Integrating Datasets and Services in the Solid Earth Domain: the EPOS case.

Rossana Paciello[1], Daniele Bailo[1], Valerio Vinciarelli[1], Riccardo Rabissoni[1], and Keith Jeffery[2]

[1]Istituto Nazionale di Geofisica e Vulcanologia
[2]Keith Jeffery Consultant

November 24, 2022

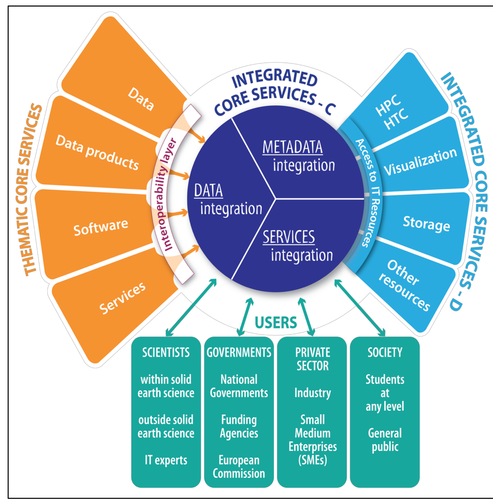
Abstract

A growing number of initiatives in disparate scientific domains arose in the last decade, with the common goal of clustering data and services resources across Europe and make them integrated, open, sharable and available through infrastructures built according to FAIR principles[1]. In this context, the European Plate Observing System (EPOS[2]), now in its Implementation Phase and soon getting to the status of ERIC, is a long-term plan to facilitate integrated use of data, data products, software, services (DDSS) from distributed research infrastructures in Europe in the Solid Earth Domain. Its innovation potential consists in the opportunity of integrating distributed heterogenous resources from European, National and institutional resource providers, and making them available in one single environment. EPOS technical architecture has three main layers: a) National Layer of data providers, that provide access to DDSS; b) community-specific, European-Wide Thematic Core Services (TCS) layer, that collect and integrate DDSS from specific sub-domains and make them available at European level; c) EPOS Integrated Core Services (ICS) system where the integration of DDSS occurs. The architecture relies on three main concepts: Metadata: it is fundamental for describing assets and resources managed by ICS. A twofold approach was used for metadata: at metadata management level, the CERIF model was used for storing all information within the system; at metadata transfer level, an extension of DCAT-AP[3] was created (EPOS-DCAT-AP[4]) to facilitate TCS metadata collection. An architectural approach based on microservices that ensures scalability, flexibility and system interoperability. A harmonization process, that focused on technical aspects like data formats and protocols to access DDSS, but also required intra-domain work on semantic interoperability that includes adoption of common standards and vocabularies. We will discuss these topics and show a demonstration of the ICS prototype. [1] https://www.force11.org/group/fairgroup/fairprinciples [2] https://www.epos-ip.org/ [3] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe [4] https://github.com/epos-eu/EPOS-DCAT-AP/

Integrating Datasets and Services in the Solid Earth Domain: the EPOS case.

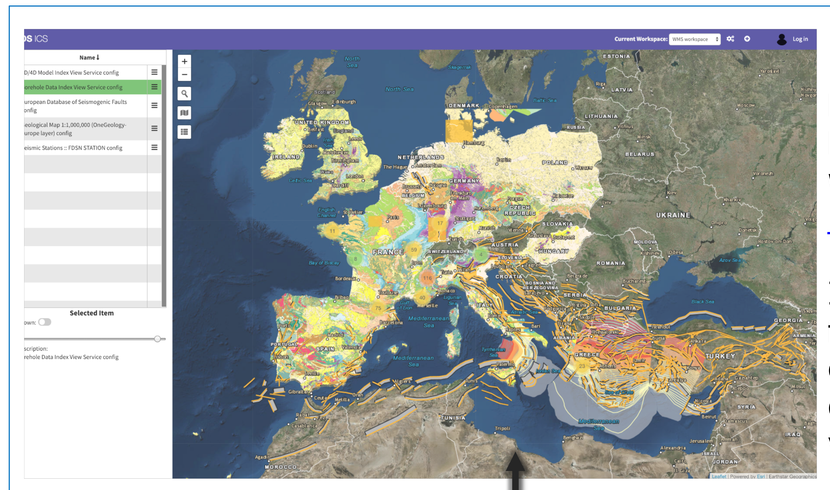
Rossana Paciello[1], Daniele Bailo[1], Valerio Vinciarelli[1], Riccardo Rabissoni[1], and the *EPOS IT TEAM*

(1) INGV – Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy



EPOS-ERIC Research Infrastructure

EPOS enables users to access and integrate data, datasets, software and services (DDSS) provided by Research Infrastructures in Europe in a **FAIR** way, and is composed by:
Thematic Core Services (TCS), international nodes integrating DDSS of same scientific domain provided by National Research Infrastructures; **Central Hub (ICS-C),** main system orchestrating resources and managing the metadata catalogue; **Distributed Services (ICS-D) [7],** i.e. distributed resources from e-Infrastructure providers (HPC, HTC, visualisation, processing etc.); **Users Access** tailored to user type, now demonstrated with a web **Graphical User Interface (GUI).**



EPOS User Interface - Data Access

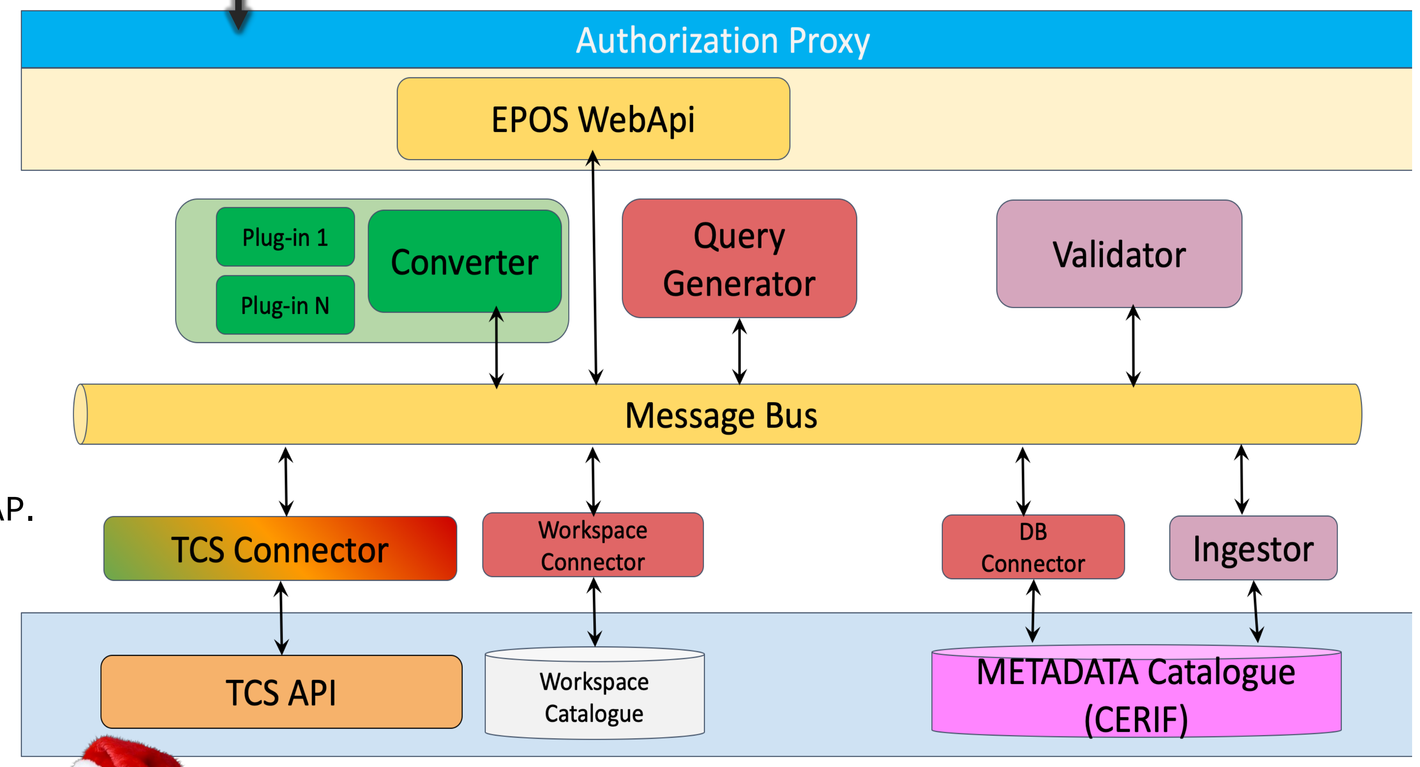
Data, data products and services integrated within EPOS system can be accessed by means of a GUI. A web demonstration is available at https://epos-ics-c-beta.brgm.fr/epos/epos-gui/master/

It includes search, filtering, download, pre-visualisation functionalities (map view) and a user workspace where configured datasets and services can be visualised or downloaded and in the future pushed to processing or visualisation facilities (ICS-D)

Integrated Core Services Central Hub | ICS- C | Microservices Approach

ICS-C adopted an architectural approach based on microservices that ensures scalability, flexibility and system interoperability. Such architecture includes components and technolgies that *enable FAIR principles to become reality.*

- **EPOS WebAPI:** it represents the entry point to access the EPOS system and provides RESTful services.
- **Message Queue/Bus:** provides an asynchronous communication protocol, enables the different components to communicate by sending and receiving messages.
- **Workspace:** this module manages results found by users and stored in their personal workspace.
- **Query Generator:** vreates SQL query from request Web API.
- **Converter:** converts an input format to a different output format, e.g. converts from CERIF model to EPOS-DCAT-AP.
- **Ingestor and Validator:** these components are used to ingest metadata into the catalogue and validate it.
- **TCS Connector:** Represents the interface to access TCS API.
- **Metadata Catalogue:** based on the CERIF [8] standard, it stores information about Data, Users, Resources, Processing models.



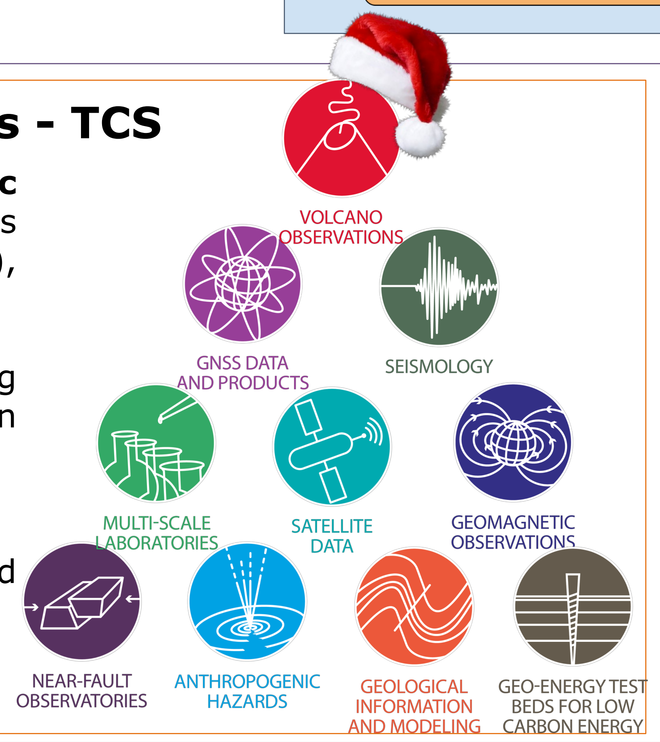
Resources Harmonization - Thematic Core Services - TCS

Each community provides access to its resources through European-wide services called **Thematic Core Services (TCS).** For each of them a governance framework and a data provision platform is established, as in the case of EIDA/ORFEUS[1] (*Seismology*), ESA GEP[2] platform (*Satellite data*), INTERMAGNET[3] (*geomagnetic observations*) European Geological Surveys[4] nodes and others[5].

TCS are characterized by enormous **heterogeneity of data types and formats**, way of accessing data and metadata, scientific methods. In order to integrate their resources in one single portal, an **harmonization process** has been carried out along three dimensions:

1. *Governance*, in order to organise communities and avoid overlappings
2. *Data and Metadata formats*, in order to decrease heterogeneity if not necessary
3. *Resources representation in a common agreed metadata standard*: EPOS-DCAT-AP [6] was selected to describe TCS assets and ingest them into the EPOS metadata catalogue (CERIF[7] based).

All TCSs provide web services / APIs which enable ICS-C to access data and metadata.



Rich Metadata Catalouge

ICS-C include a canonical metadata catalog: it is a rich superset representation of the various subset metadata standards used by the TCS allowing representation of the TCS assets in a consistent form and in a machine-readable and machine actionable way.

A twofold approach was used for metadata:

1. at metadata management level, the CERIF model was used for storing all information within the system;

2. at metadata transfer level, an extension of DCATAP was created (EPOSDCATAP [6]) to facilitate TCS metadata collection from TCS services to CERIF catalogue [8].

[1] https://www.orfeus-eu.org/
[2] https://geohazards-tep.eo.esa.int/
[3] http://www.intermagnet.org/index-eng.php
[4] http://www.eurogeosurveys.org/, http://www.onegeology-europe.org/, http://www.bgs.ac.uk/
[5] Full list here: https://www.epos-ip.org/thematic-core-service-index
[6] https://github.com/epos-eu/EPOS-DCAT-AP, - *IN31B-33 EPOS-DCAT-AP: a DCAT Application Profile for solid-Earth sciences, Wednesday, 12 December 2018 08:43 - 08:46*
[7] IN31A-05 Integrated Computing in solid Earth sciences: the case of EPOS Integrated Core Services Distributed Infrastructures, Wednesday, 12 December 2018 09:00 - 09:15
[8] CERIF: Common European Research Information Format. Now maintained by http://www.eurocris.org/

EUROPEAN PLATE OBSERVING SYSTEM www.epos-ip.org | info@epos-ip.org | epos@ingv.it

This project has received funding from the European Union's Horizon 2020 research and innovation programm under grant agreement N° 676564 - Disclaimer: the content of this poster reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.