D8.8
The FAIRness of SIOS

Work Package	WP8
Lead partner	SIOS
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	28 February 2023
Submission date	03 March 2023

Deliverable abstract

This deliverable provides the "The FAIRness of SIOS" and refers to the status of the implementation in relation to the updated, 2nd version of the implementation plan from September 2021. This includes the implementation of services for meta data and data access including machine-to-machine interfaces within the ENVRI hub, and outside, utilising e.g., vocabulary and ontology specifications. It summarizes the FAIR implementation plan, the progress on FAIRness and the remaining tasks for SIOS. The main achievements are listed and an outlook on the next 5 years is given.

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Lara Ferrighi	SIOS	01.03.2023
Contributing Authors			
Reviewer(s)	Cathrine Lund Myhre	NILU	02.03.2023
Approver	Andreas Petzold	FZJ	03.03.2023

DELIVERY LOG

Issue	Date	Comment	Author
V 0.1			

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.4471374.

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

ENVRI
FAIR

D8.8 – The FAIRness of SIOS

1   Introduction

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 (atmosphere), and SIOS (atmosphere).

D8.3 Atmospheric subdomain implementation plan describing the implementation plan of FAIRness for the atmospheric sub-domain was ready March 2020. Later there is a revised version produced, taking the implementation of the ENVRI-Hub into account. This 2nd version was finalized September 2021, and is the most recent implementation plan for WP8.

Towards the end of the project, a set of deliverables are due M50 describing the implementation of FAIRness, for each RI.
- D8.4: The FAIRness of ACTRIS
- D8.5: The FAIRness of EISCAT_3D
- D8.6: The FAIRness of IAGOS
- D8.7: The FAIRness of ICOS-atm
- D8.8: The FAIRness of SIOS

This deliverable provides the "The FAIRness of SIOS" and refers to the status of the implementation in relation to the updated, 2nd version of the implementation plan from September 2021.
Furthermore, a FAIRness assessment is performed in the first part of 2023 to monitor the complete progress over the project period using FAIR Implementation Profiles (FIPs). This is ready in April and will be reported in the Deliverable "D8.13 Atmospheric subdomain FAIRness assessment report" due May 2023.

2   Background and the starting point for SIOS

2.1   Gap analysis

The RIs of the ENVRI atmospheric subdomain conducted a comprehensive analysis of the FAIRness of their data centre, data curation and management late 2018.

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. Although SIOS does not own data, it serves as a central node where metadata records are harvested and exposed to the community. SIOS also is responsible for promoting FAIRness within the data centres, as well as for providing added value services though human and machine interfaces. An initial assessment of the FAIRness status within SIOS has highlighter some strengths, but also some weaknesses that could be improved though the ENVRI-FAIR project and served as basis for the implementation plan within WP8.

At findability level, the lack or partial implementation of metadata endpoints to be harvested was identified as a gap. Moreover, the SIOS Data Access Point did not provide a user-friendly search interface, lacking faceted search, and a limited number of controlled vocabularies.

Access to data also had some gaps as OPeNDAP protocol was not available for some partner repositories. On the other hand, the lack of a common authentication schema was not considered relevant for SIOS, as access to data is open also to anonymous users.

Interoperability for discovery metadata, although quite mature (based on ISO19115 and DIF), was still suffering from some lack of standardization and difficult translation of information between the data centres and the central node, while interoperability at the data level suffered from major gaps due to standardization of metadata. Internal vocabularies were not resolvable but provided in a public git repository.

The lack of harmonization on how to provide license information was also a clear problem, and such information was often either picked up as simple string or not harvested at all from partner repositories.

Based on the gap analysis and the 1st FIP, a detailed implementation plan was developed for SIOS.

2.2 The FAIR implementation plan for SIOS

The main implementation tasks for SIOS have been related to the harmonization, exchange and exposure of metadata, both from a machine-to-machine and a human perspective. The implementation plan, and tasks, accordingly, have been updated during the project to accommodate new requirements and development within the working groups and task forces organized within ENVRI-FAIR. The prioritized tasks for SIOS have been:

- The deployment of standardized endpoints for metadata harvesting and improving content harmonization/standardization.
- Updating the SIOS metadata model to accommodate new requirements, such as URL for licenses and use of CF/GCMD vocabularies for keywords
- Implementation of a vocabulary server to host internal vocabularies
- Major upgrade of backend and frontend solutions to include new features (e.g., faceted dynamic search) and services (e.g., link to vocabularies)
- Integration of selected data from the SIOS data portal into the WP8 science demonstrator
- Registration of SIOS services on the ENVRI catalogue of services through TF1
- Improve and disseminate knowledge within SIOS about citation best practice, FAIR vocabularies and semantic technologies

3   Progress on FAIRness for SIOS

This section includes the progress reports of implementation of FAIRness within SIOS in relation to the implementation plan and identification of potential open issues and need for changes in the implementation plans for each RI.

The symbols used are

🟢	Completed
🟡	In progress
⚫	Not started, according to plan (0)
🔴	New tasks not in the implementation plan

It should be noted that some of the tasks below (i.e., implementation of SSO, set up PythonAPI endpoint, provide service records and link data portal to vocabulary server) where not in the first version of the implementation plan, but they have been added in its revised version, according to the developments achieved through ENVRI-FAIR task forces and working groups.

Table 1: SIOS summary of implementation of FAIRness

Task and Milestones under implementation	Due month Implementation plan	2020	2022	Final
1.1 Use of PIDs throughout workflow				
SIOS: METNO supports DOI minting	1st QTR 20	🟢	🟢	🟢
1.2 Standard interfaces for (meta)data access				
SIOS: deploy pyCSW software	3rd QTR 20	🟡	🟢	🟢
SIOS: indexing of METNO data in pycsw	4th QTR 20	⚫	🟢	🟢
SIOS: impl. OAI-PMH at NILU and CNR	2nd QTR 22	🟡	🟢	🟢
SIOS: partner data in central pycsw	4th QTR 21	⚫	🟢	🟢
SIOS: translation to INSPIRE	1st QTR 22	⚫	🟢	🟢
SIOS: translation to WMO Core	2nd QTR 22	⚫	🟡	🟢
SIOS: NILU OPeNDAP implementation	4th QTR 20	🟡	🟢	🟢
SIOS: CNR OPeNDAP implementation	2nd QTR 22	🟡	🟢	🟢
SIOS: support link to WIGOS	2nd QTR 20	🟢	🟢	🟢
SIOS: update indexing engine	4th QTR 21	⚫	🟢	🟢
1.3 Data indexing in WIS and GEOSS				
SIOS: indexing in WIS and GEOSS	4th QTR 22	⚫	🟡	🟡
1.4 Common authentication schemes				
SIOS: implementation of SSO (orcID)	4th QTR 22	⚫	🟢	🟢
1.5 Service endpoint to ENVRI-hub				
SIOS: set up PythonAPI endpoint	3rd QTR 21	⚫	🟢	🟢
SIOS: provide service records	4th QTR 22	⚫	🟡	🟢
1.6 GUI recommendations				
SIOS: release of new data portal	3rd QTR 21	⚫	🟢	🟢
SIOS: link data portal to vocabulary server	2nd QTR 22	⚫	🟢	🟢
1.7 Document provenance				
SIOS: Expose provenance information	4th QTR 22	⚫	⚫	⚫
1.8 Domain vocabulary / ontology				
SIOS: Domain vocabulary / ontology	3rd QTR 22	⚫	🟡	🟢
1.9 Recommendation for licenses				
SIOS: licenses	1st QTR 21	⚫	🟢	🟢

4 Remaining tasks

In progress:
Indexing in WIS/GEOSS is still ongoing. Several harvesting tests have been made. We have been in contact with GISC operated by DWD for the CSW endpoint to be harvested using the INSPIRE profile. Some minor changes are needed from SIOS side which are ongoing. Additionally, since SIOS is an aggregator, thus exposing records that are also exposed directly from partners repository, the harvester will need to set up a filtering script to avoid metadata duplication for partner's data that are already harvested directly by WIS. Considering other related activities, we plan to have this connection established in autumn 2023.

Need revision of plans:
An example of provision of minimal provenance information for selected data at METNO has not been implemented. As SIOS does not own data, it relies on partners to provide information on data records. A separate project will start at METNO to also work on provenance, thus, due to minimum overlap in time between the two projects, activities could only be initiated. A strategy on how to make provenance metadata accessible from partner repositories, such that harvesting of relevant information can be achieved from the central node is still an issue which should be discussed within the SIOS Data Management System working group.

Not relevant anymore:
We consider that setting up an example of minimal provenance is not particularly relevant at this point, as involvement of partners require a mature state of handling, recording and provision of provenance, which is not available yet.

5 Main achievements and improvement of FAIRness for SIOS

Below is a list of the main achievements for SIOS since the beginning of the project, some are specifically related to the implementation plan in WP8, some are also connected to other activities within ENVRI-FAIR:

- Implementation of a dedicated M2M endpoint (pyCSW) for all data exposed on the SIOS data access portal providing OGC CSW, OAH-PMH, OpenSearch protocols
- Deployment and integration of OAI-PMH and OPeNDAP for participating partners in WP8
- Implementation of dynamic/faceted search for data using controlled vocabularies in the SIOS data portal (https://sios-svalbard.org/metsis/search)
- Exposure of linked-based license specification, when provided from contributing data centers, on the data portal
- Exposure of downloadable metadata files from the SIOS data access portal in different formats (original internal format, DIF, ISO, WMO, INSPIRE)
- Resolve controlled vocabularies from the SIOS data access portal directly on a vocabulary server (https://vocab.met.no/en/)
- Link CF standard names, when provided, to the NERC vocabulary server from the SIOS data access portal, and in general a major improvement in harmonizing variables and keywords
- Exposure of DOIs and citation elements on the SIOS access portal, for data with DOI assigned
- Creation of a Privacy Policy Statement which can be accessed as a banner on the SIOS data portal
- Integration of the Single Sign On (OpenID Connect) for users requiring access to internal areas or services (still this is not needed for accessing datasets).

5.1 What can we do now which we could not before advancing of the FAIRness?

Most of the achievements above mentioned have been leading to major improvements of standardization and services available within SIOS:

- Metadata records are harvestable in a unique endpoint using standardized protocols
- Metadata records are available in different standardized formats
- A vocabulary server is now available, and it is hosting vocabularies used for the internal metadata standard, including relevant mapping to other vocabularies
- Variables and other metadata values that belong to controlled vocabularies are accessible as URIs, allowing users to access the complete information about terms
- License and citations are now harmonized
- Showcase interoperability through the Science Demonstrator of WP8

5.2 What would we like to do / achieve within the next 5 years?

As the basis for higher standardization and interoperability are improved through the project, we foresee the possibility to finalize and further improve some tasks:

- Interact with WMO for exposing data from aggregators, such as SIOS, in WIS, as mentioned in section 4.
- Update the vocabulary server to Docker and allow for exposure of other vocabularies if needed
- Automatize the request for DOI at METNO
- Further improve interoperability at data level
- Further develop/improve visualization services on the SIOS Data Access Portal