DELIVERABLE

D5.4 Feasibility assessment for the establishment of a DSS access service as part of the EPOS framework covering technical integration, organization, governance and financial aspects

Work package	WP5 (NA3: Networking deep seismic sounding data and products)
Lead	UU
Authors	Kuvvet ATAKAN (UiB) Helga K.S. Indrøy (UiB) Nil Eryilmaz (UiB) Ramon Carbonell (CSIC) Irene de Felipe (CSIC) Monika Ivandik (UU) Roland Roberts (UU).
Reviewers	-
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	30/04/2019
Submission date	02/05/2019
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D5.4_Assessement_DSS_EPOS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

This deliverable is written within the framework of the project “Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe – SERA” (Project no: 730900), funded by the Horison2020, INFRAIA-01-2016-2017 Programme.

1 Introduction

Scientific research is producing a vast amount of scientific contributions published in different research journals. A significant amount of contributions of this vast number of papers are based on experimental data. This is the case of experimental seismology which is mostly responsible of generating Deep Seismic Sounding (DSS) data.

In order to consider the technical feasibility of data access and data products, all the available options have to be considered, including the infrastructure initiatives currently being developed. Thus, access to these products is provided under the requirements of the EU mandates of open access. The benefits of open access data are numerous (Fig. 1) and its advantages are deeply explored by the European initiative of the European Open Science Cloud (EOSC).

Fig. 1 Open access science policy. Increasing or busting up the benefits of national and European investment into research.

In particular, a large amount of data has been produced by the earthquake engineering and seismology research infrastructures in Europe over the last years, promoting new information technology tools for data sharing within each research community. Such web-based tools are an emerging trend to manage and share results with the scientific community, decision-makers and the general public. The key objective of this networking activity is to make this data available, so that it can be processed with the latest processing and number crunching tools and interpreted by benefiting from new approaches.

Behind the open data policy, four important concepts have to be regarded, it is not only disclosure it must be FAIR (Fig.2). The data is well described, licensed and freely available to download. Data is assured to be in an open format, so that it can be easily read and used and finally it is well managed, mining traceable and with the acknowledgements properly provided. It is mainly the concept or philosophy of making the data FAIR.

Fig 2. FAIR Data. Open data is mode that disclosure it means Findable, Accessible, Interoperable and Reusable.

Accessing data and data products generated by the Deep Seismic Sounding community can be approached in different ways, taking into account the different tools available, initiatives currently active and new developments in the horizon. The main aim is to evaluate the advantages of the available tools and go into a process of integration to obtain the overall benefits of those tools and approaches.

Accessing these experimental data, which has nurtured scientific publications, can be considered in the same way as a publication. Libraries and specifically, digital libraries have developed an important number of efficient archiving management and search tools. These are lately being busted by the EU open science mandates. Such tools not only are able to search for specific authors (in the current case would be data) but they also perform analysis on the access and origin of the search. These provide measures and statistics of the relevance of the manuscript or data. Furthermore, using access and management schemes designed within a library framework provides a persistent identifier to the DSS data and data products, a persistent identifier such as DOI and/or handle which links the data to a unique address where to find the data. Implementing data (or data products) access as a journal paper the approach would benefit from a number of utilities such as:

- SHARE: an open-source community developing tools and services to connect related research outputs for new kinds of scholarly discovery.
- BASE: the Bielefeld Academic Search Engine.

We have to consider also that global sharing of seismological data has been a long-lasting tradition tracing back to the beginning of the previous century. In Europe, this tradition is materialized in the form of two European organizations, namely ORFEUS (**O**bservatories and **R**esearch **F**acilities for **Eu**ropean **S**eismology) for seismological waveform data and EMSC (**E**uropean **M**editerranean **S**eismological **C**entre) for earthquake source parameters, that both were sustained by the community for several decades. In addition to this, recently, the ESFRI (**E**uropean **S**trategic **F**orum for **R**esearch **I**nfrastructures) initiative and the **EPOS** project (**E**uropean **P**late **O**bserving **S**ystem), provided a larger framework for the integration of all solid Earth science data into a single Pan-European e-infrastructure following the FAIR principles.

Up to now, these two adjacent scientific disciplines, earthquake engineering and seismology, have not interfaced their data structures, lacking an interoperable data-sharing structure. However, there is need to increase the interaction of the earthquake engineering and seismology communities by integrating the most important databanks and related informatics services in Europe. An efficient use of resources, know-how data sharing and the added value of bringing together the data sources and data exchange services of the two communities, will be a step forward in the benefit of a wide range of users.

In this deliverable we present a review of the current EPOS system: the main features are discussed, focusing on the architecture, content, databases, functionality and access portals. This will serve as the basis for the future roadmap for integration of DSS data and products to EPOS Delivery Framework and informatics services that will support the needs of the two research communities and beyond.

2 EPOS project and database review

This chapter reviews the current status of the implementation of the "European Plate Observing System – EPOS" (EPOS-IP – Project no. 676564, InfraDev programme - Horizon2020) and the underlying national and European level thematic service structure. EPOS architecture is explained and the main components of EPOS, the "Integrated Core Service – ICS" and the "Thematic Core Services – TCS" are elaborated. Special emphasis is given to the TCS "Seismology" and "Multiscale Laboratories". Metadata Reference Model based on the European standard CERIF (Common European Research Information Format), is described and current technical implementation procedures are explained.

EPOS is a single, pan-European Research Infrastructure plan for sharing solid Earth Science data, observations and research results. Its mission is to integrate the present and future of the European Research Infrastructures for solid Earth Science into a single, distributed and sustainable infrastructure taking full advantage of the new e-science opportunities and warranting increased accessibility and usability of multidisciplinary data. Through its IT platform, EPOS will enable sharing and integration of science data collected by the research infrastructures and facilitate common access to services from a single online environment.

The following sections present the basic features of the EPOS project and the related database. This text is written based on original EPOS documentation (e.g. Atakan and Michalek, 2017) and is similar to the descriptions given in SERA deliverable D6.4 (Atakan et al., 2018), that are updated with current developments.

2.1 EPOS: Project description and timeline

EPOS is a large-scale research infrastructure for European solid Earth science, integrating existing research infrastructures to enable innovative multidisciplinary research. It is included in the Roadmap of the European Strategy Forum on Research Infrastructures (ESFRI) in December 2008 and was prioritized for its implementation which started in 2015 and will be completed in October 2019.

The goal of EPOS is to offer tools and data to promote and facilitate innovative approaches for a better understanding of the physical processes controlling earthquakes, volcanic eruptions, unrest episodes, and tsunamis as well as those driving tectonics and Earth surface dynamics. This overarching goal will be achieved throughout the integration of existing and newly developed national and international Research Infrastructures (RIs) that provide multidisciplinary data recorded by monitoring networks, acquired in laboratory experiments, and produced by computational simulations. The establishment of EPOS will, therefore, foster worldwide interoperability in Earth sciences and provide services to a broad community of researchers. This will promote major advances in the understanding of the dynamic processes occurring in the Earth, particularly relevant in the context of geo-resources and geo-hazards.

By improving access to data and data products, together with tools for their use in analysis and modelling, EPOS will transform the European research landscape, driving discovery and developing solutions to the geo-hazards and geo-resources challenges facing European society.

The innovation potential of the EPOS infrastructure involves facilitating the integration and use of solid Earth science data, data products, services and facilities, based on distributed national research infrastructures across Europe.

The ground-breaking nature of the EPOS federated approach relies on joining the capacity of delivering high-quality standardized and multi-disciplinary data, the involvement of ICT experts in guaranteeing novel e-science opportunities and the leverage effect of user’s engagement.

EPOS is aimed to reach a broad community including the following stakeholder categories:

- Data and service providers from the solid Earth science community
- Scientific and educational community (hereinafter, the users)
- Governmental organizations
- Industry and other data and service providers

EPOS will continue to fully integrate data and service providers from solid Earth science as well as governmental organizations. EPOS will be also essential to further engage the scientific user community and other service providers and users not belonging to solid Earth science. The construction and operation of the EPOS infrastructure will allow the interaction with the general public through the novel e-infrastructure.

2.1.1 Mission and aim of the EPOS project

EPOS integrates many heterogeneous Research Infrastructures (RIs) using a novel approach based on the harmonization of existing services and components interfaces. EPOS can be described as a tool to make integrated use of data, data products, software and services (including laboratories) provided by different RIs operating in the solid Earth Sciences domain. EPOS aims to provide a simple “one-stop shop” tool by providing an integrated environment where the user can browse, preview and/or select data and then, simply download them or perform processing and modelling directly online.

The official EPOS community (Fig. 3) during the project implementation phase (2014-2019) is made up of 47 partners plus 6 associate partners (including the European Space Agency, EuroGeoSurveys, Global Earthquake Model) from 25 countries from all over Europe and several international organizations (ORFEUS, EMSC, EUREF, INTERMAGNET). However, the community contributing to the EPOS integration plan is larger than the official partnership of EPOS implementation phase project, because more countries are represented by the international organizations and within each country there are several research institutions involved.

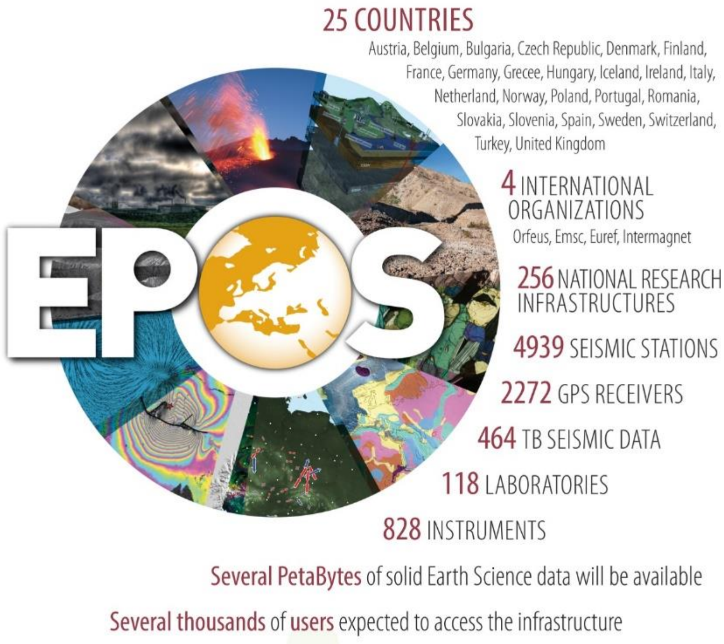

Fig. 3: The EPOS implementation phase community.

2.1.2 Basic features of the EPOS project

EPOS provides services to scientists, national and international authorities and to the general public with the vision of creating a pan-European e-Research Infrastructure for solid Earth science to support safe and sustainable society. In accordance with this scientific vision, EPOS will enable interdisciplinary and transnational use of data to broaden the horizons of research technological progress while fostering widespread awareness of issues related to sustainable and responsible use of resources. A basic feature of the EPOS system is "*the possibility of carrying on data intensive science orchestrating the distributed resources made available by EPOS data providers and stakeholders*" (Bailo and Jeffery, 2014). A key aspect of EPOS is to provide end-users with a homogeneous access to services and multidisciplinary data collected by monitoring infrastructures and experimental facilities as well as access to software, processing and visualisation tools.

EPOS provides for ten Thematic Core Services (TCS), which are responsible for integrating all data, metadata and services arriving from various national and international infrastructures. It also provides Integrated Core Services (ICS), which offer a new interface for users by adopting data access policies aligned to open science principles. The ICS makes data, services and products accessible to users in a useable form that allows innovative, disciplinary and cross-disciplinary research.

2.1.3 General timeline of the EPOS project

The general timeline of the EPOS project consists of the following four phases (Fig. 4).

- The **Conception Phase** (2002-2008),
- The **Preparatory Phase, EPOS PP** (2010-2014),
- The **Implementation Phase, EPOS IP** (2014-2019) and
- The **Operation Phase, EPOS OP** (2020 and after).

The conception phase evolved thanks to a number of European and national-funded projects and initiatives which focused on integrating and distributing digital data. The EPOS Preparatory Phase (EPOS PP) ran from 2010-2014 and it was funded by the European Commission's FP7 Work Programme. The project included 20 partners from 18 countries and one international organization (ORFEUS). Five countries were associate partners and there was one international organization (EMSC). The EPOS Preparatory Phase was key for the implementation of the pan-European research infrastructure for solid Earth science that is EPOS mission. The EPOS Preparatory Phase project had the ambitious goal of creating the conditions for the integration of existing and future national and international research infrastructures (RIs) in Europe with the final goal of improving access to data, products and services.

The EPOS Implementation Phase (EPOS IP), funded by the Horizon2020 Program of the European Commission, started in 2014 following the completion of the preparatory phase. Running until 2019, the EPOS IP project is a key step in EPOS' vision of a pan-European Earth science monitoring platform. Built on the EPOS PP project, EPOS IP will deliver a suite of domain-specific and multidisciplinary services in one platform and the legal, governance and financial frameworks to ensure the future operation and sustainability of the platform.

In order to make the services of the EPOS platform operational in 2019, after the end of the Implementation Phase, actions and support running in parallel at both national and European level are necessary. Such actions include the maintenance and operational support to Thematic Core Services from National Research Infrastructures, support from national initiatives and in-kind contributions to the implementation of TCS, funding of national and European projects in order to further develop TCS components and third-party contributions from the private and government sector. EPOS IP project supports these actions through the EPOS European Research Infrastructure Consortium (EPOS-ERIC, see Section 0) and the implementation of a technical, legal and governance, and financial framework for all its components. The IT architecture and interoperable services will also be created during the IP. The service to the user community is foreseen to start in the EPOS OP.

Fig. 4: How the EPOS project is evolving.

2.1.4 EPOS Technical, Legal, Governance and Financial structure

EPOS has established a European Research Infrastructure Consortium (EPOS-ERIC) and thus a legal entity in October 2018, with 12 countries signed as founding members.

EPOS-ERIC General Assembly (GA) is the organization's governing body. An Executive Director, supported by its Coordination Office (ECO), is directly responsible to the GA for all aspects of the EPOS activities. A Services Coordination Board representing all the TCS and the ICS will inform and advice the Executive Director in formulating and executing the EPOS Annual Work Programme. Fig5 presents a schematic illustration of the EPOS-ERIC governance structure.

The EPOS-ERIC legal seat is hosted in Rome (Italy) at the National Institute of Geophysics and Volcanology (INGV) headquarter. The ICS Central Hub (ICS-C) will be hosted in the United Kingdom (British Geological Survey - BGS) and France (Bureau de Recherches Géologiques et Minières - BRGM) with technical operational support from Denmark (Geological Survey of Denmark and Greenland – GEUS). The location of service providers in charge of data and service provision for the Thematic Core Services, are identified and will be signing contracts with EPOS-ERIC in the operational phase.

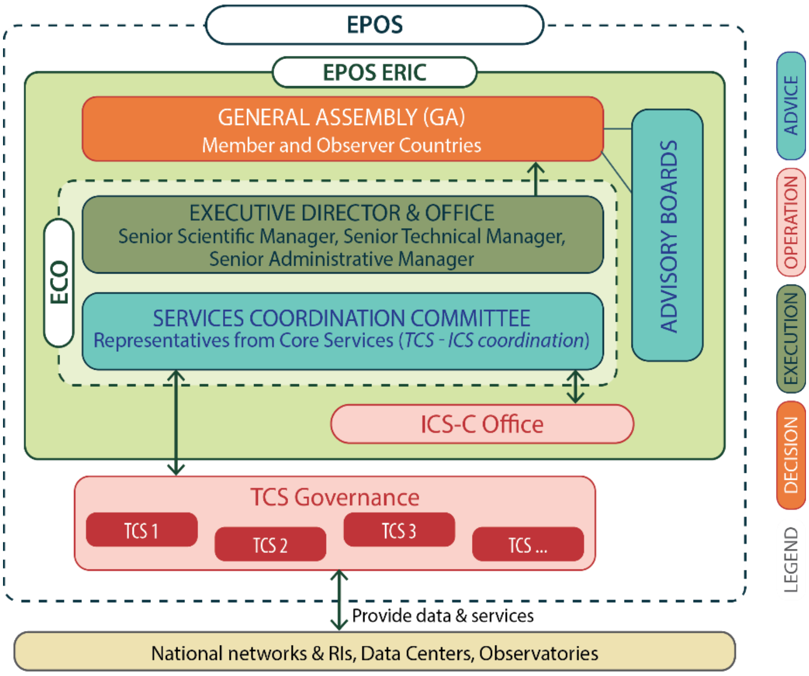

Fig. 5: EPOS-ERIC Governance Structure

2.2 EPOS Functional Architecture

EPOS functional architecture is based on a three-layer structure where the bottom layer, consisting of the national Research Infrastructures (RIs), represents the backbone of the EPOS (Fig. 6). The second layer represents the community developments at the European level, where Thematic Core Services (TCS) are organised. Currently 10 different TCSs are operating in EPOS, where some of them were established more than a century ago, whereas others were recently formed:

- TCS Seismology – WP08
- TCS Near Fault Observatories (NFO) – WP09
- TCS GNSS data and products – WP10
- TCS Volcano observations – WP11
- TCS Satellite data and products – WP12
- TCS Geomagnetic observations – WP13
- TCS Anthropogenic Hazards – WP14
- TCS Geological data and modelling – WP15
- TCS Multiscale Laboratories – WP16
- TCS Geo-Energy Test-Beds – WP17

The third layer is the pan-European level where Integrated Core Services (ICS) are offered as part of the EPOS architecture. ICS Central-hub (ICS-C) will be hosted by the Geological Surveys in UK (BGS), France (BRGM) and Denmark (GEUS) jointly. The distributed resources (ICS-D) are designed such that external resources, such as High-Performance Computing (HPC) and High-Throughput Computing (HTC), visualisation processing and analysis represents services that are decentralised and offered by third parties such as European Open Science Cloud (EOSC). Both the ICS-C and EPOS-ERIC hosting was decided by the Board Governmental Representatives (BGR), where each member country is represented by its relevant ministry, following a competitive tender process. Currently EPOS is in the final stage of the Implementation Phase (EPOS-IP) and EPOS-ERIC is now operational with a General Assembly replacing the BGR.

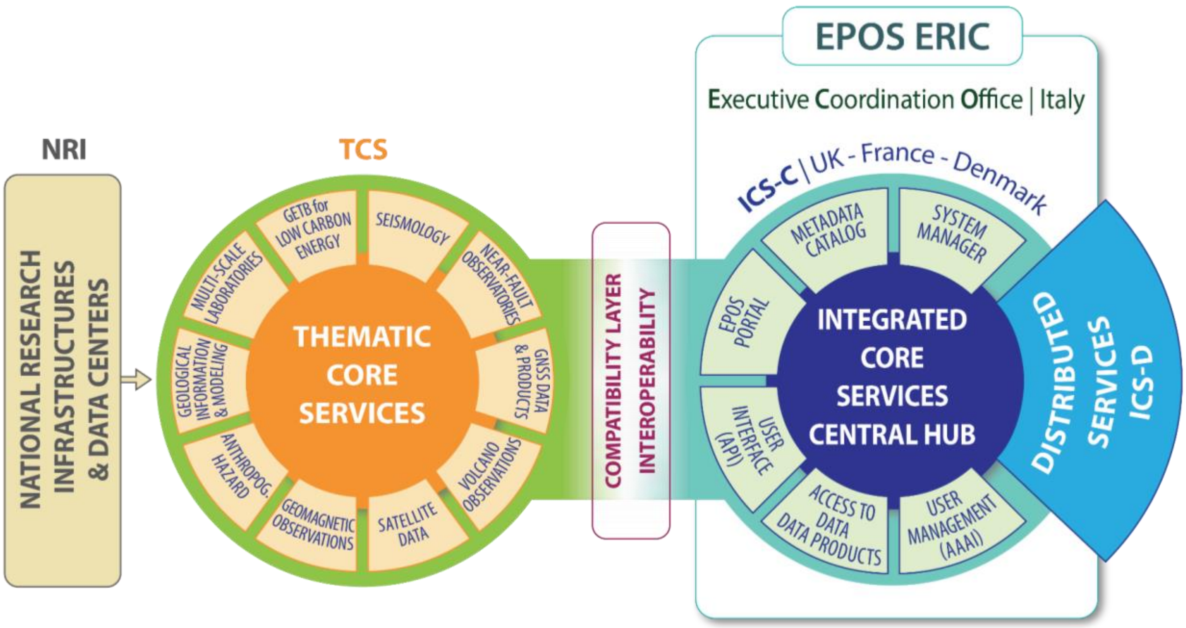

Fig. 6: The EPOS functional architecture

The National Research Infrastructures (NRIs) are research infrastructures and data centres that provide data to the Thematic Core Services. Thus, the NRIs represent the EPOS data providers that will guarantee access to quality-checked data and products and they provide services at national level.

The Thematic Core Services (TCS) are the pan-European e-infrastructures that provide and disseminate data and services to specific communities and international organisations (e.g. ORFEUS for seismology). The TCS are community-specific integration and they represent transnational governance framework with the scope of integrating the data, metadata and services arriving from various national and international infrastructures and data centres.

The Integrated Core Services (ICS) provide a new interface for users by adopting data access policies aligned to open science principles. The ICS make data, services and products accessible to users in a useable form that allow innovative, disciplinary and cross-disciplinary research. This e-infrastructure will allow access to multidisciplinary data, products (including synthetic data from simulations, processing and visualization tools), and services to different stakeholders, including but not limited to the scientific community (i.e. the main EPOS users). The key element of the ICS in EPOS will be a Central Hub (ISC-C) where users can discover and access data and data products available in the TCS and NRIs, as well as access a set of services for integrating and analysing multidisciplinary data.

2.3 EPOS Service Provisions

There are 10 work packages (WP08-WP17) forming part of each Thematic Core Services (TCS) (see 2.2). Thus, data in EPOS will be available from the disciplines that each community deals with.

In order to facilitate the integration among the communities, data have been categorized in the following levels:

- Level 0: raw data, or basic data (example: seismograms, accelerograms, time series);
- Level 1: data products coming from nearly automated procedures (earthquake locations, magnitudes, focal mechanism, shake maps);
- Level 2: data products resulting by scientists' investigations (crustal models, strain maps, earthquake source models, etc.);
- Level 3: integrated data products coming from complex analyses or community shared products (hazards maps, catalogue of active faults, etc.).

In order to increase the visibility and to enhance the dissemination of all this kind of data and to have them accessible and findable, a permanent identifier is needed together with a standardized description of the metadata. It will allow an easier interpretation and re-use of the data for the scientific community. Thus, following the data management requirements established in the framework of the EU Programme for Research and Innovation Horizon2020, datasets have to be preserved and unequivocal identified by a Digital Object Identifier (DOI). Thus, these permanent identifiers will multiply the presence of the research through web search engines, thematic websites, catalogues, open access aggregators and emerging services (Fig. 7).

Fig. 7: full range of possibilities to disseminate scientific data to academy, industry and education users (Bernal, 2019).

As a part of the Requirements and Use Cases (RUC) collection from the WPs, a specific list was prepared to include all Data, Data product, Software and Services (DDSS). This DDSS Master Table is used as a mechanism to update the RUC information as well as providing a mechanism for accessing more detailed IT technical information for the development of the Integrated Core Services (ICS) Central Hub (ICS-C). The DDSS Master Table is also used for extracting the level of maturity of the various DDSS elements in each TCS as well as providing a summary of the status of the TCS preparations for the ICS integration and interoperability.

Currently, there are 137 distinct DDSS elements implemented through 264 web-services distributed over the ten TCS. In total, there are 368 DDSS elements in the DDSS Master Table out of which 231 are declared at TCS level and planned to be included in EPOS ICS. These DDSS elements are of different maturity and are declared by TCS groups to be ready for implementation which means that the data are well described with metadata, following the standards specific for their domain and, in the best case, with some services allowing their access already. The DDSS elements differ by their complexity as well. The DDSS Master Table serves as an overview of the DDSS elements and includes most of the important information needed for further implementation and is continuously updated as the project evolves. Fig8 shows the number of DDSS elements in EPOS TCS WPs.

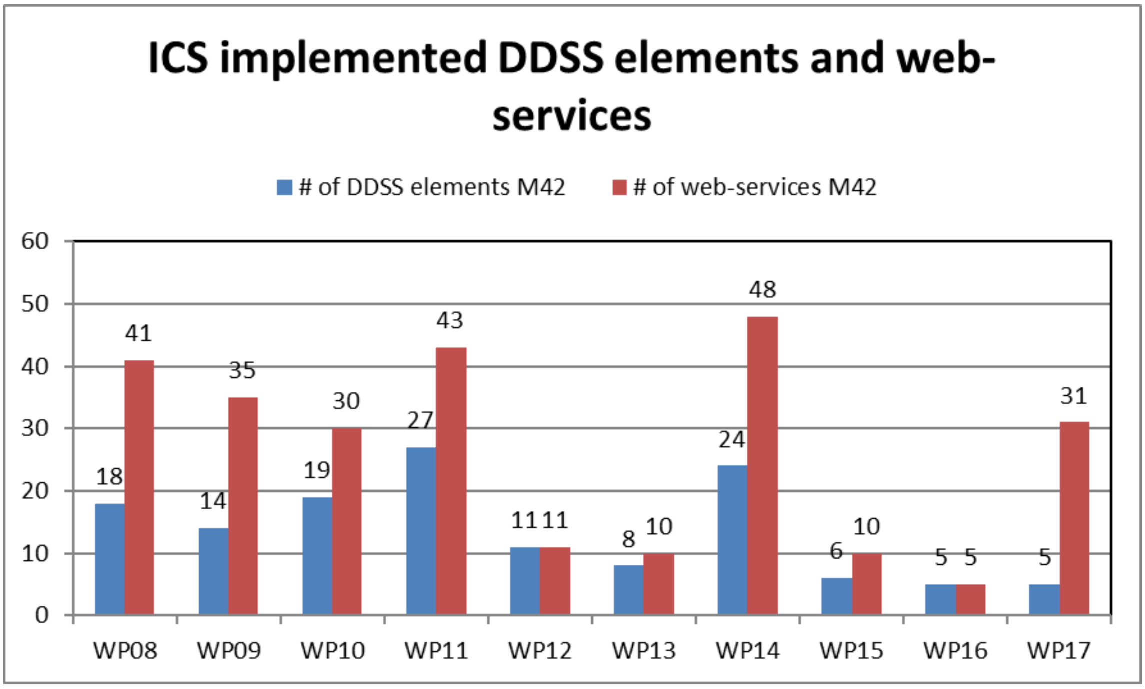

Fig. 8: Number of DDSS elements and corresponding web-services implemented in EPOS ICS (as of April 3, 2019). WP8: Seismology; WP9: Near Fault Observatories; WP10: GNSS Data and Products; WP11: Volcano Observations; WP12: Satellite data; WP13: Geomagnetic Observations; WP14: Anthropogenic Hazards; WP15: Geological Information and Modelling; WP16: Multi-scale Laboratories; WP17: Geo-Energy Test Beds for Low Carbon Energy.

2.4 EPOS Integrated Core Services (ICS) User Interface

The EPOS demonstrator has been developed during the project implementation phase. The Graphical User Interface (GUI) is currently being validated against use cases collected in the beginning of the EPOS-IP (during the Requirement and Use Cases Collection; Jan-March 2016), where each TCS group provided several user stories. The user stories were analysed, so GUI is thus connected and integrated to the ICS Architecture scheme. However, the current GUI is a still a preliminary version. The EPOS ICS GUI has the following features:

- Discovery (searching for data, persons, services, etc.);
- Workspace (saving items for later use, e.g. during the discovery);
- Visualise (plot and analyse data);
- Process (combine items from workspace into a workflow for customized processing/analysis).

The user workspace enables a user to search and browse the data, data products, software and services provided by the communities, and to perform visualisation and processing on the selected resources. It provides end-users with their view of what EPOS has to offer and it allows TCS to expose their assets for use or re-use by users. The EPOS data portal GUI is presented in Fig9.

A user can be a guest to test the functionality of the demonstrator or can log-in in order to store previous workspaces. Through the GUI, the user can search for data, persons and services by filling the form that is presented in Fig. In the lower tab, the user can see the search results that can be saved and/or analysed. The user also has an individual workspace where items can be saved for later use and the visualization features can be used to plot and analyse data. The user can also keep track of the previous workspaces where the results can be stored for future use.

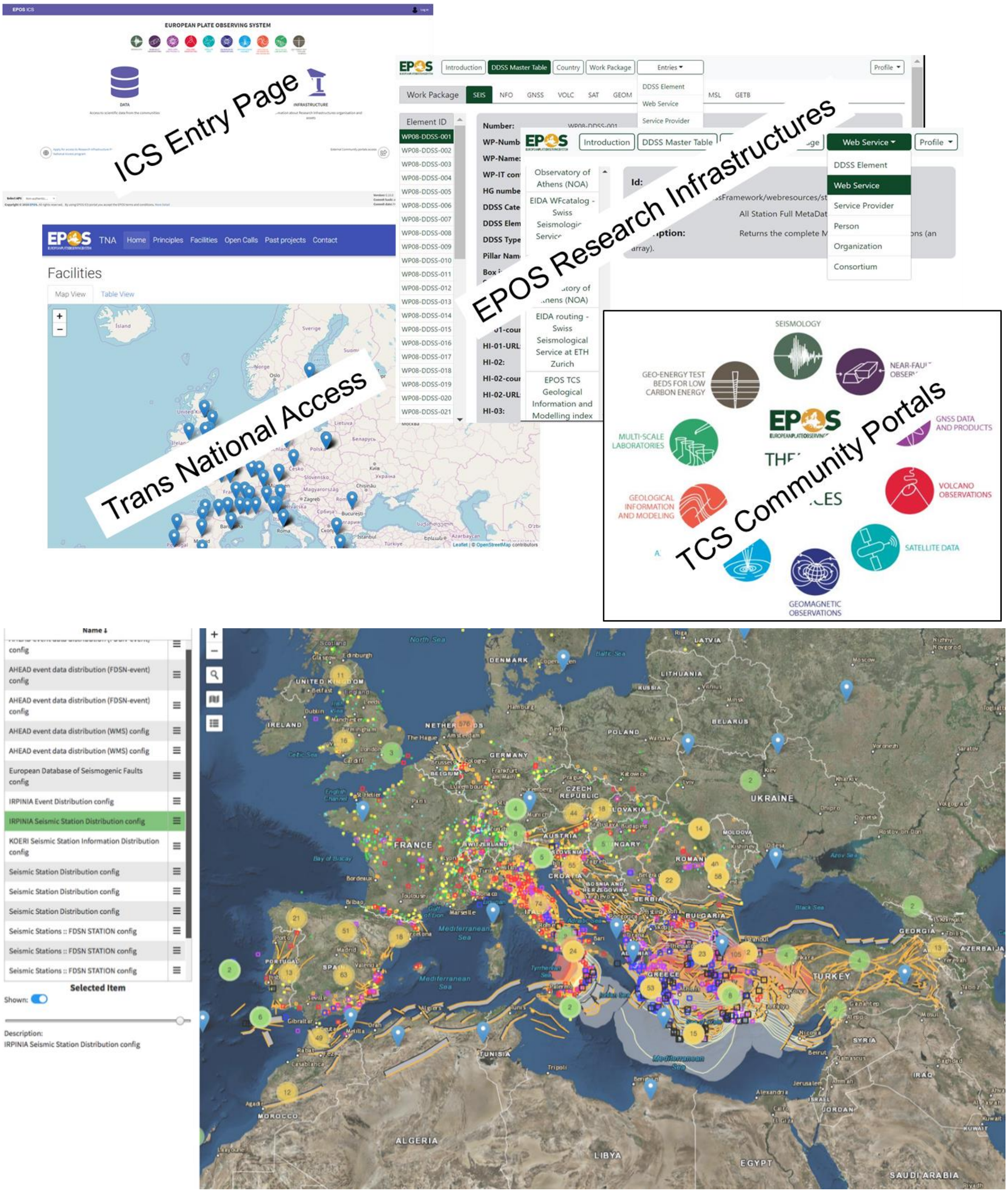

Fig. 9: The EPOS ICS user interface.

2.5 EPOS Metadata Reference Model

2.5.1 EPOS ICS Central Hub

EPOS Integrated Core Services-Central (EPOS ICS-C) is an essential component of EPOS-ERIC as it provides end-users with a view of what EPOS provide and it allows the TCS to expose their assets for (re-)use by users. Functionalities and usage of the system have been demonstrated by the EPOS Portal User Interface demonstrator.

The ICS provides the central hub (ICS-C) of the EPOS e-infrastructure ensuring interoperability between the data and services provided by the TCSs and the National Research Infrastructures (NRIs). The ICS system architecture has been designed to provide the tools to facilitate the discovery of data, data products, software and services (DDSS) and the integration of these resources to fulfil users' requests across the EPOS community.

The system architecture of the ICS is composed of several modular components which are depicted in Fig. 10 and described below.

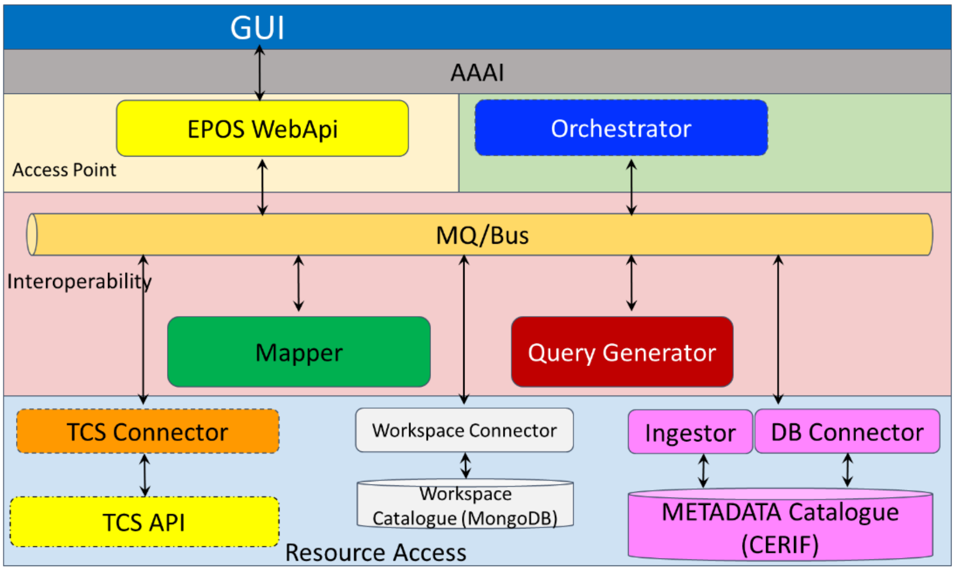

Fig. 10: System architecture of the EPOS Integrated Core Services. See text for explanation of each modular component.

Metadata Catalogue (CERIF)

It is used for storing metadata provided by the Thematic Core Services (TCS) which has been harmonised to a common standard known as CERIF (Common European Research Infrastructure Format) to facilitate efficient searching.

TCS Connector

It allows TCS Application Programming Interface (API) to convert DDSS elements to a TCS specific metadata standard to be connected into the ICS. For example, it provides wrappers to extract and store this metadata in the Metadata Catalogue described before.

Query Generator

This component is used to create an SQL query from the parameters provided by the Web Application Programming Interface (API) and to pass this to the Metadata Catalogue so that the relevant data can be retrieved via the database connector.

Mapper

The Mapper is a software that performs mapping functions between the Common European Research Information Format (CERIF) data model and other metadata schemas such as TCS specific metadata formats.

Message Queue (MQ)/Bus

This component provides a system management function by orchestrating the interaction among other modules in order to satisfy user requests.

Workflow Manager

The workflow manager executes specific workflows, for example to process data using specific software using appropriate workflow engines.

GUI (Graphical User Interface)

It provides the interface to the user and allows them to firstly query the DDSS elements of interest, and then specify appropriate workflows to process the data.

EPOS WebAPI

This component performs the connection between the GUI and other system components. For example, by formatting the query parameters input to the GUI by the user and directing these to the Query Generator.

AAAI (Authentication, Authorisation, Accounting, Infrastructure)

This module will manage and interoperate with all the major AAAI services such as SAML, OAuth, OpenID, X.509, and related products such as Shibboleth and EduGAIN.

Integrated Core Services-Central

The ICS-C (Integrated Core Services-Central) provides the following functions:

- Harvesting metadata (or accepting "push" of metadata) from the TCS;
- Authenticating a user and setting up the appropriate authorisations and accounting.
- User request expressed via the user interface. The end-user, with assistance from the ICS-C, defines the workflow of operations to be done on datasets using particular resources. Two modes are envisaged:
  - The workflow is executed with monitoring information passed to the user. This is efficient since the ICS-C can optimise the workflow deployment, but the end-user is disconnected once the execution starts.
  - The first step is executed, and the user then interacts to accept the results of that step and initiate the next successively through the workflow. This is less efficient (since optimisation of the deployment cannot be done) but the end-user controls each step.
- The results are passed back to the end-user and the results, any derived datasets, the workflow may be curated for later re-use.

2.5.2 Description of CERIF

CERIF (Common European Research Information Format) is an international standard relational data model to storage research information, allowing its interoperability. It is a reference model for the development of Research Information Systems (CRIS). The CERIF data model is presented in Fig. 11 and comprises the following three levels:

- Specification (Conceptual Level): a concept about research entities and their relationships;
- Model (Logical Level): a description of research entities and their relationships;
- Database Scripts (Physical Level): a formalization of research entities and their relationships.

The data model allows a metadata representation of research entities, their activities, research interconnections and their output as well as a high flexibility with formal and semantic relationships. It also enables quality maintenance, archiving, access and interchange of research information and supports knowledge transfer to decision makers, for research evaluation, research managers, strategists, researchers, editors and the general public.

A Research Information System (CRIS) can be implemented using a subset or superset of the full CERIF model for projects, people, organisations, publications, patents, products, services and facilities with role-based, temporally-bound relationships.

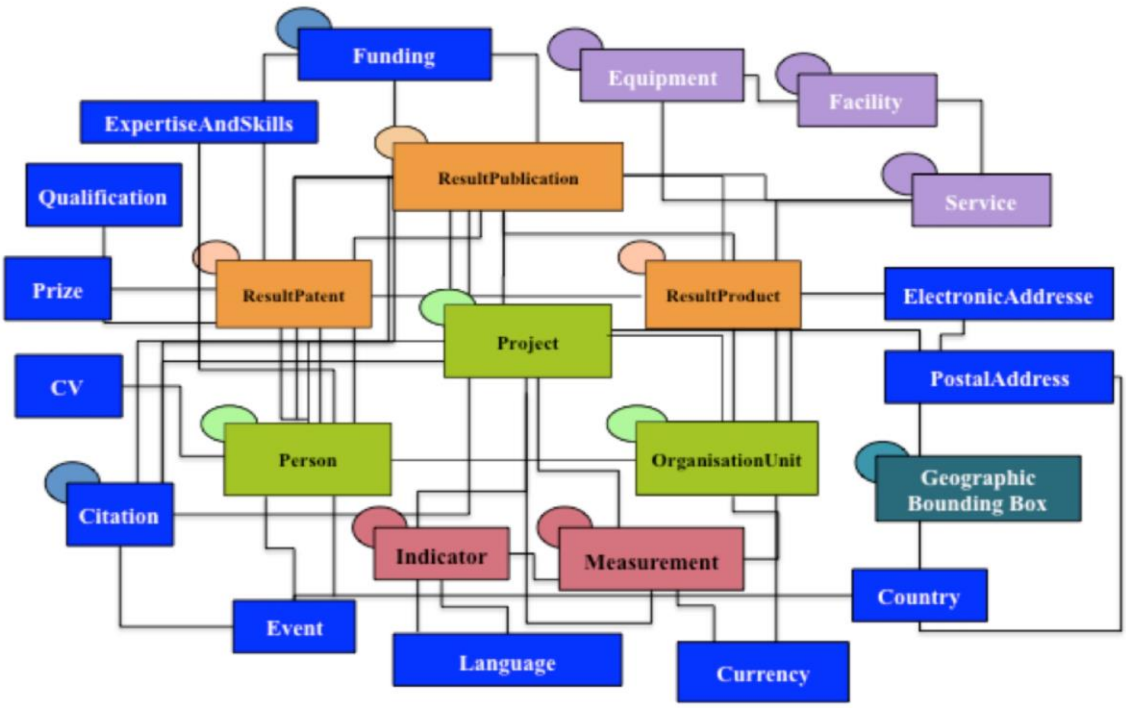

Fig. 11: The CERIF Data Model (https://www.eurocris.org/cerif/main-features-cerif)

The advantages of CERIF are the following:

- Neutral architecture;
- The data model can be implemented (relational, object-oriented, information retrieval);
- The process model can be implemented [DBMS and query (centralised or distributed), HTML web / harvesting / IR-query, advanced knowledge-based technology];
- Broad coverage: includes all aspects of RI (projects, persons, organisations, funding, publications, datasets, etc.);
- Its fine-grained structure and flexible architecture:
  - Input and output of virtually any (meta)data format that is used in the RI Domain;
  - The expression of virtually any formalized use case;
  - The ingestion of an unlimited number of controlled vocabularies;

- Linking entities: This is the key feature of CERIF. Linking entities means that most of the characteristics (attributes) of an object (entity) are not stored with the entity (in the entity table) but are expressed by means of 'linking entities' (in database terms: linking tables), allowing multiple roles/characteristics to be expressed for the same aspect. Only the absolute unique characteristics of an entity are stored in the entity table.

Presently CERIF is used as a model for implementation of a standalone Research Information System, as a model to define the wrapper around a legacy non-CERIF CRIS to allow homogeneous access to heterogeneous systems and as a definition of a data exchange format to create a common data warehouse from several CRIS.

2.5.3 Basic features of EPOS metadata reference model

The EPOS metadata reference model (also known as the Baseline) was created to aid the TCS in collection and description of their metadata. An important aspect of this abstraction-layer was to enable the ICS team to understand the available metadata elements. Core elements from common metadata standards have been included in the EPOS metadata reference model. Other elements have been applied accordingly on the needs of the EPOS community. These elements are represented through various entities as shown in Fig. 12. For each entity, a set of attributes has been specified.

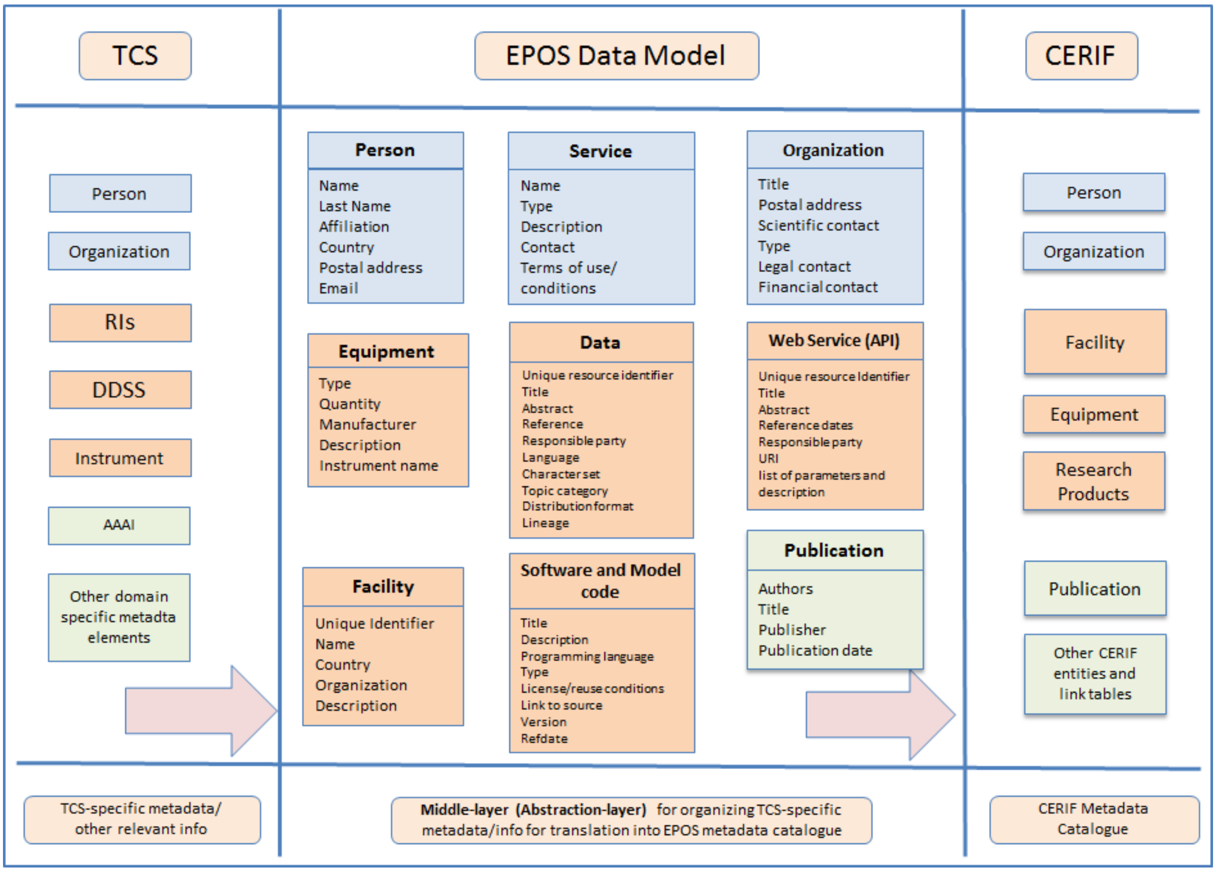

Fig. 12: The elements of the EPOS metadata reference model and mapping to CERIF

2.5.4 EPOS metadata convertor and mapping to CERIF

During the EPOS Preparatory phase, the CERIF (Common European Research Information Format) was chosen for implementation. The metadata catalogue will enable the user to perform discovery, visualization, processing and other functions. In order to manage all the information needed to satisfy user requests, all metadata describing the Thematic Core Services (TCS) through Data, Datasets, Software and Services (DDSS) should be stored into the EPOS Integrated Core Services (ICS), internal catalogue. In this context, the user will be provided with a single homogeneous access to the heterogeneous resources from the TCS. A key challenge in providing such a homogenized view is the diverse metadata profiles used within the scientific communities. Metadata established within the TCSs reflect the usage of many different standards (ISO19115/19, Data Catalog Vocabulary DCAT, Dublin Core Metadata Initiative, INSPIRE, etc.) Different interpretation of these metadata standards further increases the level of complexity. Hence, the EPOS ICS has constructed a metadata reference model (baseline) to create a level of harmonization for these different profiles.

The process of converting metadata acquired from the EPOS TCSs to CERIF is twofold (Fig. 13). Each TCS is required to map its metadata into the EPOS baseline. The mapping occurs between two metadata standards (source: TCS; destination: EPOS baseline). The EPOS baseline, which serves as an abstraction layer, has been implemented by extending the Data Catalogue Vocabulary Application Profile (DCAT-AP). EPOS ICS is responsible of ingesting the EPOS baseline format (EPOS-DCAT-AP) into CERIF. This has been achieved through custom converters made by ICS.

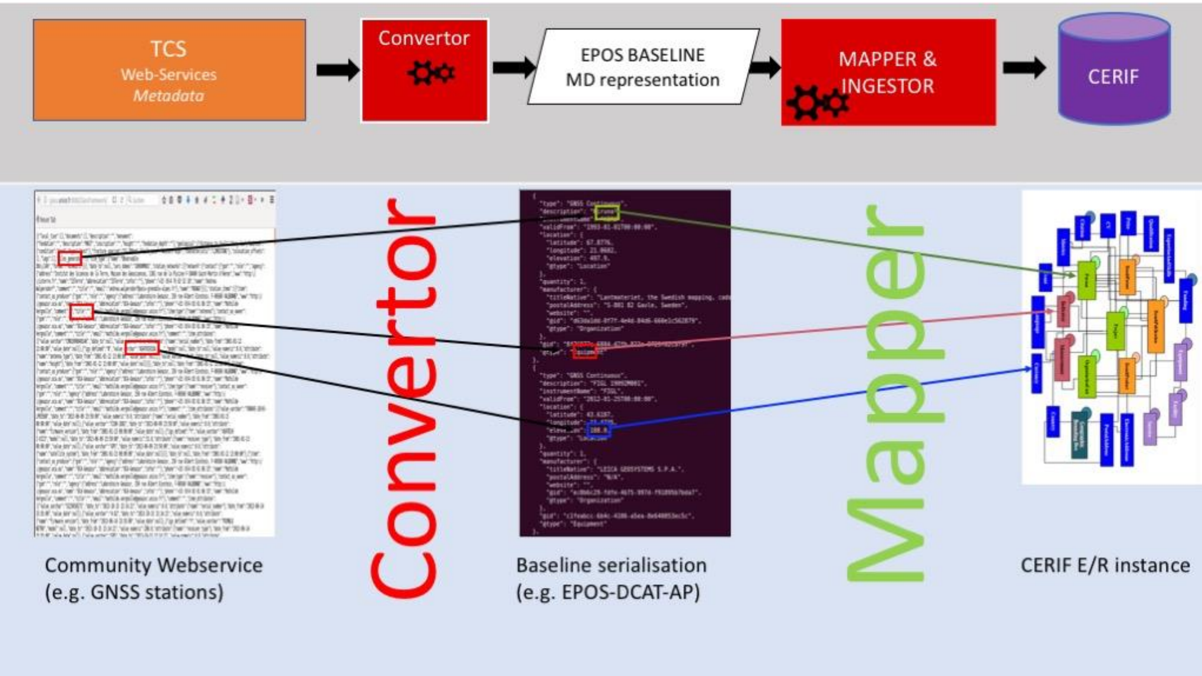

Fig. 13: Schematic representation of EPOS metadata mapping to CERIF (http://wiki.epos-ip.org/index.php/TCS_Metadata_Mapping)

2.5.5 EPOS Data Catalogue Vocabulary Application Profile (EPOS-DCAT-AP)

The EPOS ICS team has provided a standard template for serialisation of the EPOS metadata reference model in order to make a conversion feasible. The template is built upon the existing standard vocabularies, Data Catalogue Vocabulary Application Profile (DCAT-AP). DCAT-AP is used to ingest the metadata from TCS and to present data in a standardised way to machine agents willing to get information from the EPOS ICS system. Extensions have been applied to accommodate for the EPOS specific needs of entities, attributes and relationships.

Interactions with TCS have been crucial in understanding their metadata capacities and to provide the appropriate template for their conversion. In order to facilitate the process of metadata mapping, ICS has provided a UML diagram representing the EPOS-DCAT-AP (DCAT extension), along with a schema definition and an XML example for indicating expected values. A [GitHub] environment was established to provide them with the proper documentation.

It was also expected that each TCS contribute to the GitHub project by uploading their converted XML files into a specific folder. Thus, ICS and TCS could easily interact and solve mutual issues encountered within each conversion. Three EPOS-DCAT-AP entities (Person, Organisation and WebService) have been declared as ready to be used by the ICS. Hence, these are the prioritised entities that the TCSs will use in their first conversion phase. An example of an EPOS-DCAT-AP XML is presented in Fig. 14.


```xml
<?xml version="1.0" encoding="UTF-8"?>
<eposap:Epos
xmlns:adms="http://www.w3.org/ns/adms#"
  xmlns:cnt="http://www.w3.org/2008/content#" xmlns:dcat="http://www.w3.org/ns/dcat#"
  xmlns:dct="http://purl.org/dc/terms/" xmlns:eposap="http://www.epos-ip.org/"
  xmlns:foaf="http://xmlns.com/foaf/0.1/" xmlns:http="http://www.w3.org/2006/http#"
  xmlns:locn="http://www.w3.org/ns/locn#" xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:schema="http://schema.org/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
  xmlns:xml="http://www.w3.org/XML/1998/namespace" xmlns:xsi="http://www.w3.org/2001/XMLSchema-ins
  xsi:schemaLocation="http://www.epos-ip.org/ EPOS_DCAT-AP.xsd ">

  <!-- EPOS Catalog -->
  <eposap:Catalog>...
  <!-- EPOS Person -->
  <eposap:Person>...
  <!-- EPOS Organisation -->
  <eposap:Organisation>...
  <!-- EPOS Project -->
  <eposap:Project>...
  <!-- EPOS WebService -->
  <eposap:WebService>...
  <!--  EPOS Publication -->
  <eposap:Publication>...
  <!--  EPOS Service -->
  <eposap:Service>...
  <!--  EPOS Equipment -->
  <eposap:Equipment>...
  <!-- EPOS Facility -->
  <eposap:Facility>...

</eposap:Epos>
```


Fig. 14: example of an EPOS-DCAT-AP XML showing the elements and datatypes used in the schema.

2.5.6 Selected Thematic Core Services (TCS): review

Within the EPOS IP, ten different communities were identified and organized in Thematic Core Services (TCS). A Working Package is assigned to each TCS taking into account the requirements of the different EPOS communities. The TCS are the community-specific integration (table 1) and they represent transnational governance frameworks where data and services are provided to answer scientific questions. It is within the TCS where each community discusses their specific implementation, best practices and sustainability strategies as well as legal and ethical issues.

The different TCS have varying degrees of maturity in their development and it is not possible to deal with all together as if they are all equal and homogeneous. Some TCS have a very specific services architecture based on years of experience in that specific domain where as others TCS do not have a history of developing services. Some TCS have already done the effort of defining metadata standards and web services to disseminate the data but others are still in the process of undertaking such work.

According to their level of maturity, TCS will build new interoperable services by: a) using the EPOS metadata standard APIs, or b) making existing services interoperable with ICS through the use of community-accepted or standards APIs and interface to EPOS metadata.

Table 1: Thematic Core Services (TCS) implemented in EPOS

TCS	CONTENT
SEISMOLOGY	Seismic waveforms (ORFEUS) Seismological products (EMSC) Hazard and risk products (EFEHR) Computational seismology
NEAR FAULT OBSERVATORIES	NFO multidisciplinary data and products Borehole data Virtual laboratory and early warning test beds
GNSS DATA AND PRODUCTS	GNSS primary data and deprived products Processing and visualization tools
SATELLITE DATA	SAR interferograms Integrated satellite products On-line processing tools
VOLCANO OBSERVATIONS	Multidisciplinary volcanic data and products Hazard products TNA to volcano observatories
ANTHROPOGENIC HAZARDS	Data for AH episodes Multi-hazard simulator – multi-risk assessment AH data visualisation
GEOMAGNETIC OBSERVATIONS	Global and regional geomagnetic models Magnetotelluric data
GEOLOGICAL INFORMATION AND MODELLING	Geological multi-scale data Integrated geological maps Borehole visualisation
MULTI-SCALE LABORATORIES	Experimental and analogue data TNA to experimental and micro-analytical facilities
GEO ENERGY TEST BEDS FOR LOW CARBON ENERGY	Geo energy test beds Access to in-situ GETB experiments

Prior to the conversion process, each TCS has provided the ICS with a list of prioritised DDSS elements. The main purpose was to estimate the number of elements that would be ready for the validation phase, together with a promise from the TCS on what they could deliver. The ICS team has reviewed each xml file provided by the TCS by checking their syntactic validation, consistency, proper use of person- and organizational identifiers, web service link and specific attributes (domain, sub-domain, keywords, operation) required for the graphical user interface.

3 Deep Seismic Sounding Control Source Data Technical Specs.

Seismic data is stored in various different formats, which have been homogenized and harmonized by recognized international geophysical societies. For instance, the Society of Exploration Geophysicist (SEG) developed a series of formats for data storage in early 70's. These formats were soon adopted by hardware manufacturers, and broadly extended by academic and commercial seismic processing software packages. These data formats store the seismic data as time series data from one or several sensor points. Each time series is associated to the location of the sensor (seismic station) and the location of the source point (in the case of the control source). The specifications of the acquisition system and geometry are usually defined within the information stored in the headers of the files. All the formats defined by SEG include the time series (seismogram) and the information concerning the location of the source, location of the receiver and format specifications of the time series among other parameters. The most generally used formats are SEG-Y, SEG-D, SEG-B, SEG-C and SEG-2.

These formats are constantly being improved and updated through the SEG Technical Standards Committee which serves as a forum for discussion of geophysical developments in which standards for acquisition and processing of geophysical data are being identified and/or improved. When data problems of a sufficient magnitude are identified, an appropriate subcommittee is appointed to develop a set of standards and make a recommendation to the SEG Board of Directors. The committee works in partnership with a wide range of industry organizations in order to adopt the best improvement.

The most used standards are provided as annexes so they are widespread through the general geophysics community. All software packages, both commercial and academic, have reading and conversion commands, so any of the SEG formats could be a base format. The compatibility with other seismic standards used by academy, especially in earthquake seismology, is assured by software compatibility codes. This software establishes format bridges between DSS and natural source time series data.

SEG-Y file format could be the optimal choice for DSS. It has been the accepted exploration industry data format for digital seismic data since 1975. Therefore, if academically acquired normal incidence seismic reflection data is also considered to be archived and distributed with the same frameworks as the DSS, it would be an optimal format. SEG-Y is originally a magnetic tape-centric format. This format has evolved over the years to become the data format industry standard for data analysis and processing. The latest format revision of SEG-Y was released in 2002.
The academic community related to DSS data and specially the USA IRIS-PASSCAL consortia has expanded further the SEG-Y format making it more flexible. The format accommodates a variety of industry-specific metadata (table 2), for example, source sweep information and trace identification code (uphole, water break, etc.), and legacy metadata, such as the tape reel number, in the binary header within the SEG-Y file. The format proposed now includes the advantages provided by PH5 data-base related scheme.
The three types of data in table 2 can be represented and stored in SEG-Y format with the appropriate header values (Fig. 15), thus enhancing the use of this data format.

Table 2: level 1 seismic data types for DSS which can be represented by standard SEGY format definitions.

Data Type #	Data Category	Data Types	Data descriptions and comments
1	**Raw DSS data (deep seismic sounding, long range refraction, wide-angle data)**	Time series data for an individual sensor point. Associated location information, including altitude. Sensor type. Original recording equipment type and current form of storage. Station comments.	The time series ground motion data may be one component (usually the vertical), although three components are the most common. The coordinate system used should be clearly specified. For old data, where the location of the stations may be inaccurate, all available data should be included in the metadata file. When possible, it has to be included how the position was measured. Sensor type should include the sensor manufacturer and identification number, if possible, with a link to information regarding sensor performance. Tt should be specified if old data was collected on f.m. tape, or other outdated recording mode. Timing accuracy may depend on the recording mode. Sensors may be deployed on different material (e.g. bedrock or sand) and the operator may have noted noise sources nearby.
2	**Pre-process data DSS data sections**	Commonly, reduced time sections after filtering. These may be digital time series, or as images.	Information about station position has been converted to source-receiver position. The reduction velocity and filter parameters are important and should be taken into account. For image files (old data), parameters such as the form of scaling of the traces, should be noted. Some sections may use more advanced processing techniques that may be relevant, e.g. including three component information.
3	**Models derived from DSS data**	They primarily result in models of the arrival time of identified phases, producing a two-dimensional velocity model, of P- or S-waves velocities	Entities will be represented by the occurrences or mines, which could either be spatially represented by points or polygons. The data are already handled by Minerals4EU and it should be examined whether these services can simply be integrated “as is”.

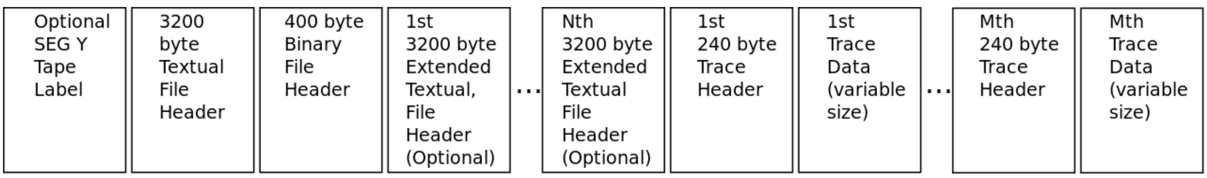

Fig. 15: Byte stream structure of a SEG-Y file, with rev 1 Extended Textual File Header records.

3.1 Description of Extended SEGY format

A SEG-Y file typically contains all the traces which recorded a given source and is commonly referred to as a shot gather. The file may also contain traces from a single receiver which recorded a large number of sources and that is known as a receiver gather. Each trace within the gather is limited to 32,767 samples.

The PASSCAL SEG-Y trace file is in brief a variation of the SEG-Y format. A PASSCAL SEG-Y trace file contains one trace and accommodates an extended duration (greater than 32,767 samples) by using some of the unspecified words within the data-trace header. Only a trace header is defined within the PASSCAL SEG-Y trace file.

Table 3. Description of the SEG-Y header data format with byte location, header word length and type, as well as definition of the information content in the header.

Byte Offset	Field Description
0 - 3	Line trace sequence number
4 - 7	Reel trace sequence number
8 - 11	Field record number
12 - 15	Field trace number
16 - 19	Energy source point number
20 - 23	Ensemble number
24 - 27	Trace number
28 - 29	Trace ID code
30 - 31	Number of vertically summed traces
32 - 33	Number of horizontally summed traces
34 - 35	Data use
36 - 39	Offset (distance)
40 - 43	Receiver group elevation
44 - 47	Source elevation
48 - 51	Source depth
52 - 55	Elevation at receiver group
56 - 59	Source elevation
60 - 63	Water depth at source
64 - 67	Water depth at group
68 - 69	Elevation and depth scalar
70 - 71	Coordinate scalar
72 - 75	X coordinate of source
76 - 79	Y coordinate of source
80 - 83	X coordinate of receiver group
84 - 87	Y coordinate of receiver group
88 - 89	Coordinate system
90 - 91	Weathering velocity
92 - 93	Sub-weathering velocity

94 - 95	Uphole time at source in ms
96 - 97	Uphole time at group in ms
98 - 99	Source static correction in ms
100 - 101	Group static correction in ms
102 - 103	Total static applied in ms
104 - 105	Lag time A, ms
106 - 107	Lag time B, ms
108 - 109	Delay recording time, ms
110 - 111	Mute start time, ms
112 - 113	Mute end time, ms
114 - 115	Number of samples
116 - 117	Sample interval, microseconds
118 - 119	Gain type
120 - 121	Gain constant
122 - 123	Early gain
124 - 125	Correlated?
126 - 127	Sweep frequency at start
128 - 129	Sweep frequency at end
130 - 131	Sweep length in ms
132 - 133	Sweep type
134 - 135	Sweep taper at start, ms
136 - 137	Sweep taper at end, ms
138 - 139	Taper type
140 - 141	Alias filter frequency, Hz
142 - 143	Alias filter slope, dB/octave
144 - 145	Notch filter frequency, Hz
146 - 147	Notch filter slope, dB/octave
148 - 149	Low-cut frequency, Hz
150 - 151	High-cut frequency, Hz
152 - 153	Low-cut slope, dB/octave
154 - 155	High-cut slope, dB/octave
156 - 157	Year
158 - 159	Day of Year
160 - 161	Hour
162 - 163	Minute
164 - 165	Seconds
166 - 167	Time bias code
168 - 169	Trace weighting for LSB
170 - 171	Geophone group number
172 - 173	Geophone group number (field)
174 - 175	Geophone group number, last trace (field)

176 - 177	Gap size
178 - 179	Over travel
180 - 185*	station name code (5 chars + 1 for termination)
186 - 193*	sensor serial code (7 chars + 1 for termination)
194 - 197*	channel name code (3 chars +1 for termination)
198 - 199*	extra bytes (2 chars)
200 - 203*	sample interval in microsecs as a 32 bit integer
204 - 205*	data format flag: 0=16 bit integer 1=32 bit integer
206 - 207*	milliseconds of second for first sample
208 - 209*	trigger time year
210 - 211*	trigger time day of year
212 - 213*	trigger time hour
214 - 215*	trigger time minutes
216 - 217*	trigger time seconds
218 - 219*	trigger time milliseconds
220 - 223*	scale factor (IEEE 32 bit float) (true amplitude = (data value)*(scale factor)/gain)
224 - 225*	instrument serial number
228 - 231*	number of samples as a 32 bit integer
232 - 235*	max value in counts
236 - 239*	min value in counts.
	* extended SEG-Y header
	All unused fields in the trace header are set to 0x00

Having the data arranged into SEG-Y files, seems to be the most reasonable arrangement as the historical character of some of the available legacy data might be in old format types. This is partly a consequence of the practicalities of the different types of data (e.g. single station *vs.* sensor arrays; single source can array of sources) and to technical and scientific developments and improvements. It is not sensible or feasible, at least in the short to intermediate term, that a common data format should be used for all types of seismic data. However, in order to have the data integrated into a homogenized EPOS system, the data should be delivered to the user in one of the standardized forms, preferably in SEG-Y standards.

In order to make old-legacy datasets available and usable, an additional layer of information (metadata) is required. This should include a text description of the information files with the information that is, currently and conventionally provided within the current revision of the SEG-Y format (table 2).

4. METADATA: DSS Data-set Description for each experiment

In comparison with earthquake time series from regular networks, DSS datasets feature significant differences mainly related to the economic cost of the data acquisition experiments. Usually, they are a couple of orders of magnitude more expensive than natural source (EQ) data. Control-source seismic data need the contribution of active sources (e.g. Vibroseis, Impact, explosions, etc.), therefore the acquisition costs are higher. These facts emphasize the importance of traceability and visibility of both the research team and the funding agencies that made the project and experiment possible.

In addition, data should have assigned a permanent identifier, and the assumption by the external user of a citation policy. These identifiers are referred to as DOI (Digital Object Identifier) or handle which are a string of numbers, letters and symbols used to permanently identify an article, document or dataset linked on the web. A DOI will help the user, reader or scientist to easily locate a document/data set as it will always refer to that article, and only to that one. In summary DOI identifies the content and provides a persistent link to its location on the WWW (Internet). The DSS dataset should have a DOI assigned to them as soon as they are available. Currently there are several agencies that provide DOI numbers. In particular DIGITAL.CSIC, the electronic resource services of the Spanish National Research Council (CSIC), makes these id numbers available.

In order to have data available, identifies and with the acquisition parameters and characteristics specified, DSS datasets should follow the DataCite Metadata Schema. It is a list of core metadata properties chosen for the accurate and consistent identification of a resource, along with recommended use instructions. These recommendations follow the FAIR principles for data, (which are also the main guidelines of the EPOS infrastructure, Fig. 2):

- Findable: Easy to find by both users and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets;
- Accessible: Stored for long term hence they can be easily accessed and/or downloaded with well-defined license and access conditions (Open Access when possible).
- Interoperable: Ready to be combined with other datasets by users as well as computer systems;
- Reusable: Ready to be used for future users and to be further processed using computational methods by research, academy and industry.

DataCite recommends a description of the research data including the following aspects:

- Context: description of the project, purpose of the research and the methodology used.
- Nature of the data, history of the data, content and structure, terminology, software, date of creation and modification, number version, responsible of the project and participants.
- File formats, structure and nomenclature of the files.
- Legal aspects, access and security policies.

The description of the data must be complete enough to be able to answer the following questions:

- Who has produced the data?
- Is the title specific enough?
- Why have the data been created?
- What limitations do the data have (for example, confidential data has been deleted)?
- How should the data be interpreted?
- Are there gaps in the data or do they give a complete view of the subject studied?
- What processes have generated the data?
- What does the data measure in the columns of the files?
- What software is needed to read the data?
- How should the data be cited?
- Can the data be reused? What license of use is assigned to the data?
- Are there more versions of the data? Where?
- Have the technical terms and acronyms to which the data refer been defined?
- Have the geographic and chronological parameters of the data been qualified?

- Are the keywords sufficiently specific to the data? Are they based on a thesaurus?
- What is the name of the research project in which the data are framed?
- Who has financed the production and management of the data?

In addition, some data repositories which have implemented the Data Seal of Approval -DSA- (such as the DIGITAL.CSIC), have a data policy that recommends the data providers and scientist to generate additional txt and/or read.me files with additional information, to increase the value of the actual data-set.
In general terms the data-set should include:

- A file template based on the DataCite schema
- Full extensive descriptions of the data (as fully as possible, not simply upload the files).
- Good metadata which should help discovery, citation and reuse
- Information about reuse conditions, the methodology and software used, technical requirements to visualize and reproduce,
- Cite the publication/s associated with the datasets
- Version/date of the dataset if there are more than one
- File naming and consistency are very important
- Include a readme file with additional information.

5. Technical possibilities: Prototype example

With the collaboration of CSIC's digital office DIGITAL.CSIC, a prototype has been built to demonstrate the possibilities of a data archiving system that includes the library designed utilities, thus treating current datasets as published papers (e.g., Bernal, 2011). The example displayed below constitutes the evidence of the technical feasibility of a data access platform which is not uniquely a web page filled up with links but a resource tool which provides each dataset a persistent identifier DOI and handle.
DSS data files are provided in compress format for data transfer proposes and each data is stored in standard SEG-Y format. Furthermore, additional text files take care of providing the description of the experiment and the location of the sources and receivers. The example provided can be accessed by searching for "doi:10.20350/digitalCSIC/8549" (Torné et al., 2018). It corresponds to a normal incidence data set acquired in the west coast of the Iberian Peninsula, in the Atlantic margin, funded by the European Community Framework program in 1993.
Once the dataset is available online, its subsequent usage can be tracked and the user can be provided with an account of the results achieved so far by the provided dataset. Furthermore, it provides access to the full dataset files in all the processing steps in which they are available including data products if they have been provided.
The Iberian Atlantic Margin (IAM) project is a multinational research programme coordinated by the Institute of Earth Sciences Jaume Almera, ICTJA-CSIC (Spain). It involves different academic institutions of United Kingdom (University of Oxford and University of Durham), France (IFREMER), Portugal (University of Lisbon and Institute of Meteorology of Lisbon) and more than 40 associated scientists of several European institutions. The project, financed by the European Community within its JOULE-Programme (Contract# JOU2-CT92-0177), aims to explore the deep continental and oceanic Atlantic margins of Iberia to understand better the formation and evolution of Atlantic type margins, outstanding structural features and potential location of hydrocarbons and other natural resources. The project was designed to acquire deep seismic multichannel data and onshore recording of wide angle and refraction data along the North Iberian margin; the West Iberia margin, including the Galicia margin, the Iberia and Tagus abyssal plain, and the continental shelf off Portugal; and the Gorringe Bank region and Gulf of Cadiz. The experimental part of the study consisted of the acquisition of over

3,700 km of near vertical incidence deep seismic multichannel reflection data. This data together with wide-angle reflection and refraction data recorded by land stations and OBS (Ocean Bottom Seismometers) has resulted in a broad range of data about the crustal structure and velocity information of the study region.

Fig. 16 shows the data portal prototype taking into account the possibilities offered by the library tools. The web portal features the seismic DSS data including normal incidence data. The project includes seismic wide-angle and normal incidence seismic reflection data. The web portal provides download access to the data files of the different transect lines, with all the explanation metadata files provided as txt and excel files with position and all other information available and required for reprocessing and interpretation. The data repository provides access to raw data as well as processed data files (stacks cross-section images of the different profiles). Note that indication on how to reference the data file is included. A short description of the project and the acquisition techniques are provided as well as the persistent identifiers (DOI and handle). Furthermore, links to the library utilities are also indicated by arrows and the statistics to measure the accessibility and download counter. Additional tools include a link to reference management software and research networks such as Mendeley; cite indexes in Google Scholar; scientific networks like ResearchGate; and business and employment-oriented website such LinkedIn. Other visualization formats are also encompassed (e.g., MARC, Dublin Core…).

In the upper part of the application shown in Fig. 16, a link to SHARE provides further statistics information about the data. Fig. 17 shows a screen view of the publications referred to and/or citing the available dataset. Other contributions such as patents, conference proceedings and software among other can be search for. In addition, SHARE allows the user to classify scientific works according to the source, date, language used, funding agency and publisher among others.

Further statistics can be consulted in the "statistics" label, for instance, the item views and downloads according to different countries and the number of downloads through time (Fig. 18).

A link to BASE (Fig. 19) provides further information about the academic resources related to the data. So, scientific publications derived can be consulted and further options are provided such as Email the reference, save a search and creating a list of favorites.

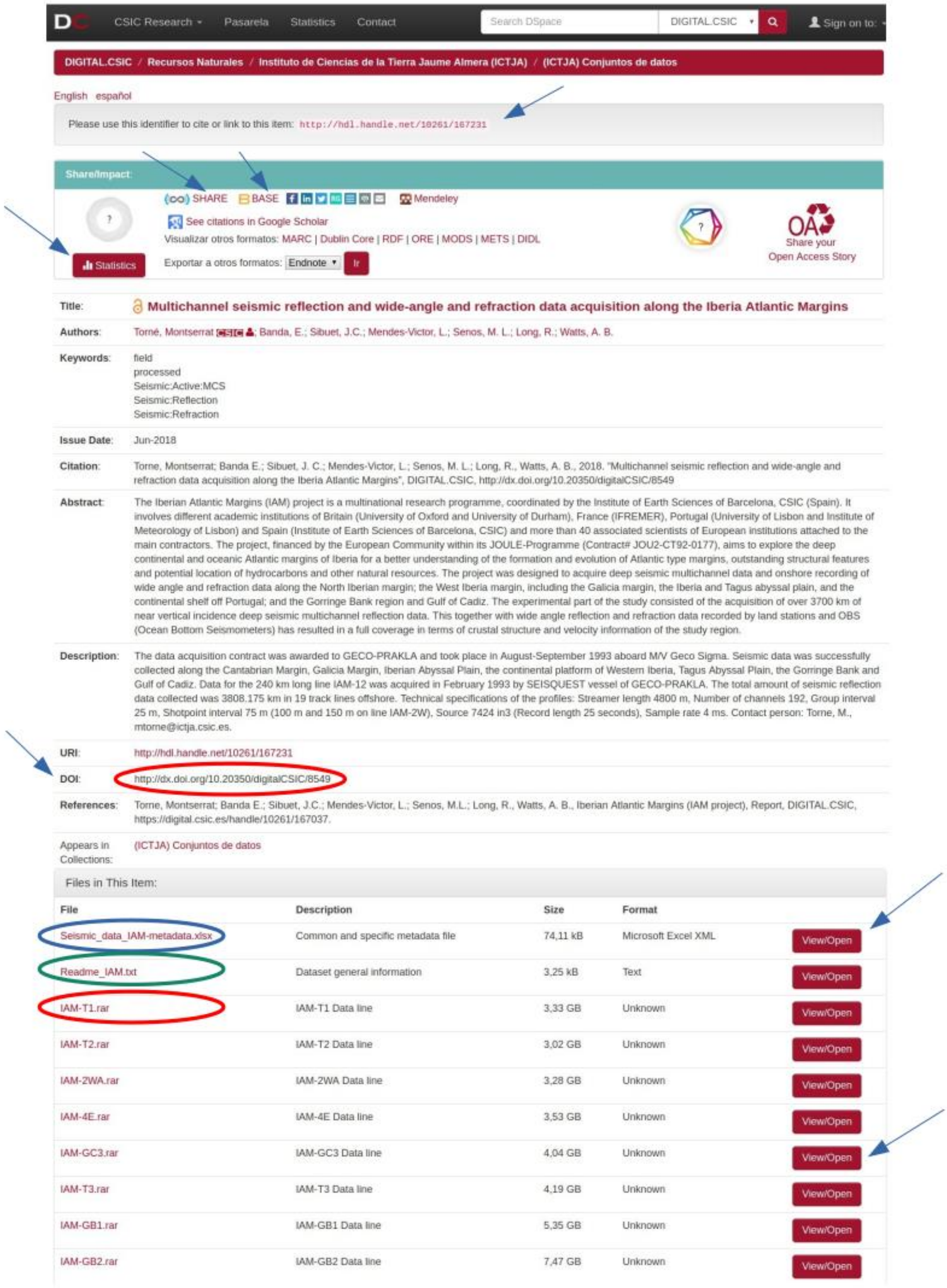

Fig. 16: prototype example of the IAM project provided by DIGITAL.CSIC.

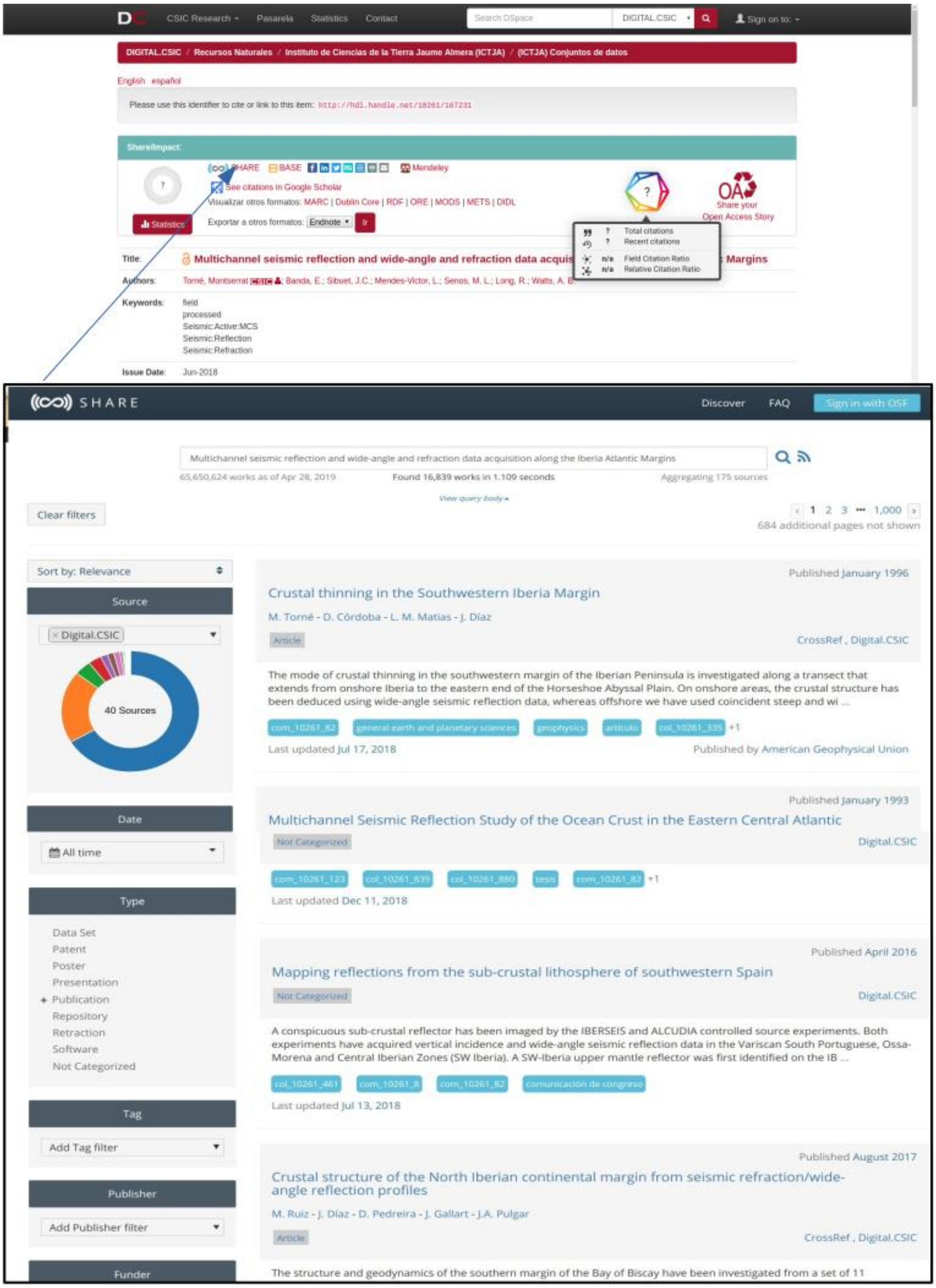

Fig 17: Link to Share that provides a link to the published material related to the dataset in question.

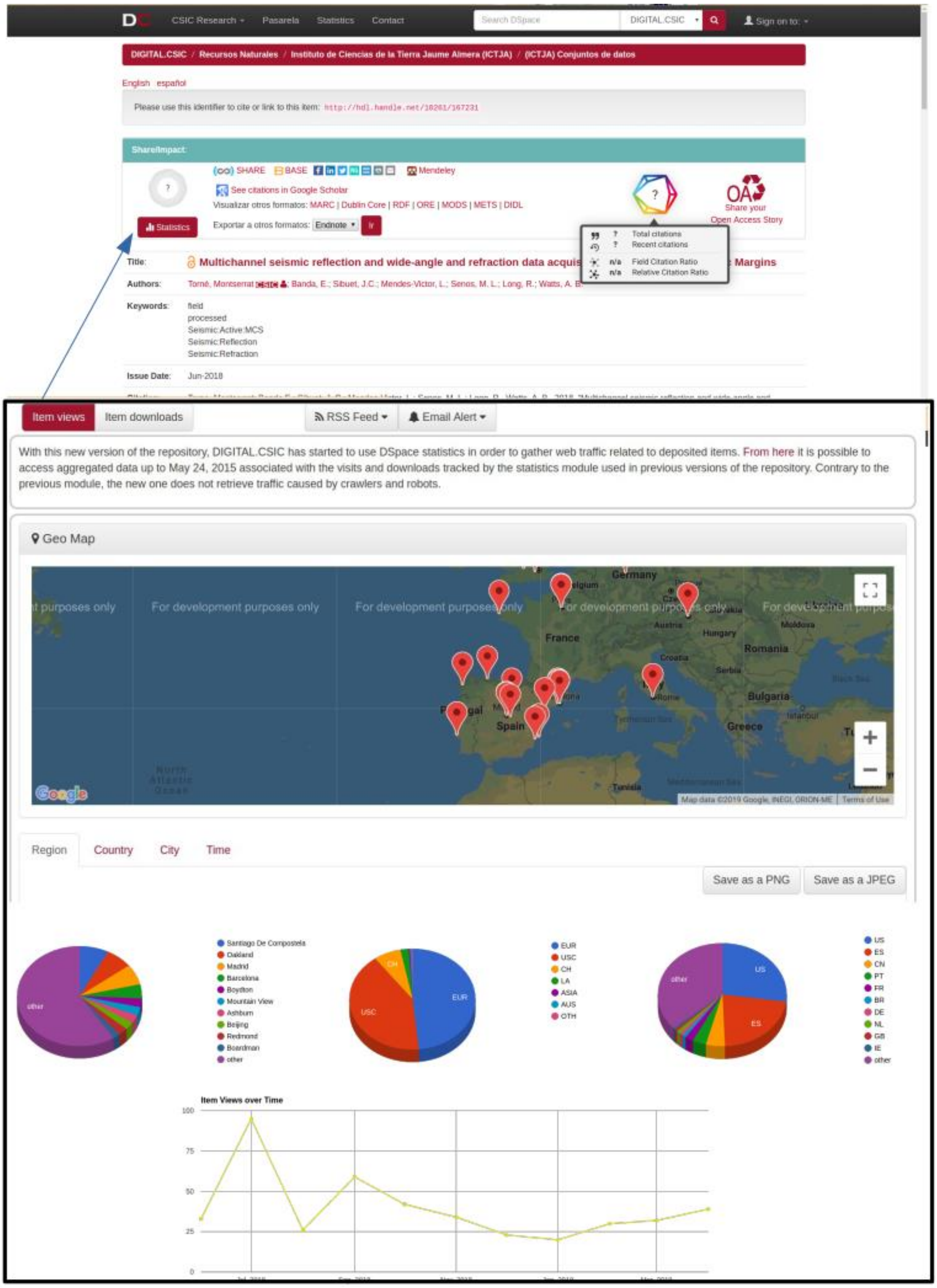

Fig. 18: Statistics of the access to the data baseline.

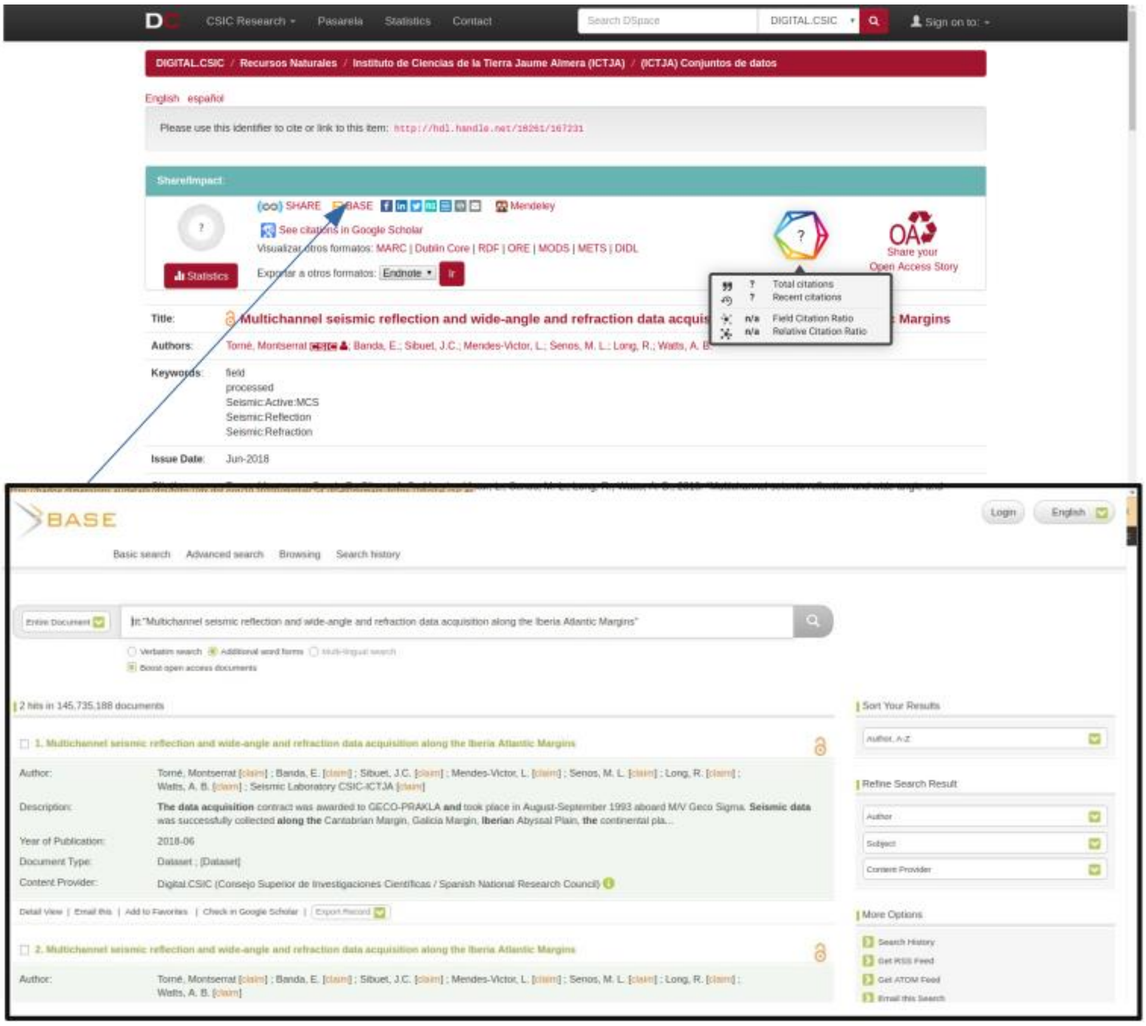

Fig. 19: : Link to Base that provides information about academic web resources.

6. Organization, governance and financial aspects

A first analysis of the DSS data site organization drives to the suggestion that each experimental dataset has to be available in an online repository for its reuse. The organizational structure that make all the EPOS goals possible (Fig. 19) should include the data repositories, as well as some representation of the User community. The data repositories would be responsible for the archiving, maintenance of the data and provide them with a permanent identifier. As currently the data is mostly legacy seismic deep seismic sounding data acquired on land, a supporting institution need to be overviewing the data managing process. This supporting institution would have the mission of assuring the persistent identifiers (DOI, handle, and the reference), as well as the supervision of the setup so that the library specs are accomplished. The entire process should be supervised by the institutional library or advisory board which would be the direct link with the EPOS infrastructure.

Currently the expenses foreseen would correspond to a budget which would need to support the Operational management which should be associated to a specific institution with a direct link to the EPOS Seismology TCS.

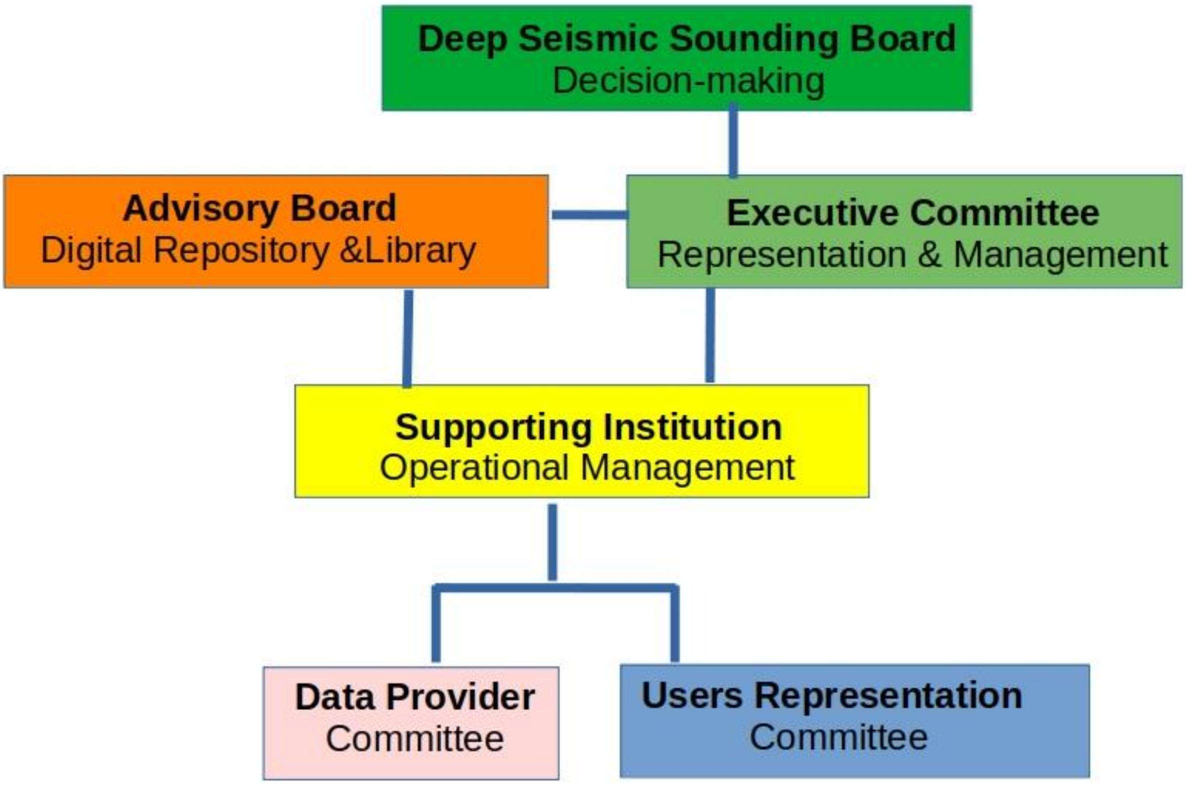

Fig. 19: Organization structure from raw data to open access data ready for a broad range of users.

References

Atakan K., Michalek J., (2017). EPOS-IP Deliverable 6.1 (M24). Description of TCS Requirements and Use Cases. European Plate Observing System (EPOS) Implementation Phase Project No. 676564, InfraDev programme - Horizon2020, 24p.

Atakan, K., Athanasopoulou, A., Avouris, N., Bousias, S., Lamata Martinez, I., Molina, J.F., Pegon, P., Sintoris, C., Tsionis, G. (2018). Review of current SERIES and EPOS databases. SERA Project Deliverable D6.4, 49p.

Bernal, I. (2011). Digital.CSIC: Making the Case for Open Access at CSIC,Serials Review, 37:1, 3-8. https://doi.org/10.1080/00987913.2011.10765338

Bernal, I. (2019). Digital.CSIC: Cómo usar DIGITAL.CSIC para aumentar la presencia de la investigación del ICM en la web. http://hdl.handle.net/10261/180683

Common European Research Information Format (CERIF), maintained by the EUROcris organisation

Jeffery K.G., Bailo D. (2014). EPOS: Using Metadata in Geoscience. Metadata and Semantics Research. Springer International Publishing. DOI: 10.1007/978-3-319-13674-5_17.

Michalek J., Tellefsen K., Atakan K. (2017). EPOS Deliverable 6.4 Regular updates of the user feed-back from the EPOS community.

Torné, M., Banda, E., Sibuet, J. C., Mendes-Victor, L., Semos, M. L., Long, R., and Watts, A. B. (2018). Multichannel seismic reflection and wide-angle and refraction data acquisition along the Iberia Atlantic Margins", DIGITAL.CSIC, http://dx.doi.org/10.20350/digitalCSIC/8549

Glossary

AAAI: Authentication – Authorization – Accounting – Infrastructure

API: Application Programming Interface

CERIF: Common European Research Information Format

DAP: Data Access Portal

DDSS: Data, Datasets, Software and Services

EPOS: European Plate Observation System

EPOS-CS: EPOS Core Services

EPOS-DCAT-AP: EPOS Data Catalog Vocabulary Application

EPOS-PP: EPOS Preparatory Phase

EPOS-IP: EPOS Implementation Phase

EPOS-OP: EPOS Operation Phase

ERIC: European Research Infrastructure Consortium

ICS: Integrated Core Services

ICS-C: Integrated Core Services Central (Hub)

ICD-D: Integrated Core Services Distributed Services

NRI: National Research Infrastructure

TCS: Thematic Core Services

RIs: Research Infrastructures

RUC: Requirements and Use Cases

SERIES: Seismic Engineering Research Infrastructures for European Synergies

SOAP: Service Object Access Protocol

TNA: Transnational Access

Contact

Project lead	ETH Zürich
Project coordinator	Prof. Dr. Domenico Giardini
Project manager	Dr. Kauzar Saleh
Project office	ETH Department of Earth Sciences

Sonneggstrasse 5, NO H62, CH-8092 Zürich

sera_office@erdw.ethz.ch

+41 44 632 9690

Project website www.sera-eu.org