

Rapporti tecnici INGV

**Integration of heterogeneous data,
software and services in Solid Earth
Sciences: the EPOS system design and
roadmap for the building of Integrated
Core Services**

393



Direttore Responsabile

Silvia MATTONI

Editorial Board

Luigi CUCCI - Editor in Chief (INGV-RM1)

Raffaele AZZARO (INGV-CT)

Mario CASTELLANO (INGV-NA)

Viviana CASTELLI (INGV-BO)

Rosa Anna CORSARO (INGV-CT)

Mauro DI VITO (INGV-NA)

Marcello LIOTTA (INGV-PA)

Mario MATTIA (INGV-CT)

Milena MORETTI (INGV-CNT)

Nicola PAGLIUCA (INGV-RM1)

Umberto SCIACCA (INGV-RM2)

Alessandro SETTIMI

Salvatore STRAMONDO (INGV-CNT)

Andrea TERTULLIANI (INGV-RM1)

Aldo WINKLER (INGV-RM2)

Segreteria di Redazione

Francesca Di Stefano - Referente

Rossella Celi

Tel. +39 06 51860068

redazionecen@ingv.it

in collaborazione con:

Barbara Angioni (RM1)

REGISTRAZIONE AL TRIBUNALE DI ROMA N.173 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale di Geofisica e Vulcanologia

Rappresentante legale: Carlo DOGLIONI

Sede: Via di Vigna Murata, 605 | Roma



Rapporti tecnici INGV

INTEGRATION OF HETEROGENEOUS DATA, SOFTWARE AND SERVICES IN SOLID EARTH SCIENCES: THE EPOS SYSTEM DESIGN AND ROADMAP FOR THE BUILDING OF INTEGRATED CORE SERVICES

Daniele Bailo¹, Rossana Paciello¹, Riccardo Rabisoni¹, Manuela Sbarra², Valerio Vinciarelli¹

¹INGV (Istituto Nazionale di Geofisica e Vulcanologia, Sezione Sismologia e Tettonofisica)

²INGV (Istituto Nazionale di Geofisica e Vulcanologia, Amministrazione Centrale - Centro Servizi Informatici)

393

How to cite: Bailo D. et al., (2018). Integration of heterogeneous data, software and services in Solid Earth Sciences: the EPOS system design and roadmap for the building of Integrated Core Services. Rapp. Tec. INGV, 393: 1-22.

Index

Introduction.....	7
1. EPOS Architecture.....	7
1.1 EPOS ICS architecture.....	9
1.2 EPOS TCS architecture.....	10
1.3 ICS-TCS integration.....	12
2. Development and interaction principles.....	13
3. EPOS Web API implementation.....	14
4. Conclusion.....	16
References.....	16
Appendix 1 - Acronyms.....	18

Introduction

Understanding Earth systemic processes is critically important to support a safe and sustainable modern society. Volcanic eruptions, earthquakes, floods, landslides, tsunamis, weather, and global climate change are all Earth phenomena impacting citizens life and the way governments and nations are organized and managed.

Accessing data and information describing such phenomena is not trivial due to the a) scattered nature of infrastructures providing access to the data and b) heterogeneity of data, protocols and methods used to obtain, analyse and process data [Bailo D. and Jeffery K.G., 2014].

In this context, EPOS¹ (European Plate Observing System) intend to provide a substantial contribution by facilitating the process of obtaining and integrating data, data products, services and software (DDSS) for Solid Earth science generated and provided by monitoring networks, observing systems and facilities belonging to European countries. EPOS will bring together Earth scientists, national research infrastructures, ICT (Information & Communication Technology) experts, decision makers, and general public to develop new concepts and tools for accurate, durable, and sustainable answers to societal questions concerning geo-hazards and those geodynamic phenomena (including geo-resources) relevant to the environment and human welfare.

DDSS provided by more than 200 Research Infrastructures will be integrated under a central, single but distributed Infrastructure – the so called Integrated Core Services – by establishing an e-platform where access, use and reuse of heterogeneous data (in terms of format, metadata and accessibility) and associated software, services and resources is easy and transparent. EPOS will use new e-science opportunities to create a pan-European infrastructure for Solid Earth science that will support a safe, innovative and sustainable society.

EPOS project has been coordinated since the beginning by INGV. The EPOS Implementation Phase (IP) started in 2014 following the completion of the Preparatory Phase (2010-2014). Running until 2019, EPOS-IP it has two major milestones. Firstly, the Implementation Phase project² which is funded under the European Commission's Horizon 2020 programme and which runs from 2015 to 2019. Secondly, the establishment of the EPOS European Research Infrastructure Consortium (ERIC) which will see EPOS becoming a legal entity³.

Key objectives of the Implementation Phase project are: implementing Thematic Core Services (TCS), the domain-specific service hubs for coordinating and harmonizing national resources/plans with the European dimension of EPOS; building the Integrated Core Services (ICS) to provide a novel research platform to different stakeholders; designing the access to distributed computational resources (ICS-D); ensuring sustainability and governance of TCS and EPOS-ERIC.

The research infrastructures (RIs) that EPOS is coordinating include: i) distributed geophysical observing systems (seismological and geodetic networks); ii) local observatories (including geomagnetic, near-fault and volcano observatories); iii) analytical and experimental laboratories; iv) integrated satellite data and geological information services; v) new services for natural and anthropogenic hazards; vi) access to geo-energy test beds.

In this context, the purpose of this report is to provide a description of the EPOS technical infrastructure as a whole, and introduce the reader to the technical details and procedures required to achieve the integration between data providers and EPOS main platform.

Readers can find an explanation of the several acronyms in the appendix at the end of this work.

1. EPOS Architecture

The EPOS architecture has been designed to organize and manage the interactions among different EPOS actors and assets [Jeffery K. and Bailo D., 2014]. To make it possible for the EPOS enterprise to work as a single, but distributed, sustainable research infrastructure, its architecture takes into account technical, governance, legal and financial issues. In this work we focus on technical aspects. The technical infrastructure is formed by three complementary layers (Figure 1):

1 <https://www.epos-ip.org>

2 <https://www.epos-ip.org/node/147>

3 <https://www.epos-ip.org/node/240>

1. The **National Research Infrastructures (NRIs)** contribute to EPOS while being owned and managed at a national level and represent the basic EPOS data providers of the DDSS. These require significant economic resources, both in terms of construction and yearly operational costs, which are typically covered by national investments that must continue during EPOS implementation, construction and operation. Data providers in this layer are independent national institutions or organizations which have their own technical solutions that may (or may not) follow international standards in providing data and data products to the community.
2. The **Thematic Core Services (TCS)** enable integration across specific scientific communities. They represent a technical and governance framework where data and services are provided and where each community discusses its implementation and sustainability strategies as well as legal and ethical issues. This is the (European) Community Layer where community standards are applied to DDSS that are relevant to the specific thematic area of concern.
3. The **Integrated Core Services (ICS)** represent the e-infrastructure that will allow access to multidisciplinary resources provided by the NRIs and TCS. These will include data and data products, as well as synthetic data from simulations, processing, and visualization tools. The ICS is composed of the ICS-Central Hub (ICS-C) and distributed computational resources including also processing and visualization services (ICS-D). ICS is the place where integration occurs. A couple of words to describe ICS-D (ICS-C is detailed in the following paragraph): ICS-D include services from external computing facilities (Figure 1). These will include HPC (High Performance Computing) machines for modelling and simulation according to the requirements of the Computational Earth Science community, and HTC (High Throughput Computing) clusters for data intensive applications such as data mining. The data workflow will be managed by EPOS ICS-C in order to provide the end user with appropriate computational services, even though actual computations will be provided by ICS-D. Additional ICS-D services will provide visualization and processing capabilities. ICS-C will have to develop provisions for communicating with these external services in a seamless manner. An example candidate ICS-D is the VERCE platform, that gained consensus and usage in the seismological community [Atkinson M., et al., 2015].

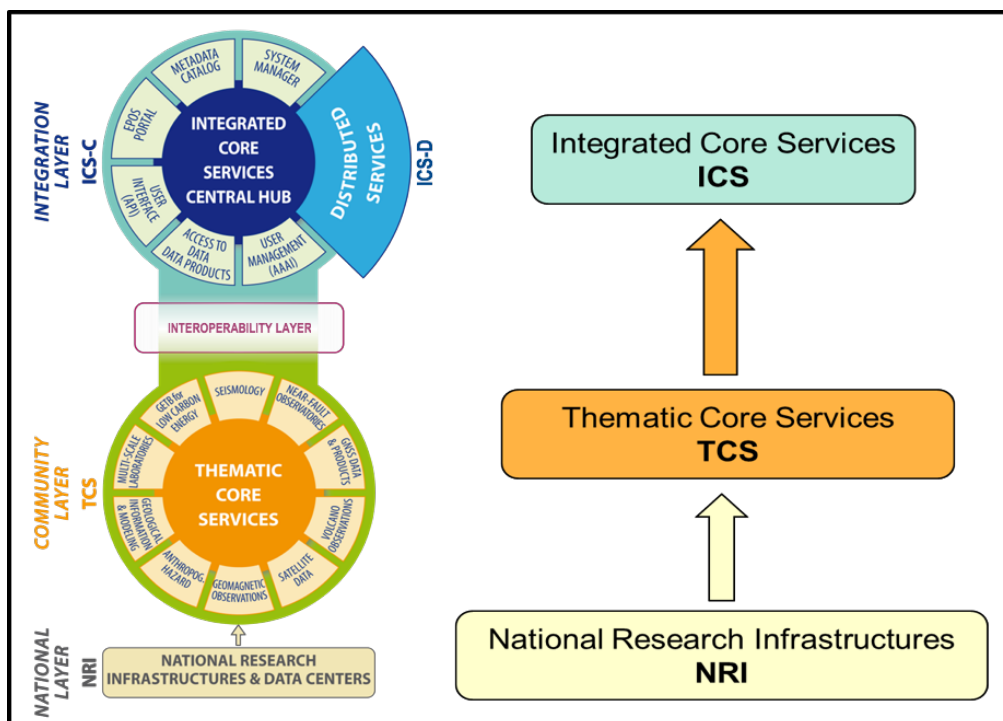


Figure 1. EPOS technical architecture. The diagram shows the three layers in which the EPOS components (institutions and services) have been organized: National Layer (represented by National Research Infrastructures), Community Layer (i.e. the Thematic Core Services), Integration Layer (Integrated Core Services) including also an Interoperability Layer.

The main concept is that the EPOS TCS data and services are provided to the ICS (see Figure 1) by means of a communication layer called the *interoperability layer*. This layer contains all the technology to integrate data, data products, services and software (DDSS) from thematic communities into the single integrated environment of the Integrated Core Services (ICS). The ICS represents the *core* of the whole e-infrastructure and those responsible for its implementation will provide the specification of the *interoperability layer*. The ICS is conceptually a single, centralized facility but in practice is likely to be replicated (for resilience and performance) and localized for particular natural language groupings or legal jurisdictions.

1.1 EPOS ICS architecture

The ICS technical architecture is made up of several, modular, interoperable building blocks (Figure 2). The various components of the ICS are now explained in more detail.

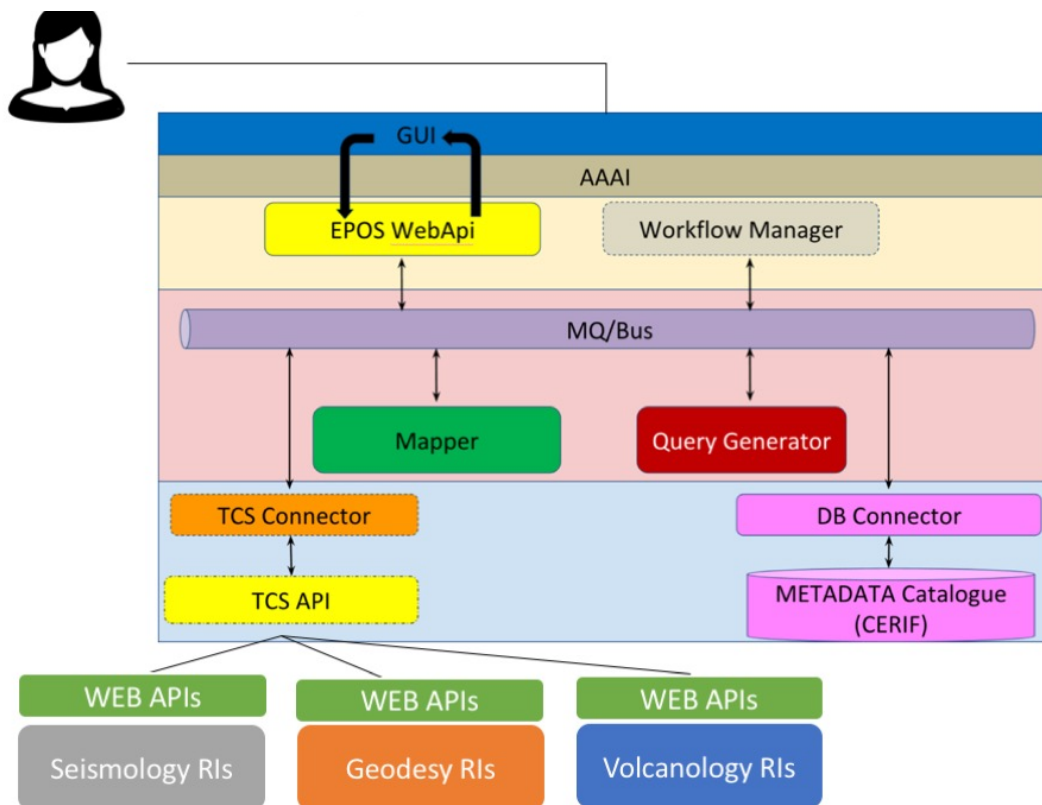


Figure 2. The EPOS ICS-C system architecture, designed and developed with the aim of integrating data and services provided by communities through community specific data centers called Thematic Core Services (TCS).

GUI: the access point by which users send requests to the system (typically, a web portal).

AAAI: Authentication, Authorization, Accounting, Infrastructure: a protocol to safely control user accesses to the system.

WebApi: it represents the entry point to access the EPOS system and provides different RESTful services. It receives the REST request from GUI, creates the message that contains the request parameters and sends it to Query Generator. It waits for an answer message from Mapper and sends it to GUI component.

Workflow Manager: handles and correctly addresses messages that are exchanged among components depending on the type of request.

MQ/Bus: Message queue is a software-engineering component used for inter-process communication.

Mapper: this component maps the result set from CERIF data model to EPOS set of pre-identified basic entities. In principle it is a converter from one metadata format to another.

Query Generator: this component receives the message from WebApi within the queue, parses it and creates the SQL query from request parameters, updates the message with the query and then sends it to DBConnector.

TCS Connector / TCS API: set of functionalities with the purpose of linking services (that TCS externally provide) with ICS-C.

DB Connector: it receives the message from Query Generator within the queue, extracts and executes the query on CERIF database. It provides in output the result set and sends it to Mapper.

Metadata catalogue: during the Preparatory Phase of the EPOS Project (EPOS-PP), metadata describing the TCS DDSS are stored using a European metadata catalogue standard, the CERIF⁴ data model [Jeffery, K., Houssos, et al. 2014] which differs from most metadata standards in that it (1) separates base entities from linking entities thus providing a fully connected graph structure; (2) using the same syntax, stores the semantics associated with values of attributes both for base entities and (for role of the relationship) for linking entities, which also store the temporal duration of the validity of the linkage. This provides great power and flexibility. CERIF also (as a superset) interoperates with widely adopted metadata formats such as DC⁵ (Dublin Core), DCAT⁶ (Data Catalogue Vocabulary), CKAN⁷ (Comprehensive Knowledge Archive Framework), INSPIRE⁸ (the EC version of ISO 19115⁹ for geospatial data) and others.

The metadata catalogue will also manage the semantics, in order to provide the meaning of the attribute values. CERIF stores the semantics in a ‘semantic layer’ referenced from the syntactic layer thus providing a single integrated semantic environment which is efficient because it uses standard IT (usually relational but all other database/processing environments may be used). CERIF interoperates with OWL (Web Ontology Language), SKOS (Simple Knowledge Organization System) and other semantic representation languages. Metadata from the communities will be mapped to the metadata catalogue in order to create appropriate links between common concepts in different disciplines. This process involves the harmonization and interoperability of the various DDSS from the different TCS through dedicated software modules.

1.2 EPOS TCS architecture

The different Thematic Core Services (TCS) have varying degrees of maturity in their development [Bailo D., Jeffery K.G., et al., 2015]. Therefore, 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. Other TCS don’t have any history of developing services. Some TCS rely on a federated architecture, others on a single sited data center.

4 <http://eurocris.org/cerif/main-features-cerif>

5 <http://dublincore.org/>

6 <https://www.w3.org/TR/vocab-dcat/>

7 <https://ckan.org/>

8 <http://inspire.ec.europa.eu/>

9 https://it.wikipedia.org/wiki/ISO_19115

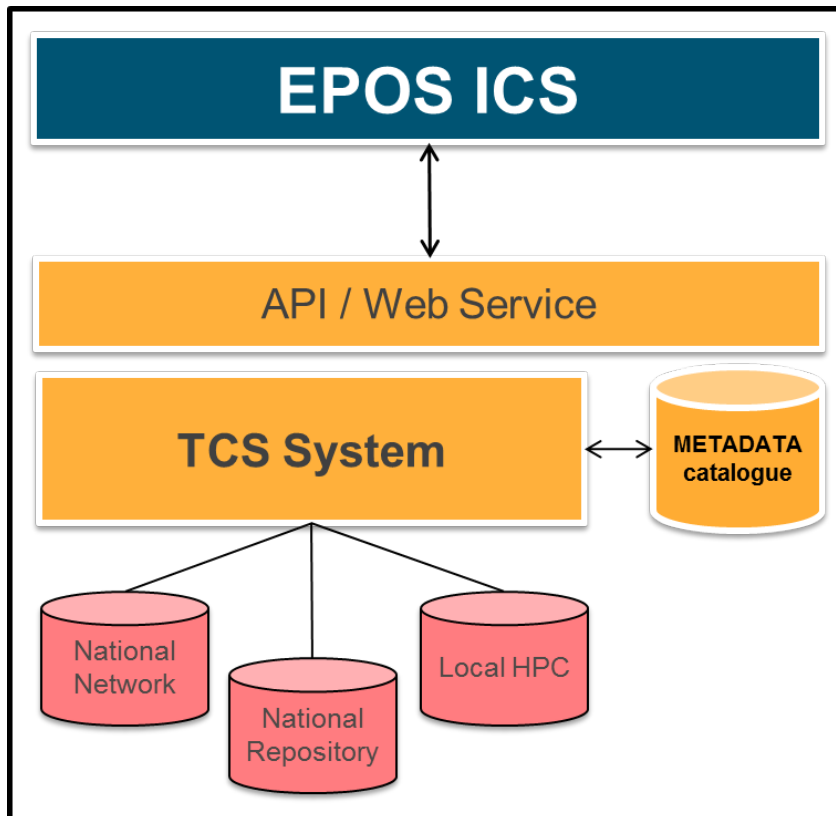


Figure 3. Generic TCS architecture. Independently of the community technical choices to federate and make the data accessible, all TCS should have web services / APIs which enable ICS to access the data.

Some have already done the effort of defining metadata standards and web services to disseminate the data. Others are still in the process of undertaking such work.

As a consequence, the scenario is very heterogeneous and includes many services with different levels of *technical maturity*.

Nevertheless, it is still possible to define a very generic architecture which, independent of the TCS level of maturity, can be applicable to any TCS (Figure 3).

The architecture is very general purpose, simple, and presents the following elements:

- a) *National Layer Research Infrastructures*: these can vary in different countries within a specific scientific community. They represent the existing DDSS.
- b) *TCS system*: this represents the e-Infrastructure for a specific scientific community. It may include the software used to federate National RIs, or the software to present results on the web (web portal).
- c) *Metadata Catalogue*: this is usually a database where each data object (e.g. file, in case of non-streamed data) is referenced and described by the metadata. It can be used to drive the Web Services.
- d) *Web Services / API*: this is the entry point where users can access the data. EPOS ICS has no particular recommendation with respect to the software used to build the TCS system. What matters is the way ICS and other stakeholders can access data. API based Web Services ensure data and metadata are accessible and discoverable by humans and machines (e.g. ICS system).

With respect to the metadata and related web services, there are two possible strategies to follow in order to achieve the ICS –TCS interoperability:

1. *Metadata dump*: a subset of the metadata from TCS copied to the ICS metadata catalogue. It guarantees that the metadata is fully managed by the ICS, and it lowers the burden of TCS in providing a highly efficient and robust system providing access to the metadata. However, it requires periodic (e.g. daily) polling/copying procedures and synchronization mechanisms must be put in place to guarantee consistency.

2. *Metadata Runtime Access*: The access to metadata is done at runtime by querying web services with the defined APIs. The APIs specification must be stored in the ICS Metadata catalogue (to enable ICS to access the system in an autonomic way). It avoids the error-prone procedure of dumping the metadata. However, it requires that TCS build very reliable and robust systems, able to manage a high number of concurrent queries as generated by users of the ICS.

In either case the conversion to CERIF needs to be done. Solution 1 is likely to be more efficient unless the data structures (not the data themselves) of a particular TCS are very fast changing. It is however likely that we need to use both strategies described above due to the high degree of variability in the TCS DDSS.

1.3 ICS-TCS integration

The joint work is carried out in the *interoperability layer*, a software/technical layer to enable communication among ICS and TCS. Once the TCS-ICS communication is established, data, metadata and services from the various communities must be managed.

This is accomplished in the ICS Central Hub (ICS-C) by means of a metadata catalogue, namely the facility which, together with other system components, manages and orchestrates all resources required to satisfy a user request. By using metadata, the ICS-C can discover data requested by a user, gain access to them, send them to a processing facility (or move the computation to the data), and perform other complex tasks. The catalogue contains: (i) technical specifications to enable autonomic ICS access to TCS discovery and access services, (ii) metadata description of the digital object (DO) with direct link to DO, (iii) information about users, resources, software, and services other than data services (e.g. rock mechanics, geochemical analysis, visualization, processing).

In the roadmap to enable data integration and ICS-TCS communication, three steps are required (Figure 4):

- (i) Metadata definition,
- (ii) web services/APIs definition,
- (iii) match and map with ICS.

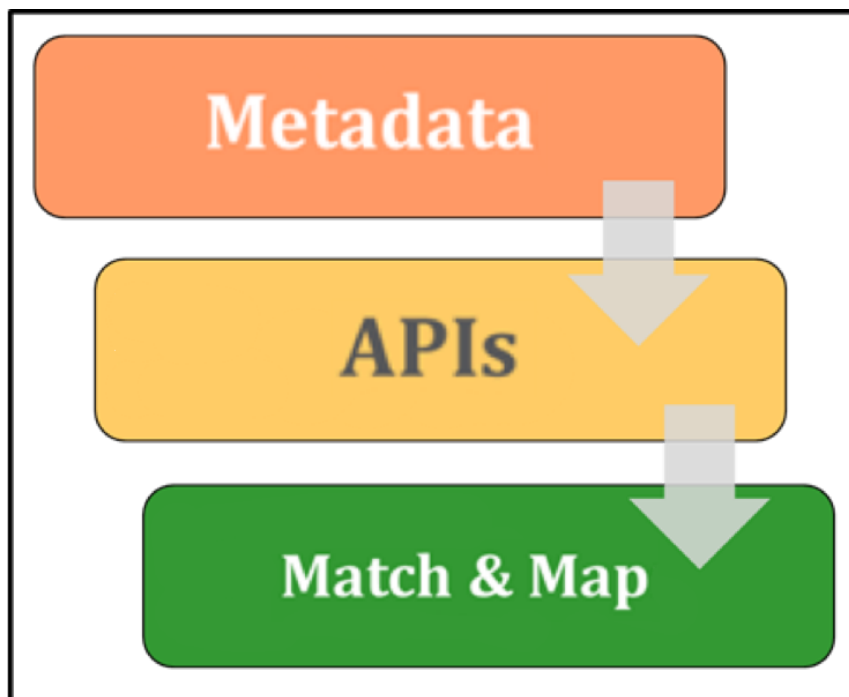


Figure 4. The three steps required in TCS-ICS common work and interaction through the interoperability layer.

Their description follows:

STEP 1: Metadata

As anticipated, it includes the definition and usage of metadata describing the community data (e.g. seismic waveforms, GPS time-series, geological maps), software (e.g. an analysis or visualization application), services (e.g. use of specialist equipment), resources (e.g. computers, instrumentation, detectors) and users (with their roles, responsibilities and authorities used for AAAI).

STEP 2: Web services for access

The heterogeneity of the EPOS community implies that TCS have different maturity levels. Some communities may only give access to metadata attached to raw Digital Objects (e.g. ftp repository containing files with metadata in the header). However the best option is to provide a software layer which implements services for (a) *Data discovery*; (b) *Data Access and Retrieval* (e.g. APIs, RESTful web services); (c) *data analysis/visualization/modelling/mining*, possibly based on existing standards.

- a. *Data Discovery services*
are used to discover, through metadata, the data of interest in remote repositories. EPOS ICS is indifferent with respect to topology (single-sited, distributed or federated repository) and access points (single or multiple) as long as clear definitions about how to access to the services are provided.
- b. *Data Access and Retrieval services*
are used to access a digital object (i.e. a file, a dataset, a document) and use it. The term access used can be thought also as a download process.
For both services the use of already existing, international, community accepted standards is highly recommended. OGC-services standards, INSPIRE based standards, Dublin Core and other international standards can be relatively easily integrated because a number of tools to manage data and metadata already exist. Reusing existing software tools will save resources for software maintenance in the long term on both TCS and ICS side.
- c. *Data analysis/ visualisation/ modelling/ mining services*
are used to extract meaning from the data using the software services provided. EPOS ICS will discover appropriate software services for the kinds of data requested by the user and (semi-automatically) compose the software stack to process the dataset(s) as required.

STEP 3: Metadata match and map

For each metadata standard from a single TCS a metadata Matching/Mapping procedure is required. This is also called the metadata assisted canonical brokering.

The purpose is matching the metadata elements of each dataset (but also for software services, users and resources as computers, detectors) at TCS level and metadata entities in the ICS metadata catalogue, and generating mappings between matched elements so that any data instances (values of attributes within a record) under one schema can be converted to instances under the other schema.

Matching involves finding corresponding attributes in the two schemas even if they have different names (especially a problem in multilingual environments) or – in the worst case – even different data types. As an example, in two different datasets both concerning observations on the earth surface, there may be attributes describing geospatial positions having different names (Latitude/Longitude versus Northing/Easting) and different accuracy and precision attributes. In order to properly match metadata elements, a common ICS and TCS technical staff shared work, is required.

2. Development and interaction principles

In the following the high-level principles for the ICS development and the ICS-TCS interactions are explained. Such principles, as described elsewhere [Bailo D., Jeffery K.G., et al., 2015], have a direct impact on the way the system is developed, i.e. in the architectural approach and in technical choices.

Co-development (ICS and TCS): the development of the ICS depends on end-user requirements and the DDSS provided by the TCS. The TCS are at different stages of maturity. For some – with little infrastructure to date - the adoption of the EPOS architecture is straightforward. For others – with several years (decades)

of infrastructural investments already existing - a jointly agreed evolutionary plan to converge to interoperability with ICS will be adopted. The EPOS approach in this context is neither top-down nor bottom-up: the main idea is to use the general architecture and follow a cooperative approach in the designing and development of the software to build the compatibility layer, which is the place where harmonizations and communications are achieved.

Such an approach also encourages EPOS to focus more on an architectural design for the system that is capable of being engineered and implemented collaboratively rather than using a top-down reference model approach.

This would facilitate cross-disciplinary access and utilization of the data for the scientific purpose.

Do not reinvent the wheel: reuse local technologies: the co-development philosophy maximizes re-use of existing software services, data availability and resources.

Do not build a supercomputer: build an integration with an ICS-D service provider.

EPOS will have some of its own computing resources in order to provide the ‘uniform view’ over EPOS entities. However, to facilitate intensive processing, EPOS will provide – subject to authorization – access to appropriate computing facilities including HPC (High Performance Computing) machines for modelling and simulation, and HTC (High Throughput Computing) clusters for data intensive applications such as data mining. Other ICS-D services will include visualization and processing services. These facilities are known collectively as ICS-D (Integrated Core Services – Distributed) and will be provided by external service providers.

Microservices approach: the Microservices architecture approach envisages small *micro* services dedicated to the execution of a specific class of tasks, which have high reliability [Newman S., 2015]. Presently, there is no precise definition of microservice architecture, and even the distinction between *service* and *component* (in a monolith) is up for debate [Stubbs J., Moreira W., et al., 2015]. In this context, EPOS will either take existing software services and ‘wrap’ them for EPOS use or build new services complying with EPOS architectural standards. The aim is to have micro services with defined interfaces which can be composed together to form a software stack able to correctly address user request to the appropriate system component.

Clear long-term technical goals, but iterative short-term approach: the overall architecture of EPOS is clearly defined and agreed through EPOS-PP. However, its realization in the implementation phase (EPOS-IP) would require a step-by-step approach to build a reliable system environment to meet the requirements of end-users and their communities. To this purpose iterative work cycles both for ICS developments and for ICS-TCS communication have been set up.

3. EPOS Web API implementation

The EPOS Web API was developed using the open source Swagger¹⁰ framework. It is a specification for documenting RESTful API formatted in JSON. We used it to specify the URL, method, and representation to describe EPOS Web API and to generate the skeleton source code for the application.

Java and Spring¹¹ open source framework allowed us to design the Spring’s classes as JavaBeans¹², and they enabled us to inject data using the setter methods of these classes. The entry point to the application is a Spring Controller and several annotations helped to map HTTP requests to methods.

Figure 5 shows how to search and display EPOS data on Swagger User Interface¹³. The shown example allows to access the information present on the CERIF database implemented with PostgreSQL¹⁴. Search is possible via spatial and temporal axes. Additionally, full text search is available. One can also specify further research parameters using different filters (e.g., datatype, domain, subdomain, keywords, etc.): these constitute a base set of parameters that are common among research areas in the domains of Solid Earth science.

10 <https://swagger.io/>

11 <https://spring.io/>

12 <https://docs.oracle.com/javase/tutorial/javabeans/>

13 <http://epos.cineca.it/webapi/swagger-ui.html#/Search>

14 <https://www.postgresql.org/>

GET /search searchGet

Implementation Notes
The sources endpoint returns all the metadata elements. The search function queries the Cerif catalog. Provides a list of resource items from the DDSS.

Response Class (Status 200)

Response Content Type application/json

Parameters

Parameter	Value	Description	Parameter Type	Data Type
page	1	actual page	query	string
resultsPerPage	25	number of results per page	query	string
bbox		spatial bounding coordinates, geographic bbox	query	string
time		Time interval, temporal bounding of the search. Start and End Time.	query	string
datatypes	webservice	datatype search.	query	string
domain		domain search; list of the work package domains specific to EPOS.	query	string
subdomain		subdomain search; list of the work package domains specific to EPOS.	query	string
keywords	fault	keywords	query	string
freetext		free text search	query	string

Response Messages

HTTP Status Code	Reason	Response Model	Headers
401	Unauthorized		
403	Error		
404	Not Found		

[Try it out!](#) [Hide Response](#)

Curl

```
curl -X GET --header 'Accept: application/json' 'http://epos.cineca.it/webapi/search?page=1&resultsPerPage=25&datatypes=webservice&keywords=fault'
```

Request URL

```
http://epos.cineca.it/webapi/search?page=1&resultsPerPage=25&datatypes=webservice&keywords=fault
```

Request Headers

```
{
  "Accept": "application/json"
}
```

Response Body

```
{
  "summary": {
    "Page": 1,
    "resultsPerPage": 25,
    "numberOfResults": 4
  },
  "results": [
    {
      "name": "European Database of Seismogenic Faults | Crustal Fault Sources - Web Feature Service",
      "id": "services.seisefaults.eu/geoserver/EDSF/crustal_fault_sources_top/wfs",
      "description": "The European Database of Seismogenic Faults (EDSF) was compiled in the framework of the EU Pro",
      "type": "webservice",
      "geometry": {
        "minElev": 0,
        "minLat": 49,
        "minLong": -12.39,
        "maxElev": 0,
        "maxLat": 30.3,
        "maxLong": 45.28
      }
    }
  ]
}
```

Response Code

```
200
```

Response Headers

```
{
  "date": "Tue, 14 Nov 2017 11:49:36 GMT",
  "server": "Apache/2.4.7 (Ubuntu)",
  "connection": "keep-alive",
  "keep-alive": "timeout=5, max=100",
  "content-length": "3838",
  "content-type": "application/json; charset=UTF-8"
}
```

Figure 5. Example of how to retrieve information from the EPOS system by means of Swagger User Interface.

4. Conclusion

The scope of this work is to describe how EPOS architecture improves and facilitates integration, access and use of Solid Earth science data, data products, services (DDSS) and software. Whereas heterogeneous communities provide resources in a scattered and non-standard manner, EPOS provides an added value on different aspects of the European-wide scenario:

- both on the organizational / governmental side, with community building and improvement of sustainability action;
- on the technical side, by promoting the adoption of common standards and providing access to heterogeneous data from a single platform in a transparent way;
- regarding the building of the so-called Integrated Core Services, it requires efforts both on the architectural side, in order to design and implement a complex centralized but distributed system, and on the domain specific side, where communities are required to provide robust systems that give access to quality metadata;
- on the hard task of the communication with the TCS, metadata must be provided by the communities, and that's how EPOS is committing itself in a strong community building action.

The diamond tip of the infrastructure presented in this work lies in the fact that it is the result of a brand-new way to approach, when dealing with Research Infrastructures development, for it is the first time that such an infrastructure is set up for Solid Earth science, and moreover because it is a process - that has started in its preparatory phase (EPOS-PP), where guiding principles that influenced subsequent technical developments have been outlined – that encouraged, given the high heterogeneity of the community in terms of technical maturity, a co-development approach that envisages a strong collaboration between ICS and TCS. This encouraged EPOS to develop a system architecture that can be engineered and implemented collaboratively, rather than using a top-down reference model approach. As a consequence, this also facilitates cross-disciplinary access and utilization of the data for the scientific purpose. This great result also constitutes the main hard point the project had to get over, of course.

This approach allowed EPOS to define a specific layer where integration occurs, i.e. the ICS system, which TCSs should interface with, and similarly, define a TCS generic architecture with core components considered a must-have to interact with the infrastructure (e.g. Web Services / API). Procedures to integrate National RI and metadata has also been outlined, as well as procedures to enable communication among ICS and TCS by means of a software/technical layer of the infrastructure known as *interoperability layer*.

ICS Architecture was then defined following the microservices approach to keep single components separated (e.g. the Mapper to match result set from CERIF to EPOS set of pre-identified basic entities).

The architecture, as presented in this work, has been implemented and tested, by providing a Graphic User Interface and a set of WEB APIs that will be finally validated within October 2017.

Future work is expected to take place on internal ICS components, like the Workflow engine, and in the interfacing with distributed ICS-D resources, for which appropriate interfaces are needed.

References

- Atkinson M., Carpené M., Casarotti E., Claus S., Filgueira R., Frank A., Galea M., Garth T., Gemünd A., Igel H., Klampanos I., Krause A., Krischer L., Hoon Leong S., Magnoni F., Matser J., Michelini A., Rietbrock A., Schwichtenberg H., Spinuso A., Vilotte J.P., (2015). *VERCE delivers a productive e-Science environment for seismology research*. In: IEEE 11th International Conference on eScience 2015 (in press).
- Bailo D. and Jeffery K.G., (2014). *EPOS: A novel use of CERIF for data-intensive science*. In: *Procedia Computer Science* (Vol. 33, pp. 3–10). 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*. In IEEE 11th International Conference on eScience 2015.
- Jeffery K.G. and Bailo D., (2014) *EPOS: Using Metadata in Geoscience*. In: Closs S., Studer R., Garoufallou E., Sicilia MA. (eds) *Metadata and Semantics Research. MTSR 2014. Communications in Computer and Information Science*, vol. 478, pp. 170–184. Springer, Cham.

- Jeffery K.G., Houssos N., Jörg B., Asserson A., (2014). *Research information management: the CERIF approach*. International Journal of Metadata, Semantics and Ontologies, vol. 9(1).
- Newman S., (2015). *Building Microservices*. O'Reilly Media.
- Stubbs J., Moreira W., Dooley R., (2015). *Distributed Systems of Microservices Using Docker and Serfnode*. In: Science Gateways (IWSG), 2015 7th International Workshop.

Appendix 1 - Acronyms

AAAI	Authentication, Authorization, Accounting, Infrastructure (in the context of architecture)
AARC	Authentication and Authorization for Research and Collaboration, https://aarc-project.eu/
API	Application Programming Interface
CERIF	Common European Research Infrastructure Format
CKAN	Comprehensive Knowledge Archive Framework
DC	Dublin Core Metadata Standard
DCAT	Data Catalogue Vocabulary
DDSS	data, data products, services and software
DO	Digital object
EPOS	European Plate Observing System
EPOS-IP	EPOS Implementation Phase
EPOS-PP	EPOS Preparatory Phase
ERIC	European Research Infrastructure Consortium
GUI	Graphical User Interface
HPC	High Performance Computing
HTC	High Throughput Computing
HTTP	HyperText Transfer Protocol
ICS	Integrated Core Services
ICS-C	Integrated Core Services - Centralized
ICS-D	Integrated Core Services - Distributed
INSPIRE	INSPIRE is a EU Directive in (May 2007), establishing an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment
ISO-19115	International standard for geospatial data
JSON	JavaScript Object Notation
NRI	National Research Infrastructures
OGC	Open Geospatial Consortium
OWL	Web Ontology Language
SKOS	Simple Knowledge Organization System
RESTful web service	Web Services based on the REST architecture (web)
TCS	Thematic Core services
URL	Uniform Resource Locator
VERCE	Virtual Earthquake and seismology Research Community in Europe e-science environment http://www.verce.eu/

Quaderni di Geofisica

ISSN 1590-2595

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html>

I Quaderni di Geofisica coprono tutti i campi disciplinari sviluppati all'interno dell'INGV, dando particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari, che per tipologia e dettaglio necessitano di una rapida diffusione nella comunità scientifica nazionale ed internazionale. La pubblicazione on-line fornisce accesso immediato a tutti i possibili utenti. L'Editorial Board multidisciplinare garantisce i requisiti di qualità per la pubblicazione dei contributi.

Rapporti tecnici INGV

ISSN 2039-7941

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html>

I Rapporti Tecnici INGV pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico e di rilevante interesse tecnico-scientifico per gli ambiti disciplinari propri dell'INGV. La collana Rapporti Tecnici INGV pubblica esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. L'Editorial Board multidisciplinare garantisce i requisiti di qualità per la pubblicazione dei contributi.

Miscellanea INGV

ISSN 2039-6651

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html>

La collana Miscellanea INGV nasce con l'intento di favorire la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV (sismologia, vulcanologia, geologia, geomagnetismo, geochimica, aeronomia e innovazione tecnologica). In particolare, la collana Miscellanea INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc.

Coordinamento editoriale e impaginazione

Centro Editoriale Nazionale | INGV

Progetto grafico e redazionale

Daniela Riposati | Laboratorio Grafica e Immagini | INGV

© 2018 INGV Istituto Nazionale di Geofisica e Vulcanologia

Via di Vigna Murata, 605

00143 Roma

Tel. +39 06518601 Fax +39 065041181

<http://www.ingv.it>



Istituto Nazionale di Geofisica e Vulcanologia