly, DOIs for secondary datasets (in ACTRIS denoted level 3 data) In situ: planned ARES: PID partially implemented: Unique PIDs are assigned to each product processed through the central processing tool A subset of ACTRIS ARES data is published with DOI on CERA CLU: planned GRES: UUID version 4 ASC: UUID version 4"

Research Infrastructure Name	Do you assign PIDs manually or automatically?	Which PID registration provider do you use?	Do you use the PID Record to store attributes about the data?	Are these repositories certified? If so, which methods are used?	Are repository policies mentioned at the website? If so, indicate the major ones.	Are your repositories registered in a registry? If so which registry?	Which persistency guarantees are typically given?
ICOS	Handle PIDs automatically, DOIs currently manually (but using custom-build user-friently web application).	DataCite for DOIs, Handle.net software server hosted by PDC at KTH.	Not for generic Handle PIDs. In case of DOIs we copy relevant attributes into the DataCite metadata catalog (schema version V4.1).	The ICOS Carbon Portal itself isn't (yet CTS) certified. However, our certification status might not be relevant, as we have outsourced our storage to data centers (EUDAT and SNIC) that are not certified.	What is a repository policy? Yes the ICOS data Policy is clear and described and available at the website, but does not cover the elements that are part of the B2SAFE policy, although we agreed with them for two replicates.	Yes, re3data.org, working on B2FIND, GEOSS, WMO WDCGG	By whom? ICOS has 20-25 year longevity goal.
IAGOS	Manually; work in progress within the French Atmospheric Cluster AERIS in order to assigne DOI and fragments automatically through the AERIS metadata repository	Datacite for DOI	Metadata for datacite mapped with IAGOS metadata	No French initiative planned (at local level in Toulouse and national level for AERIS). IAGOS will be part of it	No	No	No explicit guarantees given
EISCAT	N.A.	N.A.	N.A.	No	Yes, EISCAT "rules of the road", described in the EISCAT BlueBook	No	No
EISCAT	N.A.	N.A.	N.A	No	Yes, "These data are the intellectual property of the EISCAT Scientific Association Except where clearly noted as Common Programme (CP), use of these data is restricted to the original experimenter [...] for one year from the date of the experiment."	No	No
SIOS	Depends on the partner repository. SIOS does not mint directly, but through partner repositories.	See above answers, will depend on repository.		Some are some not, all are mandated.	Not at the SIOS website, neither is this covered by present SIOS guidelines.	Some are some not and often information is outdated. BTW, should be re3data.	Concerning what?
ACTRIS	"Several solutions for various units within the DC: ACCESS: DOIs for secondary datasets (in ACTRIS denoted level 3 data) In situ: planned automatically ARES: Automatically by script for the data processed centrally. The DOI assignment is manual CLU: planned automatically GRES: currently manually but planned automatically ASC: currently manually but planned automatically"	"ACCESS: BibSYS / DataCITE for DOIs InSitu: not clarified for general PIDs, BibSYS / DataCITE for DOIs ARES: ERA for DOIs no registration provider are used for the other PIDs , but this is planned CLU: Planned GRES: DataCITE for DOIs ASC: DataCITE for DOIs"	"ACCESS: planned In Situ: planned ARES: Yes: each dataset submitted to the centralised processing is identified by a unique PID, which is stored in a local database. Using this PID in proper database queries, it is possible to get all the information used for the processing. CLU: planned GRES: yes ASC: yes"	"ACCESS: no In Situ: no ARES: no CLU: no GRES: no ASC: no"	ACCESS: partially, data policy, data management plan, submission instructions, operating procedures In Situ: partially, data policy, data submission instructions, version control ARES: partially, data policy, data submission instructions, version control CLU: partially, data policy, version control GRES: partially, data management plan, submission instruction, data policy, … ASC: partially, data management plan, submission instruction, data policy, ...	ACCESS: re3data.org In Situ: WIS, GEOSS, re3data.org ARES: no CLU: re3data.org GRES: no ASC: no	ACCESS: implicitly, cross-funded, and RPO support In Situ: implicitly, funded by policy framework and RPO support ARES: not officially state this, but guaranties the persistency of the whole EARLINET database CLU: implicitly, funded by policy framework, and in-house support GRES: cross-funded ASC: cross-funded

Research Infrastructure Name	Which are the most popular data types used?	Which are the preferred data formats?	Do those formats include metadata headers? if so, which?	Do you provide search on data?	Did you register your schemas in a common registry?	Which metadata schemas are mostly used?
ICOS	time series, spatial raster data (3D-5D)	netcdf, csv	Yes, netcdf conforms usually to CF-1.4 conventions, CSV file have headers that are community specific and contain general metadata and column names and units etc.	ICOS doesn't provide searches inside datasets, as these typically contain mostly numbers. Searches for files containing specific variables (column names) are supported via queries to the metadata database.	No, not formally (e.g. in schemas.org). But ICOS data objects follow schemata hosted by ICOS that links to the data format specification definition in the ICOS ontology, which is openly accessible (in OWL) via the ICOS SPARQL endpoint (https://meta.icos-cp.eu/sparqlclient/).	rdf, inspire, iso19115, geo-dcat
IAGOS	Scientific data time series in popular scientific formats	NetCDF and NASA Ames (ASCII CSV)	Yes same metadata in the headers of the two formats describing the dataset (same as in the metadata repository + columns names, units, etc.) Homemade header	Yes through the web portal for individual users. Multiple criteria: time period, geographic area, variables Restful web services in progress for machine readable access Thredds access for WCS and OpenDAP in progress	No Format described on the IAGOS data portal: http://www.iagos-data.fr/#DataFormatPlace:	Use of a pivot format (JSON) developped by the French atmospheric cluster AERIS. No schema published, tools to convert to standards schemas like ISO 19115 Dupplicate into geonetwork, format ISO 19115, INSPIRE
EISCAT	Binary data	.mat (version 4)	Yes, instrument parameter block	No	No	SQL database
EISCAT	Plots of data	ASCII text files (preferred by users), other formats exist (HDF, Cedar)	Yes, if chosen by the user	Yes	No	Madrigal's own metadata schema
SIOS	Heterogeneous, moving towards standardisation for new data, but standard depends on discipline.	NetCDF/CF, JPEG, CSV, JSON/GeoJSON, ...	Wherever applicable we recommend NetCDF/CF served through OPeNDAP.	Yes, discovery metadata are checked, enriched and served through a SolR based engine.	Not structurally.	Assuming this related to discovery metadat, GCMD DIF, ISO19115 with specific requirements on terminology for keywords and URLs.
ACTRIS	ACCESS: implicitly, cross-funded, and RPO support In Situ: time series, ARES: height profiles and time series CLU: time-height profiles and time series GRES: time series, totoal columns and height profiles ASC: Chamber experiments, IR spectra, mass spectra	ACCESS:defined by primary repository In Situ: EBAS NASA Ames 1001, netCDF-CF ARES: netCDF CLU: netCDF-3 and netCDF-4 GRES: NASA Ames 2110, GEOMS data format ASC: EDF (ASCII format) for simulation chambers data JCAMP-DX (ASCII format) for IR and mass spectra Text format for mature data	ACCESS:defined by primary repository In Situ: yes, all metadata are stored in header ARES: yes metadata are stored in the headers CLU: yes, metadata stored in header GRES: yes, all metadata are stored in header ASC: yes, all metadata are stored in header	ACCESS: yes, structured search on discovery metadata In Situ: yes, structured search on discovery metadata ARES: yes data can be explored and searched by variables CLU: not yet GRES: yes data can be explored and searched by variables ASC: yes data can be explored and searched by variables	ACCESS: defined by primary repository In Situ: not yet ARES: not yet. CLU: not yet GRES: not yet ASC: not yet.	ACCESS: defined by primary repository In Situ: ISO19115, WIS profile ARES: ISO 19115-2, NCML, JSON CLU: community-based (CF compliant) GRES: ISO19139 ASC: ISO19139

Research Infrastructure Name	Are all categories used in the schemas defined in open registries?	How is provenance included?	Are PIDs included in the metadata description?	What is the primary storage format for metadata?	Which are the export formats supported?	Which metadata exchange/harvesting methods are supported?
ICOS	The ICOS-specific ontologies are openly available as linked open URLs, e.g. via the ICOS SPARQL endpoint (https://meta.icos-cp.eu/sparqlclient/).	We follow a simplified PROV-O and track lineage of data objects. The versioned metadata store keeps track of all metadata updates. However, information on data processing steps etc is not yet described in the ICOS data model.	Yes	As assertions in the form of RDF triples	Metadata export formats include json, xml, turtle, txt, html (all available via content negotiation of dataset landing pages).	SPARQL open endpoint, landing pages contain a subset in content negotiable formats (see previous question).
IAGOS	Not all some registered vocabularies recommanded by ISO and INSPIRE are used home made vocabularies not registered (see semantic section)	not yet included	DOI of the dataset in the metadata planned for instruments and platforms (already available but no PID used)	JSON (in MongoDB database) Dupplicated into AERIS geonetwork	JSON, XML (ISO 19139), RDF, PDF through geonetwork http://catalogue2.sedoo.fr/geonetwork/srv/eng/catalog.search#/metadata/575882c0-64ce-4648-bb19-00030d5d63af	CSW through geonetwork
EISCAT	No	Invluded as text files in info directories	N.A.	Text files	tar	N.A.
EISCAT	No	Not for present data, but added manually into older datasets	N.A.	Text files	tar	xml were provided through the EC FP7 project ESPAS.
SIOS	Identified in guidelines which are internal or provided upon request. System setup is still in progress.	We are waiting for PROV-O, within NetCDF/CF this is currently mapped in history attribute.	yes, it is the id or an alias to the id of the dataset.	XML, harvest multiple forms, transforms into unified and then ingest in SolR.	XML	For export OAI-PMH (but not supporting everything yet due to the harvest process), OpenSearch in process, OGC CSW planned when time.
ACTRIS	ACCESS: t.b.d. In Situ: as long as defined vocabulary is available ARES: no CLU: no GRES: use of the GCMD vocabularies ASC: use of the GCMD vocabularies	ACCESS: not yet In Situ: not yet ARES: Provenance of the data is described through attributes about DataOriginator DataProvider DataProcessor and similar new fields included in the new data format we have almost implemented. These fields are set up following the requirements for the ESA validation datasets. CLU: in global attribute fields GRES: not yet ASC: not yet	ACCESS: no In Situ: no ARES: yes. Typically netCDF global attributes are used for that: Measurement_ID= "20120710po00", HoI_systemID=2, etc CLU: no GRES: yes ASC: yes	ACCESS: relational database In Situ: relational database ARES: relational database CLU: relational database GRES: mongoDB database ASC: mongoDB database	ACCESS: none, defined by primary repository In Situ: XML, metadata header, CSV ARES: netCDF header, JSON , XML CLU: some metadata available via JSON GRES: JSON, XML ASC: JSON, XML	ACCESS: none In Situ: OAI-PMH ARES: THREDDS (ISO 19115-2, NCML) JSON on the new interface CLU: none yet GRES: CSW containers ASC: CSW containers

Research Infrastructure Name	Do you have a local search engine?	Do you support external search engines?	Do you make statements about access policies in your metadata?	Is your metadata machine actionable?	How is authentication done?	Do you maintain an own user database?	Do you use ORCID in your AAI?
ICOS	Yes, see https://data.icos-cp.eu/portal/ (data) and https://meta.icos-cp.eu/sparqlclient/ (metadata).	Yes, through access to our open SPARQL endpoint. (We do not operate an OAI-PMH endpoint, though…)	Access and licence information is provided at the Carbon Portal, and specifically to users during data download.	In principle yes, as the metadata follows OWL ontology which is exposed in adherence with Linked Data principles.	The Carbon Portal operates its own AAI service (CPAuth), which apart from local username/passwords also allows logins via eduGain and OAuth (ORCID id, Facebook)	Yes, users can register at the Carbon Portal and store voluntarily-provided profile information including individual settings for services (e.g. data cart contents). We also use logins to control access to specific local services, such as metadata modification, access to VREs and on-demand computation etc.). In support, we provide a trivial username/encrypted password database for users of password authentication.	Yes, ORCID ids are supported - see above.
IAGOS	Not local but metadata included in the French Atmospheric cluster AERIS metadata repository (geonetwork) with search engine : http://catalogue2.sedoo.fr/geonetwork/srv/eng/catalog.search#/home	No but metadata is harvestable through CSW (tested and demonstrated with B2find during ENVRI+)	yes	no	Homemade system with local user database (mongodb) SSO implementation planned in 2019 in the frame of AERIS (ORCID, shibboleth, etc.)	yes	In progress and should be done in 2019
EISCAT	No	No	No	No	Based on IP address	No	No
EISCAT	Yes	Yes (also via ESPAS)	No	Yes, through madrigal API https://www.eiscat.se/madrigal/	Open, e-mail address is provided by the user	Yes, the e-mail addresses are stored	No
SIOS	Using SolR, not public available. Human interface available at https://sios-svalbard.org/metadata_search	Not yet, this is next step once UID/PIDs are properly handled to avoid duplicates.	Licence is supported and recommended in SIOS Data Policy	Depends on host repository	Free and open currently, working on EduGain as SSO, OAUTH supported by some data centres.	For users coming through the SIOS web portal and requiring specific services.	No
ACTRIS	ACCESS: http://actris.nilu.no/ In Situ: http://ebas.nilu.no ARES:The ACTRIS dataportal is a metadata search engine working on the topical datacenter CLU: The ACTRIS dataportal is a metadata search engine working on the topical datacenter GRES: https://en.aeris-data.fr ASC: https://data.eurochamp.org/	ACCESS: no In Situ: yes e.g. THREDDS server ARES: yes e.g. THREDDS server CLU: no GRES: no ASC: no	ACCESS: no In Situ: no ARES: no CLU: no GRES: no ASC: no	ACCESS: no In Situ: no ARES: no CLU: no GRES: no ASC: no	ACCESS: relay of repository authentication In Situ: home-made, web-interface login ARES: CAS server (including Google authentication) CLU: none yet GRES: rely on the OAuth2 service provided by ORCID ASC: rely on the OAuth2 service provided by ORCID	ACCESS: no In Situ: yes ARES: yes CLU: no GRES: no ASC: no	ACCESS: no In Situ: no ARES: no CLU: no GRES: yes ASC: yes

Research Infrastructure Name	What is the major access technology supported?	How is authorisation done?	Which specific licenses do you use for your data?	Are metadata openly available?	Do you use or provide specific DMP tools?	Do you apply special data publishing steps?	Do you apply special data processing steps?
ICOS	HTTP GET method	For data download/access, we use API tokens, or email address/password. For access-controlled services, login via the cpauth service is required, with some operations or functionality requiring the user to be listed in service-specific configuration files.	In most cases, CC4-BY. Some data however are CC-zero.	Yes	DMP tools allow to describe the DMP, they do not make them. We used DMPonline, EU template, up to midterm	ICOS is all about curation from data generation through QC, processing to cataloguing and dissemination/publishing.	ICOS observation data are processed and quality controlled at our Thematic Centres, before they are uploaded to the Carbon Portal. Here, extreme care is taken to preserve all ingested data objects in a binary-exact form and store indefinitely. To ensure fixity can be proven, data object checksums are evaluated on the fly during ingestion, and the Handle PID suffix is calculated based on this checksum.
IAGOS	http get	Stored in local user database: - login/password for the web portal - tokens for the rest services (in progress)	IAGOS license: http://www.iagos-data.fr/#CMSConsultPlace:DATA_POLICY change for a CC will be discussed	yes via CSW through AERIS metadata repository (geonetwork) also accessible throught the AERIS metadata catalogue (https://en.aeris-data.fr/catalogue)	No DMP yet	yes, automatic metadata conversion from IAGOS/AERIS pivot format to datacite metadata schema	yes. Manual and automatic data qualification (from raw to final data). Automatic Level 2 data merging and formatting before curation. automatic processing for Level 4 products production as soon as L2 data available.
EISCAT	http	Country connected to the IP	EISCAT rules of the road	Yes	No	No	Yes
EISCAT	http	None	EISCAT rules of the road	Yes	No	Data validated by expert	Yes
SIOS	Please clarify question	Answered above	CCBY is recommended, but depends on type of data	answered above	No but if people ask either EasyDMP or DMPOnline.	No, this is handled at the scientist/repository level, but the recommendation from SIOS is to have DOIs for future KPI reporting.	Depends on the dataset
ACTRIS	ACCESS: ? In Situ: ? ARES: ? CLU: ? GRES: ? ASC: ?	ACCESS: password In Situ: password ARES: password CLU: n/a GRES: password ASC: password	ACCESS: none yet In Situ: none yet ARES: none yet CLU: none yet GRES: none yet ASC: none yet	ACCESS: no In Situ: no ARES: no CLU: no GRES: no ASC: no	ACCESS: no In Situ: no ARES: no CLU: no GRES: no ASC: no	ACCESS: n.a. In Situ: check semantic structure, sanity and consistency of metadata and data, manual inspection (please give more examples what is meant here). ARES: published 2000-2015 data on CERA, providing them all required information, like metadata on authors, variables, datasets, accuracy report and other info CLU: n.a. GRES: check the sanity and consistency of metadata and data, manually and/or automatically ASC: check the sanity and consistency of metadata and data, manually and/or automatically	ACCESS: n.a. In Situ: yes ARES: Centralised data processing from raw data to advanced data level CLU: yes GRES: yes (home-made processing chain of treatment) ASC: n.a

Research Infrastructure Name	Do you apply workflow frameworks for processing your data?	Do you use distributed workflow tools? if so, which?	Do you offer other type of support or analytics services?	Do you offer data products in your RI?	Do you use semantic vocabularies from generic vocabularies, ontologies, etc.? If so point to the registries.	Do you use discipline specific vocabularies, ontologies etc.? If so point to the registries.
ICOS	Yes, mostly custom workflows (carried out at the ICOS Thematic Centres.)	Yes and no; we do not use specific workflow tools (like Taverna), but we extensively use pseudo-standardised scripts to e.g. instantiate Virtual Machines for HTC computation & storage.	Yes, VREs through Jupyter Lab/Notebooks	Yes, see a description at https://www.icos-cp.eu/dataproducts	Yes, rdf, prov, foaf, … See e.g. http://purl.org/dc/elements/1.1/; http://purl.org/dc/terms/; http://www.w3.org/ns/prov#; http://www.w3.org/2002/07/owl#; http://www.w3.org/2000/01/rdf-schema#	Yes, wds (world data system) and http://www.w3.org/2003/01/geo/wgs84_pos#
IAGOS	use of framework Apache camel	no	no	yes elaborated products (Level 4) climatologies (Level 3)	GEMET (https://www.eionet.europa.eu/gemet/en/themes/) INSPIRE Data Themes (https://www.eionet.europa.eu/gemet/en/inspire-themes/)	NetCDF-CF convention for the names of the variables: http://cfconventions.org/standard-names.html
EISCAT	No	No	Yes, for example an online analysis tool developed through ENVRIplus	Yes	No	No
EISCAT	No	No	Yes, plotting etc.	Yes	Yes, in the ESPAS project	Yes (specified through ESPAS) https://www.espas-fp7.eu/portal/browse.html#ontology
SIOS	Depends on the dataset	Not yet in a structured approach, Jupyter is used at individual level, looking into structured support, also experimenting with Galaxy.	Not yet	Please define the concept... We do serve analysed satellite imagery, numerical simulations etc. The concept of a product is vague and not uniform, we consider everything a dataset with specific features.	CF, GBIF, WIGOS, WMO, OSGEO, ...	See above
ACTRIS	ACCESS: n.a. In Situ: yes ARES: yes CLU: yes GRES: no ASC: n.a	ACCESS: n.a. In Situ: no ARES: no CLU: no GRES: no ASC: n.a	ACCESS: n.a. In Situ: RRT data production, submission interface, QC tools ARES: data processing, the QC tools, the feedbacks to users, submission interface and the DOI assignment CLU: NWP model evaluation service GRES: not yet ASC: data generation tools, modelling tools	ACCESS: n.a. In Situ: yes ARES: yes CLU: yes GRES: yes ASC: yes	ACCESS: no In Situ: CF: http://cfconventions.org/, ISO19115, ACDD ARES: CF: http://cfconventions.org/ CLU: CF: http://cfconventions.org/ GRES: CF: http://cfconventions.org/,ISO19115, ISO19139 ASC: CF: http://cfconventions.org/,ISO19115, ISO19139	ACCESS: no In Situ: CF ARES: CF CLU: CF GRES: CF ASC: CF

Research Infrastructure Name	Do you use project defined vocabularies, ontologies, etc.? If so point to the registries.	Do you believe that your data is Findable (F)? if not, indicate where you see major gaps.	Do you believe that your data is Accessible (A)? if not indicate where you see major gaps	Do you believe that your data is interoperable (I)? if not indicate where you see major gaps	Do you believe that your data is re-usable (R)? if not, indicate where you see major gaps
ICOS	Yes, ICOS (domain) specific, e.g. BADM, WMO GAW. See http://meta.icos-cp.eu/ontologies/cpmeta/ !	Yes	Yes	Starting to be interoperable... Need to work more on our data model, applying relevant standards (including vocabularies) and making sure the definitions & attribute names we use are registered.	Mostly, yes - but we need to include more provenance information in the metadata (including more links to ICOS observation & data processing protocols).
IAGOS	Home made vocabularies for platform, instruments and variables names based on GCMD Not properly published yet (work planned in the frame of AERIS)	yes	yes	partially need to use standard vocabularies or publish own vocabularies in order to provide full interoperability interoperable standard services implementation in progress (OpenDAP, WCS)	no need to improve and add new metadata: provenance, etc. and also improve the data versioning system for history
EISCAT	No	To some degree, gaps are in the lack of PIDs and metadata registry	Yes, some work needed on AAI for special cases	No, the only standardised metadata is the time of the experiment	Yes, but gaps exists in standardisation
EISCAT	Yes https://www.eiscat.se/madrigal/index.html	To some degree, gaps are in the lack of PIDs and metadata registry	Yes	Yes, within the instruments covered by madrigal	Yes, following the madrigal community standards
SIOS	Not yet, we try to avoid this, but may have to address it	Not all yet as this is work in progress, but quite much yes.	Those that are F yes	Some within the meteorological/oceanographic/... domain, but much not yet	Same as above, depends on host repository and discipline.
ACTRIS	ACCESS: no In Situ: no ARES: for the variables not covered yet by CF convention we are applying nomenclature as agreed within ACTRIS CLU: no GRES: no ASC: no	CCESS: defined by primary repository In Situ: partly ARES: yes through standard tools/protocols CLU: partly GRES: partly ASC: partly	ACCESS: yes In Situ: yes ARES: yes CLU: partly GRES: partly ASC:partly	ACCESS: defined by primary repository In Situ: partly ARES: partly CLU: partly GRES:partly ASC: partly	ACCESS: defined by primary repository In Situ: partly ARES: version controlled database and a new data format reporting many info for traceability ready April 2019. CLU: partly GRES: partly ASC: partly