D8.3
Atmospheric subdomain implementation plan

Work Package	WP8
Lead partner	NILU
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	30 April 2020
Submission date	28 April 2020

Deliverable abstract

The RIs of the ENVRI atmospheric subdomain have conducted a comprehensive analysis of the FAIRness of their data management. The identified gaps considered most important from a user perspective concern the following FAIRness functions:

- Findable: globally unique identifier; indexed in searchable resource.
- Accessible: (meta)data retrievable by standardised protocol
- Interoperable: common vocabulary
- Re-usable: established license, documented provenance, (meta)data meets community standards.

The tasks for implementing these functions were grouped by the following criteria:

1. Importance of the function for the user
2. Maturity of the function in the RI
3. Accordingly, the tasks were grouped into 3 categories:
4. Immediate implementation: the function implemented is highly important for the user, the implementation plan is consolidated.
5. Immediate implementation planning: the function implemented is highly important for the user, the implementation plan needs to be refined, know-how at the RIs needs to be established.

Task for second half of project: the function implemented is considered important, but implementation can be postponed to allow focussed work on the highly important FAIRness functions.

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
Authors	Markus Fiebig Cathrine Lund Myhre Damien Boulanger Léonard Rivier Alex Vermeulen Ingemar Häggström Anders Tjulin Lara Ferrighi Simo Tukiainen	NILU NILU CRNS UVSQ ULUND EISCAT EISCAT METNO FMI	31 March 2020
Reviewer(s)	Zhiming Zhao, Markus Stocker	UvA, TIB	31 March 2020
Approver	Andreas Petzold	FZJ	28 April 2020

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1	31 March 2020		Markus Fiebig

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. Introduction & Procedure

The ENVRI-FAIR project's objective is to implement "FAIRness" for data produced in the European Research Infrastructures (RIs) organized in the Environmental Research Infrastructures (ENVRI) community, in order to make them ready for connecting to the European Open Science Cloud (EOSC). In this context, "FAIR" is an acronym comprising the aspects of "Findable", "Accessible", "Interoperable", and "Reusable" as specified by the FORCE11 community.

ENVRI-FAIR WP8 organises and conducts this implementation work for the community of ENVRI RIs in the atmospheric subdomain, comprised of the RIs ACTRIS, EISCAT, IAGOS, ICOS, and SIOS. This deliverable constitutes a plan for implementing data FAIRness functionality in the data production of the atmospheric ENVRI RIs. It is based on assessments of the current data handling functionality in relation to the requirements defined by the FAIRness principles by each of these five RIs.

These assessments were conducted on several occasions:

- By means of a questionnaire during the writing process of the ENVRI-FAIR project proposal
- During a workshop under the ENVRI-FAIR kick-off meeting in January 2019.
- During a workshop of ENVRI-FAIR WP8 11-12 June 2019 in Amsterdam
- During a virtual WS 11-12 December 2020
- in ENVRI-FAIR Deliverable D5.1 (Requirement analysis, technology review and gap analysis of environmental research infrastructures).

This deliverable condenses the identified FAIRness issues common to all or most atmospheric ENVRI RIs into WP8 implementation tasks for the first part of the project.

The timing of the internal milestones and deadlines in this deliverable do not take into account possible changes and delays forced by the world-wide corona virus situation. As the situation is 20 March 2020, we do think this work can proceed as planned in this deliverable.

2. Task List

The implementation plan consists of tasks organised into three groups:

1. Essential tasks where the specification is mature and which are considered very important as to begin implementation immediately.
2. Essential tasks which are considered very important, but where the specification is in need of clarifications before implementation can begin. Further clarification of specification, i.e. implementation planning, should start immediately.
3. Tasks which are considered relevant for implementing data FAIRness, but are considered less important than the tasks in the first 2 groups. These tasks are scheduled for the second half of the project.

2.1. Tasks for immediate implementation

The implementation progress of the tasks in this group is monitored by WP8 Task 8.4 (Implementation of FAIR roadmap in the Atmospheric subdomain RI) under the responsibility of project partner NILU. Each RI in the atmospheric ENVRI RI is individually responsible for working on these tasks.

RI implementation responsible persons: Damien Boulanger (IAGOS), Markus Fiebig (ACTRIS), Ingemar Häggström (EISCAT), Alex Vermeulen (ICOS), Lara Ferrighi (SIOS)

2.1.1. Consolidation of consistent use of PIDs throughout data production workflow

The work on this task is based on ENVRIplus deliverable 6.1 on data identification and citation, which constitutes analysis, specification, and implementation plan for this task. In short, Digital Object Identifiers (DOIs) are to be used for final data products, whereas Persistent Identifiers for eResearch (ePIC) or equivalent handle PIDs are to be used for most of the other identifiable items in the data production workflow (software, data pre-products, instruments, …). ePIC and Handle PIDs combine the advantage of a central register, i.e. expected long-term availability, with low costs. Contact point for further details on the implementation is Maggie Hellström in ENVRI-FAIR WP5.

2.1.1.1. ACTRIS Implementation Plan - PID

Despite consisting of 5 thematic data centre units with variable specific data production workflows, the ACTRIS data centre (DC) will make use of a homogeneous system of PIDs to identify all entities involved in data production:

- **Digital Object Identifiers (DOIs)** for all final data products
- **Handle PIDs**, either by RI's own handle server or provided by ePIC, for data pre-products, software, QA / QC documents.
- **Open Researcher and Contributor ID (ORCID)** for humans involved in data production
- **PIDs for organisations and instruments** still to be determined in consultation with experts in ENVRI-FAIR WP7, and this part of the work is in the next step (under section 2.3).

Milestones in implementing PIDs in ACTRIS DC:

- July 2020: determine solution for handle PIDs.
- December 2020: implement primary DOIs for all final data products in primary repositories with homogeneous granularity per repository, resolving contributions of individual principal investigators (PIs) and research performing organisations (RPOs), thus facilitating accounting of data use by means of DOIs.
- June 2021: implement PIDs for data pre-products.
- December 2021: implement ORCIDs and PIDs for organisations.

2.1.1.2. EISCAT_3D Implementation Plan - PID

EISCAT is implementing the metadata specification, using the exposed DataCite metadata fields. This includes assigning a DOI for each Data Collection and PIDs for the Data Sets therein. At least for higher level data sets, the B2Handle system seems most suited.
To secure provenance documentation, the data sets will be referenced by PIDs of the underlying data and the processes done on it, including both software PIDs and hardware PIDs. This is planned for all levels of data, from the lowest voltages samples to the final data products.

2.1.1.3. IAGOS Implementation Plan - PID

All IAGOS published data sets (from data processing level 2 to level 4) have DOIs minted by the IAGOS Data Centre. IAGOS will subscribe to ePIC in order to be able to mint ePIC PIDs for all entities managed by the Data Centre, including data sets under level 2, instruments, documents, etc. So far, those entities have internal identifiers.
After registration with ePIC, the data production workflow will be updated to integrate the minting of PIDs for each entity, involving:

- Assigning PIDs to all existing entities: fine data sets (i.e. time series for a flight) including version control, sensors, QAQC documents, etc.
- Updating the workflow to automatically assign PIDs to new entities
- Having all members of the IAGOS team with an ORCID and integrating the information into the data production workflow

This task is mandatory to be able to implement the provenance in the IAGOS workflow. Thus this work will be jointly done with to the task "Consistent documentation of provenance throughout data production workflow".

Milestones:

- June 2020: ePIC registration and ORCID for all IAGOS team members
- September 2020: definition of the new data production workflow with PID management
- December 2020: consistent use of PIDs and new data production workflow implemented

2.1.1.4. ICOS Implementation Plan - PID

ICOS has already implemented from its onset minting of Handle PIDs in the repository for all data objects, including raw data. PIDs are assigned during the machine to machine ingestion of the data objects or the creation of a collection. Next to this ICOS also mints DataCite DOIs for either the individual data objects of higher level data products and/or collections of data objects. A DOI minted by ICOS always refers to the PID on the data object or collection and shares the same landing page. All data is available through HTTP GET under the ICOS CC4BY licence.

All landing pages are human and machine readable and provide metadata in a number of standard metadata exchange formats that are selected through content negotiation for machine to machine communication or by selecting the respective link on the landing page by the human user. Provenance and other metadata of the landing pages is generated dynamically from the ICOS RDF triple store.

2.1.1.5. SIOS Implementation Plan - PID

SIOS does not assign or manage PIDs for the contributing centres, which have full ownership of the datasets provided via the SIOS data portal. On the other hand, SIOS recommends having DOIs in place, and the different data centres are aligning to this recommendation. In this context, SIOS is also working on identifying core data, i.e. with particular relevance for the key questions within Arctic research, which shall require a DOI.

2.1.2. Common standard interfaces for metadata and data access

Efficient machine-to-machine (M2M) interfaces to both metadata and data are a prerequisite for virtually any services processing RI produced data further, e.g. indexing portals, visualisation services, or virtual research environments (VRE) offering data analysis tools. Depending on the nature of the service, two types of interfaces are in use:

- Interfaces based on existing standards, offering a high degree of generic reusability, but limited functionality.
- Interfaces based on Representational State Transfer (REST), which offer more functionality, but currently are not standardised, albeit being readily documented, and thus offer little reusability.

WP8 partners agree that atmospheric ENVRI RIs need to offer both types of interfaces to cater to different types of users. This section deals with implementation of standardised M2M interfaces for metadata and data, which offer a high degree of maturity.

Concerning standardised metadata M2M interfaces, WP8 considered the following:

- **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface serving metadata in ISO19115 format with WMO Information System (WIS) profile.**
- Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface serving metadata in ISO19115 format with Infrastructure for spatial information in Europe (INSPIRE) atmosphere profile.
- **Open Geospatial Consortium (OGC) Catalogue Service for the Web (CSW)**
- OGC Sensor Observation Service (SOS)

The two options marked in bold font are the ones agreed within WP8 (in contact with WP7 and WP5) due to their widespread use and associated advantages of re-using implementations. Users of these protocols include, among others, WMO Global Atmosphere Watch World Data Centres and GEOSS.

Concerning standardised data M2M interfaces, the workshop considered the following:

- Open-source Project for a Network Data Access Protocol (OPeNDAP), a protocol for accessing structured data though the web. Allows for sub-setting and streaming of data, thus well suited for web applications.
- OGC Web Coverage Service (WCS), for accessing structured data, allows sub-setting, transfers of files for client-side processing.
- OGC Web Mapping Service (WMS), for accessing structured data, provides rendered map products of gridded data.

Each of these interfaces is targeted to different types of data. Thus, depending on the nature of the data produced by the RI, any of the interfaces mentioned above may be used.

This task will also comprise the interoperability for metadata and data exchange within the ENVRI atmospheric subdomain.

2.1.2.1. ACTRIS Implementation Plan - machine-to-machine interface

The ACTRIS data centre (DC) consists of 5 distributed topic DC units: for data from surface in situ stations; surface-based aerosol remote sensing data; surface-based cloud remote sensing data; surface-based trace gas remote sensing; atmospheric simulation chambers. While the topic DC units serve as primary data repositories, discovery and access to ACTRIS data is provided by a common portal, in an integrating DC unit. This DC unit also hosts a common, continuously updated ACTRIS DC metadata repository for discovery of data through the portal. The internal metadata profile of this repository will cover items contained in common standardised, domain-specific profiles, but will go beyond these to cover topic specific use metadata.

ACTRIS metadata will be served by the ACTRIS DC metadata repository connected with the ACTRIS data portal. This will include M2M interfaces, both classic standardised and REST ones, e.g. an OAI-PMH interface serving ISO19115 metadata with WMO profile. The served metadata records will include a direct link to the data, where the access protocol will depend on the data type (OPeNDAP, OGC-WCS).

Milestones:

- First version of ACTRIS DC metadata repository available (October 2020).
- OAI-PMH interface serving ISO19115 metadata with WMO profile for ACTRIS data available, will include links for data access. (January 2021).

2.1.2.2. EISCAT_3D Implementation Plan - machine-to-machine interface

EISCAT metadata access

EISCAT plans to make metadata harvestable in formats including DataCite and possibly other standards. The existing EISCAT archive of analysed data (https://portal.eiscat.se/madrigal) is in the process of migration to a new version, Madrigal 3, with HDF5-based archival. An OAI-PMH server for harvesting metadata will be added to Madrigal 3 and B2FIND will harvest its metadata. The Madrigal database has API script examples available for several high level languages (python, Matlab and IDL). Similar routines will then be added to the DIRAC portal for EISCAT 3D data and the DIRAC file catalogue will be extended accordingly.

EISCAT data are the property of EISCAT Scientific Association and low level data should not be copied outside the EISCAT e-infrastructures without consent. Future EISCAT 3D users will use DIRAC to search and stage data, submit analysis jobs and retrieve the results. All metadata and approved data products are generally open for any non-military and non-commercially use. We also consider using notebooks for users to upload their own analysis software. At present, the protocol for these interactions needs to be decided.

Milestones:

- June 2020: DataCite metadata attributes defined for level 3 data
- September 2020: DataCite metadata attributes defined for level 1/2 data
- November 2020: Metadata ingestion into B2FIND
- December 2020 : Interface OAI-PMH to the DIRAC file catalogue

2.1.2.3. IAGOS Implementation Plan -- machine-to-machine *interface*

Metadata Access

IAGOS metadata are available through a Geonetwork server (http://catalogue2.sedoo.fr/geonetwork/srv/eng/catalog.search#/home) that serves metadata , in ISO19115 format with INSPIRE profile through:

- CSW interface with endpoint: http://catalogue2.sedoo.fr/geonetwork/srv/eng/csw-iagos?SERVICE=CSW&REQUEST=GetCapabilities

- OAI-PMH interface with endpoint: http://catalogue2.sedoo.fr/geonetwork/srv/eng/oaipmh

IAGOS is currently updating its metadata model to implement the WIGOS metadata profile and is also implementing a web REST API in order to serve metadata in JSON format for IAGOS partners and users.

Data Access

IAGOS will implement a THREDDS server that will allow to serve IAGOS data through standardized protocols:

- OPeNDAP service for all data sets (time series and gridded data)
- WCS and WMS services for gridded data sets

IAGOS is also implementing a web REST API in order to serve data sets to users who need machine-to-machine access. Data will be delivered in the same formats as on the IAGOS Data Portal (NetCDF and Nasa Ames) and in JSON.

Milestones:

- June 2020 : implementation of WIGOS metadata profile
- September 2020 : REST API for metadata access
- December 2020 : THREDDS server implemented with access to all IAGOS products
- December 2020 : REST API for data access

2.1.2.4. ICOS Implementation Plan -- machine-to-machine interface

The gaps identified for improvement of the fairness of the ICOS repository concern mainly small details and refinements that nevertheless are important for implementing future machine to machine interoperability. Until now the main emphasis in the development has been on pure functionality through a robust, flexible and modern core (back-end), and achieving a high scalability and performance of the system. Next steps are on core functionalities for the users, for access, preview and attribution of the data providers.

Planned improvements and time schedule:

1. **Metadata provision**:
   - Metadata tags on landing pages following schema.org to enable for Google data search harvesting (9-2020)
   - Implement OAI-PMH export of metadata to feed higher level data collection metadata into GEOSS (12-2020)
   - Publish the ICOS ontology as controlled vocabulary (12-2020)
   - Implement a ISO19135 XML format landing page (12-2020) (implementing WIS at the same time as WIS is a weak subset of ISO19135)
   - Extend the metadata with related documents, datasets, images, software and other resources (6-2021)
   - Add license information as resolvable URL on the landing pages (4-2021)
   - Extend the provenance metadata to the full ICOS processing chain (7-2021)
2. **Data access**
   - Implement a ENVRI standard RESTful API for data subsetting (12-2021)
   - Implement the export of ICOS timeseries as NetCDF-CF files (12-2020)
   - Updating the current THREDDS (possibly ERDDAP for compatibility with the marine sub-domain) service to the latest version and connecting this to the ICOS metadata and data objects (6-2021)
3. **User interface**
   - Increased search capabilities for spatial extent, conditions and free text keywords (9-2020)

2.1.2.5. SIOS Implementation Plan -machine-to-machine interface

Metadata access: SIOS is currently not exposing metadata harvested from the contributing centres. Two possibilities will be considered, either provide a list of OAI-PMH endpoints from the contributing centres, or provide a common endpoint when the implementation of pyCSW will be up an running. Implementation of pyCSW has already started and it is expected to be ready for testing in the near future, i.e. in summer 2020.
Data access: OPeNDAP is available for some contributing centres, some have already started implementing the protocol which should be finalized during 2020. Additional data centres are also planning full implementation shortly. The SIOS Data Management Working Group is continuously monitoring the progress.

2.1.3. Indexing of data resources in WIS, GEOSS

For meeting the FAIR requirement of "Findable", data needs to be discoverable in relevant data search portals.
The following data discovery portals were considered:

- **Global Earth Observation System of Systems (GEOSS).**
- **WMO Information System (WIS)**
- Google Dataset Search: considered not specifically relevant to atmospheric community, but relevant to other users, can be considered at a later stage.

- European Open Science Cloud (EOSC): currently no mature data search interface available
- WMO Integrated Global Observing System (WIGOS): still under development.

The WP8 workshops decided to focus on connecting to the portals displayed above in bold font due to the combination of relevance to the atmospheric community and maturity.
This task is connected to ENVRI-FAIR WP3, which in Task 3.3 works on the liaison with international stakeholders.

2.1.3.1. ACTRIS Implementation Plan -link to WIS, GEOSS

ACTRIS will serve metadata by standardised interface, OAI-PMH, and in ISO19115 format with WMO profile. These are exactly the metadata exchange specifications for the WMO Information System (WIS). ACTRIS will therefore connect to WIS by setting up a link to its closest Global Information System Centre (GISC) in WIS.
WIS is continuously harvested by the GEOSS. Thus, by linking to WIS, ACTRIS data will be discoverable in GEOSS as well.

Milestones:

- Metadata link between ACTRIS metadata repository and closest GISC in WIS established (July 2021).

2.1.3.2. EISCAT_3D Implementation Plan -link to WIS, GEOSS

Work at EISCAT has been directed towards the ESPAS ontology, http://ontology.espas-fp7.eu/vocabs, as the GEOSS fields have not been sufficient. The plan is to revisit the proposed ones. It has to be taken into account that use of Google Search is impossible for the EISCAT Associates in China.

Milestones:

- June 2020: ESPAS ontology entered into datasets.
- October 2020: Define an atmospheric parameter set for EISCAT_3D suitable for WIS and GEOSS
- December 2020: links with WIS established for the existing atmospheric data sets

2.1.3.3. IAGOS Implementation Plan -link to WIS, GEOSS

IAGOS serves metadata through standard interface such as CSW and OAI-PMH and will soon implement the WIGOS metadata profile. These are the metadata exchange specifications for the WMO Information System (WIS).
As WIS is continuously harvested by GEOSS, by linking IAGOS to WIS, IAGOS data will be discoverable in GEOSS as well.

Milestones:

- June 2020: implementation of WIGOS metadata profile
- December 2020: link with WIS established

2.1.3.4. ICOS Implementation Plan -link to WIS, GEOSS

The ICOS repository is fully searchable through an open SPARQL endpoint that also allows to access the complete ontology in OWL. Furthermore, there are several graphical web interfaces for human search and access to the data. Currently there is no connection to WIS for greenhouse gas observations. WMO collects greenhouse gas observations through the World Data Centre for Greenhous Gases (WDCGG) and maintains station metadata trough the OSCAR database. The WDCGG does not have a machine-to-machine interface for metadata and data exchange. It requires at the moment that individual PIs enter manually the metadata and data into their portal and then the data is processed in several manual steps for publication. The WDCGG is still far from being a FAIR repository and therefore publication on WDCGG has little added value. A request to allow ICOS in the role of recognized WMO GAW contributing network to submit station metadata to WMO OSCAR on behalf of the station PIs is pending.

The GEOSS portal allows upload of metadata in CKAN standard, which will be implemented by ICOS the end of 2020.

2.1.3.5. SIOS Implementation Plan -link to WIS, GEOSS

This task is related to the finalization of pyCSW. Once this is done, we will connect through WIS primarily, and directly to GEOSS for some datasets. We need however to create guidelines on this to avoid multiple paths for the data in question.

2.1.4. Common use of authentication schemes

The authentication scheme is a subject that clearly goes beyond the sub-domains as it is very technical and not specific for the services. It also is a common requirement for all RIs. As it is deeply connected to the architecture of the systems it has a strong influence on workflows and security of the systems. As many RIs already have systems that are (partly) operational, changes in this basic functionality are not easy, risky and can easily turn out to be expensive. However, from the perspective of user experience a single sign on for all services of ENVRI will be very valuable.
The authentication, authorisation and identification system will be studied in ENVRI-FAIR in a separate task force that will report at the end of 2020 with recommendations for the whole domain. Although waiting for this report carries the risk that for some RIs stages are passing where decisions are made that are difficult to revert, we consider that it is nevertheless essential to agree on a common approach for the domain, also connecting to the developments in the framework of EOSC and making use of the efforts and experiences in the framework of the AARC and AARC2 projects.
Meanwhile, the RIs in the atmospheric subdomain will follow a strategy of non-regret where operational choices are made to support the current workflows and daily work in the RIs. Adding or further building on support for Oauth2 in the AAI systems will then allow for a transition to more harmonized approaches in the second half of the project that is transparent to the users. Implementing Oauth2 and linking to login through ORCiD will allow to make to use of the ORCiD institutional login through SAML and at the same time will promote the use of ORCiD for identification of persons.
The tiered approach follows two consecutive steps where step 1 is implemented in 2020 and step 2 follows in the course of 2021 if this step is approved by the ENVRI community.

1. User authentication via Open Researcher and Contributor ID (ORCiD)
2. The Blueprint Architecture developed by the H2020-funded initiative Authentication and Authorisation for Research and Collaboration (AARC), which connects to ORCiD

This task is to link to the respective efforts in ENVRI-FAIR WP5, which defines interoperability requirements for the whole of ENVRI-FAIR.

2.1.4.1. ACTRIS Implementation Plan -authentication schemes

In the ACTRIS Data Centre (DC) as distributed data centre, the authentication schemes need to be aligned with national policies and infrastructures of the contributing partners.
In this setting, ACTRIS will focus on ORCiD as common a authentication solution across the RI.

Milestones:

- ORCID implemented for authentication across ACTRIS DC (December 2021)

2.1.4.2. EISCAT_3D Implementation Plan -authentication schemes

EISCAT is working on EGIcheckin for authentication. EGIcheckin brokers a selection of Idps, like EDUgain, B2access, ORCiD, and other social identities such as QQ and WeChat. In addition, there will be a local login for users outside of the general Ipd services, and for a second level authentication for wizard/expert users.
For authorisation, it is planned to setup a local ldap server, using as much as possible automatic selection procedures from the Idp provisions. For some users it is needed to add EISCAT specific rules,

2.1.4.3. IAGOS Implementation Plan -authentication schemes

Currently IAGOS has a homemade authentication system. Users have a IAGOS account and use a simple login/password connection. A new version of the IAGOS data portal is currently under development. In this frame a new authentication / authorization system will be implemented.

A solution developed by the French Atmospheric Data and Services cluster AERIS and based on the tool Keycloak will be used. It will allow to provide Single Sign On system through ORCID, eduGAIN or a local AERIS users database. It will be easy to connect to other similar system if needed.

Milestones:

- December 2020: AERIS authentication implemented in new IAGOS Data Portal and then ORCID and eduGAIN authentication operational

2.1.4.4. ICOS Implementation Plan -authentication schemes

ICOS already supports SAML and OAuth2 for login, next to a local login on the basis of the user email address that can be mixed and matched. When the choice for a common AAI scheme is made, it will be integrated in the current system in 2021.

2.1.4.5. SIOS Implementation Plan -authentication schemes

The SIOS data portal gives free and open access to data without any authentication needed, if not for further processing of data using local infrastructures. Authentication in this case is done within the content management system (Drupal) in use. SIOS is currently not prioritizing this task, although the implementation of EduGain as SSO is looked at with interest. SIOS will participate into the dedicated task force (AAAI) to closely follow the discussion on the topic.

2.2. Essential tasks for immediate implementation planning

For each of the tasks in this section, a working group is set up to specify implementation planning.

2.2.1. Domain vocabulary / ontology for observed parameters, discovery and use metadata

The ontology for the ENVRI atmospheric domain should reflect the time independent inner logic of properties to be organised, rather than referring to other organisational structures which have a tendency to change with time. It should re-use existing vocabularies wherever possible. The first action items will consist of:

- Defining the ontology scope. Which properties should be covered, taking into account size of and resources for the task.
- Mapping of existing vocabularies, assess their suitability for being used in the ontology, identify gaps in vocabularies.

This task will connect to ENVRI-FAIR WP5, which works on common requirements for (meta)data services and cataloguing.

Lead: Markus Fiebig (ACTRIS)
Participants: Lara Ferrighi (SIOS), Barbara Magagna (ENVRI-FAIR WP5), Damien Boulanger (IAGOS), Claudio D'Onofrio (ICOS), Giuseppe d'Amico (ACTRIS), Vincent Douet (ACTRIS), Bénédicte Picquet-Varrault (ACTRIS).
Aim to be ready with implementation plan: 31 April 2020 (MS45, MS8.3)

2.2.1.1. Work Plan

Terminologies follow a hierarchy of terms, starting from controlled lists of vocabulary, via glossaries (lists of terms with definitions), taxonomies (adding the hierarchy of terms), thesaurus (setting the terms in a relation) to an ontology (strict order and relation of terms). In its constituting telecon, the WG concluded that the largest need in the atmospheric subdomain concerns harmonised vocabulary and glossaries for documenting data with machine-actionable, rich use metadata, documenting provenance, but also for enabling semantic search. Organising terms following higher semantic principles would be beneficial, but controlled lists of vocabulary and glossary should be prioritised.
The working group screened existing, domain specific controlled vocabularies for their potential use and extension in the WP8 ontology WG. The vocabulary defined for the WIGOS Metadata Record (WMDR) as part of the WMO Integrated Global Observing System (WIGOS) appears to be a good starting point. The work will involve the following steps:

1. Map the existing metadata items and vocabularies used in the WP8 RIs onto terms used in WIGOS.

2. Identify gaps in both metadata items and controlled lists used as values for metadata items.
3. Define new metadata items and controlled list items where needed.
4. Work with WMO towards adopting upgraded vocabulary

2.2.1.2. Timeline

1. Finish mapping between existing RI vocabulary and WIGOS and identify gaps (June 2020).
2. Propose new metadata and controlled list items needed to fill gaps (January 2021)
3. Preliminary outcome of negotiations with WMO on adopting additions and modifications (June 2021).

2.2.1.3. Interaction with ENVRI-FAIR WP5 Expert Environment

The WG has constant interaction with WP5 through WG members representing WP5, which actively guide the ontology building process.

2.2.2. Consistent documentation of provenance throughout data production workflow

Provenance entails making each execution of an RI data production workflow traceable, i.e. a data point can be traced back to the time of measurement. This task will identify a provenance scheme suitable for atmospheric ENVRI RIs, and will develop a concept for implementing it. The first measure will be one or several training events for RI provenance focal points. These will then work on the provenance implementation in one or several workshops.
This task is to link to the respective efforts in ENVRI-FAIR WP7, which targets implementation support across ENVRI-FAIR.

Lead: Damien Boulanger (IAGOS)
Participants: Barbara Magagna (ENVRI-FAIR WP5), Lara Ferrighi (SIOS), Alex Vermeulen (ICOS), Markus Fiebig (ACTRIS), Ingemar Häggström (EISCAT), Markus Stocker (ENVRI-FAIR WP7), Doron Goldfarb (ENVRI-FAIR WP7).
Aim to be ready with implementation plan: 31 October 2020.

2.2.2.1. Work Plan

The working group had its first meeting in November 2019 when experts from WP7 shared their knowledge about provenance. It allowed all the RIs to understand the requirements for provenance implementation in their data production workflow.
The first action has been to assess the level of implementation in each RI. The group concluded that, with the exception of ICOS, almost all RIs lack provenance implementation so far.
The requirements for provenance implementation are:

- The use of PIDs in each RI data workflow as all entities and agents that are part of the provenance need to be referable. This part is linked to and dependent on the task for immediate implementation “Consolidation of consistent use of PIDs throughout data production workflow”.
- The use of a standard provenance model. All RIs agreed to implement the PROV standard developed by W3C, more precisely the PROV-O implementation (i.e. RDF implementation of the PROV Data Model).
- To have a triple store to manage and query the provenance information.
- The integration of provenance management in each RI data workflow.
- Provide the provenance information to the users. At least two RIs want to provide the information in NetCDF data files. The solution of NetCDF-LD is currently under investigation. The group needs to determine how to provide the information on landing pages. Different user’s profiles have to be considered: data managers will use the complete provenance data for workflow management (machine readable), scientific users will need a filtered subset of this data (human readable).
- Ontologies can be used to name entities, agents and activities that are part of the provenance workflow. It has been decided that for starting, each RI will choose the vocabularies they want. The group will interact later with the Working Group on Ontologies.

In order to gain expertise on provenance, a work plan has then been defined for each RI as follows:

1. Describe its data production workflow using graph representation in order to have a schema for provenance information. Provenance information will be data instantiated from executing the workflow schema.
2. Implement PROV-O on several workflow elements for testing purposes.
3. If necessary, implement PROV-Template to describe all the provenance workflows within an RI and be able to control the provenance profiles for each dataset. The tool developed during ENVRIplus will be used.
4. Implement a triple store. This part isn't mandatory as some RIs can choose to use tools such as ProvToolbox in Java or Prov Python for PROV-O data generation. Although the use of a triple store is recommended as it will allow for instance to manage provenance data version control and querying.
5. Provide the provenance information to the users through landing pages of data sets and/or in data files.

2.2.2.2. Timeline

The working group is having monthly telecons since the November 2019 and will continue until October 2020. More meetings might be scheduled if needed. This task is linked to the immediate implementation of consistent PID use and of the work done in Task Force 4 on Triple Stores.
The group will check the existence of training material on provenance. The monthly meetings will allow to check the progress status of each RI.
Demonstration cases will be proposed in order to share expertise. The ENVRI-FAIR Knowledge database UI (https://envri-fair.github.io/knowledge-base-ui/) shows the repositories that lack provenance information (R1.2) and links to the corresponding demonstrators. We will use this tool to present our demonstration cases. The demonstrators will be extended, e.g. include also PROV-Templates and SPARQL queries on provenance information.
In October 2020, it is expected that each RI will have enough technical expertise to implement provenance in their data workflow. A short document will be delivered, including all the requirements and recommendations identified by the group. A demonstration on a chosen use case will also be available for the ENVRI community, for training purpose for instance.

2.2.2.3. Interaction with ENVRI-FAIR WP7 Expert Environment

WP5 and WP7 experts are active members of the working group and at least one representative always attends the monthly meetings. The lack of knowledge and experience of almost all RIs regarding provenance has been filled by presentations and demonstrations by the experts. This group is also connected to the Task Force 4 of WP5.

2.2.3. Recommendations for licenses on metadata and data

This task comprises work on an overview of licenses and policies currently in use in the atmospheric ENVRI RIs for both data and metadata. In collaboration with WP4, the task is to condense this information into a recommendation for license use. Here, the aspect of licensing metadata, in addition to the data, is to be explored, and the options and implications considered.
This task is to link to the respective efforts in ENVRI-FAIR WP4, which works on common FAIR Policies.

Lead: Leo Rivier (ICOS)
Participants: Cathrine Lund Myhre (ACTRIS), Damien Boulanger (IAGOS), Lara Ferrighi (SIOS), Ingemar Häggström (EISCAT).
Aim to be ready with implementation plan: 31 May 2020.

2.2.3.1. Work Plan

ENVRI FAIR Milestone 34 entitled “"Existing data policies and licenses in use in the Atmospheric subdomain" was produced. It provides access to reference material for both data policies and licenses used by the 5 Atmosphere RIs ACTRIS, EISCAT, IAGOS, ICOS-Atm, SIOS. It shows that data policies exist for all. However, only ICOS operates with operational license provision for data distribution.

Policy harmonization has an imminent political dimension to it. This is to be dealt in ENVRI FAIR WP4. However here we recommend considering the already widely used Creative Commons CC license. In 2017, over 1,4 billion creative commons licenses were at work. This set of licenses is characterised by it ease of use. CC licenses have a three layer structure: a legal code, a human readable code and a machine readable layer. This last layer is implemented by having all the HTML provided by the CC licenses chooser be automatically annotated with metadata in RDFa format. CC licenses provide a large spectrum from public domain to right reserved licenses. This is implemented by a concise set of icons symbolising different rights, that can be combined. See table below:

Icon	Right	Description
	Attribution (BY)	Licensees may copy, distribute, display and perform the work and make derivative works and remixes based on it only if they give the author or licensor the credits (attribution) in the manner specified by these. Since version 2.0, all Creative Commons licences require attribution to the creator and include the BY element.
	Share-alike (SA)	Licensees may distribute derivative works only under a license identical ("not more restrictive") to the license that governs the original work. (See also copyleft.) Without share-alike, derivative works might be sublicensed with compatible but more restrictive license clauses, e.g. CC BY to CC BY-NC.)
	Non-commercial (NC)	Licensees may copy, distribute, display, and perform the work and make derivative works and remixes based on it only for non-commercial purposes.
	No Derivative Works (ND)	Licensees may copy, distribute, display and perform only verbatim copies of the work, not derivative works and remixes based on it. Since version 4.0, derivative works are allowed but must not be shared.

A practical and operational implementation of CC4.0 can be seen with the example of ICOS data https://www.icos-cp.eu/data.
However, it is recommended against using Creative Commons licenses for software. Instead, we encourage the use of one of the very good software licenses which are already available. Licenses listed as free by the Free Software Foundation and listed as “open source” by the Open Source Initiative could be considered.

2.2.3.2. Timeline

Follow progress timeline in WP4 and WP5 for their recommendation on respectively technical and political aspects of licensing and feedback towards the Atmosphere domain.

2.2.3.3. Interaction with ENVRI-FAIR WP4/5 Expert Environment

See last paragraph.

2.2.4. Semantic search for atmospheric ENVRI RI user interfaces

This task is on an implementation strategy for semantic search interfaces in data portals operated by the atmospheric ENVRI RIs. This includes:

- Describing and reviewing the state of the art in semantic search interfaces
- Describing typical use case scenarios and their demands
- Reviewing and giving recommendations on implementation technologies.

This task is to link to the respective efforts in ENVRI-FAIR WP7 which supports these implementation efforts.

Lead: Lara Ferrighi (SIOS)
Participants: Paul Eckhardt (ACTRIS), Richard Rud (ACTRIS), Oleg Mirzov (ICOS), Guillaume Brissebrat (IAGOS), Ingemar Häggström (EISCAT), Xiaofeng Liao (WP7)
Aim to be ready: 31 May 2020.

2.2.4.1. Work Plan

The working group will put effort on gaining knowledge on the topic of semantic search to be able to have some expertise within each RI. More specifically, the focus will be on the different conceptual and technical components that relate to the subject.
A shared Google drive folder has been created and will include several documents, both for internal use and for working on the tasks and deliverable.
The aim is to deliver a document in which we present a short background of the topic and review of existing approaches/efforts in the international community; propose and describe one use case; summarize and suggest a possible implementation to support semantic search within each RI and possibly across RI.
The topic of semantic search is strongly related to other working groups and task forces within the project, thus recommendations and deliverables from other groups should be pointed to in the deliverable. Working group participants will join other relevant working groups and task forces and report back on useful insights during the virtual meetings.

2.2.4.2. Timeline

The working group is having telephone conferences approximately every 4/5 weeks. It has been established in November 2019, thus expecting to have a total of 6/7 virtual meetings. More tight meetings might be expected during April and May 2020 to finalize and review the documents to be delivered.

2.2.4.3. Interaction with ENVRI-FAIR WP7 Expert Environment

Interaction with experts of WP7 is vital for this working group, both for the lack of previous experience within the participants and for technical guidance and harmonization of approaches with other domains. WP5 and WP7 will shed light upon conceptualization of RDF and linked data, as well as on the implementation of triple stores and SPARQL endpoints.

2.2.5. Improve Graphical User Interfaces

A Graphical User Interface (GUI) provides an intuitive way for humans to communicate with a computer system. The same system could be possibly accessed via text-based user interfaces, such as computer terminal, but these kinds of methods typically require deeper knowledge of the underlying system and are generally more suitable for machine-to-machine communication or to serve expert users. A GUI is a better option for users who want to access the offered services without necessarily knowing any details of the back-end API.
The FAIR principles promote data usage by requiring data to be findable, accessible, interoperable and reusable. Having a user-friendly and attractive looking GUI for data portal improves the findable and accessible parts from the human perspective. The interface for humans should not be neglected as it is the key component to engage new users and to serve occasionally visiting ones (see, e.g., https://www.rd-alliance.org/system/files/RDA%20DDP%20IG%20Use-Cases%20Requirements%20and%20Best%20Practices.pdf).

This working group aims to analyse and refine the current GUIs of the research infrastructures (RI) in the atmospheric domain. For any RI, a successfully established data portal is one of its most critical components. It is not enough to produce high quality data products, we have to be able to provide first-class interfaces for humans to access the data as well. This will make our data more FAIR, and by doing so increases an important key performance indicator: the number of users accessing our services.

Lead: Simo Tukiainen (ACTRIS)
Participants: Simo Tukiainen (ACTRIS), Lauri Kangassalo (ACTRIS), Cathrine Lund Myhre (ACTRIS), Markus Fiebig (ACTRIS), Richard Rud (ACTRIS), Claudio Dema (ACTRIS), Giuseppe D'Amigo (ACTRIS), Guillaume Brissebrat (ACTRIS), Jerome Tarniewicz (ICOS), Damien Boulanger (IAGOS), Ingemar Häggström (EISCAT), Lara Ferrighi (SIOS), Markus Stocker (ENVRI-FAIR WP7), Xiaofeng Liao (ENVRI-FAIR WP7)
Aim to be ready: 31 May 2020.

2.2.5.1. *Work Plan*

The concrete work plan consists of the following steps:

- Analysis of the existing WP8 RI GUIs. In particular, their
  - Technology
  - Design
  - Features
  - Performance
  - Pros / Cons
- The actual GUI revisions. The working group can utilize, for example:
  - Demonstrations
  - Feedback
  - Peer support
- Common web-services and modules
  Where suitable, develop common web-services and/or modules available for processing, visualizing, searching and/or filtering of data e.g. through Jupyter notebook environments

This work plan can be just a general framework because, in practice, the different RIs involved have different maturity levels concerning their GUIs. Some have highly advanced GUIs while others have just started to build them. However, because all RIs have fundamentally the same objective, to provide quality controlled atmospheric data products for end users, this working group is a useful forum to exchange information, share best practices and get feedback. The working group is mainly targeted for the developers actually implementing the GUIs.

2.2.5.2. *Timeline*

The group will have a remote meeting every two or three weeks since the end of February 2020. The first few meetings are reserved for overviewing the current status of the different GUIs, as they present their technology and design choices. In the forthcoming meetings, we will monitor the progress of the different RIs, as they work to improve their existing GUIs. This work will continue until October 2020 when many RIs are expected to have demonstrated significant progress. A report about results will be delivered. This report will include an analysis of the initial status of the different GUIs, recommendations about the features, options, design etc. for a data portal GUI, and overview of the progress made during the project.

2.2.5.3. *Interaction with ENVRI-FAIR WP7 Expert Environment*

WP7 experts are heavily involved in this working group and their expertise is crucial in order to have successful GUI improvements.

2.3. Tasks for second half of the project

The following tasks are scheduled for the second half of the project:

- **Common metadata standards and interfaces for use** of metadata
  Current metadata M2M interfaces focus on discovery metadata, i.e. metadata elements needed in searches for data. Use metadata comprise items such as uncertainties, data quality and associated measures, operating procedures, etc. This task links to ENVRI-FAIR WP5.
- **Machine-readable license and attribution metadata.**
  This task is to specify how license and attribution metadata are to be stored and served in order to make them machine-readable. This functionality is a prerequisite for a license broker service, i.e. a service that determines licensing options for a data product based on the licenses of the data serving as input for generating the product. This task links to ENVRI-FAIR WPs 4 and 5.
- **Common strategy for structured search interfaces, including common base set of searchable items**
  Structured search interfaces are preferred by expert users with a high knowledge about the data they are searching. This task is to give guidelines for a common architecture of structured search between atmospheric RIs.

- **Traceable post-production user feedback services**
  Data users routinely find issues with data they are analysing. This task is to work on an inter-RI service for collecting this feedback. This task links to ENVRI-FAIR WP7.
- **Data indexing in further data portals.**
  This addresses data portals considered not mature or relevant enough for the first round of implementation (Google, EOSC, WIGOS).
- **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.**
  Builds on the previous task for RESTful APIs for metadata and data. This task links to ENVRI-FAIR WP5.
- **PIDs for organisations and instruments to be determined in consultation with experts in ENVRI-FAIR WP7.**

3. Summary

The RIs of the ENVRI atmospheric subdomain have conducted a comprehensive analysis of the FAIRness of their data management. The identified gaps considered most important from a user perspective concern the following FAIRness functions:

- **Findable**: globally unique identifier; indexed in searchable resource.
- **Accessible**: (meta)data retrievable by standardised protocol
- **Interoperable**: common vocabulary
- **Re-usable**: established license, documented provenance, (meta)data meets community standards.

The tasks for implementing these functions were grouped by the following criteria:

1. Importance of the function for the user
2. Maturity of the function in the RI

Accordingly, the tasks were grouped into 3 categories:

1. Immediate implementation: the function implemented is highly important for the user, the implementation plan is consolidated.
2. Immediate implementation planning: the function implemented is highly important for the user, the implementation plan needs to be refined, know-how at the RIs needs to be established.
3. Task for second half of project: the function implemented is considered important, but implementation can be postponed to allow focussed work on the highly important FAIRness functions.

The Gantt-chart below illustrates the timeline of the implementation. Both implementations and task specific, detailed implementation plans will be verified by experts of WPs 5 and 7.

	Project Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
	Quarter in Year	1st QTR 19			2nd QTR 19			3rd QTR 19			4th QTR 19			1st QTR 20			2nd QTR 20			3rd QTR 20			4th QTR 20			1st QTR 21			2nd QTR 21			3rd QTR 21			4th QTR 21			1st QTR 22			2nd QTR 22			3rd QTR 22			4th QTR 22		
Task																																																
	Milestone																																																
1.1 Use of PIDs throughout workflow																																																
	ACTRIS: determine PID solution																																															
	ACTRIS: implement primary DOIs																																														
	ACTRIS: PIDs pre-products implemented																																														
	ACTRIS: ORCiD and org. PIDs implemented																																														
	EISCAT: PIDs pre-products implemtd.																																														
	IAGOS: ePIC registration, internal ORCiD PIDs																																													
	IAGOS: workflow with PIDs, definition																																														
	IAGOS: workflow with PIDs, implemented																																														
1.2 Standard interfaces for (meta)data access																																															
	ACTRIS: metadata repository, 1st version																																														
	ACTRIS: metadata M2M interface available																																														
	EISCAT: DataCite metadata defined (Level 3)																																														
	EISCAT: DataCite metadata defined (Level 1&2																																														
	EISCAT: metadata ingested in B2FIND																																														
	EISCAT: OAI-PMH interface available																																														
	IAGOS: WIGOS metadata implemented																																														
	IAGOS: REST API for metadata access implemntd.																																														
	IAGOS: THREDDS server for data access implmntd.																																														
	IAGOS: REST API for data access implemented																																														
	ICOS: metadata for Google search published																																														
	ICOS: OAI-PMH metadata server implemented																																														
	ICOS: publish ICOS controlled vocabulary																																														
	ICOS: implement WIS metadata provision																																														
	ICOS: implement link to external documents																																														
	ICOS: implement license as resolvable URL																																														
	ICOS: full provenance metadata implemented																																														
	ICOS: REST API for data subsetting implemented																																														
	ICOS: NetCDF-CF data export implemented																																														
	ICOS: THREDDS server updated																																														
	ICOS: increased search capabilities in interface																																														
1.3 Data indexing in WIS and GEOSS																																																
	ACTRIS: link to GISC in WIS implemented																																															
	EISCAT: ontology updated																																														
	EISCAT: set up WIS compatible metadata																																														
	EISCAT: link to WIS established																																														
	IAGOS: WIGOS metadata implemented																																														
	IAGOS: link to WIS established																																														
	ICOS: link to GEOSS established																																														
1.4 Common authentication schemes																																															
	ACTRIS: ORCiD authentication implemented																																														
	IAGOS: ORCiD / eduGAIN authentication avail.																																														
	ICOS: common authentication implemented																																														
2.1 Domain vocabulary / ontology																																															
2.2 Documentation of provenance																																															
2.3 Recommendations for licenses																																														
2.4 Semantic search																																															
2.5 Graphical user interfaces																																														