SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D6.4 Review of current SERIES and EPOS databases

Work package	WP6 (NA4: Networking experimental seismic engineering databases (SERIES))
Lead	Joint Research Centre (JRC)
Authors	Kuvvet ATAKAN (UiB) Adamantia ATHANASOPOULOU (JRC) Nikolaos AVOURIS (UPAT) Stathis BOUSIAS (UPAT) Ignacio LAMATA MARTINEZ (JRC) Javier Francisco MOLINA (JRC) Pierre PEGON (JRC) Christos SINTORIS (UPAT) George TSIONIS (JRC)
Reviewers	[Name, Institution]
Approval	Management Board
Status	Draft
Dissemination level	Public
Delivery deadline	12.04.2018
Submission date	12.04.2018
Intranet path	DOCUMENTS/DELIVERABLES/SERA_D6.4_Review_of_current_SERIES_and_EPOS_databases

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. Main objective of this deliverable is to review the existing EPOS service provisions and SERIES databases and provide a comparative assessment.

A review of the current SERIES (**S**eismic **E**ngineering **R**esearch **I**nfrastructures for **E**uropean **S**ynergies) and EPOS (**E**uropean **P**late **O**bserving **S**ystem) platforms developed by the earthquake engineering and solid Earth science communities in Europe, respectively, are presented. The document includes a summary of the two projects and a review of the main features of the two systems, focusing on the database content, structure, functionality and access portals. A special emphasis is given to the Thematic Core Service (TCS) on Seismology of EPOS due to its relevance to SERA project. Results from this deliverable will further be used for outlining a roadmap for integration of earthquake engineering and seismological data using the infrastructures provided by the SERIES and EPOS platforms.

It is noted that resulting form a better knowledge of the EPOS and SERIES projects, the deliverable needed to review also the two approaches for data sharing and integration (governance, architecture, collaboration), and not be only limited to a review of current SERIES and EPOS databases.

1 Introduction

Over the years, a large amount of data has been produced by the earthquake engineering and seismology research infrastructures in Europe, together with 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.

In a world-wide scale, collaborations in earthquake engineering lack a common interoperability framework, resulting in tedious and complex procedures to integrate data and results. Up to now, the most significant effort in Europe towards the interoperability of earthquake engineering experimental data was the SERIES project (Seismic Engineering Research Infrastructures for European Synergies) in the period 2009 - 2013.

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 (Observatories and Research Facilities for European Seismology) for seismological waveform data and EMSC (European Mediterranean Seismological Centre), that both were sustained by the community for several decades. In addition to this, recently, the ESFRI (European Strategic Forum for Research Infrastructures) initiative and the EPOS project (European Plate Observing System), provided a larger framework for the integration of all solid Earth science data into a single Pan-European e-infrastructure.

Up to now, the two adjacent scientific disciplines, i.e. the 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, i.e. SERIES and EPOS (and in particular the ESPO community on seismology). An efficient use of resources and know-how sharing along with the added value of bringing together the data sources and data exchange services of the two communities will be a step forward in the provision of tools and knowledge for the benefit of a wide range of users.

This deliverable presents a review of the current SERIES and EPOS systems: the main features of the two are discussed, focusing on their architecture, content, databases, functionality and access portals. It is noted that resulting form a better knowledge of the EPOS and SERIES projects, the deliverable needed to review also two approaches for data sharing and integration (governance, architecture, collaboration), and not be only limited to a review of current SERIES and EPOS databases.

The review will serve as the basis for the future roadmap for integration of earthquake engineering and seismological data 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 two Thematic Core Services (TCS) in "Seismology" and in "Multiscale Laboratories". Metadata Reference Model based on the European standard CERIF (Common European Research Information Format – https://www.eurocris.org), is described and current technical implementation procedures are explained.

The European Plate Observing System (EPOS)[1] is a single, pan-European Research Infrastructure plan for sharing solid Earth Science data, observations and research results. Its mission is to integrate the existing and future advanced 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 in summary the basic features of the EPOS project and the related database. The review has been based on the available information in the EPOS portal (https://www.epos-ip.org/), EPOS deliverable D6.1-M24 "Description of TCS Requirements and Use cases" (Atakan and Michalek, 2017) and the publications by Bailo et al (2016) and Bailo et al. (2015). A full list of references to EPOS is provided at the end of the document.

2.1 EPOS: Project description and timeline

The European Plate Observing System (EPOS – www.epos-eu.org) 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 trans-national 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 georesources and geohazards.

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.

[1] https://www.epos-ip.org/

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 user 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.

Following an initial four years (2010-2014) with a Preparatory Phase project (EPOS-PP) funded by EU-FP7, EPOS is now in its Implementation Phase (EPOS-IP) which is funded by Horizon2020 Program of the European Commission for the period 2015-2019. After its construction during the implementation phase, EPOS will enter its Operational Phase from 2020 onwards.

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. It is EPOS project ambition 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 (see Figure 1) during the project implementation phase (2014-2019) is made up of 47 partners plus 6 associate partners from 25 countries from all over Europe and several international organizations (ORFEUS[2], EMSC[3], EUREF[4], INTERMAGNET[5]). 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 (see https://epos-ip.org).

[2] https://www.orfeus-eu.org/
[3] https://www.emsc-csem.org/#2
[4] http://www.epncb.oma.be/
[5] http://www.intermagnet.org/index-eng.php/

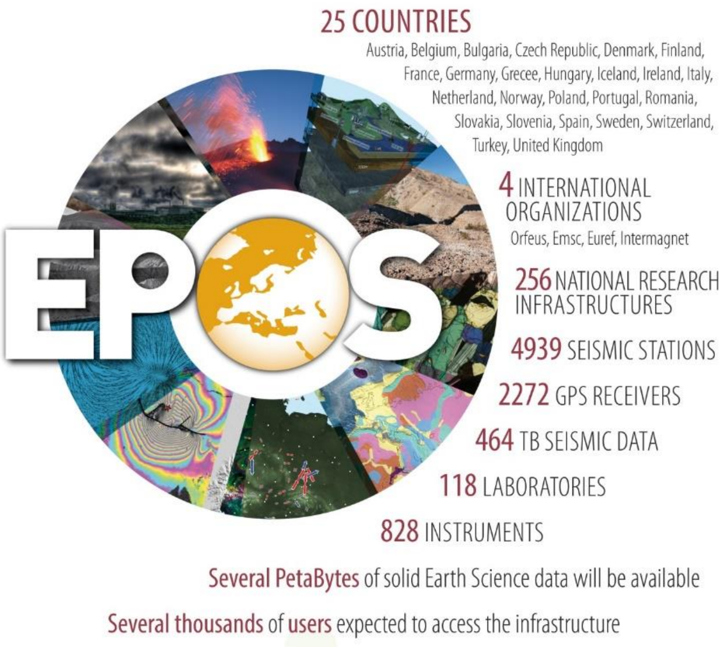


Figure 1: The EPOS implementation phase community (https://www.epos-ip.org/about/who-makes-epos/community)

2.1.2 Basic features of the EPOS project

EPOS provides services to scientists, national authorities and society at large 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 environments. 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 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 make 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.

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

The EPOS Preparatory Phase (EPOS PP) ran from 2010-2014 and it was funded under 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) started in 2014 following the completion of the Preparatory phase. Running until 2019, the EPOS IP project is a joint project of 47 partners, 6 associate partners (including the European Space Agency[6], EuroGeoSurveys[7], Global Earthquake Model[8]) and several international organizations (ORFEUS, EMSC, EUREF, INTERMAGNET) for a total of 25 countries involved. The EPOS IP project is a key step in EPOS' vision of a pan-European Earth science monitoring platform. Building on the EPOS PP project, it 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 will support these actions with the establishment of the EPOS European Research Infrastructure Consortium (EPOS-ERIC, see Section 2.1.4) agreed by the EPOS Governmental Representatives 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 Operation phase (OP).

2.1.4 EPOS Technical, Legal, Governance and Financial structure

It is foreseen that by 2018 EPOS will become a European Research Infrastructure Consortium (ERIC) and thus a legal entity (EPOS-ERIC). The ERIC legal framework will provide EPOS with a legal personality and extensive legal capacity recognised in all EU Member States and the flexibility to adapt to the specific requirements of each infrastructure.

EPOS will meet statutes and other commitments necessary for the legal establishment of the ERIC. The ERIC-mandated elements are central to the success of EPOS: a General Assembly (GA) of members will be the organization's governing body. The current Board of Governmental Representatives (BGR) will transform into the GA once the ERIC will be operational. An Executive Director, supported by its Coordination Office (ECO), will be directly responsible to the GA for all aspects of the EPOS activities. A Services Coordination Board (SCB) representing all the TCS and the ICS will inform and advice the Executive Director in formulating and executing the EPOS Annual Work Programme. Figure 2 presents a schematic illustration of the EPOS-ERIC governance structure.

The EPOS-ERIC legal seat will be hosted in Italy (Rome) at the National Institute of Geophysics and Volcanology (INGV[9]) headquarter. The ICS Central Hub (ICS-C) will be hosted in the United Kingdom (British Geological Survey - BGS[10]) and France (Bureau de Recherches Géologiques et Minières - BRGM[11]) with technical operational support from Denmark (Geological Survey of Denmark and Greenland –

[6] https://www.esa.int/ESA
[7] http://www.eurogeosurveys.org/
[8] https://www.globalquakemodel.org/
[9] http://www.ingv.it/en/
[10] http://www.bgs.ac.uk/
[11] http://www.brgm.eu/

GEUS[12]). The location of service providers in charge of data and service provision for the Thematic Core Services, will be decided during the validation phase that started on October 2017.

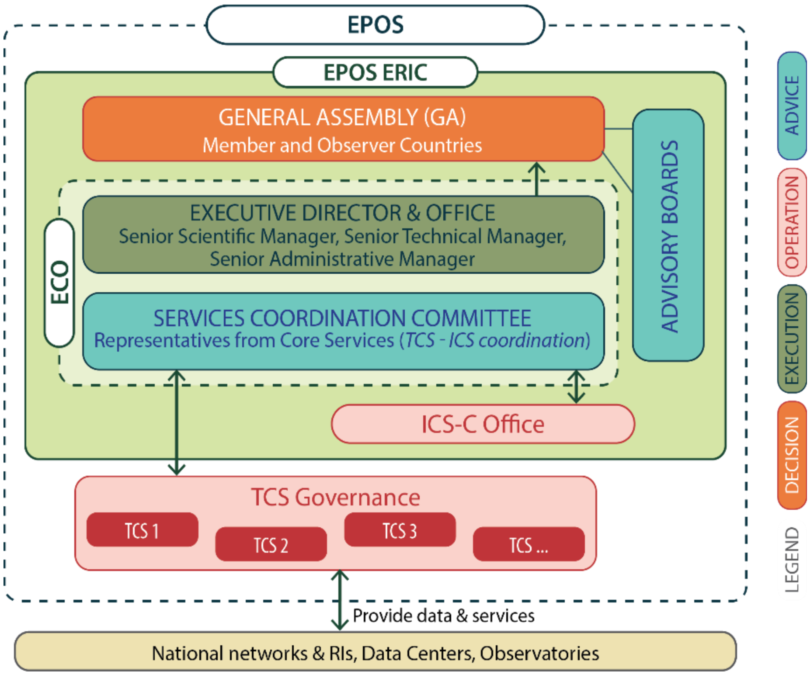


Figure 2: EPOS-ERIC Governance Structure (K. Atakan)

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 (Figure 3). The second layer represents the community developments at the European level, where Thematic Core Services (TCS) are organised. Currently 10 different TCSs are operational in EPOS with varying degree of maturity. Some the TCS communities are quite mature and were established more than a century ago, whereas others were recently formed. The 10 TCSs are:

- 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 and the uppermost 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-

[12] http://www.geus.dk/UK/Pages/default.aspx/

Throughput Computing (HTC), visualisation processing and analysis represent services that are decentralised and offered by third parties such as European Open Science Cloud (EOSC).

The governance of EPOS is based on establishing an EPOS-ERIC (ERIC: European Research Infrastructure Consortium, a legal organisational entity above the jurisdiction of the member states in EU), which will be hosted by Italy (INGV). Both the ICS-C and EPOS-ERIC hosting was decided by the Board Governmental Representatives where each member country is represented by its relevant ministry, following a competitive tender process. Currently EPOS is in the middle of its Implementation Phase (EPOS-IP) and has applied to the European Commission officially for establishing ERIC

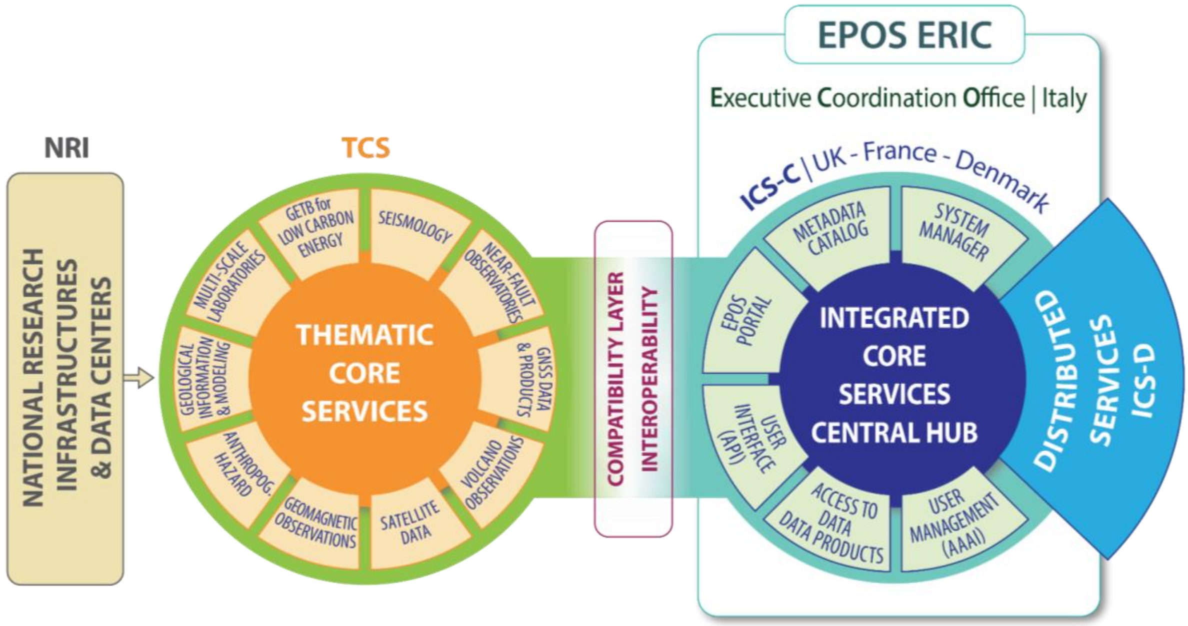


Figure 3: The EPOS functional architecture

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 to integrate 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

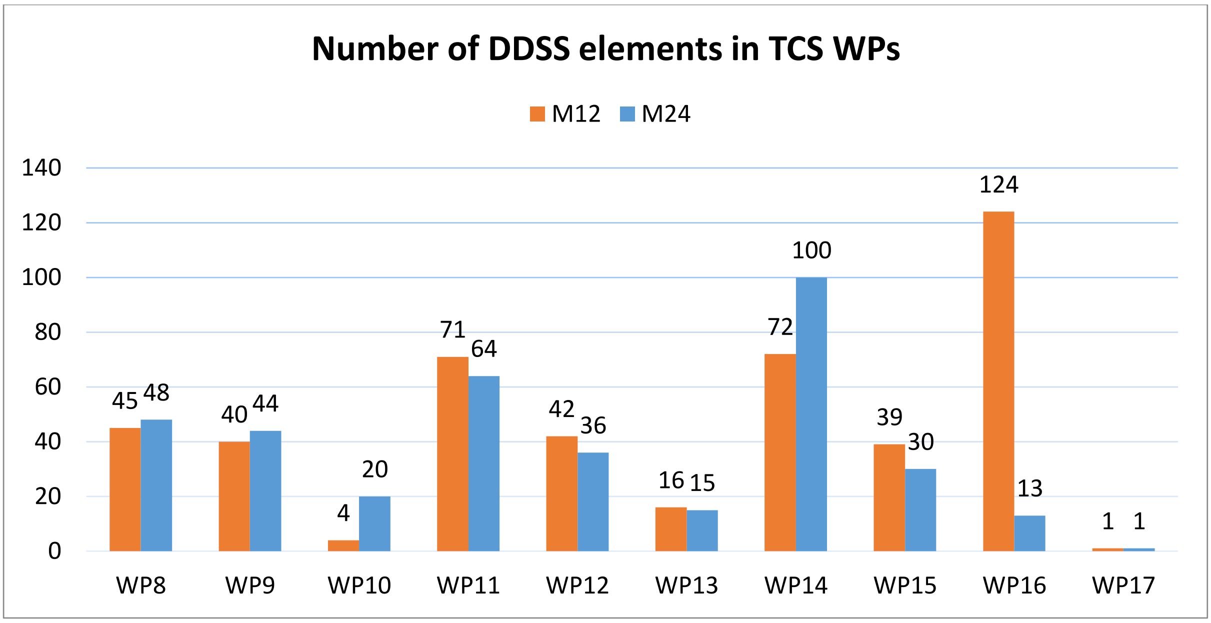
There are 10 work packages (WP8-WP17) creating the Thematic Core Services (TCS), grouped by a specific topic, namely: 1) Seismology, 2) Near Fault Observatories, 3) GNSS Data and Products, 4) Volcano Observations, 5) Satellite Data, 6) Geomagnetic Observations, 7) Anthropogenic Hazards, 8) Geological Information and Modelling, 9) Multi-scale laboratories and 10) Geo-Energy Test Beds for Low Carbon Energy. 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 had 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, shakemaps);
- 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.).

As a part of the requirements and use cases (RUC) collection from the TCS 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 372 DDSS elements in the DDSS Master Table. These DDSS elements are of different maturity and about 122 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. Figure 4 present the number of DDSS elements in EPOS TCS WPs.



Note: 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.

Figure 4: Number of DDSS elements in EPOS TCS WPs (Atakan and Michalek, 2017)

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)[13] was newly designed and currently is 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, and the new GUI reflects the basic general features requested by the TCSs. The GUI is a draft version that is not fully working, but it is properly connected and integrated to the ICS Architecture scheme.

Currently, 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 Figure 5.

[13] The GUI is currently hosted at: nodedev.bgs.ac.uk/epos/epos-gui/master/

A user can be a simple 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 Figure 5. 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.

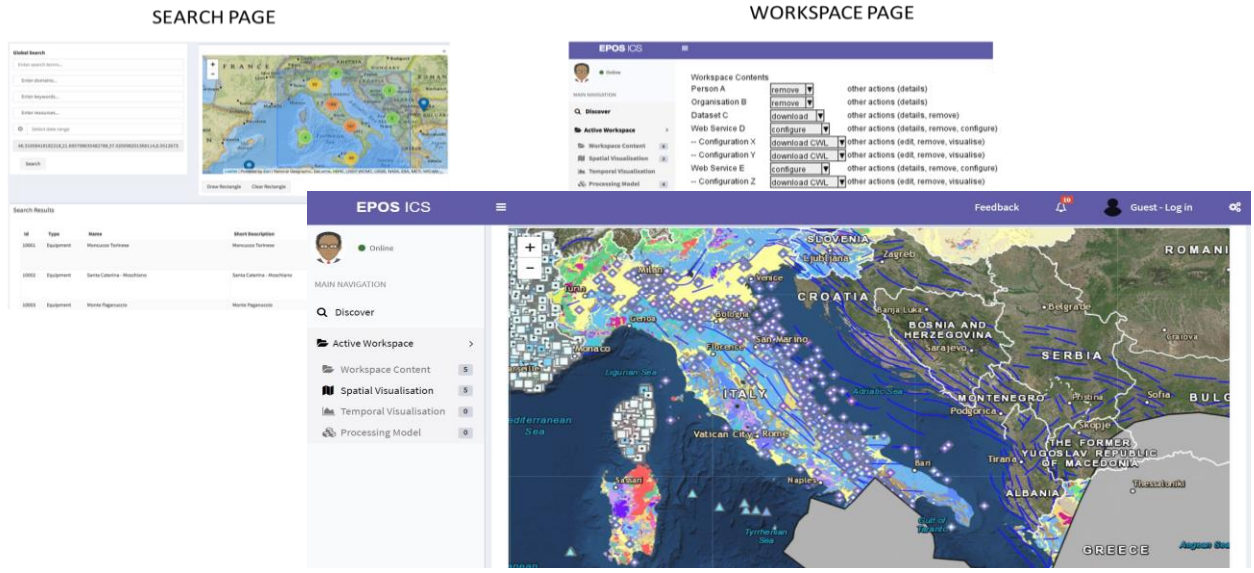


Figure 5: The EPOS ICS user interface

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. It is the heart of EPOS; it provides end-users with a view of what EPOS has to offer 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 (see Section 2.4).

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 Figure 6 and described below.

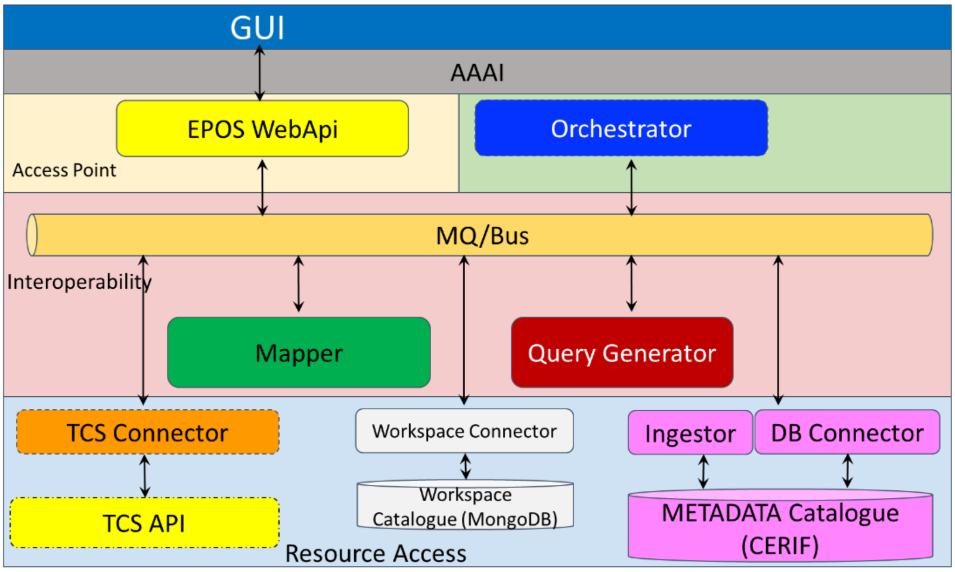


Figure 6: System architecture of the EPOS Integrated Core Services (https://www.epos-ip.org)

Metadata Catalogue

It is used for storing metadata about the various DDSS elements 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) developed for converting DDSS elements to a TCS specific metadata standard to be connected into the ICS, for example by providing 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/Bus

This component provides a system management function by orchestrating the interaction between the 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.

EPOS WebAPI

This component performs the connection between the graphical user interface (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

GUI (Graphical User Interface)

It provides the interface to the user (see 2.4) and allows them to firstly query the DDSS elements of interest, and then specify appropriate workflows to process the data.

Integrated Core Services-Central

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

- Harvesting of 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 for storage and interoperability of research information. It is a reference model for the development of Research Information Systems (CRIS). The CERIF data model is presented in Figure 7.CERIF comprises of 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, interconnections (research) and their output (results) as well as high flexibility with formal (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.

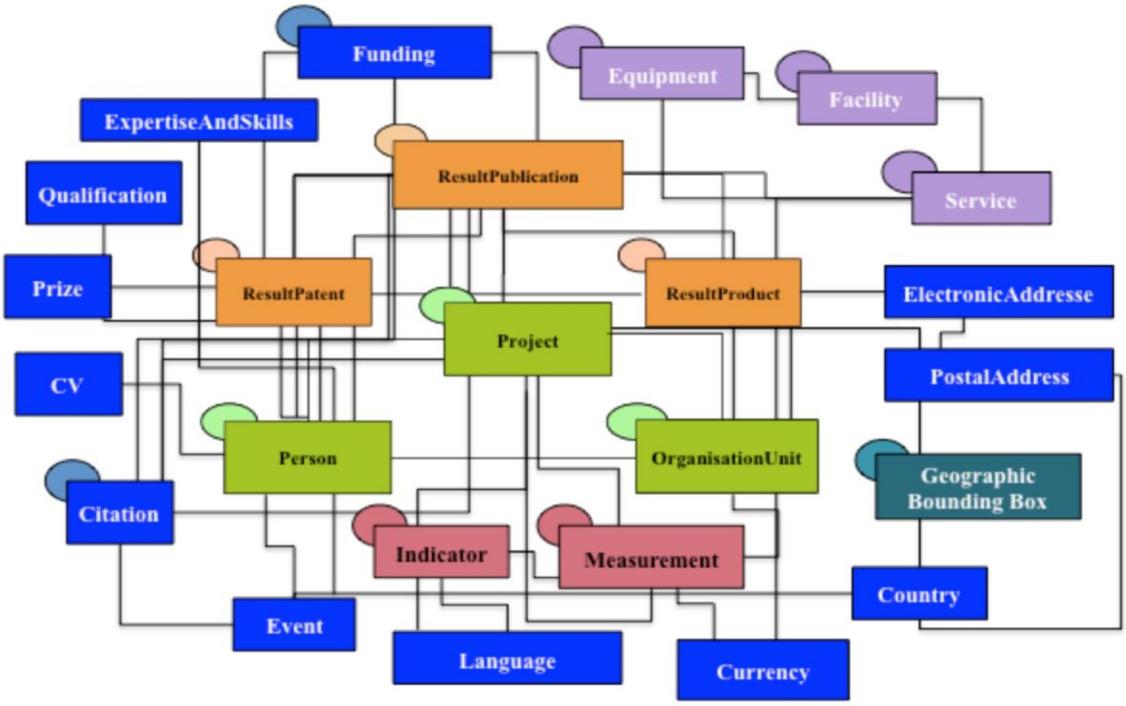


Figure 7: The CERIF Data Model (https://www.eurocris.org/cerif/main-features-cerif)

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 expressed through linking '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 (CRIS), 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 TCSs 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 the various entities shown in Figure 8. For each entity, a set of attributes has been specified.

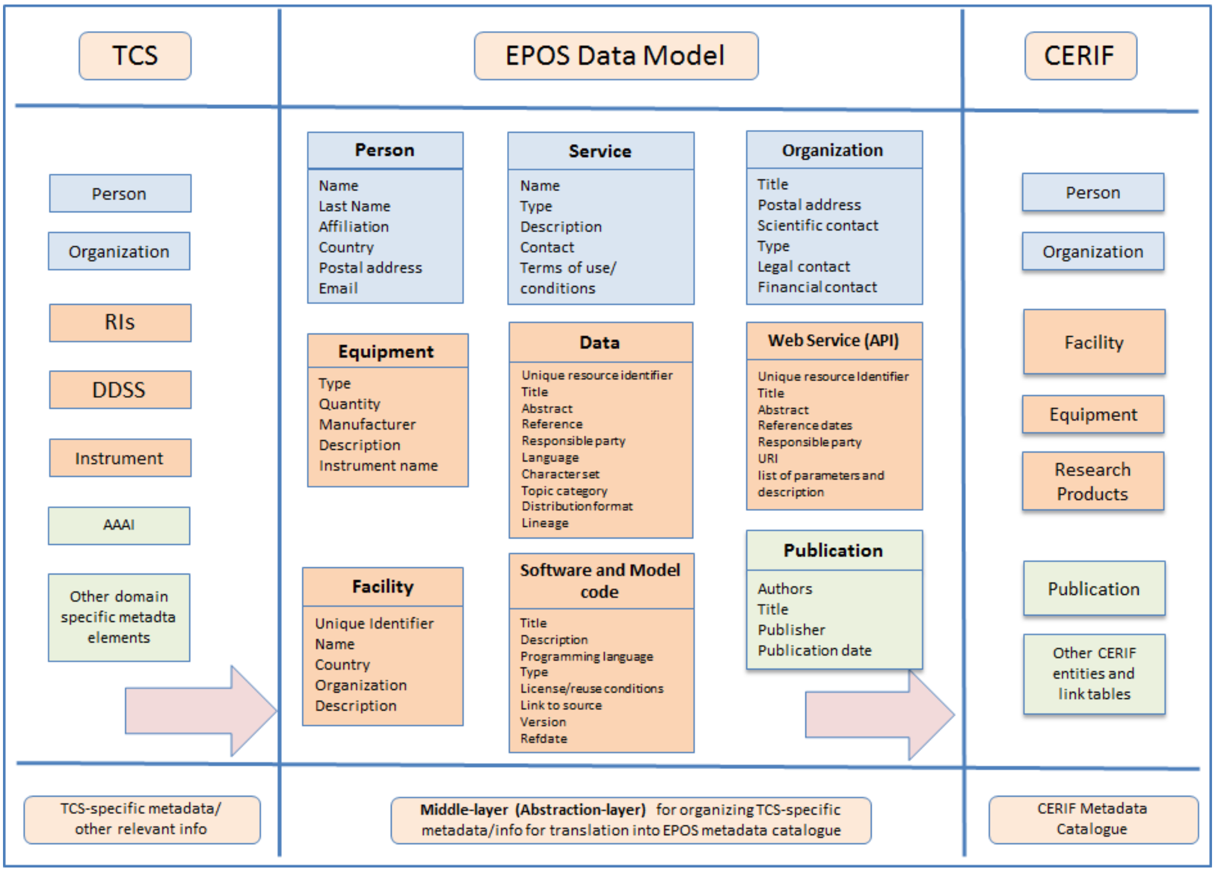


Figure 8: The elements of the EPOS metadata reference model and mapping to CERIF (K. Atakan)

2.5.4 EPOS metadata convertor and mapping to CERIF

During the EPOS-PP project, the CERIF (Common European Research Information Format) was chosen for implementation. The metadata catalogue will enable the user to perform discovery, some 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 TCSs. 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, DCAT[14], Dublin Core[15], INSPIRE[16], etc.) Different interpretation of these metadata standards further increases the level of complexity. Hence,

[14] Data Catalog Vocabulary
[15] Dublin Core Metadata Initiative (http://dublincore.org/)
[16] Infrastructure for Spatial Information in Europe (https://inspire.ec.europa.eu/)

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 (see Figure 9). 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.

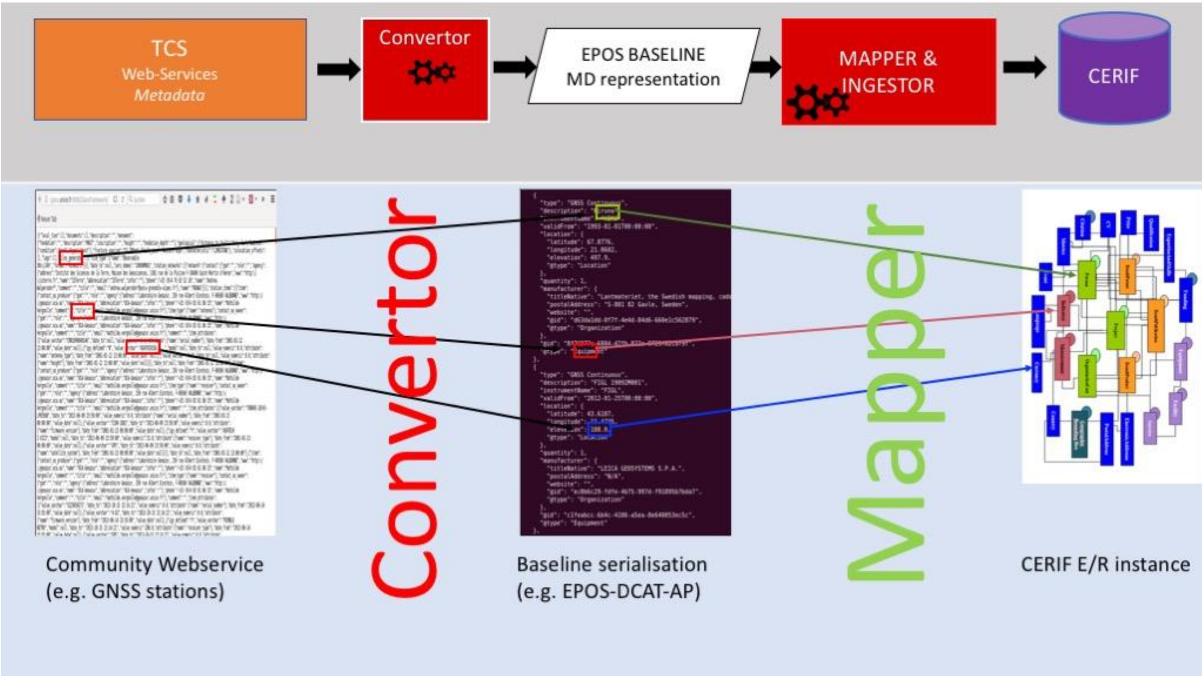


Figure 9: Schematic representation of EPOS metadata mapping to CERIF (http://wiki.epos-ip.org/index.php/TCS_Metadata_Mapping)

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, so as 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 TCSs 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[17] environment was established for providing 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 dedicated folder on GitHub. Thus, ICS and TCS could easily interact and solve mutual issues encountered within each conversion. Three EPOS-DCAT-AP entities (Person, Organisation and

[17] https://github.com/epos-eu/EPOS-DCAT-AP

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 Figure 10.

```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>
```

Figure 10: An example of an EPOS-DCAT-AP XML showing the elements and datatypes used in the schema.

2.6 Selected Thematic Core Services (TCS) review

Within the EPOS IP, ten different communities were identified and organized in as many Thematic Core Services (TCS) and Working Packages, taking into account the requirements of the different EPOS communities. The TCS, as explained before, are the community-specific integration (e.g. seismology, volcanology, geodesy, experimental laboratories, etc.) 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 list of TCS is presented in Table 1.

The different TCS have varying degrees of maturity in their development and it is not possible to deal with TCS 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.

TCs, according to their level of maturity, 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

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

2.6.1 EPOS TCS Seismology

EPOS-Seismology relies heavily on the community-driven European level institutions that are already established, namely : a) Observatories and Research Facilities for European Seismology (**ORFEUS**) that provides access to earthquake waveforms; b) the European-Mediterranean Seismological Centre (**EMSC**) that determines the principal parameters (epicentre, depth, magnitude, focal mechanisms, etc.)

of major seismic events; c) the European Facilities for Earthquake Hazard & Risk (**EFEHR**) that provides access to data, models, tools and expertise relevant for assessment of seismic hazard and risk in Europe.

EPOS TCS Seismology will improve and extend these existing services, producing a single framework which is technically, organisationally and financially integrated with the EPOS architecture. It is planned that the TCS Seismology will be operational in 2019 providing virtual access to a set of seismology data and services for users, such as scientists, engineers, public managers, citizens, scientists, covering infrastructures, data, products, hazard and information services.

Each of the existing European level institutions already has its own governance structure (or is developing it), which addresses how member organizations are represented, how they participate in the decision making processes, and how duties and responsibilities are assigned. Thus, EPOS-Seismology builds its internal organization on the three European seismological institutions (ORFEUS, EMSC, EFEHR). An EPOS-Seismology consortium is currently under development and will provide overall governance and coordination among the three organisations.

All the national member institutions of ORFEUS, EMSC and EFEHR are key contributors to EPOS and main providers of data and products to EPOS-Seismology. Here, the national infrastructures play a crucial role as providers of data and products, which are then distributed through the common European level services.

Data and services expected to be available through EPOS Seismology are the following:

- Seismic waveforms and metadata from permanent and temporary networks (including strong-motion data) and from ocean-bottom seismometers; derived parametric data (e.g. acceleration parameters for engineering) and metadata.
- Seismological products: authoritative earthquake locations and magnitudes; bulletins; earthquake catalogs (including historical); moment-tensors; shaking and damage maps; seismic source models; site response data.
- Earthquake hazard and risk data and products: tectonic fault maps and models; geotechnical, geological and site conditions inventory; tools for processing/analyzing/interpreting building/infrastructure weakness; exposure and vulnerability data and models for building/infrastructure risk assessment; hazard maps; risk maps & scenarios.
- Virtual Access to computational platform/s Massive-data mining, data-intensive processing, visualization, processing (synthetic data from 3D Earth simulations).

2.6.2 TCS Multi-scale laboratories

TCS Multi-scale laboratories include a wide range of world-class experimental laboratory infrastructures such as high pressure-temperature rock and fault mechanics, analogue modelling and paleomagnetic laboratories. The objective of the TCS Multi-scale laboratories is to develop a coherent and collaborative network of European solid earth science laboratories. It is foreseen that the TCS will offer coordination of the laboratories' network, data services, and trans-national access to experimental facilities.

At present, most data produced by the various laboratory centres and networks are available only in a graphical or sampled form in publications. The complete dataset and many data remain inaccessible and usually requires at least a direct interaction with the authors, may have a specific format preventing easy sharing, lack of traceability for intellectual property and appropriate citation and/or if not poorly preserved unfortunately.

The TCS Multi-scale laboratories will collect and harmonize available and emerging laboratory data on the properties and process controlling rock system behaviour at multiple scales, in order to generate products accessible and interoperable through services for supporting research activities into Geo-resources and Geo-storage, Geo-hazards and Earth System Evolution. The Multi-scale laboratories infrastructures are grouped into four main application areas: Analogue modelling of geologic processes, Paleomagnetism, Rock physics and high pressure-temperature laboratories and Analytical laboratories. The TCS will bring forward data products in the form of published experimental research data.

TCS Multi-scale laboratories is also working to establish access rules to the several solid Earth experimental laboratory facilities in Europe. The groups and lab facilities will be made visible to the community in the EPOS Portal, creating new opportunities for synergy, collaboration and innovation, in a framework of trans-national access rules. Through this system, researchers and research teams across Europe will have the opportunity to present project proposals and be selected to perform experiments at key EPOS Multi-scale laboratory centres. Two Transnational Access (TNA) pilot calls for research projects were open in 2018; in the second TNA call, 28 applications were submitted for access to 13 solid Earth laboratory facilities across Europe.

3 SERIES project and database review

The 4-year project SERIES (**Seismic Engineering Research Infrastructures for European Synergies**) started in 2009 and addressed the fragmentation and sub-optimal use of European research infrastructures by creating a consortium of key actors in Europe's seismic engineering research. The Consortium comprised essentially all major experimental research infrastructures in Europe in the fields of structural or geotechnical earthquake engineering, covering in a complementary way the full range of seismic testing techniques and capabilities. In addition, the Consortium included three private industrial partners with large experience and expertise in seismic applications.

The SERIES project mission was to overcome the fragmentation that characterized the European earthquake engineering community by bringing together Europe's research infrastructures in structural and geotechnical earthquake engineering into a coherent and sustainable platform of co-operation. A major objective was to create a culture of collaboration and integration. To do so, a concerned programme of networking activities and a coordinated transnational in-person access of users was established.

One part of the Networking Activities aimed at facilitating the exchange of data and data communication among research infrastructures in Europe providing both off-line access to data by means of a database and on-line access by means of telepresence implementation that allowed collaborative decision making during experimental test campaigns.

The platform of co-operation between the Research Infrastructures (RIs) comprised of (i) a corporate web-portal[18] as the central contact point for SERIES and the main reference point for RIs in earthquake engineering in Europe, during the project and afterwards. The portal contains, among others, education and dissemination material, repository of scientific knowledge (including the one generated during the SERIES) and, most importantly, access to the distributed database; (ii) a distributed database of experimental information, where the data are saved at the individual facility and a communication protocol ensures their transfer to the end user in a common language and format.

The following sections present in summary the basic features of the SERIES project and the related database. The review has been based on the available information in the SERIES portal (www.series.upatras.gr) and deliverables D2.1 "Distributed DataBase: Review of beneficiary current data format – Specifications for the common data format and Data Access Portal" and D2.3 "Preliminary version of Distributed Database and Data Access Portal". A full list of references to SERIES is provided at the end of the document.

3.1 SERIES: The Project and the Database

3.1.1 Aim of SERIES

The SERIES Project aimed at bridging the two gaps of Research and Technological Development (RTD) in experimental earthquake engineering and structural dynamics:

- between Europe and the US or Japan,
- between European countries with high seismicity but less advanced RTD infrastructures on one hand and some more technologically advanced but not so seismic Member States on the other.

The SERIES mission was supported by integrating the entire European RTD community in earthquake engineering via:

[18] www.series.upatras.gr

1. A concerted program of Networking Activities, fostering a sustainable culture of cooperation among all research infrastructures and teams active in European earthquake engineering:

- A distributed database of test results, pooling data from the beneficiary research infrastructures and others, accessible and maintained by a virtual research community after the project's end,
- Telepresence and geographically distributed concurrent testing at the research infrastructures,
- Standards, protocols and criteria for qualification of RTD infrastructures in earthquake engineering,
- Enhancement of human resources by training new users and beneficiary technical/research personnel in courses on good practices in operation and use of research infrastructures,
- Co-coordination and collaboration with national, European and international related initiatives and support to the deployment of global approaches to research in earthquake engineering;
- Dissemination to the entire European S/T community of earthquake engineering via all relevant national, European or international organisations, networks or bodies,
- Clustering and coordinated actions amongst related European and national projects,
- International Workshops and other targeted actions, to integrate the earthquake engineering community of the highly seismic regions of the Balkans and Turkey.

2. Co-ordinated Transnational Access of Users to a world class portfolio combining:

- EU's four largest earthquake Shaking Tables, each one with diverse capabilities: the TAMARIS laboratory of CEA/Saclay (FR), the EUCENTRE/TREES Lab in Pavia (IT), LNEC in Lisbon (PT) and the Bristol University Earthquake and Large Structures Laboratory (UK),
- EU's largest Reaction Wall and Pseudodynamic testing facility (ELSA) at the JRC, Ispra,
- Unique Centrifuge Test facilities at LCPC in Nantes (FR) and Cambridge University (UK).

3. Joint innovative Research toward new fundamental technologies and techniques promoting efficient and joint use of the research infrastructures, in three areas where the beneficiaries excel at world level:

- Concepts, technical requirements and prototyping for new-generation electrodynamic actuators (including coupling with hydraulic ones) for high-performance, enhanced-quality real-time testing,
- Instrumentation and sensor techniques for improved sensing and test control. Dedicated software for data collection, processing and communication, serving current needs for model calibration and interpretation of structural response. Use of data assimilation and model updating to develop virtual models of the equipment-specimen system, in combination with recent advances in control, to reduce calibration pre-tests, optimise instrumentation and improve the quality results,
- New capabilities and techniques for experimental study of soil-structure-interaction and seismic wave propagation phenomena, currently insufficiently covered by experimental research infrastructures at world level.

The work plan of the SERIES Project comprised a set of intertwined and synergistic networking, transnational access and joint research activities. One of the project's main goals was to establish a seamless and sustainable platform of co-operation between the European research infrastructures in earthquake engineering, developing synergies and complementarities between them and fostering their joint development in terms of performance and access.

3.1.2 Partners

SERIES Project was a 23-member Consortium of partners, as listed in Table 2.

Table 2: The SERIES Project Consortium

PARTNER	COUNTRY
UNIVERSITY OF PATRAS	Greece
ARISTOTLE UNIVERSITY OF THESSALONIKI	Greece
COMMISSARIAT ÉNERGIE ATOMIQUE (CEA)	France
CENTRO EUROPEO DI FORMAZIONE E RICERCA IN INGEGNERIA SISMICA (EUCENTRE)	Italy
GEODYNAMIQUE ET STRUCTURE	France
TECHNICAL UNIVERSITY OF ISTANBUL	Turkey
INSTITUTE OF EARTHQUAKE ENGINEERING AND ENGINEERING SEISMOLOGY (IZIIS) MK	FYROM
EUROPEAN LABORATORY FOR STRUCTURAL ASSESSMENT (ELSA), JOINT RESEARCH CENTRE (JRC)	European Commission (Italy)
BOGAZICI UNIVERSITY - KANDILLI OBSERVATORY AND EARTHQUAKE RESEARCH INSTITUTE (KOERI)	Turkey
INSTITUT FRANCAIS DES SCIENCES ET TECHNOLOGIES DES TRANSPORTS, DE L'AMENAGEMENT ET DES RESEAUX	France
LABORATÓRIO NACIONAL DE ENGENHARIA CIVIL (LNEC)	Portugal
MIDDLE EAST TECHNICAL UNIVERSITY (METU)	Turkey
NATIONAL TECHNICAL UNIVERSITY OF ATHENS (NTUA)	Greece
P&P LMC SRL	Italy
TECHNICAL UNIVERSITY ‘GHEORGHE ASACHI’ OF IASI RO	Romania
UNIVERSITY OF CAMBRIDGE	U.K.
UNIVERSITY OF LJUBLJANA	Slovenia
UNIVERSITA DEGLI STUDI DI NAPOLI FEDERICO II	Italy
UNIVERSITÄT KASSEL	Germany
UNIVERSITÁ DEGLI STUDI DI TRENTO	Italy
UNIVERSITY OF BRISTOL	U.K.
UNIVERSITY OF OXFORD	U.K.
VCE HOLDING GMBH (VCE)	Austria

3.2 The SERIES Distributed Database

One main component of this the SERIES platform of co-operation is the distributed database. In light of the differences in institutional practice within the SERIES consortium, a virtual databases was developed instead of a centralised repository. The virtual database provides access to multiple distributed sources of information by using a single, centralised gateway. The experience to the end user is though similar to accessing a single data depository.

The SERIES distributed database was created to store experimental information, whereby the data stay at the individual facility and a communication protocol ensures their transfer to the end user in a common language and format. It contains experimental data and all supporting documentation: data generated by the research infrastructures during the project (transnational access included), past data from the research infrastructures and from literature (converting them to the database format) and new data uploaded in the future. It provides real-time access to data generated during experimental campaigns and on-line access and interaction through telepresence and distributed testing. It is also used for storing test results, pooling data from the beneficiary research infrastructures and others and it is accessible and maintained by a virtual research community.

The creation of the distributed database aimed to improve the dissemination and use of experimental results and to foster the impact of earthquake engineering research on practice, innovation and earthquake risk mitigation. This required harmonization and unification of the European databases in earthquake engineering and the possibility of accessing, through a unified portal, the data stored at different database nodes which can dialog with the central portal using a common communication protocol. The SERIES distributed database also encouraged the automation of data processing by providing systems that treat information in a standard way and store it in a formal, common format. The distributed database enhanced the networking of European research infrastructures by improving their capacity for data exchange, sharing and access, on-line (for telepresence or distributed testing) and off-line (by uploading and/or downloading from a repository). A broad and solid base for the calibration of numerical models is achieved by enriching the database with data already available with the project beneficiaries or elsewhere.

The main characteristics of the SERIES distributed database are discussed below:

- It is **decentralised** – the database has no centralised repository to store data herein every institution in the virtual database is a different node acting as a source of information. Nodes are responsible and ultimate owner of the data they produce and decide what to share.
- **It provides a centralised access** – the access to every node is the virtual database is provided by a single Web interface (also called Data Access Portal). This single access is provides a single route to the data and enables operations (such a searching) over the data as a whole instead of accessing each node's data separately.
- The centralised access performance is enhanced by a **central site**, which hosts the Data Access Portal (DAP) and provides facilities to increase performance of the data access. Thus, the central site stores metadata of every node's data and when an end user access the DAP, it is not necessary to connect to every node in the virtual database but to use the local metadata information.
- The virtual database is **node-oriented** – the primary component is the node and not the central site. Nodes are the providers of information and they are managed and controlled by their respective institutions. The central site does not provide a method to submit data to the repository. On the contrary, it is the node that provides the necessary mechanism to store and access its data.
- The virtual database architecture is **service oriented** – this means nodes provide information by following a Simple Object Access Protocol (SOAP). SOAP is an architectural paradigm in distribute systems that focuses on organizing and utilizing distributed capabilities of heterogeneous systems under the control of different owners (MacKenzie et al., 2006). This reference model allows interoperability between different systems.

3.2.1 SERIES Distributed Database General Architecture

The SERIES distributed database general architecture is comprised of two main parts.

- The **Central Site:** Central Site holds the Virtual Database and the Data Access Portal (DAP). Physically, the SERIES central site is located, maintained and controlled by the University of Patras (Greece).

- The **Remote nodes**: The remote nodes or remote laboratories, or partner nodes, host experimental results in their local repositories. Physically, the remote nodes are located in research institutions around Europe (U.K., Italy, France, Greece, Portugal, etc).

The typical distributed architecture uses a Client-Server model. In this model, one computer is the server and provides services to the rest of the computers, known as clients. Server and clients are usually connected through a network (see the schematic representation of the distributed network architecture in Figure 11).

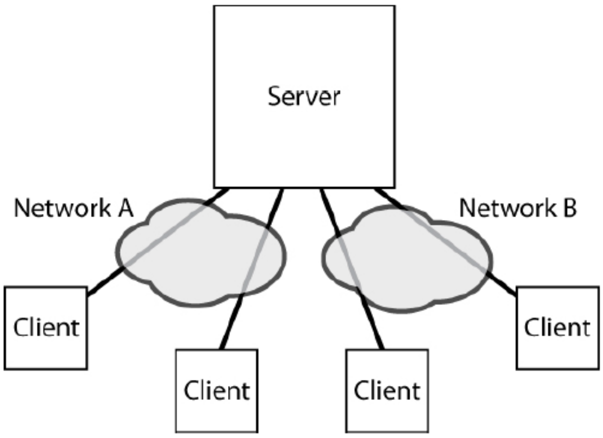


Figure 11: Typical distributed network architecture

In order to reduce the total load that falls on the server, as it has to deal with many clients, there have been implemented Client-Side (C-S) technologies to resolve this issue. By applying C-S technologies, the SERIES Central Site can perform in two ways, as a server and as a client:

- The **SERIES Central Site as a Server** provides services to the laboratories. The laboratories start the communication and they are responsible for keeping the Central Site updated. If a laboratory updates its repository but does not send the information to the SERIES Central Site, the Central Site will not know that its data is out of date (Figure 12).
- **As a Client, the SERIES Central Site** is provided with services from each laboratory. The laboratories now act as servers. The Central Site is responsible for keeping the Distributed Database updated by requesting information from all the laboratories (Figure 13).

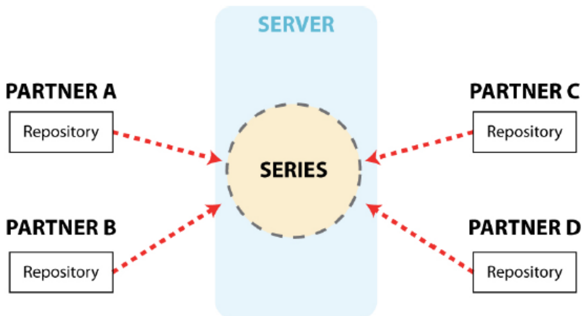


Figure 12: Partner repositories communicate with the SERIES central server

As described before, when a laboratory updates its data and does not send these information to the Central Site, the distributed database will not be updated properly. In order to resolve this updating task issue, a policy has been resolved. Updating checks could be done:

- At agreed times, for example: weekly, at midnight every day, every 2 hours, variable and depending on an updating history (more often for more active laboratories), etc.
- Every time that a user of the Central Site access to specific information of some laboratory project.
- When a laboratory explicitly requests an update.

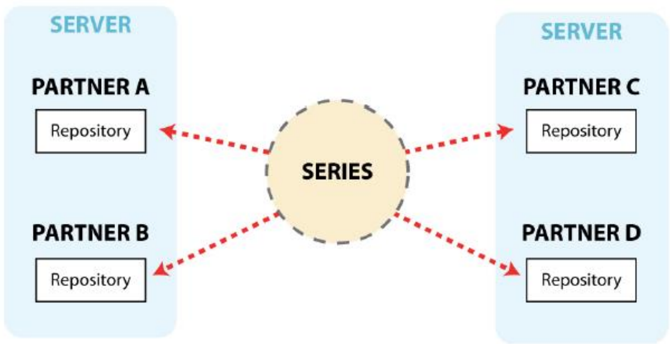


Figure 13: SERIES Central Server acting as a client, requesting information from the partner repositories

A global scheme of the SERIES architecture that was described before is presented Figure 14.

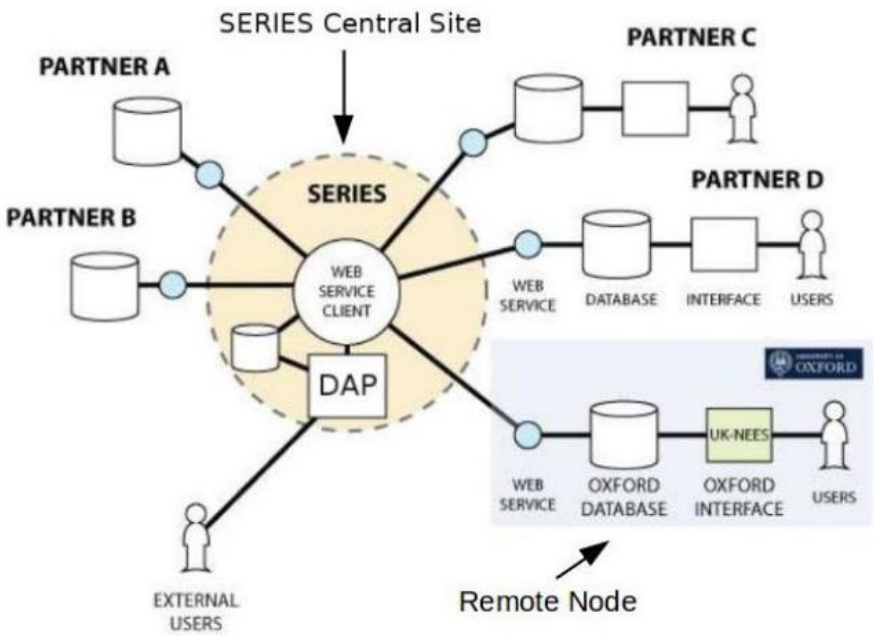


Figure 14: The SERIES Global Architecture Scheme

The rationale of this architecture is to provide a single point of access (the Central Site) to research data that are stored in Remote Nodes. The Remote Nodes are hosted, maintained and controlled by each research institution that is a partner in the SERIES network. In this scheme, each partner controls which sub-view of their repositories is made available for retrieval through the SERIES Central Site. Each SERIES partner makes available to the Central Site metadata that describe the research data that the partner wishes to publish through the SERIES Central Site. The resulting superset of a) the metadata hosted in the Central Site and b) the research data hosted in each remote node is called the "Virtual Database".

The SERIES Distributed Database consists of three main layers on the global schema, as presented in Figure 15, namely

- Central Site layer, with the Distributed Database and the DAP.
- Agreement layer, which every partner should conform to.
- Partner layer, with the various systems and repositories of each partner.

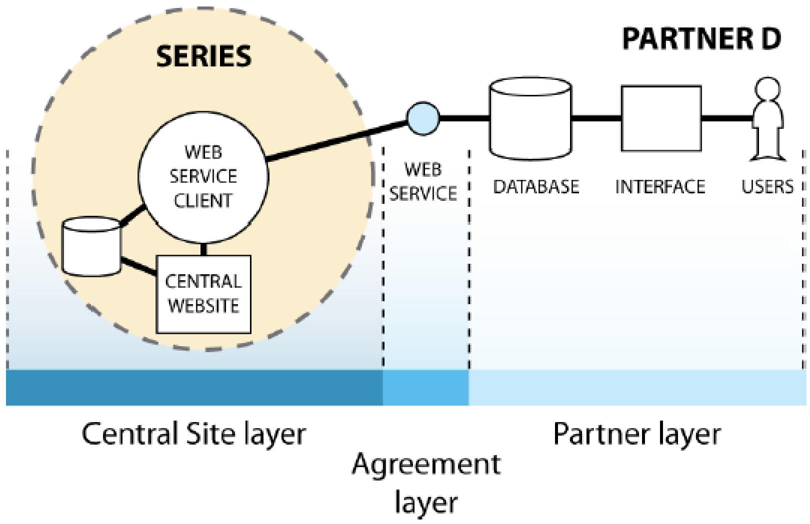


Figure 15: SERIES distributed database schema: the three main layers

3.2.2 The Central Site layer

The main purpose of the Central Site layer is to provide access to a repository implemented as a "Distributed Database". The structure can be divided in three main parts, as seen in Figure 16:

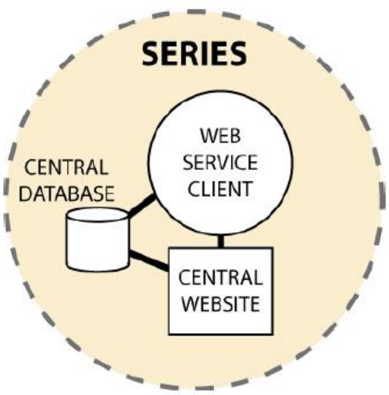


Figure 16: The SERIES central site layer structure

- Central Database, is the component that best resembles a Distributed Database. It stores information and allows first level data consultation.
- Web Service (WS) client, which connects with all partners' WS in order to obtain the data for the Distributed Database.
- Central Website or Data Access Portal (DAP), an interface to consult the Distributed Database.

In detail:

Central Database

The Central Database stores searchable metadata that describe the research data (cardinal information such as project name, project acronym etc.) that are stored in each remote node databases and metadata that describe the remote nodes. The metadata that are stored in the Central Database that describe the partners consist of partner information (Partner name, Address of partner's web service, Administrative account details, Technical contact details, Security aspects, updating parameters, Date of joining SERIES).

The Central Database can be considered to be like a cache of all partners' repositories. The Central Database should store cardinal information and all searchable data, if possible. Allowing this data storage as a cache, an external user can access the first-level information very quickly. Without cardinal information being cached by the central database, the central site would need to connect to partner repositories in order to attain this information, and the system would appear to be very sluggish to the user.

Web Service client

The Web Service in the Central Site is in charge of connecting with all partners in order to get the information that feeds the Distributed Database. It translates all the received information, coming in a common agreed format, to the data for the Central Database. As long as partners implement a Web Service consumer that complies with the WS specification, the platform and programming language that are employed are of no consequence. One of the benefits of Web Services is this freedom to choose. The remote node's Web Service Consumer provides a single point through which the remote node's data are accessible. Because of this, the implementation technology of the remote node repository is opaque to the Central Site and is of no consequence to the SERIES network.

The WS client is the core in the Central Site. The Central Database is a sleepy repository and the Central Website is user-driven. In other words, the database is just a program with no initiative, it just answers user's requests and only triggers when the user needs some information. The Central Website creates petitions only as a result of user action requests. On the other hand, the WS can suddenly wake up to connect to all partners' databases to check their status, see if it can reach a partner DB and then send an email automatically to their administrators if there is a problem to warn them of the situation.

The WS client-Central Database communication runs in one way to get location information about the partners from the Database and, in the other, to store the repository data that comes from the partners.

The Central Website-WS client communication exists to satisfy external user requests such as, for example, advance searching.

Central Website (Data Access Portal DAP)

The main purpose of the Central Site is to provide access to the "Virtual Database" through the Data Access Portal (DAP). The DAP provides a unique access point for external users to access the information of the Virtual Database. Thus, there is only a single interface to the Virtual Database, in the form of the single DAP Website. All the information that an end user receives or downloads through the DAP appears as though it comes from a single point, even if it actually comes or is downloaded from different sources. The DAP has direct access to the Central Database.

The Central Website or DAP provides a unique access point for external users to consult the Distributed Database and access its information. From a user point of view, there is one interface, one single Website, and all the information received or downloaded seem as though it came from the same place, even if it actually comes or is downloaded from different sources.

The DAP has direct access to the Central Database and is developed considering not only the access to the Distributed Database, but also possible future services. In that way, for instance, the DAP could be used not only as a Repository Portal but as a Testing Portal.

3.2.3 The Agreement layer

The agreement layer specifies the contract between the Central Site and partners that is required for successful communication with each other in a uniform, standard way. This layer is the language format that allows the partners to speak and understand one another. Every partner uses its own language format but when communicating with other partners, a common format is adopted using the same grammatical rules in order to provide communication.

The "contract" in the agreement layer specifies these grammatical rules. Partners agreeing to and complying with the contract will be understood by the Central Site. Technically speaking, the contract should define the services provided, by means of:

- operations that can be called;
- messages to be exchanged for each operation;
- data types of the attributes of the messages.

It is the responsibility of every partner to implement the operations defined in the contract and make sure that this implementation works properly according to the contract. Although a single contract is mentioned, actually there is one contract for each partner. These contracts define the services that the partner provides. The initial contract to be implemented by the partners will be referred as the "common contract" in comparison to the common contract "copy" that actually exists in every node. Partner contracts must be a mirror "copy" of the common contract. This common contract should be fully implemented by each partner, although they could also extend it with new services. This can be seen in Figure 17.

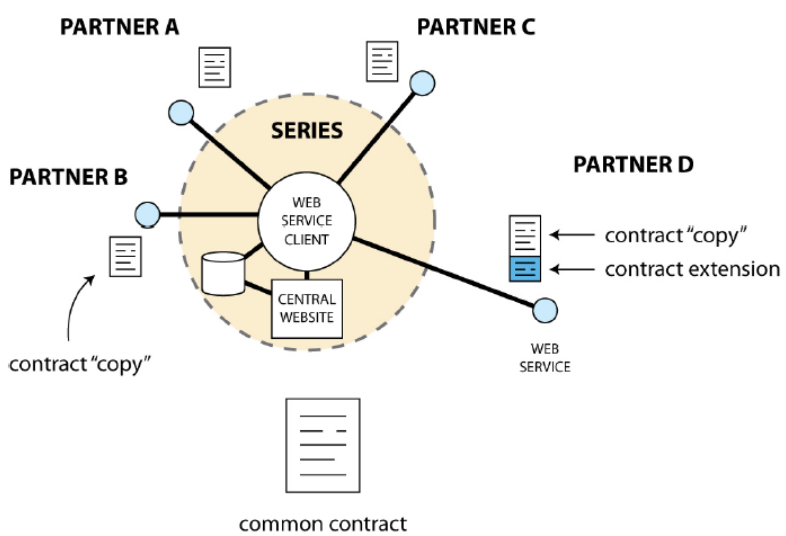


Figure 17: The SERIES agreement layer structure

3.2.4 The Partner layer

Of the three layers (i.e. the central site layer, the agreement layer and the partner layer), the partner layer has the highest level of freedom. Some SERIES' partners had already a defined structure during the project's implementation phase. It is evident that each partner may use different operating systems (Windows, Linux, etc.) and might have a different technology in its corresponding database (Oracle, MySQL, etc.). Web Services were created to deal with such heterogeneous maps. The Service Oriented Architecture (SOA) was designed to connect machines regardless of the operating system, CPU or application. The SERIES partner layer structure is schematically presented in Figure 18.

The Web Service of the Central Site is developed in a programming language where the communication between the partner and the Web service works. This provides a huge flexibility and a very wide range of possible configurations on the partner's side.

While the partner repositories can be implemented in different ways, the recommended option is to use a database. It does not matter if the information is stored in many different sources, as long as all the relevant data is correctly collected by the Web Service. Regardless of how the repository is designed and implemented, it is very important that every object in the repository has a unique ID within its scope. This ID should not vary or otherwise some mechanism is required to translate between the old and the new. The reason for this is that the Central Site may not recognize modifications and reference out-of-data Ids.

As the Client Site can also work as a service, it is suggested to store the Central Site location information (for example, IP address and other service provider data) within the local repository. This prevents one from having to hardcode the local programs or the WS consumer code. If a partner has no repository structure at all, it is suggested that a new database is implemented with SERIES functionality. For existing repositories, two main options can be considered:

- Keep the current repository and perform the "translation" for SERIES as close as possible, via the partner's Web Service.
- Develop a new repository with SERIES functionality and perform a migration from the old repository to the new one.

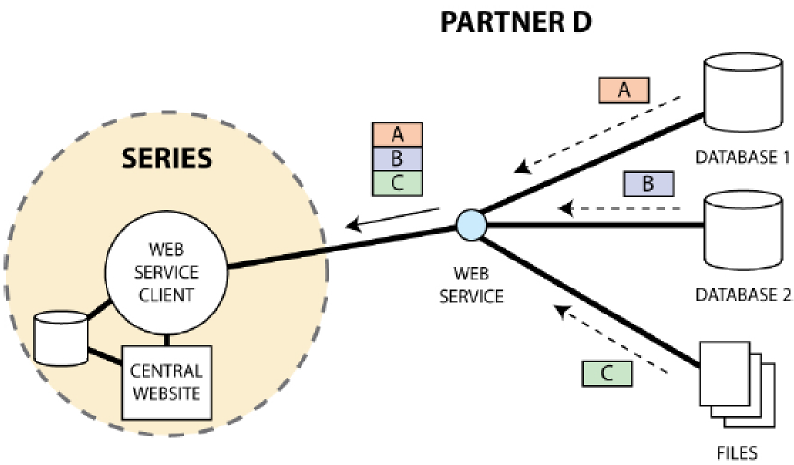


Figure 18: The SERIES partner layer structure

3.3 SERIES Exchange Data Format

In the section 3.2 of this deliverable the SERIES Distributed Database was presented.

The external user can connect to the SERIES Central Web site where various functionalities giving access to data are found. Some of these data will be stored locally to the central site, while some other data will be kept locally at the partner's repository but would be accessible by means of Web Services (WS). Since the partners do not store their own data in the same way, a common "vocabulary" must be carefully designed to access the data. In addition, the external user will obtain his data in a uniform manner, encapsulating values and context. These elements are called the Exchange Data Format (EDF).

The EDF solves the problem of partners storing their own data in their databases in an unstructured way. The EDF assumes a twofold role:

- For the laboratories that already have a database, it is the format in which their data and information is made public. Therefore, they will not have to change their databases as long as they can provide the requested data.
- The prototype database that will be implemented at all the SERIES nodes that are still missing a DB.

There is a hierarchy that was implemented in the SERIES Exchange Data Format for the creation of the prototype database. The reason of this hierarchy was to provide a general naming scheme to simplify central site searches. The hierarchy is presented in Figure 19.

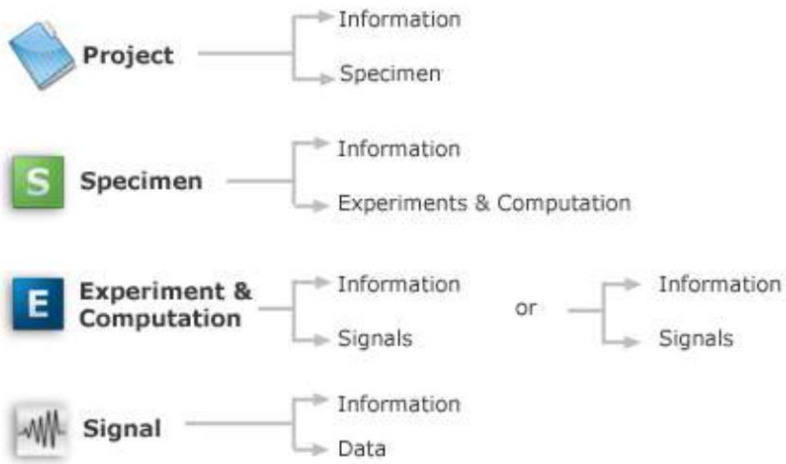


Figure 19: SERIES Exchange Data Format hierarchy

3.3.1 Project level

At the top of this hierarchy we find the 'Project' level. The 'Project' level consists of two parts, the 'Information' and the 'Specimen'. The 'Information' part is expanded in several parts as we can see in Figure 20.

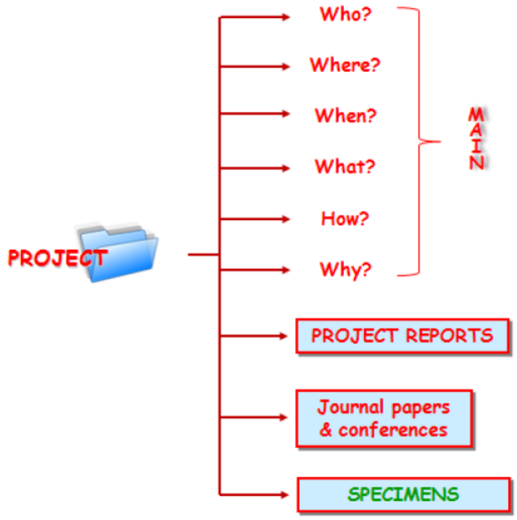


Figure 20: Expanded view of the Project->Information level of the hierarchy

Main

The first part is the 'main' part which is consisted of fields with data that contain the basic information about a project. This part contains the tables with the following data:

- *Name of the project*: includes fields for the expanded title of the project, its acronym and its sponsor. These three fields are uploaded and searchable.
- *Who is the manager of the project ('Who'):* In this field both the coordinator (s) of the project (*Principal investigator*) and the corresponding responsible(s) in the facility (*Local co-investigator*) are specified.
- *Where the project took place ('Where'):* The name and the place of the infrastructure and the facilities used during the tests are specified. Each institution should define the list of facilities they provide in order to avoid multiple naming for the *same* facility.
- *Duration of the project ('When'):* This field defines the starting date, and the ending date in case of closed projects.
- *What is the main focus of the project ('What'):* The main focus of the research project is indicated; a list of keywords to define the research areas is provided. The proposed list is quite generic but gives an idea of the ambition of the project.
- *Basic keywords that relates to the project ('Keywords')* The key actions performed in the test campaign are here listed.
- *Description of the project ('Why'):* This field contains an executive summary that describes the project, its objectives and methodologies. The role of the executive summary is not to justify the research, but to give the opportunity to use as many words as possible to qualify the experimental program and for the information searcher to clearly understand the content. It is intended that a dynamic search at the SERIES Central site will allow retrieving a project from the text of its summary.

Project reports

The second part of the 'Project' level is the 'Project report'. In this field a list of pdf documents is contained which may include:

- *preliminary* reports written before the tests to present the objectives of the project and the planning of the test campaign, and in which manner.
- *on-going* reports written during the experimental campaign,
- *final* reports with the outcomes of the projects.

Reports should describe objectively the test campaigns; interpretations of the obtained results are left to conference and journal papers. The report in PDF is considered to be accessible by users from the partner's site and the report in its original format remains private.

Journal papers & conferences

The third part of the 'Project' level is the 'Journal papers & conferences'. This field contains a list of links to journals or conferences web sites.

Specimen list

The last part of the 'Project' level is the 'Specimen list'. This field contains the specimens tested within the context of the project, specifying their main elements and materials.

3.3.2 Specimen level

At the second position of this hierarchy we find the 'Specimen' level. Multiple specimens can be tested within a single project. Two cases are possible:

- A project foresees to test more than one physical structure (a short bridge pier and a tall one; several masonry structures made by different kinds of clay; etc.).
- It is also possible to test the same structure but in different "states": for example, the structure in its original state and then with different types of retrofitting.

The 'Specimen' level is expanded in several parts as we can see in Figure 21.

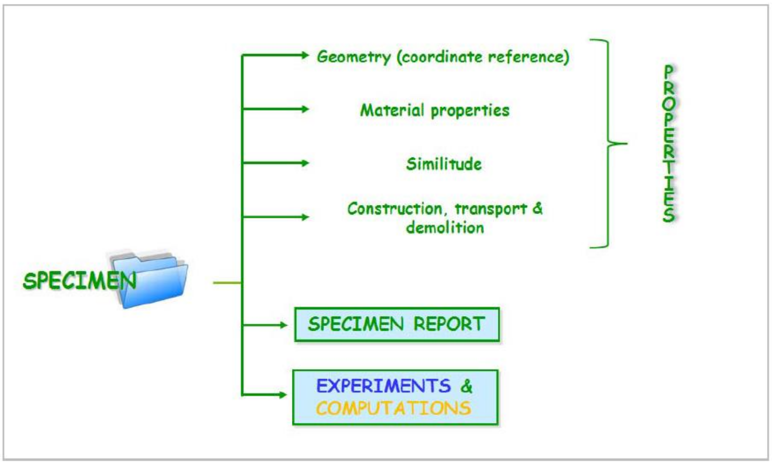


Figure 21: Expanded view of the Specimen level

Properties

The first part is the 'Properties' part which is consisted of fields with data that contain the basic information about a specimen. This part contains the tables with the following data:

- *Geometry:* In this part the maximum dimensions of the specimen are specified. A comprehensive description of the geometry and dimensions of the model scale tested in the laboratory is reported in pdf documents that provide all the necessary information for external users to adequately model the specimen; these documents show also the geometry of the facility and the location of the specimen in the facility.
- *Material properties:* In this part the materials listed in the Specimen list should be fully characterized by indicating their mechanical parameters: *nominal* values and, when they are determined, *actual* ones.
- *Similitude:* In this part a large number of experiments are performed on scaled specimens. Although scaling the geometry of a structure implies also the scaling of other quantities such as density, loads, time, etc, **only the primary scaling ratios** need to be defined in the scaling table. The other quantities are then derived according to some fundamental equations (dynamic equation, Darcy's law, etc).
- *Construction, transport & demolition:* The fourth part of the properties at the specimen level is the 'Construction, transport & demolition' part. In this part construction, transport and demolition of the specimen are usually documented by means of photos and short documentation.

Specimen report

The second part of the 'Specimen' level is the 'Specimen report'. In this part the results of all the experiments performed on a specimen are collected in a specific ***specimen report***. Preliminary, on-going and final reports may be produced, that is specified in the *status* field. In the preliminary report, the design philosophy followed for the specimen is described.

Experiment/Computation log

The last part of the 'Specimen' level is the 'Experiment/Computation log'. In this part the database collects the list of experiments and/or computations performed. The date, type, and input of the test are specified.

3.3.3 Experiment & Computation level

At the third position of this hierarchy we find the 'Experiment & Computation level' level. The 'Experiment' level is expanded in several parts as we can see in Figure 22.

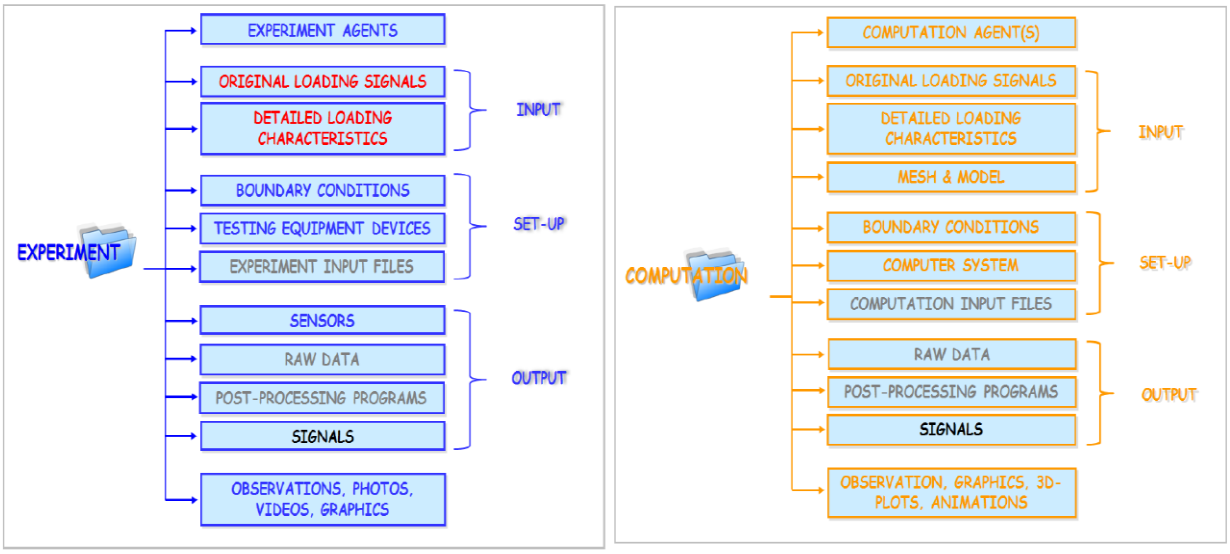


Figure 22: Expanded views of the Experiment (a) and Computation (b) levels

In case of a physical experiment, the same specimen is usually subjected to several types of experiments that differ by the type of load imposed, by the location of the loading and/or by the sensors configuration. Each of these series of tests, called an experiment, will produce different results for different scopes. Some folders, those indicated in the figure with the grey colour, are just recommended for internal use of the facility/user generating the data. For each experiment, the following fields are specified.

Experiment & Computation agents

The first part of the 'Experiment & Computation level' level is the 'Experiment agents'. In this part the personnel performing the listed experiment is indicated.

Original loading signals

This part describes the first *input* of the experiment. In this part the original time-histories and the information on the inputs used on the different experiments are collected. The original signals are preserved by giving some information on their nature: *natural* (accelerogram), *natural-normalized* (natural accelerogram normalized in the intensity), *naturalmodified* (natural accelerogram modified according to Eurocode, for example), or *generated/generated normalized* (for accelerogram generated according to Eurocode or as for cyclic tests, often performed in the preparatory phase) that can also be normalized in intensity.

Detailed loading characteristics

This part describes the second *input* of the experiment. In this part all the information that characterizes the experiment is collected.

Mesh and model

This part describes the third *input* of the experiment only for the 'Computation' level. In this part the assumptions made on the numerical model concerning the elements modeled (beam, column, etc.) and the type of loading (nodal, uniform, excitation, etc.) are reported.

Boundary conditions

The fourth part of the 'Experiment & Computation level' level is the 'Boundary conditions'. In this part are collected (a) the information on how the boundaries of the structure are connected to the testing facility and, (b), in the case of pseudo-dynamic experiments, the location of the actuators. Drawings, photos or reports may be available.

Testing equipment devices

The fifth part of the 'Experiment & Computation level' level is the 'Boundary conditions'. In this part the devices and their characteristics are listed.

Experiment input files

The sixth part of the 'Experiment & Computation level' level is the 'Experiment input files'. In this part the files that define the configuration of the testing system are collected. This folder is recommended for internal use only.

Sensors

The seventh part of the 'Experiment & Computation level' level is the 'Sensors'. In this part the following folders exist for collecting the output of tests: *sensors* (that record the signals), the *raw data* and the *treatment programs* from which *signals* are derived. Photos, videos, etc., can as well be considered as generated by sensors.

Raw data

The eighth part of the 'Experiment & Computation level' level is the 'Raw data'. In this part the direct output from the data acquisitions is contained.

Post-processing programs

The ninth part of the 'Experiment' level is the 'Post-processing programs'. In this part the treatment programs used internally for converting raw data into meaningful measurements are stored (for example Volts converted in meters following a calibration curve). Other programs used for data processing of original signals may be stored (for instance, identification of the modes and related equivalent damping).

Signals table

The tenth part of the 'Experiment & Computation level' level is the 'Signals table'. In this part signals from experiments are stored. Signals are defined by two variables: experiment and sensors.

- If a signal is issued from a direct measurement, the relationship with the sensor is obvious and should be maintained.
- If the signal results from data processing (for instance modal frequency, target displacement for a PsD algorithm, inter-storey drift, etc.), the link with the sensors is complex and cannot be expressed by means of a one-to-one relationship.

Each experiment has a sensors table, and a signals table which usually has more lines.

Observations, photos, videos and graphics

The last part of the 'Experiment & Computation level' level is the 'Observations'. In this part the observations, graphics, photos and videos of tests are stored.

- Observations are multi-media resources preserving all the information reported about a test (preliminary campaign, observations during and after the end of the test), such as scanned original handwritten notes.
- Graphics are multi-media resources summarising the results of the test in graphical form.
- The multi-media resources of Photos & Videos may be stored in different ways. For instance, photos may be either conventional or, when used for photogrammetry, time-synchronized. Video, on the other hand, can be generated by telepresence.

3.3.4 Signal level

At the last position of this hierarchy we find the 'Signal' level. The value of the signal is given by two vectors, one defining the laboratory time sequence (which can be common for blocks of signals coming from the same DAQ), and another vector storing the effective value.

3.4 Since also multi-media resources have the same time-stamp, the synchronization of all resources is possible (DAQ, but also all photos, videos, etc., coming either from measurements or telepresence).

SERIES Data Access Portal

The SERIES Virtual Database can be accessed by an external user at the Data Access Portal (DAP) at http://www.dap.series.upatras.gr/. The interface simulates that of the SERIES portal with the difference of a left column, which actually presents a breakdown list of available test results from the laboratories participating in SERIES (the presentation order is selected by the user after the Laboratory name, the Project name or the date). Information about the SERIES database and its format (EDF) as well as the user's manual is available at this level (see Figure 23).

Figure 23: SERIES Data Access Portal (DAP): the tree-based representation of published data

Navigation around all available data produced by SEIRES laboratories and flagged by them as "public", is open without restriction at any level. The information offered may be characterised as general (information about the project and contributors or detailed (when referring to *Specimen*, *Experiments*, *Computation* or *Signal level*).

Actual data of any type may be freely downloaded for all public project data. If a project is to be assess only by SERIES partners, downloading required users authentication (managed at the Data Access Portal). Nevertheless, regardless of the data type being downloaded from the database, acceptance of the Terms and Conditions displayed is a prerequisite. The statement declares that all intellectual property rights in the data, including but not limited to, copyright and database rights are vested in their respective right holders.

The DAP is further equipped with a *Search* functionality which performs a key-word-based search. The keywords forming the basis for the search are presented in categories according to the level they belong to.

4 Conclusions: a comparative assessment of the EPOS and SERIES

4.1 EPOS and SERIES platforms comparison

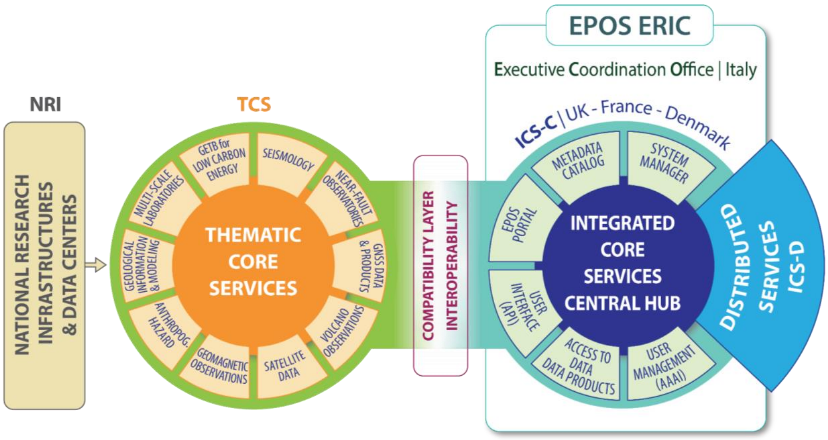
Previous Sections 3 and 4 have discussed the main characteristics of the EPOS and SERIES projects and the respective databases. Based on the review, a comparison of the two databases (common aspects, similarities and differences) is presented in the following sections.

A similarity is the principal mission of the two projects – i.e. the development of interoperable data-sharing structures for the respective scientific communities and the provision of a single tool to make integrated use of data and data products provided by different NRIs in Europe. However, the nature of the two projects is different. EPOS targets for an integration of heterogeneous data coming from several communities in solid Earth Science into a single and distributed infrastructure and facilitating access through a single online environment. On the contrary, SERIES is a domain-specific infrastructure (representing the Earthquake Engineering community) that stores data in independent, distributed sources and provides a single uniform user interface to access them.

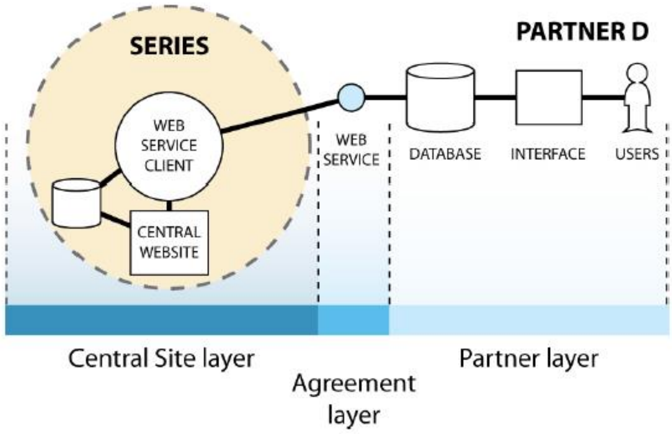
In terms of architecture and structure, similarities and differences can be observed. For the sake of convenience, the diagrams of the corresponding architectures of EPOS and SERIES, already presented in the previous sections, are repeated in Figure 24.

In both platforms' architecture, data are received from external data providers/centres. The external data centres (i.e. the data providers) share existing data with the corresponding central access point. In the case of EPOS, the external data centres are the 'National Research Infrastructures and Data Centres'. In the case of the SERIES platform, the data providers are the 'partners' which send their data to update the SERIES central database (i.e. the metadata that are cached into the Central Site database). A notable difference though, stemming from the wide range of scientific domains that relate to the EPOS project (e.g. seismology, multi-scale laboratories etc.), is that an intermediate layer exist in EPOS between the data providers and the central database, namely the Thematic Core Services. Further, even though both systems have data providers that provide data, the SERIES Central site has tight relation with its providers (i.e. the nodes) and data exchange, where as the EPOS central entity is not attached to any specific data in advance.

A notable difference between the two architectures relates to the Thematic Core Services layer in EPOS. In this regard, SERIES being a thematic community service (i.e. the Earthquake Engineering community) can be comparable to a given TCS (or an underlying community developed service) where the relevant data and products are already integrated and made available through a data gateway (the SERIES Data Access Portal). Similar TCS level (or sub-TCS) data gateways exist in EPOS - in the case of Seismology, comparable data gateways are ORFEUS/EIDA and EMSC. The EPOS Thematic Core Services (TCS) are responsible for integrating the data, metadata and services arriving from the various nodes (national and international infrastructures and data centres). The Thematic Core Services is the point in the EPOS architecture where the communities participating in the project consolidate their data so that they conform to the EPOS-DCAT-AP specification. The SERIES architecture is not designed to cover such a wide range of scientific domains as it was developed for the needs of the earthquake engineering research community. Thus, in comparison to EPOS, the SERIES architecture does not need the extra layer of the EPOS TCS. In SERIES, the partners submit their data to the Central Site in the SERIES Exchange Data Format.



(a)



(b)

Figure 24: (a) The EPOS platform architecture, (b) The SERIES platform architecture

Further, both systems are based on a Service-Oriented Architecture (SOA). EPOS requires that each type of DDSS (data, data products, services and software) that is provided by the data providers (NRIs) should be accessible by API-based web services. This happens at the TCS layer. Each scientific community exposes their metadata through web services in the TCS layer. Similarly, in SERIES, each partner exposes their data through a web service towards the Central Site.

The retrieval of the data between the two platforms' architecture has some similarities. In the EPOS architecture, in the 'Compatibility Layer' the data incoming from the TCS is converted to the 'EPOS-DCAT-AP'[19], [20], format (which is an extension of the DCAT-AP), so that the EPOS ICS can retrieve data from the TCS. Then, the ICS converts the incoming data from the EPOS-DCAT-AP to the CERIF format. In SERIES, the data that is sent to the Central Site must conform to the specification of the Exchange Data Format (also called SERIES Common Format).

Although the architecture of the compatibility layer (i.e. machine-to-machine interactions through relevant APIs between distributed web-services and a centralised hub) is similar in both EPOS and SERIES platform, the retrieval of the data between the two architectures contains some important differences. For the EPOS architecture, the compatibility layer between the Thematic Core Services and the EPOS-ICS performs the mapping between the EPOS-DCAT-AP format and the CERIF format (see

[19] Previously the format called 'EPOS Baseline' was used in this layer. EPOS Baseline is replaced by EPOS-DCAT-AP

[20] An example is provided in https://github.com/epos-eu/EPOS-DCAT-AP/blob/master/examples/EPOS-DCAT-AP_example.xml

Figure 25). On the other hand, SERIES Central Site receives data via web services in the Exchange Data format (EDF).

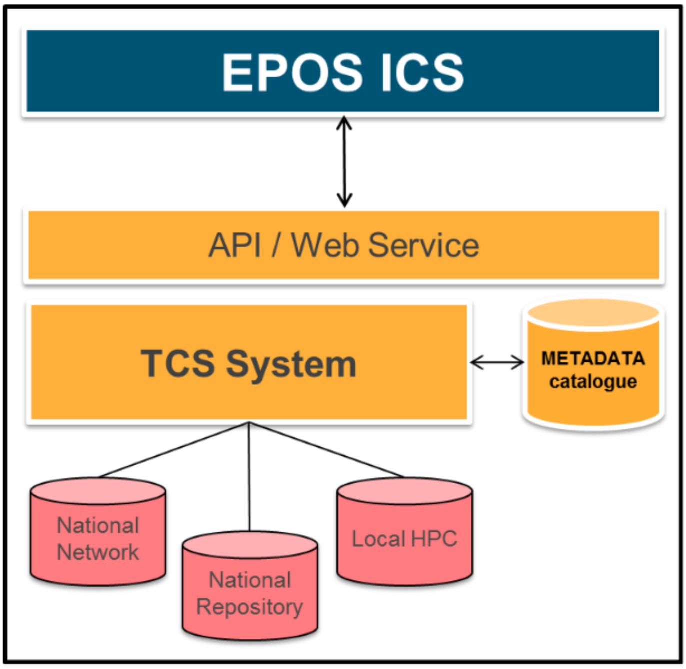


Figure 25: Generic TCS architecture. The EPOS ICS accesses the data through API-based web services that are provided by each TCS system

With regards to the data models, a direct comparison is not straightforward due to the differences and specificities in the nature and architecture of each platform. The EPOS metadata model (EPOS-DCAT-AP) follows a different approach that the SERIES Exchange Data Format (EDF). Figure 26 shows the entities that are represented by the EPOS metadata model, while Figure 27 represents the organization of the entities in the SERIES EDF.

SERIES is a domain-specific infrastructure that stores data in independent, distributed sources and provides a single uniform user interface to access them. Any new node that is added in the system, does not result in any change in the central site. The SERIES EDF follows a hierarchical organization with four level entities: Project, Specimen, Experiment & Computation and Signal. On the other hand, EPOS is a “hyper-data” provider for multi-disciplinary data. The data provided are “heterogeneous”, having one format per thematic domain (i.e. per TCS). In the EPOS-TCS data model, the entities are Person, Organization, RIs, DDSS, Instrument, AAAI, Other.

The two models are comparable in a syntactic level. Although the organization of the elements varies, it should be possible to establish a correspondence from the SERIES EDF towards the EPOS-DCAT-AP without significant loss of information, for example via a convertor. In the semantic level this might not be so straightforward and in fact there might be a mismatch between the earthquake engineering community and the communities covered by EPOS in what is represented with terms such as ‘data-point’ or ‘signal’. In addition, there may be a need for harmonizing the domain specific vocabularies and onthologies.

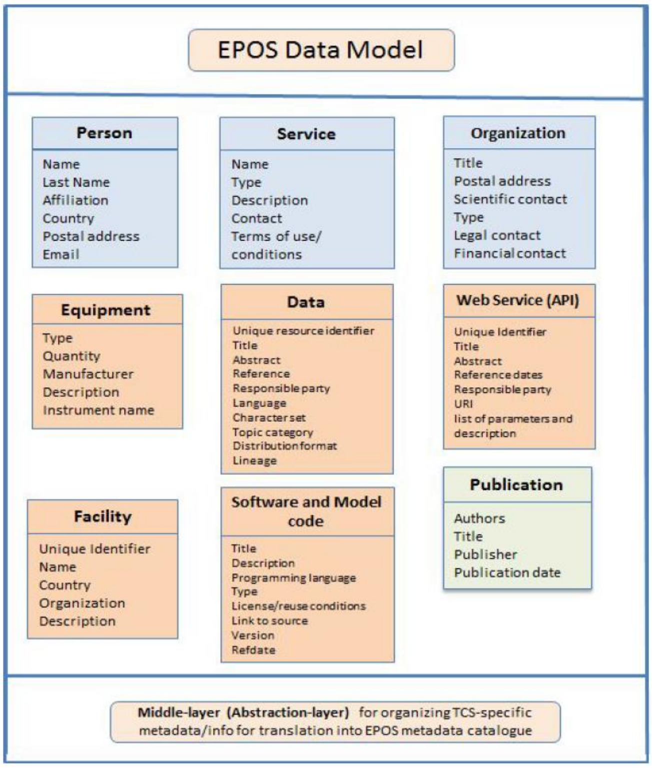


Figure 26: The EPOS metadata model

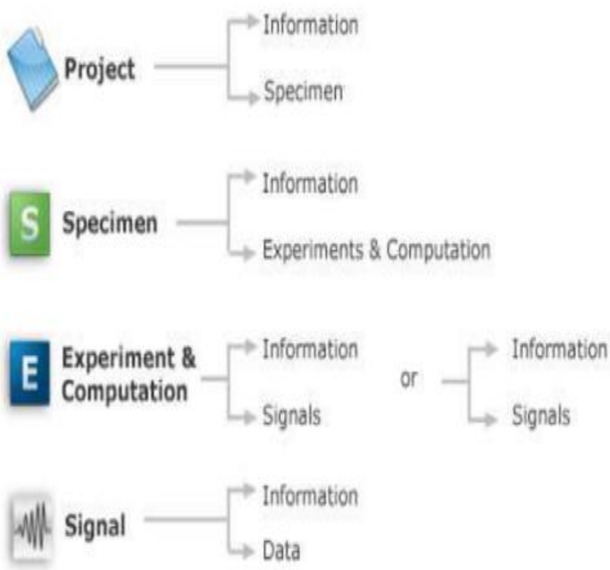


Figure 27: The SERIES metadata model

4.2 The way ahead towards the road map for integration of EPOS and SERIES data and access services

There is no single road map to be proposed for the integration of databanks and access services from the EPOS and SERIES platforms, supporting data and service-sharing between the earthquake engineering and seismology research communities so as to facilitate the interoperation of the two communities. But, at this stage, an obvious approach is to consider the SERIES from the point of view of the EPOS architecture, using the existing platforms for data sharing because the seismology community is already integrated in EPOS as a TCS (Thematic Core Services). Such an approach would also provide SERIES additional possibilities to interact with the other solid-earth science communities in EPOS and support data and services exchange with them. In terms of the EPOS architecture, the earthquake engineering community could be regarded as a new EPOS community, i.e. a TCS. The SERIES Data Access Portal (SERIES DAP) could serve as the domain specific data gateway in the same way as most of EPOS TCS have (or are developing) domain specific data portals.

The SERIES data model (i.e. SERIES EDF) can be viewed as one of several domain specific data models, concerning the scientific community of earthquake engineering. If one follows this line of thought, SERIES could provide metadata to EPOS by implementing a convertor from SERIES EDF to EPOS-DCAT-AP at the level of the SERIES Central Site. No major redesign of the EPOS or the SERIES architecture would be necessary to achieve this conversion, apart from consideration of semantic issues as mentioned in the previous section. However, development would be required in both sides: SERIES would need to create some Web Services in the Central Site to provide data and EPOS to create the components that will call the services and deliver data to the data consumers.

The success of the above mentioned possible approach is bound to the support and features that EPOS project will provide to the different communities participating as TCS when the platform will be fully developed.

Other opportunities for data integration between the two communities might already exist. For example, SERIES could use data services of EPOS as a client rather than as a server, while looking for direct interoperability opportunities with some of relevant TCS (in particular TCS Seismology).

Further details of the roadmap for integration of the databanks and access services from the earthquake engineering and seismology, i.e. SERIES and EPOS projects, respectively will be elaborated in a subsequent deliverable (D6.5 "Roadmap for the integration of data banks and access services from the earthquake engineering [SERIES] and seismology [EPOS] research infrastructure") foreseen in April 2019. Deliverable D6.4 will propose a road map on the basis of the review presented in this document, the identified research needs in each discipline and in cross-discipline issues, the latest developments in the projects and the requirements of data users beyond the research infrastructures themselves.

References

Bailo D., Jeffery K.G. (2014) EPOS: a novel use of CERIF for data-intensive science. Procedia Computer Science 33. DOI 10.1016/j.procs.2014.06.002

Bailo D., Jeffery K.G., Spinuso A., Fiameni G. (2015) Interoperability oriented architecture: the approach of EPOS for solid Earth e-Infrastructures. *2015 IEEE 11th International Conference on eScience*. DOI 10.1109/eScience.2015.22

Bailo D., Ulbricht D., Nayembil M.L., Trani L., Spinuso A., Jeffery K.G. (2016) Mapping solid earth Data and Research Infrastructures to CERIF. *13th International Conference on Current Research Information Systems*, CRIS2016, Scotland (published in Procedia Computer Science 106 (2017))

Bosi A., Taucer F., Pegon P. (2012) Distributed data bases, data communication and access, distributed testing and PsD transnational access test campaigns within the SEIRES Project. *15th World Conference on Earthquake Engineering*, Lisbon.

Bosi A., Kotinas I., Lamata Martinez I., Bousias S., Chazelas J. L., Dietz M., Hasan R., Madabhushi G., Prota A., Blakeborough A., Pegon P. (2015). The SERIES Virtual Database: Exchange Data Format and Local/Central Databases. *Experimental Research in Earthquake Engineering*. Springer International DOI 10.1007/978-3-319-10136-1_3

Common European Research Information Format (CERIF), maintained by the EUROcris organisation http://www.eurocris.org

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.

Fardis M.N. (2012) Seismic Engineering Research Infrastructures for European Synergies (SERIES*). 15th World Conference on Earthquake Engineering*, Lisbon.

Ioannidis I, Fidas C., Avouris N., Bousias S., Lamata Martinez I., Pegon P., Bosi A. (2011) SERIES Deliverable D2.3. Preliminary version of Distributed Database and Data Access Portal. [http://www.series.upatras.gr/sites/default/files/Deliverable_D2.5._public.pdf]

Ioannidis I, Lamata Martinez I., Pegon P., Fidas C., Bousias S., Avouris N. (2012) SERIES Deliverable D2.5. 2nd version of Distributed Database and of Data Access portal including user manual, documentation and guidelines. [http://www.series.upatras.gr/sites/default/files/Deliverable_D2.3.pdf]

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.

Lamata Martinez I., Ioannidis I., Fidas C., Willians M. S., Pegon P. (2015). The SERIES Virtual Database: Architecture and Implementation. *Experimental Research in Earthquake Engineering*. Springer International DOI 10.1007/978-3-319-10136-1_3

Lamata Martinez I., Pegon P., Bosi A., Dietz M., Bousias S., Fidas C., Wenzel H (2010) SERIES deliverable D2.1: Distributed DataBase: Review of beneficiary current data formats - Specifications for the common data format and Data Access Portal [http://www.series.upatras.gr/sites/default/files/pri/drafts/Deliverable2_1_0.pdf. Accessed April 2018]

MacKenzie C.M., Laskey K., McCabe F., Brown P.F., Metz R. (2006) OASIS reference model for service oriented architecture 1.0 [http://docs.oasis-open.-org/soa-rm/v1.0/soa-rm.pdf]

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

SOAP (2013) Simple Project Access Protocol. http://www.w3.org/TR/soap. Accessed April 2018.

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 Catalogue 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