D10.4
REPORT ON IMPLEMENTATION ACTIVITIES

Work Package	10
Lead partner	UKRI
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	2020-12-31
Submission date	2020-12-29

Deliverable abstract

D10.4 "Report on Implementation Activities" describes the work towards making the RIs in the solid earth subdomain FAIR in terms of availability of assets.

In this deliverable we therefore describe for each subsequent task, the work done to address the gaps identified by T10.1 in D10.1 and the implementation plans of T10.2 in D10.2.

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	Keith G Jeffery	UKRI	2020-12-08
Contributing Authors	Daniele Bailo	INGV	2020-12-08
	Jean-Baptiste Rocquencourt	BRGM	2020-12-08
	Tor Langeland	NORCE	2020-12-08
	Alessandro Spinuso	KNMI	2020-12-08
	Luca Trani	KNMI	2020-12-08
	Michele Manunta	CNR	2020-12-08
	Ivan Rodero	EMSO	2020-12-08
Reviewer(s)	Maggie Hellström Erwann Quimber	ICOS (ULUND) IFREMER	2020-12-17 2020-12-17
Approver	Andreas Petzold	FZJ	2020-12-29

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.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

D10.4 - REPORT ON IMPLEMENTATION

1 Introduction

The main goal of this deliverable is to report about the implementing FAIR principles in the Solid Earth domain community. FAIR principles have the merit of creating a common background of knowledge to engage communities in providing data in a standard way thus easing interoperability and data sharing. However, they do not explicitly refer to methodologies nor actual technical activities that RIs might adopt in order to implement software and technologies actually delivering FAIR data.
In this deliverable we therefore describe the technical activities to be undertaken to fill the gaps identified in D10.2 (the roadmap) [1] following the analysis of FAIRness (D10.1) [2].

2 Overview

Research Infrastructure (RIs) implementers setting up or upgrading an existing RI usually follow a system development life-cycle (SDLC) process[1]. In the current deliverable, an SDLC inspired by the waterfall development model and by earlier work by the INGV team[3] was proposed, which encompasses the following steps: a) *analysis*, including use cases and requirements collection; b) *design*, including architecture design and identification of architectural components matching requirements, c) *implementation*, through software developments and adoption of suitable technologies, d) *test*, e) *operation and maintenance*. Here we consider steps from b) to d) and the work is following a rapid systems development method based on pitches or short sprints of work within the Shape Up method [4].

Tasks 10.1 and 10.2 (M12 and M18 respectively) provided the analysis of FAIRness and the roadmap for implementation respectively. Here we report progress through Tasks 10.3-10.7 covering the various areas of work within EPOS ranging from the Integrated Core Services – Central (ICS-C) (T10.3) to the Integrated Core Services – Distributed (ICS-D) (T10.4: where a prototype has been developed in the seismology community) to three of the TCS domains – those involved in ENVRI-FAIR - covering seismology, satellite and marine (T10.5-10.7) (Figure 1).

[1] http://docshare01.docshare.tips/files/3771/37719429.pdf

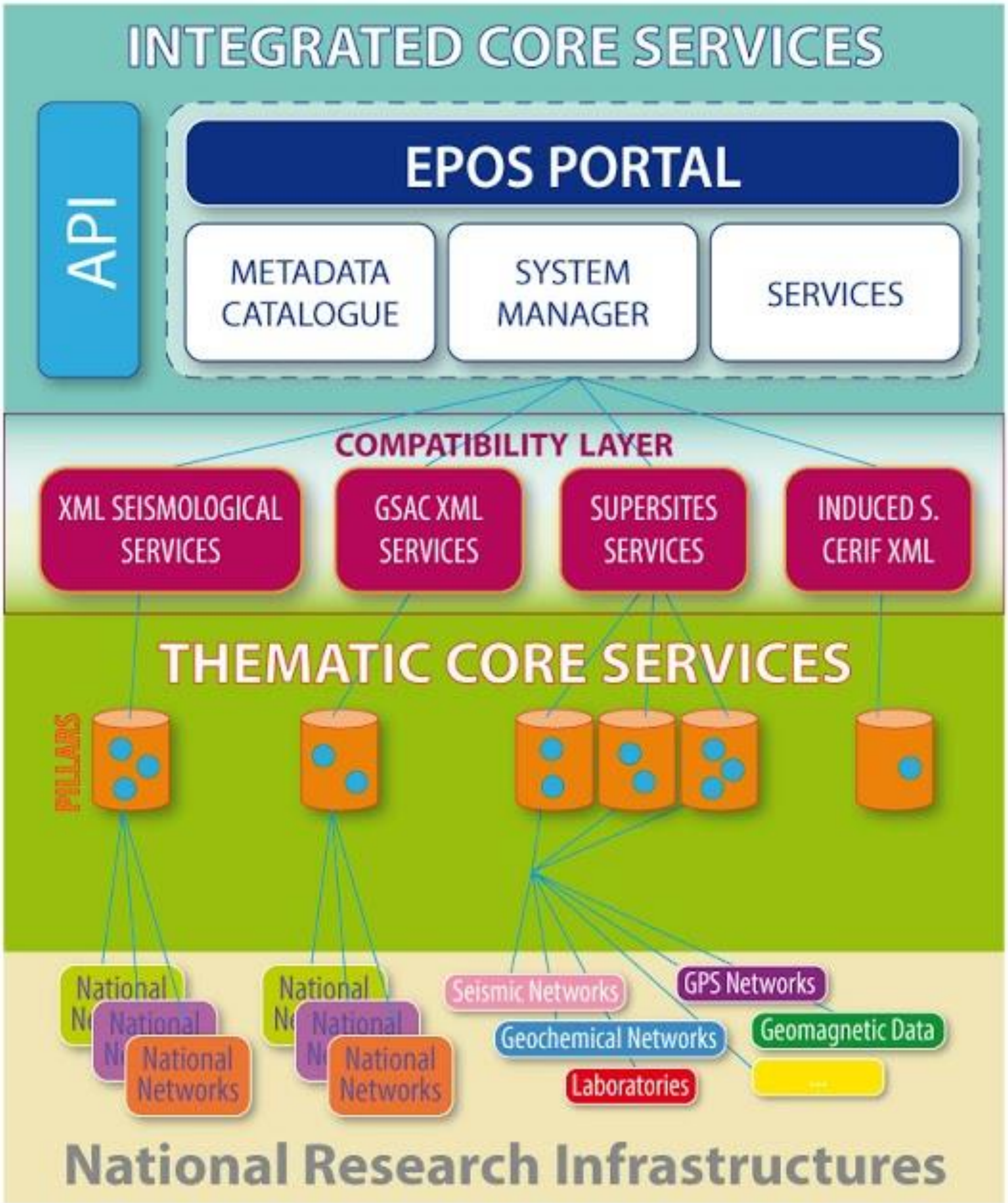

Figure 1: The Structure of EPOS

3 Task 10.3 ICS-C

This chapter describes the steps to improve – through implementation - FAIRness in ICS-C. ICS-C is the central hub or EPOS consisting of the rich metadata catalog of assets stored in CERIF and software to manage: user interaction; access to the metadata catalog; dispatch of requests to asset suppliers; and maintenance of the metadata catalog by the TCS asset suppliers.

In general ICS-C is FAIR; it complies with the Force-11 Principles as follows:

FAIR PRINCIPLE	EPOS-ICS
TO BE FINDABLE:	
F1. (meta)data are assigned a globally unique and eternally persistent identifier.	ICS-C catalog has a unique EPOS persistent resolvable Identifier for each record and also supports federated (role-based, temporal duration limited) IDs related by semantic relationships to the core ID.
F2. data are described with rich metadata.	ICS-C catalog uses CERIF[2] (a EU recommendation to Member States) as the backend model which is demonstrably a rich superset of other metadata standards, thus encouraging improved searching, contextualisation, accessibility, interoperability and re-use. ICS-C uses EPOS-DCAT-AP[3] based on DCAT v1[4] for the metadata ingestion.
F3. (meta)data are registered or indexed in a searchable resource.	ICS-C catalog is a registered searchable resource. Certification via CoreTrustSeal[5] is being investigated.
F4. metadata specify the data identifier.	ICS-C metadata records not only specify the asset identifier but may provide a qualified reference (I3) to it with role and temporal validity
TO BE ACCESSIBLE:	
A1 (meta)data are retrievable by their identifier using a standardised communications protocol.	ICS-C provides this facility with both API and GUI
A1.1 the protocol is open, free, and universally implementable.	ICS-C provides this facility through API and GUI
A1.2 the protocol allows for an authentication and authorisation procedure, where necessary.	ICS-C provides authentication aligned with AARC/GEANT blueprint architecture[6]. Various authorisation schemes exist and work on harmonising is ongoing
A2 metadata are accessible, even when the data are no longer available.	ICS-C provides for "tombstone metadata" also linked with curation and provenance information
TO BE INTEROPERABLE:	
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	ICS-C has the capability by providing JSON API serialization. The API is an extended version of Open Search. CERIF provides n-tuples of base entities and link entities thus providing a fully connected graph structure. While CERIF can be (and has been) implemented in several language environments, in EPOS we use classical n-tuples to ensure referential and functional integrity in a closed-world environment so permitting deduction and induction.
I2. (meta)data use vocabularies that follow FAIR principles.	CERIF- used in ICS-C - has a semantic layer for vocabularies, imported from various sources and following FAIR principles since the semantic layer of CERIF is a mirror image of the syntactic layer in properties such as integrity.

[2] https://www.eurocris.org/services/main-features-cerif
[3] https://www.epos-eu.org/representing-cross-disciplinary-knowledge-solid-earth-sciences-epos-dcat-ap
[4] https://www.w3.org/TR/2020/SPSD-vocab-dcat-20200204/
[5] https://www.coretrustseal.org/
[6] https://connect.geant.org/2019/04/30/the-final-aarc-blueprint-architecture-a-community-first-approach

FAIR PRINCIPLE	EPOS-ICS
I3. (meta)data include qualified references to other (meta)data.	As indicated under F4, CERIF provides rich qualified references or minimally references with referential and functional integrity
TO BE RE-USABLE:	
R1. meta(data) have a plurality of accurate and relevant attributes.	As indicated under F2, CERIF provides for rich metadata with many attributes, stored in n-tuples connected in a fully-connected graph structure using semantic and temporal relationships (qualified references). EPOS-DCAT-AP – as the intermediate format for ingestion - supports progressively the richness of CERIF.
R1.1. (meta)data are released with a clear and accessible data usage license.	CERIF has the capability to store licence documentation (including appropriate links) and also to provide a licence name for searching. However, work is ongoing to translate licence information to authorisation keys for access.
R1.2. (meta)data are associated with their provenance.	CERIF comes with inbuilt provenance because of the temporal attributes in the role-based linking relationships (qualified references) and the EPOS catalog is populated appropriately
R1.3. (meta)data meet domain-relevant community standards.	In EPOS 17 different community metadata standards have been converted to CERIF to provide a harmonised catalog, so domain-relevant standards are preserved in terms of content, but in a harmonised form with richer syntax and semantics. The exchange format for ingestion used is EPOS-DCAT-AP which was created – and is being updated incrementally - to match CERIF richness as EPOS communities improve their metadata. The API answers to a request is with JSON which is a good starting point for interoperation.

Table 1: FAIR Principles and EPOS Conformance

As indicated from the above table, there are several areas of ongoing work that are being addressed:

- Identifiers
- Repository Certification (CoreTrustSeal)
- Authentication and authorisation
- Vocabularies and vocabulary tool
- Licence translation for autonomic access (links with authentication)
- Extending EPOS-DCAT-AP and CERIF model to accommodate additional metadata needed

3.1 Identifiers

In the geoscience domain many identifier systems are in use including DoI, URI, UUID, community specific identifiers, locally unique identifiers. Moreover, different communities have different identifier systems used within their community (commonly internationally) and thus following standards outside of EPOS. Thus, EPOS has to utilise the federated ID mechanism of CERIF to accommodate different IDs for the same digital object – but commonly related to access in different roles or by different communities.

Work is ongoing to identify the different identifier systems in use and their degree of FAIRness. Some are reported in following chapters.

3.2 Repository Certification (CoreTrustSeal)

There is some advantage in registering the EPOS ICS-C catalog with CoreTrustSeal. Some EPOS asset suppliers have already such approval. However, the effort involved to fulfil the requirements is not inconsiderable and so a cost-benefit analysis is planned. The major benefit is assurance for a user, the major disbenefit is the resource required to achieve certification.

3.3 Authentication and Authorisation

Within the EPOS IT Board development environment (Shape-Up, which require development activities to be structured as 'pitches' with clear specifications and resources) there is a pitch on authorisation. It has collected information from the TCS and documented the landscape. It has already agreed that we use AARC/GEANT authentication mechanisms with tokens, evolving together to the OpenID standard. Thereafter, the intention is to utilise metadata information concerning the rights and responsibilities of the user and the availability conditions of the asset to provide authorisation keys to be encrypted into the authentication token. This work is ongoing.

3.4 Vocabularies

In link with other communities in ENVRI and wider, and in order to facilitate cross-domain search, providing a consistent way of managing vocabularies prevents communities redoing the work done by others. This boils down to choosing the right tools and defining the policies of managing vocabularies: namespace, workflow and responsibilities for vocabularies. Being pragmatic, we will reuse the one used in the TCS geology community[7]).

The tool has been installed and we are currently deploying a registry tool for defining and publishing controlled vocabularies. It is a linked data-based registry tool called UkgovLD[8], already known and used by many authoritative organizations. It provides a web API + a GUI for creating, managing, publishing, and querying (via LD-API[9] or SPARQL[10]) registers. A register is a container that allows linking a vocabulary to its metadata. Thus, it allows managing the whole workflow and life cycle of a vocabulary and its entities.

The main vocabulary type managed in this tool are Simple Knowledge Organisation System (SKOS)[11] thesauri (Concept Schemes or Collections). SKOS provides a simple semantic framework for semantics of concepts, i.e., their definitions as well as the hierarchy and relations that organise them. In addition, other registry features are included in the tool: for instance, it allows a nesting structure so vocabularies can be organized under containers. Since it is a linked data-based tool, it provides direct dereferenceable identifiers (HTTP URIs) for all the managed vocabulary and vocabulary concepts. In addition to URI, direct access to vocabularies content, this tool provides a SPARQL endpoint service for accessing the data in another standard way (SPARQL), which allows exploiting the full potential of semantic definitions and relations. The developments related to this tool are still active, allowing it to benefit from new features such as register of register federation for example.
The tool has been installed and will soon be open with an EPOS-ERIC based URI.

[7] http://data.geoscience.earth/ncl/
[8] https://github.com/UKGovLD/registry-core
[9] https://json-ld.org/spec/latest/json-ld-api-best-practices/
[10] https://www.w3.org/TR/rdf-sparql-query/
[11] https://www.w3.org/2004/02/skos/

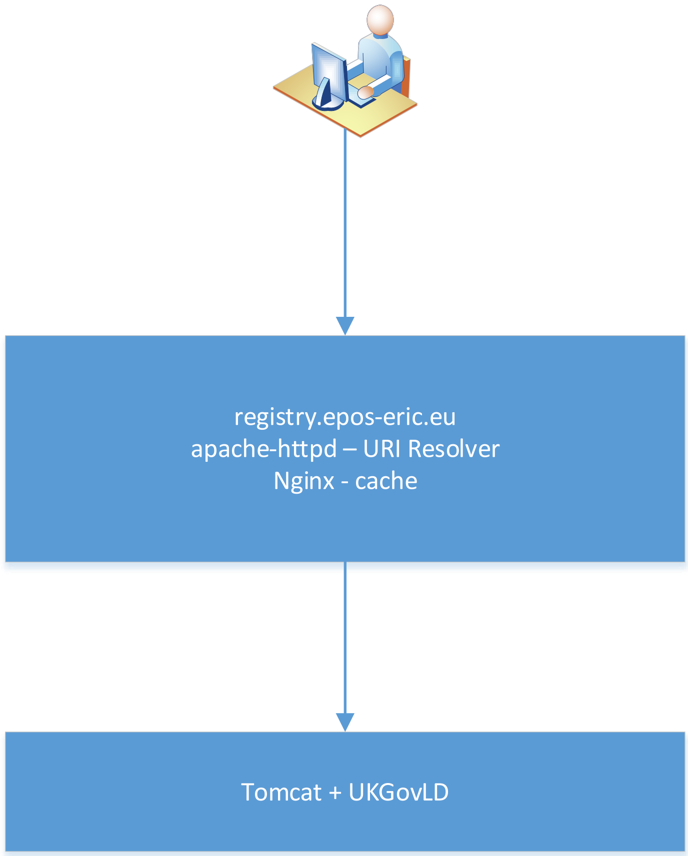

Figure 2: Vocabulary tool technical architecture

The current plan and state is as follows:

- Define the use of a vocabulary tool → done
- Start the implementation within EPOS landscape including translation into and ingestion into CERIF → todo
- Identify the common guidelines for vocabulary→ 80% complete
- Refine the guidelines according to EPOS landscape → todo
- Advertise this tool and guidelines to the different bodies → introduction done in ICS-TCS meeting

3.5 Licence translation for autonomic access (links with authentication)

As indicated above licence information can be stored (a) as a document (with a link from the metadata record); (b) as a licence name. However, the real requirement is for the implied access rules to digital objects of the licence to be encoded such that they can be used within an authorisation framework. Work on licence policy is ongoing in the EPOS Policy Group. Each asset is to be licensed, according to the wishes of the owning organisation. However, this defines the scope of asset use, it does not necessarily indicate the authorisation constraints concerning the asset related to (classes of) users. The authorisation pitch (mentioned above) will investigate how best to represent licence intent in metadata and hence through to authorisation mechanisms.

3.6 Extending EPOS-DCAT-AP towards the CERIF model

EPOS-DCAT-AP is used as an intermediary format for ingestion of metadata records, providing a target for conversion from the many metadata formats used by asset suppliers. Once metadata from many asset suppliers are in EPOS-DCAT-AP it is easier to convert to the richer CERIF standard used for the EPOS metadata catalog.

Work has been done in extending EPOS-DCAT-AP to align with the CERIF model implementation, first to cover many metadata elements not available in DCAT v1 but available within CERIF and required by the asset suppliers and users, coloured gold in (Figure 3).

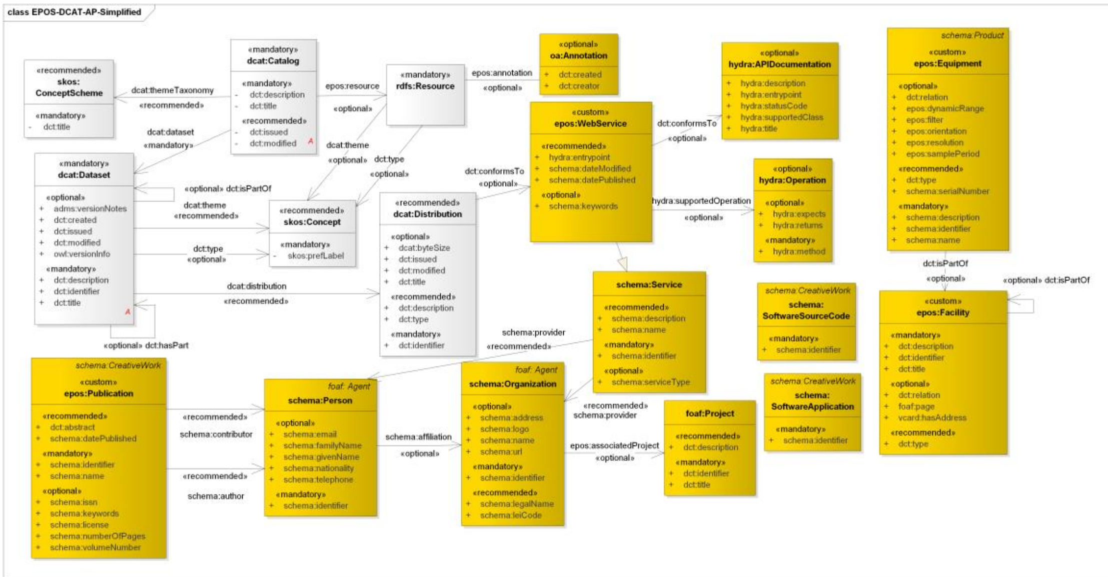

Figure 3: Extending DCAT v1 to EPOS-DCAT-AP

This (Figure 3) demonstrates clearly how much DCAT had to be extended (to EPOS-DCAT-AP) to approach the richness of CERIF – and then only for a limited set of entities common to the asset suppliers of EPOS. Current work is extending EPOS-DCAT-AP even further to support the CERIF entities Facility and Equipment, required for several purposes including Trans-National Access (TNA). We intend to demonstrate this extension with EPOS communities involved in ENVRI-FAIR. Further extensions are likely to be needed in EPOS-DCAT-AP to manage the metadata required as defined in the areas of work above. There is consequential work on ingestion from EPOS-DCAT-AP into the CERIF metadata catalog of ICS-C.

As EPOS wants to improve its FAIRness, we intend to study a potential convergence with DCATv2. However, since EPOS is quite mature in its model, this convergence will undergo study to match the EPOS Delivery Framework, as we cannot break the actual features of EPOS.

3.7 Search API

The Search API has been defined based on open search with additional endpoints.
However, there are already existing ISO standards that could cover the needs of EPOS and enhance EPOS FAIRness.
Indeed, by reusing an already standardised API this would greatly facilitate the usage by any user: human or machine.
We aim to find potential better suited API endpoint with the following constraints:

- Have to be standardised: DESIRABLE
- Opensource tools already existing : DESIRABLE
- Output in preference by descending order: DCAT-AP/json, geojson, json: MUST
- Can be extended to support CERIF/XML: DESIRABLE
- Content negotiation for the previous model and serialization: DESIRABLE

So far, we have found these potential API offerings: STAC, OGC API Records. This list has to be extended, as we do not want to miss any relevant API standard. We will have to see where we could partially or totally reuse an existing API framework. But also, as EPOS is quite mature in its

development, it should be investigated how and to what extent the EPOS experience would benefit these API communities by providing them EPOS use cases in order for them to enhance their API.
As such a presentation of EPOS has been done in OGC meeting for this community to take into account EPOS uses case which comes with the richness of the whole Earth Science community. This provoked interest and some interest in cooperation.

4 Task 10.4 - ICS-D

4.1 Introduction

ICS-D is the term for the distributed ICS services. They are to be managed from the ICS-C portal environment. The overall planned architecture for the life of a workflow in EPOS is indicated by the following diagram (Figure 4). The idea is to support a Virtual Research Environment (VRE) to allow a researcher or other end-user to compose workflows to satisfy their requirements easily, accurately and optimally. The workflows are likely – after collecting and contextualising assets - to involve analytics, simulations and visualisations.

However, in order to gain relevant experience to ensure the architecture is applicable pilot projects have been done (a) external to ENVRI-FAIR in the VRE4EIC project[12] (where ENVRIPlus[13]was also involved) and where a general workflow deployment mechanism was provided, tested using TAVERNA[14]; (b) in other EC projects concerning workflow deployment rather than composition where EPOS was represented namely PaaSage[15] concerning optimal deployment based on computing resources) and MELODIC[16] (optimising based also on data location and locality).

This chapter describes a major effort being made in the seismological community to provide an ICS-D based on the use of Jupyter. It outlines performed and planned work on the implementation of FAIR principles for the EPOS ICS-D services. These services will form a VRE for data analysis and visualisation, which can be launched from a data-discovery session in the EPOS ICS-C data portal.

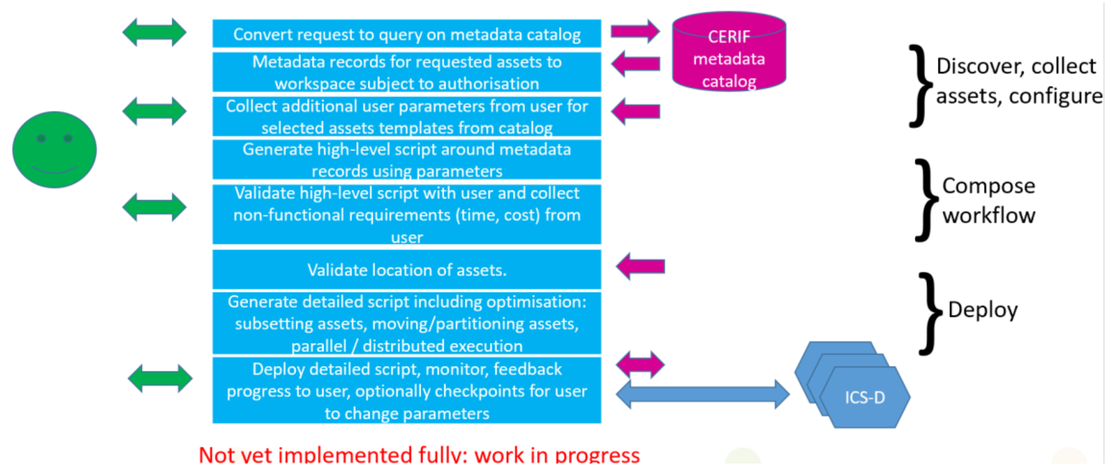

Figure 4: General EPOS Workflow

12 https://vre4eic.ercim.eu/
13 https://www.envriplus.eu/
14 https://taverna.incubator.apache.org/
15 https://paasage.ercim.eu/
16 https://h2020.melodic.cloud/

4.2 Components

The ICS-D to be demonstrated in this project consists of the following components:

- SWIRRL API[17]. This is a framework used for executing workflows for staging required data and for starting the specific ICS-D, in this project specifically Enlighten-web and Jupyter notebook.
- Enlighten-web[18] – This is a web application for interactive visual analysis. In this project we will implement a visualisation workflow that prepares data from the ICS-C workspace and that uses Enlighten-web for visualising the data.
- Jupyter notebook – This is a web programming environment for analysis. It will be made available in the same manner as described for Enlighten-web.
- Hosting facility - Deployment of the ICS-D services at the EPOS e-Infrastructure.

4.2.1 SWIRRL

SWIRRL is a web service API that allows Science Gateways to easily integrate data analysis and visualisation tools in their websites and re-purpose them to their users. A comprehensive overview of the system is shown in (Figure 5) below. The API deals, on behalf of the clients, with the underlying complexity of requesting and organising resources in a target cloud platform hosting the Kubernetes container-orchestration system. By combining storage and tools, such as Jupyter notebooks[19] and Enlighten-web, implemented as containerised services, the API creates dedicated working sessions on-demand.

Thanks to the API's workflow execution endpoint, which spawns job executing Common Workflow Language (CWL)[20] workflows for data staging and batch processing, SWIRRL sessions can be populated with raw data of interest collected from external data providers. Staged data is considered immutable with identifiers from SWIRRL. In the occurrence of updates, older versions are preserved.

The system is designed to offer customisation and reproducibility. Notebooks in the Jupyter environment can be further customised with additional or updated libraries, and the provenance of such changes is automatically captured. The recording of provenance is performed for each method of the API's affecting the session.

SWIRRL generates and stores provenance data thanks to two dedicated components: the PROV-Template Catalogue and the Neo4j database. Both are integrated as microservices and exposed through the API (Figure 5).

[17] https://gitlab.com/KNMI-OSS/swirrl/swirrl-api
[18] https://demonstrator.webfarm.cmr.no/covid19/doc/enlweb/html/filtering.html
[19] : https://gitlab.com/KNMI-OSS/swirrl/jupyterswirrlui
[20] https://www.commonwl.org

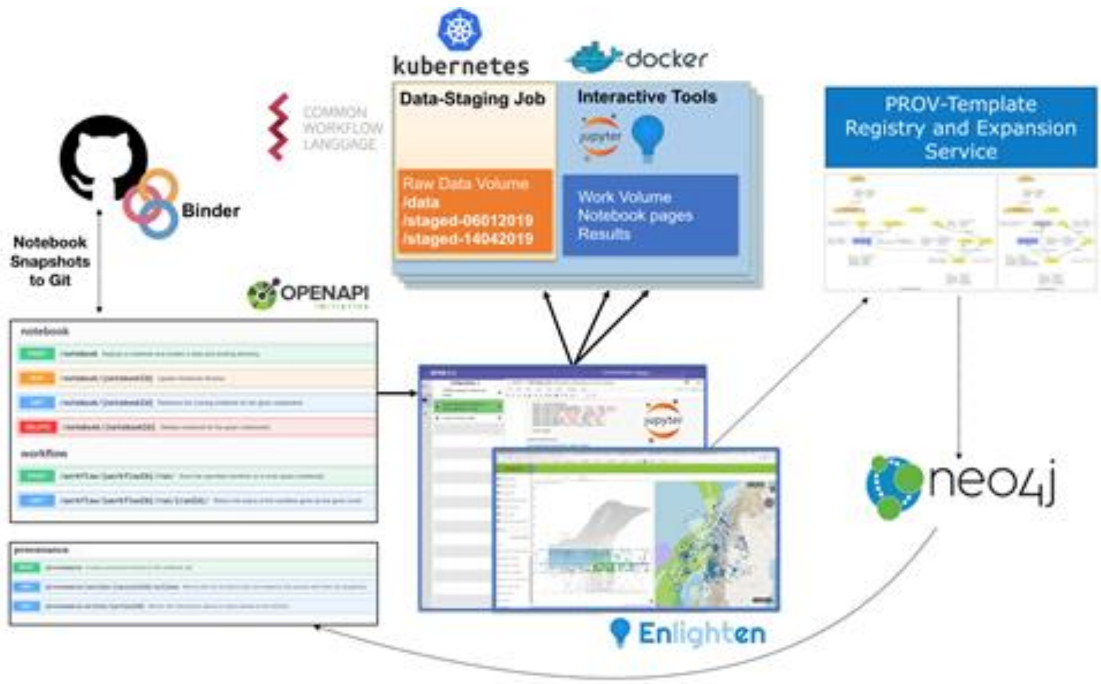

Figure 5: SWIRRL overview

4.2.2 Enlighten-Web

Enlighten-web facilitates interactive visual analysis of large multidimensional data sets. Enlighten-web uses the Cesium map engine[21]. The user can use brushing and linking to explore complex data sets to discover correlations and interesting properties hidden in the data. In this project, an Enlighten-web Docker image has been delivered to KNMI for their integration in the workflow. We have done some adaptations of the Enlighten-web code for this purpose.
The Kubernetes cluster for running ICS-D services has been deployed at BGRM, reusing the Infrastructure as a Code approach from EPOS.

4.3 Filling the Gaps

We have performed a FAIR assessment and gap analysis with regards to FAIRness of the ICS-D components. Based on the gap analysis, tasks were defined for filling the identified gaps. These tasks are described in D10.2 (Roadmap for implementation for implementation of concepts of FAIR concepts)[1]:

- Define metadata for services (Jupyter Notebook as a service and Enlighten-web visualisation service)
- Define metadata for service instances (Done for Jupyter Notebook)
- Use case driven integration of the ICS-D with the ICS-C
- Authentication/Authorization
- Licensing
- Implement SWIRRL API methods offering restoring actions
- Define use cases

The following sections summarizes the status and plans on each of these tasks.

[21] https://cesium.com/

4.3.1 Define metadata for services

The EPOS-AP vocabulary is described22 . An EPOS-DCAT-AP SHACL file that can be used for validating EPOS RDF graphs is also provided[23].
Below is the tentative plan for the remaining work.

- Specify what is needed (regardless of vocabulary) for giving an adequate description (metadata) of the ICS-D services and how to use them. In this process, we will need input from expertise on metadata in the EPOS consortium. E.g., we must assess:
  - What level of information is needed. SWIRRL API/endpoints (e.g., for starting Enlighten or Jupyter instances) could be described in detail, but is this useful?
  - How to describe interactive services like notebooks and visualisations?
  - Do we have two classes, namely tools (including Enlighten, Jupyter notebook, integrated workflows including microservices for converting data) and infrastructures (SWIRRL)?
- When we have decided on the metadata required for describing the ICS-D services, these requirements must be compared with the capabilities in EPOS-DCAT-AP to Identify and report gaps for describing ICS-D services.
- The identified gaps in EPOS-DCAT-AP must then be filled. This is outside the scope of task 10.4 and must be addressed by the metadata expertise in the EPOS consortium.
- When EPOS-DCAT-AP has been extended with required definitions, we can define metadata for the Jupyter Notebook and Enlighten-web services. The metadata must be described in EPOS-DCAT-AP ttl files. The EPOS DCAT-AP Metadata Editor[24] can be used for this purpose.

4.3.2 Define metadata for service instances

As mentioned above, SWIRRL generates and stores provenance data thanks to two dedicated components: the PROV-Template Catalogue and the Neo4j database. Entities with metadata describing Jupyter Notebook Instances are uniquely identified and persisted. Remaining work on this task is thus mainly targeted on provenance data for Enlighten-web.
Below is the tentative plan for remaining work.

- Define SWIRRL templates for the Enlighten-web service. We will re-use and adapt templates used for Jupyter Notebook instances. Moreover, new templates can be designed aiming at achieving specific FAIR objectives for Enlighten-web instances.
- To enable provenance tracking and snapshots for the Enlighten-web service, input data and visualisation specifications must be available on files. Enlighten-web visualisation specifications are stored in json files. So, it remains to provide microservices for storing results of data searches from ICS-C workspace as Enlighten-web input data on files. This will enable SWIRRL to stage immutable data.
- Define metadata for describing Enlighten-web visualisations. We must assess if existing namespaces can be used. Specific metadata for the actual visualisations can be extracted from Enlighten-web native plot specifications in JSON format and translated to relevant vocabularies in EPOS-DCAT-AP.

4.3.3 Use case driven integration of the ICS-D with the ICS-C

There has been some initial discussion on this task. The overall idea is that, having collected together the relevant assets in the workspace, the user through the ICS-C triggers the workflow. However, ICS-C needs to keep track of the workflow execution and deal with both parameterising the workflow before deployment and dealing with the workflow results returned to ICS-C. Basically, the agreed approach is that the ICS-D shall provide the ICS-C with the following information to be used in the metadata:

- URL including search parameters (data source)
- URL to describe the data format returned by the URL
- Link to tools that are compatible with the data format

---

22 https://www.epos-eu.org/epos-dcat-ap
23 https://github.com/epos-eu/EPOS-DCAT-AP/blob/EPOS-DCAT-AP-shapes/epos-dcat-ap_shapes.ttl
24 https://epos-eu.github.io/DCAT-Metadata-Editor/

ICS-C is triggering and controlling the workflow. It is the ICS-C responsibility to arrange the workflow based on the understanding of the ICS-D API provided by this metadata information from ICS-D to ICS-C.

4.3.4 Authentication/Authorisation

There is ongoing work within the EPOS central (ICS-C) development programme on authentication involving gathering requirements from the TCS communities and designing a mechanism to extract from metadata the authorisation parameters and encode them as keys on authentication tokens (see under Authentication in Section Task 10.3 ICS-C) .

SWIRRL does not support Authentication yet. This is a task that will take into account the authentication technology used by the clients adopting it. However, SWIRRL supports delegation of authentication (i.e., confirming user identity) to GitHub. This allows users to store the snapshots generated by SWIRRL to their repository but does not provide full authentication as needed for authorisation in an EPOS context.

4.3.5 Licensing

The intention of EPOS is to make software available as open source with a suitable licence and also to license binaries in a similar way. A Policy Group is working on this currently, alongside management of a liability disclaimer and informed consent to terms and conditions of use, cookies and personal data privacy. Meantime, users can make their analysis performed on SWIRRL which is publicly available on GitHub.

4.3.6 Implement SWIRRL API methods offering restoring actions

SWIRRL allows users and software clients to update Jupyter services with additional libraries. These updates are traced in provenance and can be restored on-demand.

The API also supports the production of Snapshots for Jupyter workspaces that are directly stored on GitHub accounts. Snapshots include software and data (or means to access the data) and can be re-deployed onto third parties infrastructure and tools, such as mybinder.org. We will also implement the possibility to restore full snapshots into SWIRRL itself.
Remaining work includes implementing the interactive components that allow users to control the restoring of updates and the production of snapshots. These are developed as jupyter lab extensions installed automatically in the notebook instances managed by SWIRRL (Figure 6).

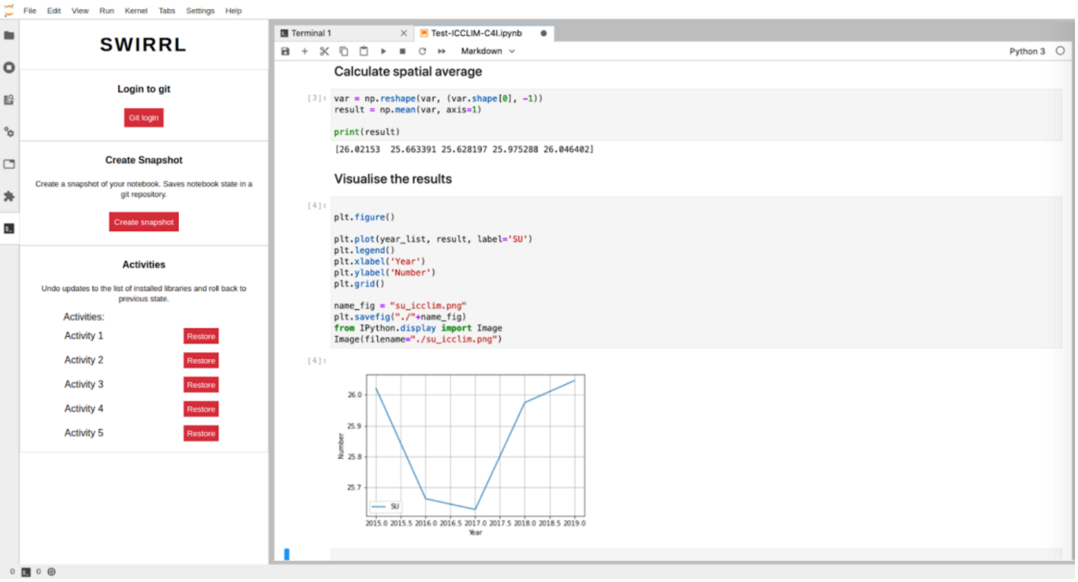

Figure 6: Mock-up of Jupyter lab extensions enabling SWIRRL controls over restore actions, snapshots and visualisation of the provenance data.

4.3.7 Use case for test/demonstration

This task is closely linked to task force 6 (TF6) of WorkPackage 5 (WP5) of ENVRI-FAIR. We will need to define a use case specifically for the ICS-D. Below is an outline of a possible use case:

- Collect datasets into ICS-C workspace.
- Add ICS-D services (e.g., Jupyter Hub with pre-installed ObsPy, Enlighten visualisation, etc.) into ICS-C workspace.
- Build workflow in ICS-C. ICS-D services have predefined options/methods for how they can be used within the workflow.
- Deploy workflow to a dedicated cluster (yet another ICS-D) and run.
    - fetch data for each TCS service based on user defined parameters
    - convert TCS payload to specific format (FEATHER ?)
- Demonstrate how the SWIRRL API works as a generic service for fetching data
- Combine and stage the files onto a dedicated environment, launch Jupyter Notebook through SWIRRL API
- Prepare data for visualisation in notebook
- Launch Enlighten through the SWIRRL API.
- Create snapshot of the environment on Github for later use/sharing.

5 Task 10.5 – Seismology

This chapter reports the activities undertaken in Task 10.5 and focuses on the application of the FAIR principles (Table 1) to seismological data and products. The goal is to develop awareness and share knowledge about FAIR in the seismological domain, and to research effective approaches to help establish the FAIR principles into existing practices and methods. We describe the approach chosen, the roadmap defined, and the activities planned to address the challenges for the implementation of FAIR principles in the EPOS Seismology community.

5.1 EPOS Seismology and ORFEUS-EIDA

EPOS Seismology builds on three large European infrastructures as its fundamental pillars: ORFEUS for seismic waveform data and products, EMSC for earthquake information, and EFEHR for seismic hazard and risk information [5] (Figure 7).

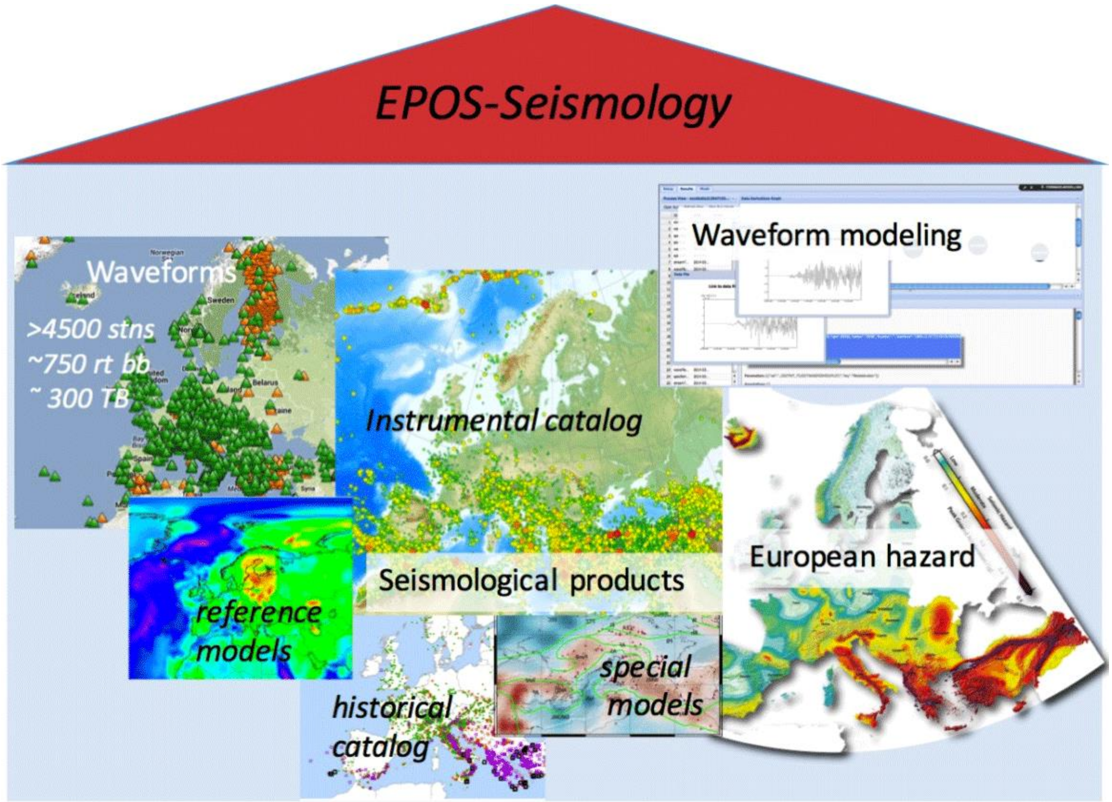

Figure 7: Overview of seismology in EPOS

In this project the seismological community is represented by ORFEUS-EIDA. ORFEUS is the non-profit foundation that coordinates and promotes digital, broadband seismology in the European-Mediterranean area. EIDA is the European Integrated Data Archive infrastructure within ORFEUS to provide access to seismic waveform data holdings.
ORFEUS-EIDA provides seismologists with high quality data and works regularly on improving the provisioning of services and the data [6].

5.1.1 Seismic waveform data

Seismic waveforms play a fundamental role in seismology as they are the primary data underpinning higher level products. For this reason, our FAIR analysis targeted seismic waveform data. Seismic waveforms are continuously collected from sensors, they are acquired and archived in geographically distributed data centres and provisioned to users in a variety of methods (e.g., webservices, portals and APIs). ORFEUS-EIDA coordinates such processes by defining data management policies, developing and contributing to international standards for data, metadata, services and products. In the context of EU projects ORFEUS-EIDA collaborates with e-infrastructure providers to build innovative services for data provisioning, analysis and computing.

5.2 FAIR Assessment

We performed a FAIR assessment targeting a specific seismological dataset category, i.e., seismic waveform data. We started with an analysis of the current status of seismic waveform data, metadata, data services and tools. By adopting the methodology described in D 10.2, we addressed systematically aspects of the FAIR principles and for each one we: a) set feasible goals for the targeted level of FAIRness, b) identified the related gaps and c) defined activities to bridge those gaps.
In the following sections we summarise and discuss the results of our analysis by addressing the different dimensions of FAIR (Table 1).

5.2.1 Findable

Our initial focus was on the findability of seismic waveform data. In particular, we considered the current adoption of Persistent Identifiers (PIDs) associated with seismic waveforms in the context of the ORFEUS-EIDA community. Our goal is to reach the broadest possible adoption of PIDs within ORFEUS-EIDA data centres. Ideally, we would aim at a harmonised and consistent way to associate and manage PIDs across the whole ORFEUS-EIDA. The current landscape includes data centres that have been pioneering PID solutions in projects such as EUDAT[25] and EOSC-hub[26] [7]. Those delivered in some cases operational products and services. In particular, the combination of B2Handle and B2SAFE were successfully adopted for minting PIDs and achieving long-term preservation of large seismic waveform archives. However, the impact of such an integrated solution on the operational infrastructure of some data centres is considered too heavy in terms of resources and capacity required. For this reason, a more flexible and lightweight solution is desirable. For instance, in some cases data centres expressed the need to define and manage their own identifiers within their own environment (implying the need for federation to make them universally unique). A discussion has been initiated and will be continued in order to achieve the highest level of harmonisation possible but at the same time taking into account the specific needs. A metadata catalogue has been identified as a key component to achieve such a goal, namely WFCatalog [8] a specific catalog for waveforms in seismology. WFCatalog is an ORFEUS-EIDA standard and it is operated in all major data centres in Europe. WFCatalog contains detailed and rich descriptions that enable discovery of and support access to seismic waveform data. The WFCatalog data model includes PIDs as metadata features, however, as PIDs are not adopted and implemented by all the ORFEUS-EIDA data centres they are often not populated in the catalogue. WFCatalog offers functionalities to cover most of the findability aspects, activities are required to implement those functionalities in a harmonised way across ORFEUS-EIDA.

5.2.2 Accessible

At present seismic waveforms are made available and delivered to users via community data services that adhere to international standards (e.g., FDSN[27]). ORFEUS-EIDA offers tools, portals and catalogues to enable interactive and/or automated discovery and access[28] . Adjustments and extensions are required in the current access mechanisms in order to increase the FAIRness level e.g., by supporting PID-based queries and by defining and establishing policies for long-term metadata management.

5.2.3 Interoperable

The interoperability aspect requires work on the definition and establishment of a common seismological vocabulary. This activity has been initiated in the EPOS-IP project[29] by engaging representatives of the community and by providing them with a framework for collaboration and knowledge sharing in order to discuss and reach agreements on definitions of concepts [9]. A mechanism to represent such agreed definitions has been provided by adopting SKOS and Linked Data integrated in EPOS-DCAT-AP. However, the establishment of a common vocabulary is a long-term, ongoing activity that requires clear processes and governance to oversee and manage authoritative definitions. In this context our goal is to continue the work in order to achieve a first set of agreed definitions that could be made available in the current seismic metadata catalogue and services. We plan progressively to reach an increased level of FAIRness by extending WFCatalog and its API within the EPOS framework.

5.2.4 Reusable

With respect to reusability our goal is the integration of computation and data supported by rich provenance that would eventually enable reproducibility. When considering the current status of the community this is a quite ambitious but feasible goal. It builds on activities initiated in projects such as

[25] https://www.eudat.eu/
[26] https://www.eosc-hub.eu/
[27] Federation of Digital Seismograph Networks https://www.fdsn.org/
[28] https://www.orfeus-eu.org/data/eida/
[29] https://www.epos-eu.org/about/epos-implementation-phase/epos-implementation-phase-project

EOSC-hub, EPOS-IP, DARE[30] and continued in Task 10.4. In particular, we are investigating mechanisms to containerise for deployment using e.g., docker[31] and properly describe processing and analysis steps. Provenance will be adopted to link computational steps with the original data descriptions and their identifiers. Moving computation to the cloud, supported by mechanisms for authentication and accounting and by efficient data staging services is being evaluated in different contexts by a number of ORFEUS-EIDA data centres. A major challenge is to find a sustainable solution that could be adopted and shared by a broader number of data centres.

5.3 Implementation plan

The challenges identified in the FAIR assessment have been translated into an implementation plan. Below we describe the activities that have been initiated and/or planned.

5.3.1 Definition and implementation of policies for PIDs

Broader adoption of at least one mechanism to mint and manage PIDs. B2Handle seems to be the right candidate for a common solution but customised implementations will be accepted as long as they comply with the agreed policies. A discussion has been initiated to evaluate the requirements in EPOS. Furthermore, a tool and a schema based on RDA's output to assign PIDs to instruments[32] have been evaluated. The CERIF metadata catalog of EPOS supports federated IDs, where each ID may be role-based and related (by semantic and temporal period relationships) to other IDs. This also allows EPOS to manage communities and asset suppliers utilising different PID systems.

To support the management and implementation of policies we will leverage RuleManager[33] – a framework for the automated management of configurable policies that has been developed and successfully piloted in the context of the EOSC-hub project. We will continue the developments for its operationalisation and broader deployment in EPOS Seismology (Figure 8).

30 https://www.epos-eu.org/dare-tame-extremes
31 https://www.docker.com/
32 https://www.rd-alliance.org/groups/persistent-identification-instruments-wg
33 https://www.eosc-hub.eu/keywords/epos-orfeus-cc

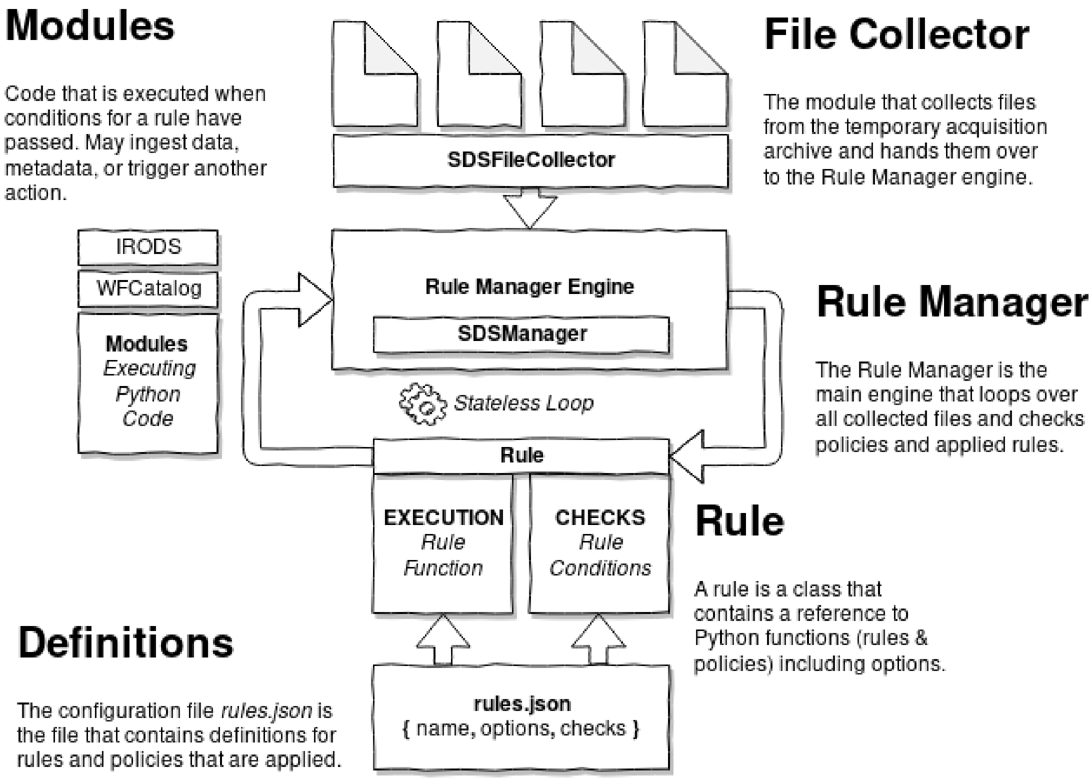

Figure 8: Use of RuleManager

5.3.2 Population of PIDs in WFCatalog

This EPOS metadata catalogue will act as primary source for discovery of and access to rich seismic waveform metadata including PIDs Rule Manager has been tested to support the population process of the PIDs in the EPOS metadata catalog.

5.3.3 Vocabulary

Continuation of the work on a **common seismology vocabulary** and interoperability of concepts and definitions with WFCatalog. This task will include work on the data model and the API. Users will be able to request semantic description in JSON-LD format. The descriptions will be compliant with EPOS-DCAT-AP

Implementation of a use case addressing computation of seismic waveform data on the cloud supported by the SWIRRL API. The preparatory work has been initiated by deploying the SWIRRL API and will be continued in collaboration with T10.4.

6 Task 10.6 - Satellite Earth Observation community

6.1 Introduction

This chapter describes the activities developed in Task 10.6 relevant to the implementation of the FAIR principles for the Earth Observation products of the EPOS TCS Satellite Data. The goal is to develop awareness and share knowledge about FAIRness in the satellite Earth Observation domain to foster the application of the FAIR principles into existing practices and methods. This task has an effective link with the results of previous and ongoing activities carried out in the context of other projects such as

EOSC-hub, EPOS-IP, EPOS-SP and OpenAIRE-Advance[34]. Last but not least, the task has to tackle the integration with the Copernicus DIAS[35] environments.

To produce an efficient framework to address the FAIR aspects in the different phases of the product lifecycle such as generation, curation, computation, dissemination and publication, the task addresses the following implementation activities:

- FAIRness gap analysis and implementation;
- Integration and exploitation of DIAS and EOSC computational services;
- Enhancement of AAAI systems for Earth Observation RIs .
-

6.2 Gap analysis and implementation of FAIR principles in the EPOS TCS Satellite Data

The FAIRness gap analysis was developed by considering the Data, Data Products, Software and Services (DDSS) which had been provided by the EPOS TCS Satellite Data (SATD) community and that were validated and included in the EPOS products portfolio. In particular, we focused on the DDSS referred to as “Line of sight displacement time series”. This DDSS is the most complex and complete among all the services deployed by the TCS SATD and it contains all the main FAIR issues of the TCS. Accordingly, this DDSS represents the touchstone of the TCS SATD FAIRness maturity.

The analysed DDSS is released with two components:

- The data, embedded in a csv (ASCII) file, represented a sparse matrix of points located on the Earth surface. Each point is represented by a coordinate triplet (latitude, longitude and altitude), a predetermined number of parameters coming from the data processing, and a time-series of displacement values. The time-series does not have a fixed size, i.e., it can change dataset by dataset, because the number of samples depends on the number of satellite acquisitions used (processed) to generate the final product.
- The metadata, formatted in a xml file, represented according to the ISO19115[36] standard.

The gap analysis was carried out by critically analysing the several components of the TCS SATD with respect to the requirements of the FAIR principles. Such an analysis identified the following gaps that impact on different FAIR principles (Table 1).

Issue	Description	Impact on
PID	No PID system is used and applied	F1, F3 A1
Data and Metadata Life	No policy has been agreed on the lifecycle of data and metadata. Metadata preservation is not guaranteed.	A2
Vocabulary	A preliminary vocabulary has been drafted but the activity is not complete	I2

Table 2: Gaps Related to FAIR Principles

On the basis of the gap analysis, a list of activities to be carried out has been drawn up and jointly agreed to form a roadmap for their implementation.

[34] https://www.openaire.eu/advance/
[35] https://www.copernicus.eu/en/access-data/dias
[36] https://www.iso.org/standard/53798.html

6.2.1 Application of a PID system

This is a long-term activity since it does not have a trivial solution. The adoption of a PID system has a strong impact on the financial and technical sustainability of the RIs. Indeed, a large number of satellite DDSS are live products, because they are regularly updated each time a new satellite image is available (in same cases several acquisitions per week are available). The preservation of all generated products (instead to have a single product regularly updated) could become unmanageable because each product is quite large (up to several gigabytes) and the disk space would exponentially increase. However, the metadata of previous versions could be preserved to provide appropriate provenance. Moreover, even if several PID systems are available (e.g., DOI and Handle) the services related to PID management have a cost that has to be supported by the RIs. A PID system suitable for the satellite community has yet to be found, also investigating its impact on the financial sustainability.

The activities to adopt a PID system within the satellite EO community started by analysing guidelines, best practices, and solutions developed by Space Agencies and other research groups. In this context, particularly useful has been the work done by Data Stewardship Interest Group (DSIG)[37] (formerly known as the WGISS Archive Task Team) of the CEOS-WGISS working group[38]. This team deals with data archiving as well as data and associated knowledge consolidation and valorisation aspects.

The DSIG recently released a document entitled Persistent Identifiers Best Practice[39] focused on providing recommendations and best practices on the use of Persistent Identifiers to Earth Observation mission data, allowing globally unique, unambiguous, and permanent identification of a digital object. In particular, the report addresses several use case scenarios, tailored for the Earth Observation community, some of which perfectly match with the EPOS TCS SATD context. The implementation activities are going ahead by building some effective pilots for the PIDs in the satellite community.

6.2.2 Data and Metadata Lifecycles

This is a short- to medium-term activity; the metadata preservation has no strong impact on sustainability and is technically manageable. It needs an agreement at RIs level to be implemented in the Data Management Plan.

6.2.3 Vocabulary

This is a dynamic action. The building and updating of the vocabulary is an ongoing activity. Once the vocabulary is consolidated, it will be documented and resolvable using globally unique and persistent identifiers.

6.3 DIAS and EOSC computational services

In addition to the above-mentioned activities, this task has to address the integration with and exploitation of DIAS and EOSC computational services. The integration of satellite RIs with Copernicus DIAS and EOSC computational services represents a challenging activity. In particular, DIAS, the cloud providers specifically selected and developed with the support of ESA and DG-GROW for Copernicus users, seems to be the more suitable environment where satellite processing services can be effectively implemented. Unfortunately, in order to efficiently exploit such an environment, one needs to thoroughly investigate the technical and financial sustainability, i.e., the suitability and robustness of the offered services, as well as their medium- to long-term competitiveness with respect to similar solutions deployed by the commercial sector, have to be accurately evaluated. In this framework, the analysis of the DIAS solutions successfully started, and it is currently ongoing. In particular, CNR integrated its processing chains on the virtual environment provided by Onda[40] (one of the DIAS providers) and carried out a 10-month analysis of its performance and characteristics. The analysis is almost over, and the results are really promising.

---

[37] https://ceos.org/ourwork/workinggroups/wgiss/preservation/
[38] https://ceos.org/ourwork/workinggroups/wgiss/
[39] http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS_DSIG_Data%20Stewardship%20Reference%20Model%20White%20Paper_v1.0.docx
[40] https://www.onda-dias.eu/cms/

6.4 AAAI system of Earth Observation RIs

The enhancement of the AAAI system of Earth Observation RIs has been recently tackled. The TCS SATD has a unique interface with the EPOS ICS-C represented by the Geohazards Exploitation Platform (GEP)[41], a cloud-based platform developed with the support of European Space Agency (ESA)[42]. GEP is an interoperable platform that is queried by EPOS ICS to retrieve data and metadata; the RIs of the TCS SATD are integrated within GEP and their products and services can be retrieved through GEP. The AAAI system of GEP has been successfully integrated with ICS and users registered within the EPOS management system can easily access the TCS resources. It is worth noting that the TCS AAAI is integrated with other systems largely used in the satellite EO community, such as ESA and EOSC AAAI systems (Figure 9).

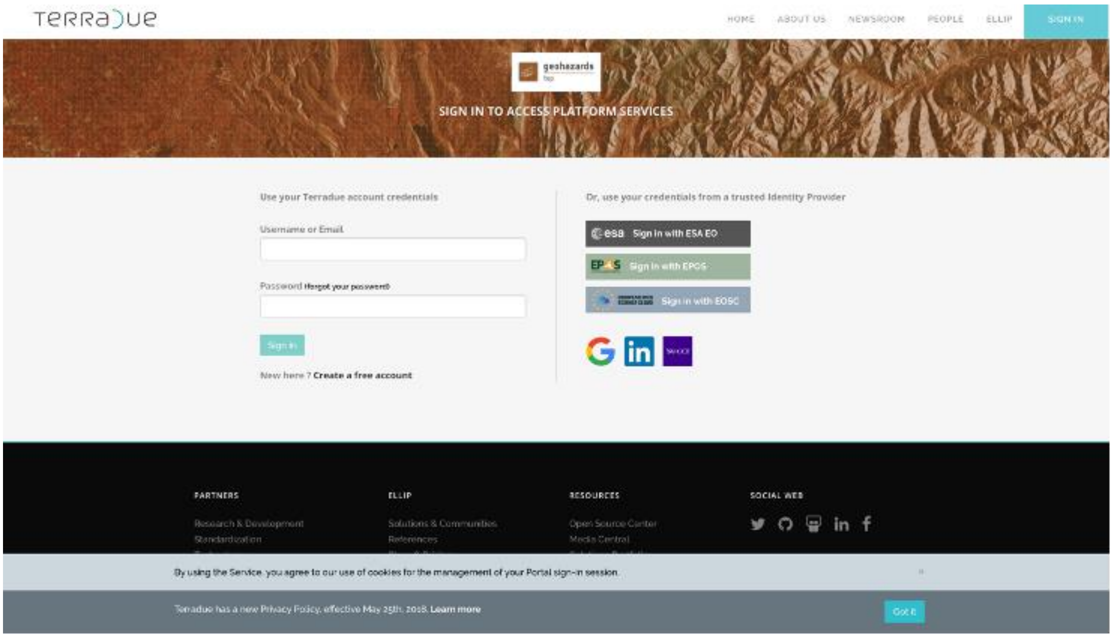

Figure 9: Login Page of Geohazards Exploration Platform

7 Task 10.7 - Marine

This chapter describes the state ad plans for T10.7 covering the solid-earth aspects of the marine domain. The prime objective of task 10.7 is to increase the still limited interoperability of geophysical data/metadata of EMSO ERIC[43] and EPOS. The adoption of common standards for data and metadata will support scientists in the joint use of land and marine data. The task gives a particular focus to the enrichment of metadata of data and sensors of Ocean Bottom Seismometers (velocity-meter and accelerometers) / Hydrophones / Magnetometers (OBS/H/M), portable modules used to set-up networks in marine areas unreachable by land networks or to extend land networks out into marine areas.

Implementation activities include engagements with stakeholders involved in the data pipeline, as well as enriching metadata to improve data documentation and ensure their re-use. A shared workflow (data curation, long-term preservation) that was not ensured across EPOS and EMSO ERIC, is being analysed and agreed for OBS/H/M data to make them findable and accessible over the long term. This task is expected to improve the range of data products and the adoption of FAIR principles over the data pipeline.

[41] https://geohazards-tep.eu/#!
[42] https://www.esa.int/
[43] http://emso.eu/

7.1 State of FAIR implementation

EMSO ERIC regional facilities do have an essential role in delivering seismic data to broad seismological and geophysical communities, to national agencies and other stakeholders. In fact, because of the still uneven geographical distribution of EMSO marine facilities[44], regional scale seismological and geophysical studies of the Mediterranean and North-East Atlantic shall become possible from the regular joint use of EPOS and EMSO data. Seismological data at several EMSO regional facilities have been utilised where standardised seismological data flow to national data centres with links to civil protection and to international seismological agencies exist. This integration helps to improve the reliability of the localization of the seismicity, especially those events occurring in marine costal and open sea areas.

The data workflow includes interactions with EPOS through ORFEUS-EIDA. However, the data workflow implemented by the EMSO across regional facilities is not standardised yet. Further, the level of adoption of FAIR principles varies depending on the regional facility. These interactions with EPOS require an extensive standardisation of the acquisition and validation process from the sensor level to the recorded metadata, including data format, data transmission protocols, data archiving platforms in order to be effective in interoperation. Additional service components are based on dedicated methods and software for integrating data from the different observation systems of the facilities and retrieving the basic standard earthquake parameters.

In order to improve the interoperability between EMSO and EPOS, it is crucial for EMSO to develop a harmonisation abstraction layer (Figure 10). We investigated the process and key features for such a harmonisation based on the FAIR principles. A summarized assessment is provided below.

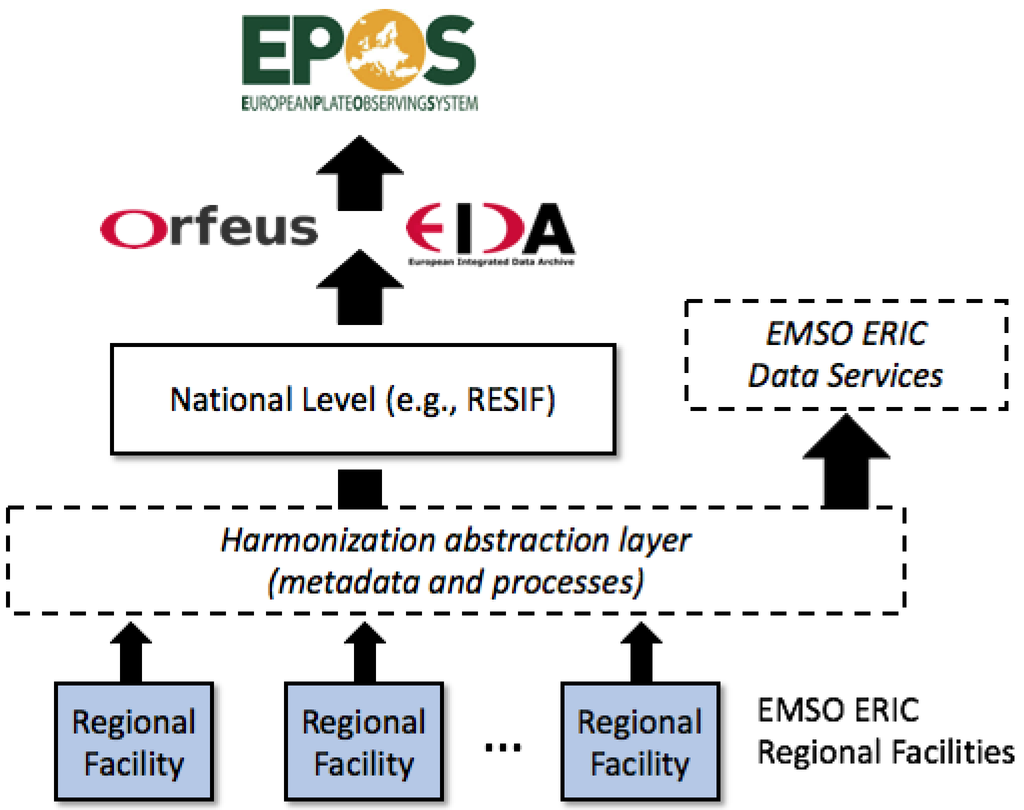

Figure 10: Seismic Data Flows from EMSO to EPOS with Proposed Harmonisation Layer

[44] http://emso.eu/observatories/#map

7.1.1 Findable

While in EMSO some data is already findable through standardised mechanisms (e.g., national agencies and ORFEUS-EIDA), not all regional facilities use the same data discovery mechanism. Our goal is to establish the most effective tools to harmonise and enrich metadata to enable discovery and integration into EPOS while providing visibility of EMSO contributions to EPOS.

7.1.2 Accessible

Current access to EMSO seismological data includes different mechanisms such as those from RESIF (French seismological network)[45] and EIDA to distribute data and metadata according to the standards of the International Federation of Digital Seismograph Networks (FDSN), central control systems or specialized applications. The harmonisation of mechanisms is essential for accessibility. Current efforts are focused on better integration with ORFEUS-EIDA and harmonised access interfaces across EMSO regional facilities.

7.1.3 Interoperable

While EPOS represents the front-end in the primary sub-domain, an essential activity within the EMSO back-end is the standardisation of processes to ensure interoperability between regional facilities and with other key stakeholders such as EPOS. EMSO will investigate the adoption of standardised seismological vocabularies at the harmonisation abstraction layer as they are established.

7.1.4 Reusable

Harmonised and enriched metadata are expected to improve current documentation processes and ensure re-use. A significant challenge is establishing an agreed workflow between EMSO and EPOS (ORFEUS-EIDA), which is necessary to enhance the data curation process.

The implementation of technical activities for the adoption of FAIR principles is preceded by an analysis of available EMSO sources of seismic-related data. In this analysis, we identified:

- regional facilities delivering seismic data;
- types of data produced at each of the regional facilities;
- current processes for providing data and metadata, including ongoing interactions with ORFEUS-EIDA

Additionally, regional facilities currently not interacting with ORFEUS-EIDA have explored the requirements and interfaces for engaging with ORFEUS-EIDA.

Based on the FAIR assessment and gap analysis, our roadmap for technical activities and implementation focuses on enriching metadata and establishing an agreed workflow with ORFEUS-EIDA to enhance the integration and improve the visibility of EMSO contributions through its regional facilities.

8 Conclusion

EPOS started in ENVRI-FAIR with an architecture designed already for FAIR since EPOS had been represented in the Force-11 discussions on principles and subsequently in the RDA FAIR Data Maturity Model Working Group[46] discussing criteria for assessing FAIRness. EPOS has also been in touch with projects such as GO-FAIR and FAIR's FAIR.

Furthermore, EPOS initiated the metadata catalog using CERIF with services as the content in order to intercept the developing strategy for catalogs in EOSC. OF course, the catalog may also contain metadata on datasets, data products, software and other assets.

This deliverable defines the detailed implementation pathways for the various parts of the EPOS – and EMSO – environments to achieve improved FAIRness for end-users. Starting from a base that was

[45] http://seismology.resif.fr/
[46] https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

already FAIR, the improvements are to allow for more sophisticated FAIR implementation, moving towards autonomic mechanisms and not relying solely on human interventions.

Each task is now prioritising the actions required as described in this deliverable in order to produce a roadmap for implementation in the remainder of the ENVRI-FAIR project. During this process dependencies will be identified (within and across tasks) and cost-benefit will be assessed leading to a prioritisation of actions. Achieving ever-improving FAIRness is a continuous effort. How much of the activity identified in this deliverable within the resources and timescale of ENVRI-FAIR means to be seen. The development teams are also evolving: new technologies become available and need evaluation, and skills are enhanced. The solid-earth community is very diverse in: domains of interest; knowledge and skill; current use of standards; degree of development of community cohesion; level of IT support available and utilised; richness of the system(s) and services offered and need/appetite for interoperation.

It is expected that the roadmap of developments emerging from the activities documented in this deliverable will improve all aspects of solid-earth information in FAIRness and readiness for interoperation within the solid-earth domain, across ENVRI and extending to EOSC.

9 Impact on the Project

Having identified gaps in FAIRness in T10.1 (D10.1)[2] and constructed an implementation plan in T10.2 (D10.2)[1] the more detailed plan for implementation is provided in this deliverable.

It provides information for other project members in how EPOS and EMSO (solid earth subdomain) are improving FAIRness and provides a basis for discussions on best practice. Moreover, if all subdomains harmonise to a rich level of FAIRness – such as that already achieved and being improved within the solid-earth domain - then discovery, contextualisation, access, interoperability and re-use across the RIs of ENVRI becomes a possibility.

The ENVRIFAIR WorkPackage (WP5) Task force 1 (TF1) (catalog) has already endorsed the approach of building a rich metadata catalog for ENVRI-Hub using the EPOS technology. The WP5 TF2 (AAAI) is evolving towards harmonisation in parallel with the EPOS approach. Nonetheless, other TFs are proposing an approach to metadata based on triplestores, Resource Description Framework (RDF)[47] and SPARQL for the catalog, or even eschewing a central catalog of metadata. The WP5 team has yet to have a detailed discussion on this issue. The EPOS point of view is that – for the metadata catalog - n-tuples are better than triples for expressivity, formality (referential and functional integrity) and performance. However, (a) in parts of EPOS e.g., the prototype work on ICS-D in section (Task 10.4 - ICS-D) work on triplestores is ongoing – in this case related to provenance; (b) in the VRE4EIC project the EPOS team demonstrated conversion of CERIF to RDF for the purposes of offering interoperability to systems choosing to use triplestores instead of n-tuples.

The list of developments to improve further FAIRness outlined in this deliverable is formidable. FAIRness is a journey, and it is uncertain how much of this development work can be achieved within the resources and timescale of ENVRI-FAIR.

10 Impact on Stakeholders

Stakeholders of the solid earth subdomain – researchers, government agencies, commercial organisations, educators, interested citizens - already enjoy findability, accessibility, interoperability and re-use through a portal (ICS-C) with a rich metadata catalog structured with rich syntax and declared semantics and supporting referential and functional integrity. Thus, the stage is set for reliable interoperation (in the widest sense) among RIs in the subdomain. This is already an advantage to stakeholders in the solid earth subdomain and in the EPOS-SP project additional stakeholders are encouraged to participate in EPOS to improve sustainability. This includes commercial organisations and requires not only FAIRness but also relevant quality services with resilience. The solid earth subdomain offering, integrated into the ENVRI-Hub should provide a much wider group of stakeholders

[47] w3.org/RDF/

across environmental science with appropriate services and asset access. Populating the EOSC asset catalog from the ENVRI-Hub will reach an even wider group of stakeholders providing them with the services and assets to approach a wide range of environmental problems.

11 References

[1] Bailo,D et al. D10.2: Roadmap for Implementation of FAIR Concepts. June 2020. https://envri.eu/wp-content/uploads/2020/08/ENVRI-FAIR_D_10-2.pdf retrieved 2020-12-18

[2] Rabissoni,R et.al. D10.1: TECHNICAL ANALYSIS AND DEFINITION OF IMPLEMENTATION COMPONENTS FOR FAIR IMPLEMENTATION OF RIS IN THE SOLID EARTH SUBDOMAIN June 2020. https://envri.eu/wp-content/uploads/2020/06/ENVRI-FAIR_D_10-1_Technical-analysis-and-definition-of-implementation-components-in-the-solid-Earth-subdomain.pdf retrieved 2020-12-18

[3] Bailo,D et.al. Perspectives on the Implementation of FAIR Principles in Solid Earth Research Infrastructures Front. Earth Sci., 31 January 2020 | https://doi.org/10.3389/feart.2020.00003

[4] Singer, R. (2020). Shape Up. Stop running in circles. Basecamp, 176pp. https://basecamp.com/shapeup/shape-up.pdf . Accessed on 2020-12-xx.

[5] Haslinger, F. and Consortium, E. S.: Staying fair while being FAIR - challenges with FAIR and Open data and services for distributed community services in Seismology, EGU General Assembly 2020, Online, 4–8 May 2020, https://doi.org/10.5194/egusphere-egu2020-18847, 2020

[6] Cauzzi, C., Bieńkowski, J., Custódio, S., Evangelidis, C., Guéguen, P., Haberland, C., Haslinger, F., Lanzano, G., Meier, T., Michelini, A., Ottemöller, L., Pedersen, H., Quinteros, J., Sleeman, R., Strollo, A., and Trani, L.: ORFEUS Services for Coordinated High-Quality Seismic Waveform Data Access in Pan-Europe, EGU General Assembly 2020, Online, 4–8 May 2020, https://doi.org/10.5194/egusphere-egu2020-8389, 2020

[7] Trani, L., Fares, M., Pereira Zanetti, J.P., Quinteros, J., Triantafyllis, N. (2019). Building the EPOS-ORFEUS Competence Center in EOSC-hub. In Geophysical Research Abstracts, volume 21. URL https://meetingorganizer.copernicus.org/EGU2019/EGU2019-16013.pdf

[8] Trani, L., Koymans, M., Atkinson, M., Sleeman, R., Filgueira, R., 2017. WFCatalog : A catalogue for seismological waveform data. Comput. Geosci. 106, 101–108. doi:10.1016/j.cageo.2017.06.008

[9] Trani, L., Atkinson, M., Bailo, D., Paciello, R., Filgueira, R. (2018). Establishing Core Concepts for Information-Powered Collaborations. Futur. Gener. Comput. Syst. 89, 421–437. doi:10.1016/j.future.2018.07.005

12 Annex A – Glossary

AAAI : Authentication, Authorisation, and Accounting Infrastructure

AARC/GEANT : Authentication and Authorisation for Research and Collaboration / Gigabit European Academic Network

API : Application Programming Interface

CERIF: Common European Research Information Format

DCAT: Data Catalog Vocabulary

EPOS-DCAT-AP: DCAT Application Profile for EPOS

DDSS : Data, Data products, Software and Services

DIAS : Data and Information Access Services

EOSC : European Open Science Cloud

EPOS Strategic Plan : Defines the activities of EPOS-ERIC

GEP : Geohazards Exploitation Platform

GUI : Graphical User Interface

ICS-C : Integrated Core Services - Central Hub

ICS-D : Distributed Integrated Core Services Distributed

PID : Persistent Identifier

SDLC: Software Development Life Cycle

TCS : Thematic Core Services