REPORT2014

ICT BOARD REPORT:

Authors: Daniele Bailo, Keith Jeffery, Damian Ulbricht, Frieder Euteneuer

Table of contents

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

2

1. The EPOS architecture

The EPOS architecture has been designed to properly organize and manage the interactions among different EPOS actors and assets. To make 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. Four complementary elements form the infrastructure:

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. These have a significant economic value 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.

2. The **Thematic Core Services (TCS)** constitute the community-specific integration. They represent a 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.

3. The **Integrated Core Services (ICS)** represent the novel e-infrastructure consisting of services that will allow access to multidisciplinary data and data products, and synthetic data from simulations, processing, and visualization tools. The ICS will be composed by the ICS-Central Hub (ICS-C) and distributed computational resources (ICS-d). ICS are the place where integration occurs.

4. The **Executive and Coordination Office (ECO)** is the EPOS Headquarter and the legal seat of the distributed infrastructure governing the construction and operation of the ICS and coordinating the implementation of the TCS.

The European Research Infrastructure Consortium (ERIC) has been chosen by the Board of Governmental Representatives as the legal model for EPOS and used in designing the Governance Model. This includes a General Assembly of members and an Executive Director, supported by a Coordination Office. The funding model as it has been designed will support the sustainable construction and operation of the whole EPOS enterprise. The model includes complementary funding sources for each of the key EPOS elements.

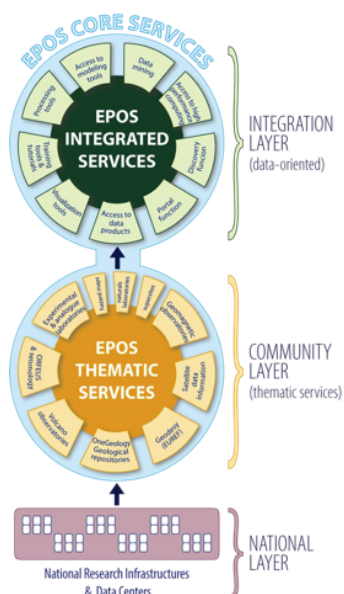Figure 1 describes the EPOS technical structure organised in three layers.

Figure 1: EPOS technical architecture. The diagram shows the three layers in which the EPOS components (institutions and services) have been organized: National Layer, Community Layer, Integration Layer.

Figure 2 illustrates the topology and the relationships among the EPOS elements. During the first phase of the EPOS-ERIC, the Executive and Coordination Office (ECO) and the Integrated Core Services Central Hub (ICS-C) compose the EPOS-ERIC.
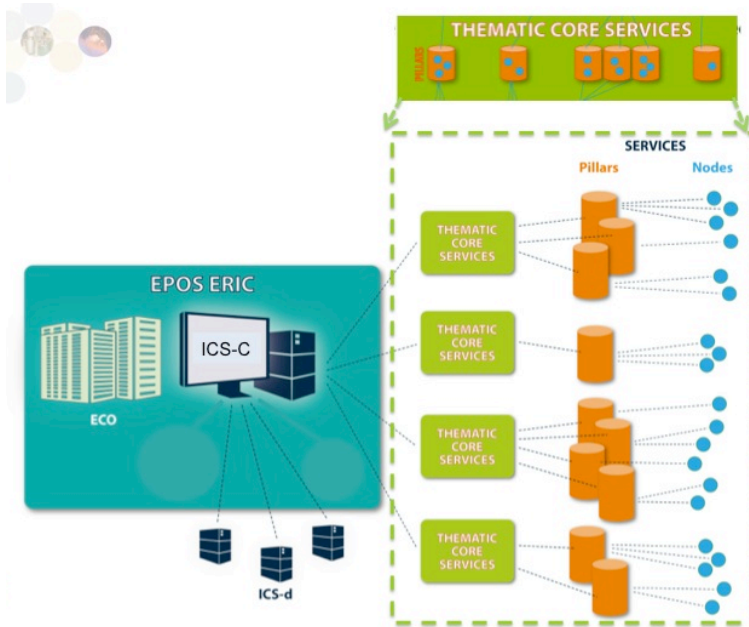

Figure 2 - The EPOS architecture designed for Phase 1. NRIs are not included in the figures because they contribute to EPOS, but they do not belong to EPOS design

PILLAR: *the environment where different sets of data, products and services, directly related to a specific community, can be coordinated, managed and organized through distributed/single nodes.*
NODE: *represents a distributed network of infrastructures or a single research infrastructure (RI) delivering data, products and/or services.*
SERVICES *are IT solutions and tools for data management provided to the scientific community through multiple or individual nodes within a pillar.*

The existing NRIs for solid Earth science in Europe, which currently support the EPOS integration plan (listed in the RIDE database, www.epos-eu.org/ride/), generate data and information and are responsible for the operation and maintenance of instrumentation in each country. The TCS bring together these data and products from national RIs and make them available for domain-specific community integration and to the ICS. The TCS represent the framework where the community organizes the delivery of data and data products and identifies suitable implementation plans for providing access to new data products and novel services to users. The TCS are a governance framework for each EPOS Community and they guarantee access to data and **services** through different **pillars** and **nodes**. The **ICS** will provide access to multidisciplinary data, data products, and synthetic data from simulations, processing and visualization tools. The key element of the ICS is the **central hub (ICS-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. This **single-sited e-infrastructure** will include the EPOS portal and its key functions: the metadata catalogue, the system manager, the Application Programming Interfaces (API) (or equivalent) to interoperate with TCS and ICS-d, and the services that will allow data discovery, interaction with users as well as access, download and integration of data. The ICS-C will also provide access to distributed computational resources for visualizing, processing and modelling data and data products. These distributed computational resources form the **distributed ICS (ICS-d)** and include access to supercomputing and facilities as well as to visualization, processing and modelling tools.

EUROPEAN PLATE OBSERVING SYSTEM phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

4

2. Integrated Core Services

2.1 e-Infrastructure design

The main concept is that the EPOS TCS data and services are provided at the layer where the integration occurs, that is to say by and into the ICS (see Fig.1). This is achieved by means of a communication layer called the *compatibility layer*, as shown in the functional architecture (Fig. 3). This layer contains all the technology to integrate data, data products and services (including software services and access to resources) from many communities into the single integrated environment of the Integrated Core Services (ICS). Therefore, the ICS, being devoted to the real integration of data, data products and services, represent the "core" of the whole e-infrastructure. 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. In the actual design, ICS are made up of several, modular, interoperable building blocks (as shown in the top part of Fig. 3):
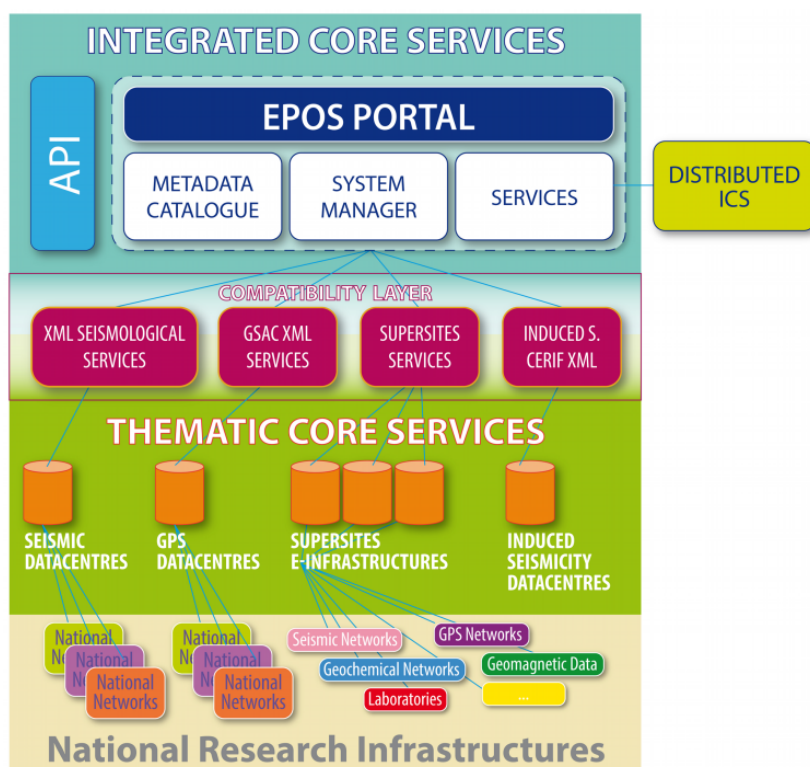

Figure 3: EPOS functional Architecture, describing the technical functional components of EPOS. It specifies for each layer the ICT modules and their function. At the ICS layer it is fundamental to understand the design of the integrating e-Infrastructure.

Metadata catalog is the key component. It contains all the information that the system might be willing to deal with describing in a digital representation the objects of the EPOS e-Infrastructure including data, software, users and resources such as computers, detectors or laboratory equipment. It uses CERIF (Common European Research Information Format, a recommendation to the EU Member States) as a tool to harmonize information on research. CERIF describes datasets but also software, services, users and resources such as computers, datastores, laboratory equipment and instruments. This catalog requires dynamic maintenance. This can be done either by human or by automated means (the latter is recommended and we are working on such services). Automation naturally is depending on the technologies implemented at the TCS layer, to which the metadata catalog will connect through the *compatibility layer*. This is a typical task accomplished by the System Manager.

System Manager can be considered as the "intelligence" of the system and is basically middleware which manages the whole metadata system. The System Manager takes advantage of the information contained in the catalog (which is the "knowledge base" of the system) and makes proper decisions according to: (i) user requests, (ii) metadata contained in the EPOS metadata catalogue covering users, resources, software services and data. Therefore, in an EPOS context, this is the place where the brokering techniques – but driven by the metadata mad explicit in the catalog rather than by program code hidden in the software – will be effective.

EPOS Portal and API are functional blocks dedicated to the interaction between users (both human and machine) and the system. The portal deals with the interaction between a human user and the system. A generic user will be enabled to perform actions like: (i) data/data products/sensors/facilities discovery; (ii) data/data products download; (iii) data /data products analytics; (iv) data/data products visualization; (v) data/data products modeling and simulation. However, this portal functionality is not sufficient because EPOS wants to (i) be interoperable with other systems, (ii) be compliant with major European standards, (iii) deliver a high quality service that enables a user to perform programmatically some actions. A locus dedicated to machine-machine interaction is therefore needed. This is exactly the Application Programming Interface (API), which includes a set of native functions enabling a machine use of the EPOS system, as for instance RESTful queries of the type:

GET */entity?data_types=[seismwav,GPS,satdata]&lat=45.5345&lon=16.334&startime="datetime".*
For this latter purpose, the reliable and fully CERIF compliant CERIF-XML standard is used at the present stage in the form of a RESTful service that can be queried to obtain XML-formatted metadata.

Services interface module includes all the software and interfaces required to connect outsourced resources as, for instance, linkage to HPC centres or workflow management infrastructure (e.g. VERCE www.verce.eu).

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

6

2.2 Data Model

The metadata catalog is the key to the implementation and is dependent on the metadata model. Metadata can be viewed or considered in two dimensions:

1. Metadata to describe the objects of the "EPOS ecosystem" in such a way that the descriptions (including restrictions of usage) can be used by the middleware and by application software;
2. Metadata for to provide the appropriate level or depth of metadata for the required task or processing.

2.2.1 Metadata to describe EPOS ecosystem objects

This dimension of the metadata concerns the objects of the EPOS "ecosystem": these are classified into users, services (including software), data and resources (computing, data storage, instruments and scientific equipment) as shown in Fig. 4.
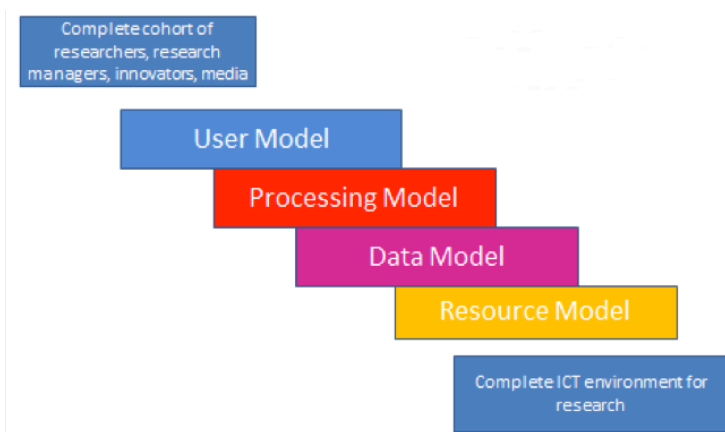

Figure 4: The four models to describe the EPOS ecosystem objects

The ***User Model*** describes how a subject (human user, but also a program, or a process) can interact with the EPOS e-infrastructure and determines the design of the EPOS web-portal. This is important to ensure all kinds of people, regardless their location, language, expertise, permissions, responsibilities, authorities and disabilities (or differently-abled abilities e.g. when driving), can easily access and use the system. Therefore, it will provide the technical information to ensure users' security, privacy and trust through its identification, authentication, authorization and accounting (IAAA). IAAA are based on the data policy and access rules describing the degree of openness of the information, data usability, data ownership, and the stakeholders' metric aimed at analysing the impact, influence, engagement, exchanges and ethical risks associated to each user category and the possible utilization of EPOS data and services

The ***Processing Model*** is by far the most difficult to create because it has to include sufficient information on how the system performs the calculations and visualization on the different data available within EPOS to allow composition of appropriate software services to accomplish the application requirement. It provides the core information of ICS, in particular the know-how on integrating data and data products beyond the simple data mining and data archiving currently distributed

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

7

but still available (with very different services, access and policies) at a community and national level. The model also has to describe the rights and responsibilities of users of the software services including (if appropriate) costs.

The **Data Model** serves to describe the data and availability and their associated detailed, domain-specific metadata in order to allow the user to find them, work with them (integrate) and download them or utilize them in a composition of software services. Again the rights asserted over the data are recorded with access to licence information. EPOS is working with others on ways to encode licence information such that it can be processed automatically by middleware to ensure the end-user does not violate any rights inadvertently[1].

The **Resource Model** is a technical description of the physical resources owned by the data providers (i.e. national RIs and Thematic Core Services) that are available for EPOS integration and of those owned outside the EPOS delivery framework that will provide specific IT services (ICS-d, see the processing model). This model is needed to provide the description of the organization of the facilities in both their hardware and software components that will guarantee: (i) data repository facilities (including long-term preservation and provenance as well as discoverability), (ii) data processing, (iii) simulation and modelling and (iv) visualization (rendering). For each category the model provides a detailed technical IT description with specifics on how to ensure a sustainable and efficient connectivity and therefore to allow the user to reach their content (the data) or to use them (processing and visualization).

2.2.2 Metadata depth for the required task
The other dimension of metadata provides the appropriate level or detail of metadata information which enables a user to perform actions and functions over data and data products. This model, the so-called three-layer model[2] is structured as follows (see also Fig. 5):

1. The discovery layer, utilising the capability to generate from the underlying contextual layer – providing DC, DCAT, INSPIRE (to integrate with other existing datacentres utilizing these standards) and both CKAN and eGMS (the latter two to encourage particularly integration with government open data (data.gov) sources);
2. The contextual layer, using CERIF from which the discovery level metadata standards can be generated and which also points to individual datasets or services to metadata in the detailed layer;
3. The detailed layer, which includes detailed metadata standards usually domain or sub-domain specific for each kind of data (or software, computer resources or detectors/instruments) to be (co)-processed. For example, detailed metadata for a dataset may include the database schema (giving attributes, types etc.) but also - for each attribute – precision and accuracy.

---

[1] Vasily Bunakov, Keith Jeffery: Licence Management for Public Sector Information Proceedings Conference for E-Demoracy and Open Government (CeDEM) 2013 Donau University Krems 2013 pp 292-302 ISBN 978-3-902505-30-9

[2] Jeffery, K., Asserson, A., Houssos, N., & Jörg, B. (2013). *A 3-Layer Model for Metadata.* Proc. Int'l Conf. on Dublin Core and Metadata Applications 2013, 3–5.
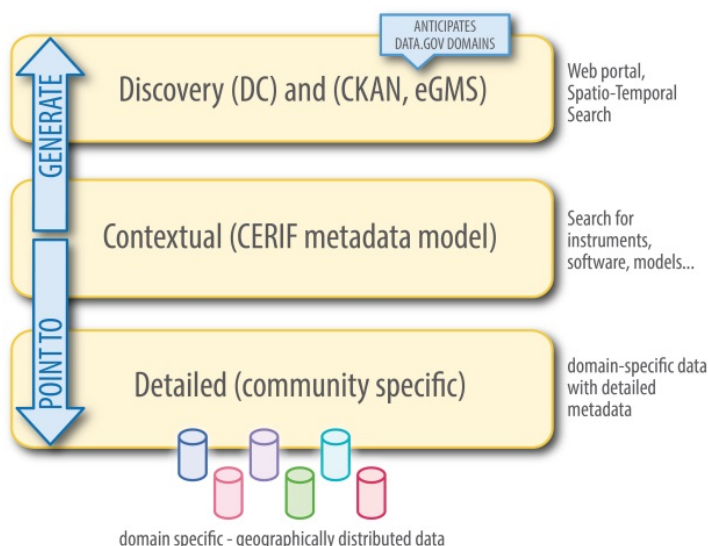
Figure 5: Three-layer Metadata Architecture

2.2.3 CERIF model in the context of EPOS

The core of the proposed structure is the contextual layer, which is built following the concepts and guidelines of the euroCRIS organization www.eurocris.org. CERIF was originally designed for research information interoperability. It covers as base entities *persons, organisations, projects, products* (including datasets and software), *publications, patents, facilities, equipment, funding* and more. The novelty of CERIF is in the linking entities; these describe the relationships between instances of base entities. Each linking entity consists of the keys of the two base entities being related, a role (the relationship) and both start and end date/time (thus providing temporal information for the period of validity of the assertion implicit in the instance of the linking entity). A role might be 'author' in the linking entity between person and publication; it could equally well be 'editor' or 'reviewer'. In this way CERIF respects both referential and functional integrity which is an advantage over simpler metadata schemes and the reason why CERIF was chosen for EPOS. For example, in CERIF a person exists independently of whether they are employed, an author, a copyright holder or a software developer – all of which might be coincident. Metadata schemes which, for example, have person as an attribute in an instance of a publication violate functional integrity and may – dependng on cardinality – violate referential integrity.

CERIF also provides for multilinguality; each text string can have multilingual variants. Additionally CERIF has a semantic layer – interoperable with the well-known W3C standards OWL and SKOS – which has base entities and linking entities. The base entities are classification schemes (rather like namespaces) and classification (i.e. terms). The linking entities allow terms from different classification schemes to be related – like a thesaurus or ontology. This flexibility is important and having the semantic layer integrated within the model – and using conventional data processsing technology - makes for efficient processing.

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

9

However the original CERIF design was aiming to describe all aspects of research process which ended up with the publication, product or patent as a final product. Within the context of EPOS, the product of research is a more complex object, a product which can potentially include very diverse kind of data and data products.

Some developments in this area to extend CERIF were already started. In 2013 a data extension proposal for CERIF for the purposes of EPOS was initiated by investigating CKAN, DCAT and eGMS and was guided by a draft proposal of the Jisc-funded CERIF for Datasets (C4D) project (cerif4datasets.wordpress.com). In EPOS, as a comprehensive list of all the possible data products would have been very difficult, if not impossible, to draft, a categorization has been carried out taking into account previous work done on levels of data products from NASA (science.nasa.gov/earth-science/earth-science-data/data-processing-levels-for-eosdis-data-products), Interface Region Imaging Spectrography (IRIS) (www.lmsal.com/iris_science/doc?cmd=dcur&proj_num=IS0076&file_type=pdf), UNAVCO (pbo.unavco.org/data/gps) and others including the particle physics community and datacentres in the natural environment communities.

This work yielded the EPOS data levels categorisation:

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

The extension to CERIF proposed by C4D – and now adopted in the current version of the CERIF model as published - is able to handle all the data encompassed by this categorisation.

2.2.4 Data discovery at EPOS ICS

The three-layer metadata structure can effectively represent and manage the levels of commonality among all the metadata describing datasets provided by the data providers (or similarly for users, software services and resources). Discovery level is somewhat abstract but provides a target list of potentially relevant data (or software, resources) while contextual level provides the lowest common metadata across all domains and allows an end-user or software to assess the relevance and quality of the data for the purpose at hand.

The community specific metadata (lowest level) is hence not ingested by the system as a whole: only a subset of it is mapped into the Metadata Catalogue – sufficient for (a) providing the end-user or software with information to characterise the object(s) being accessed and assess their quality related to the purpose at hand; (b) to provide the acces route (usually an API). However in order to have a reliable access to the local data, the Integrated Core Services had to set up efficient communication mechanisms into the so called *compatibility layer*. This layer makes possible the linkage between ICS and TCS (and when required institutional RIs) thus enabling discovery and integration

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

10

capabilities. The thematic core services (TCS) are developed independently by their respective communities and in order to provide data and metadata to the ICS (but more in general to be interoperable at international level) they provide software interfaces to access their systems, usually just end-user services to discover appropriate datasets or software and –in some cases – limited processing.

To fetch and discovery the desired data and metadata, ICS can then: (i) access to TCS web-services, (ii) access TCS generic APIs, (iii) link directly to datasets and ingest the metadata by means of some automated process or – if it is an infrequent access channel - manually. The automated process is a conversion driven by an earlier metadata matching and mapping exercise which can be manual (once and thereafter automated using the convertor driven by the mapping parameters) or part-automated using one of the techniques mentioned in the introduction.

To enable such a communication (compatibility layer) a new entity was introduced in the CERIF scheme – the *cfServiceInterfaceDescription* – and the *cfService* was used with a special meaning: the entity is supposed to store information about the webservice or API providing data. The purpose of these two new proposed entities is to store all the information necessary to enable the system to connect to the desired service and map the metadata of interest into the *cfResulProduct* entity.

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

11

3. The EPOS prototype

The prototype, together with the design of the EPOS ICS, is the final and fundamental deliverable of the EPOS Preparatory Phase. The prototype, indeed, validates the theoretical e-architecture of the design, demonstrating its feasibility. The prototype is available at the web address http://epos.cineca.it/

3.1 Prototype development

3.1.1 Graphic User Interface (GUI)

For the development of the prototype a software developed in collaboration with VERCE project was used. The prototype is programmed in Java/javascript and uses GeoEXT and the Sencha framework to organize the user interface.

The application integrates data from several different resources using different techniques. A detailed description of those different resources will be given in the following chapters. The main visualisation element, the map of the application, is composed of multiple layers of OGC Web Map Services (WMS) with adjustable transparency.

Depending on the selected products, a search request to the application triggers search requests to the EPOS extended CERIF catalogue. Additional queries are sent to community specific web services that return data in domain specific formats. The data are parsed and stored on client side.

The data with spatial information are automatically visualised on the map. Filters can be applied after the information is returned from all services. For example, for seismic event data, filtering through some geodetic parameters– as crust thickness of the Earth – can be applied. The filtered information is organized in a tabbed view and spatially characterized products are displayed on the map (Figure 6).
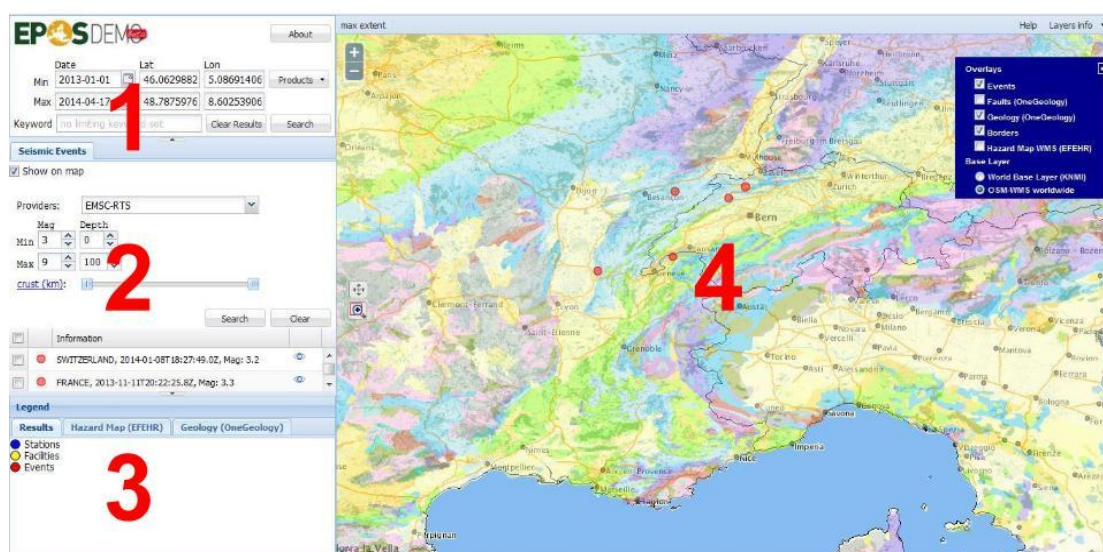

Figure 6 The Prototype web interface

The user interface of the prototype is a Javascript application that is displayed with an internet browser. It collects and displays information of the extended CERIF catalogue and the result of embedded services. The browser window is splitted in four areas

1. Area (1) in the upper left is used to generate a search request to the system. The user selects here EPOS-products from the products menu and generates a search request by entering information that should be common to all selected products. This common information could be spatial coverage, temporal coverage and keywords for the free text search.
2. Area (2) in the middle left shows the results in tabs. There is one tab per product group. The visualization of on the map can be toggled here. If applicable, further information can be entered to refine a search on the selected product group. If a result entry of a result tab is chosen, the associated marker of the map (Area (4)) is selected and displays full information in a popup window.
3. Area (3) in the left bottom is used to show a legend for the information displayed on the map.
4. Area (4) holds a map, which initially shows the European region. Visual information is organized in several layers. On the map the location of search results are displayed by using markers of different colour. By selecting a marker the associated result tab entry in Area (2) is selected and further information is displayed in a popup window.

3.1.2 System Manager Software

The development of the prototype started with the contribution of several institutions and RIs from the Seismology, Geology and Geodesy domain. The different service-providers are listed in figure 7, together with the used specific connection technique.
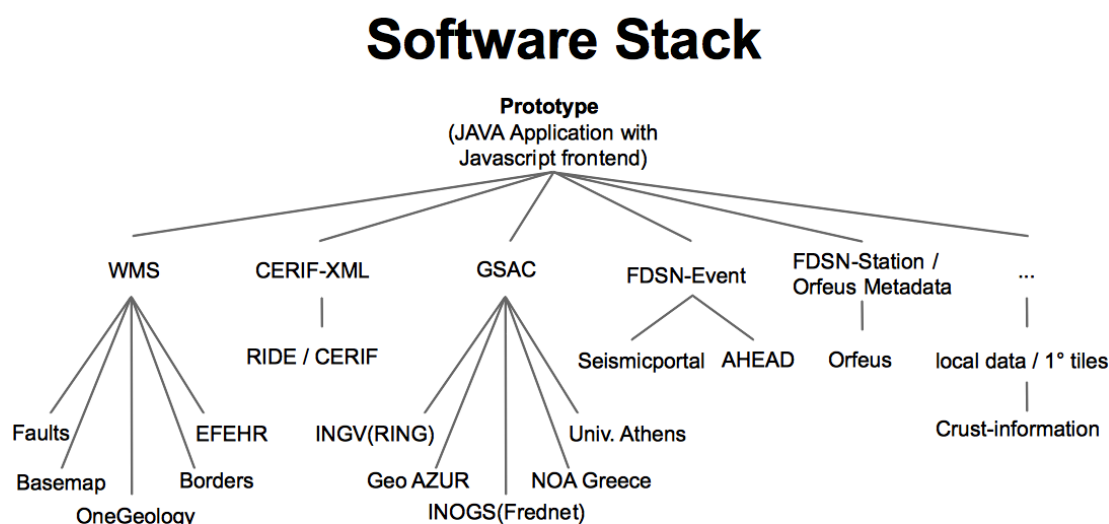

Figure 7: The prototype software stack

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

13

The *EFEHR* project (European Framework for Earthquake Hazard and Risk) offers WMS with Hazard and Risk maps for Europe as products of seismological analysis, which are implemented as Map layer in the demonstrator.

The European network of geologic institution *OneGeology* – Europe is offering a common representation of the European geology as WMS, which has also been implemented in the demonstrator. The Basemap, national borders and a representation of Faults are also implemented as map-layers. Information, which is stored in the CERIF based central metadata catalogue, like person, RIs, etc., are accessed via the CERIF-XML services.

The project *Orfeus* (Observatories and Research Facilities for European Seismology) is offering a full list of station information of the EIDA initiative (European Integrated Data Archive) via the FDSN-Station Webservice. Furthermore, metadata of seismic waveforms and waveform data streams from Orfeus stations are provided.

A list of seismic events is offered by *EMSC* (European-Mediterranean Seismological Centre) via the FDSN-Event Webservice. In addition, the *FDSN-Event Webservice of AHEAD* (European Archive of Historical Earthquake Data) provides data for historical seismic events. Currently waveform data is provided via FDSN-Dataselect for stations of the Orfeus network.

A special filtering option, at least for seismological events, is the limitation of the result list for specific range of crustal thickness. Due to the lack of Webservices, which provide such information, the data are taken directly from a local file with 1° tiles over whole Europe, i.e. europe is devided in squares with an edge length of one degree. The Crust-Information have been taken from public domain scientific publication[3] in order to demonstrate the possibility of integrating scientific published products. This is an example of how additional services could be included in the general ICS framework, if provided at TCS level.

GNSS data are integrated into the prototype via webservices of the GSAC repository software. Through GSAC station metadata and geodetic dataset files are made accessible.

The application is developed in the Model View Controller (MVC) architectural pattern using Javascript with the frameworks Sencha Ext, GeoExt, and Openlayers. The directory structure of the project follows the recommendations suggested by Sencha, i.e. depending on the tasks executed the files are in the directories "app/view/", "app/controller", or "app/store". A class chart visualizing dependencies and interactions can be downloaded as PDF[4] or graphML[5] to do additional layout tasks with yEd Graph Editor.

---

[3] Laske et al. (2013) "Update on CRUST1.0 - A 1-degree Global Model of Earth's Crust"
[4] PDF: https://drive.google.com/file/d/0B1zhFTnN5cUPZElNSlZ5b2Q5QlE/view?usp=sharing

[5] graphML: https://drive.google.com/file/d/0B1zhFTnN5cUPOTk4a2FVa1NKQVU/view?usp=sharing

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

14

3.1.3 Metadata Catalogue

The CERIF metadata catalogue holds information about Research Infrastructures (RI), persons, assets, data, budget, and services, as described in a previous section. It was built a) migrating data from the RIDE database, b) manually entering new records and c) automatically retrieving results of interest. To store the information provided by RIDE, the vocabulary of the CERIF schema had to be modified to store spatial information of equipment and facilities.

In addition to the information originating from RIDE, information about webservices operated by a EPOS TCSs are stored in the CERIF catalogue. A general description od the new entities is provided in a previous section. The detailed description of this new implementation of CERIF, however, are out of the scope of this document: a brief description can be found in Appendix I. Further**,** for the purpose of better handling metadata for integration, a set of common metadata elements of the services included in the demonstrator was created: this work, an undergoing process that is subject to change within the first half of the Implementation Phase when new services will be made available to the ICS, is also reported in **Appendix II.**

3.2 Implemented resources

In this section we provide a list of the resources included so far in the EPOS Demonstrator. Such resources include a number of different institutions and data centres across Europe.

3.2.1 Base Maps

KNMI

- Demonstrator Resources
  - Base Map http://geoservices.knmi.nl/cgi-bin/worldmaps.cgi?

OSM-WMS

- Demonstrator Resources
  - OSM-WMS worldwide (http://www.osm-wms.de) http://129.206.228.72/cached/osm?

3.2.2 Seismology

Waveform Data

ORFEUS

- Waveform Data
- Contact person: Luca Trani, KNMI, Luca.Trani@knmi.nl
- http://www.orfeus-eu.org/
- Demonstrator Resources
  - Station Metadata http://www.orfeus-eu.org/fdsnws/station/1/query?
  - Orfeus Metadata for Waveforms http://www.orfeus-eu.org/odcws/wfmetadataselect/1/query
  - Waveform Data http://www.orfeus-eu.org/fdsnws/dataselect/1/query?

EIDA

- Waveform Data

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

15

- Contact person : Angelo Strollo, GFZ, strollo@gfz-potsdam.de
- http://www.orfeus-eu.org/eida/
- Services:
  - EIDA (Waveforms + Metadata)
    - Unstandard Webservice from EIDA http://eida.gfz-potsdam.de/webdc3/ offering metadata about all EIDA stations and wavforms of all stations.
  - EIDA Nodes: providing all metadata but only Data of certain Node
    - ORFEUS complete set of EIDA metadata via FDSN Webservice and only ORFEUS Waveforms
    - GEOFON only GEOFON metadata + waveforms (via FDSN)
- Demonstrator Resources
  - see ORFEUS (station metadata)

SISMOS
- Waveform Data -Historical Data
- http://sismos.rm.ingv.it/

http://sismos.rm.ingv.it/

SMdB
- Waveform Data
- Contact Person: Lucia Luzi, INGV, luzi@mi.ingv.it

Event Data

EMSC
- Earthquake Products
- Services described on homepage
- http://www.emsc-csem.org/#2
- Services:
  - http://www.seismicportal.eu/jetspeed/portal/web-services.psml
  - Events (SOAP – QuakeML); [1]
  - Events (REST – QuakeML/JSON); [2]
  - Waveforms/Seismolink (SOAP & REST – WSDL); [3]
- Demonstrator Ressources
  - FDSN-EVENT http://www.seismicportal.eu/fdsnws/event/1/query?

Further Products

AHEAD
- Earthquake Products - Historical Data
- Contact Person: Mario Locati, INGV - mario.locati@mi.ingv.it
- http://www.emidius.eu/AHEAD/
- Online portal available; data linked to publications; historical data from 1000-1899
- Services:
  - http://www.emidius.eu/AHEAD/main/
- Demonstrator Ressources
  - FDSN-EVENT http://www.emidius.eu/services/fdsnws/event/1/query?

EFEHR

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

16

- Hazard & Risk
- Contact Person: Philip Kästli, ETH, kaestli@sed.ethz.ch
- http://www.efehr.org:8080/jetspeed/portal/
- Services described online, as maps and descriptions
  - http://www.efehr.org:8080/jetspeed/portal/services.psml;jsessionid=F9B7 F47FABDEF4071F782DBC35E0E27E
  - Documentation – WADL; (Web Application Description Language)
    - human readable: http://appsrvr.share-eu.org:8080/share/documents/wadl_doc.html
    - machine readable: http://appsrvr.share-eu.org:8080/share/hazardmap_models.wadl
  - Request URLs :[base-url]/[product]?lat=[latitude]&lon=[longitude] → for products: (curve, map, spectra) and base-url:
    - http://appsrvr.share-eu.org:8080/share/
    - Request for design-spectra: [base-url]/ design-spectra/design_spectra_elastic_ec? lon=[longitude]&lat=[latitude]&spectratype=[smallmag|bigmag]&imt =[SV|SA]&design-acceleration=[acceleration]
- SHARE Results Portal:
  - http://www.efehr.org:8080/jetspeed/portal/Hazard[products].psml (Portal Structure Markup Language)
- Offers sticking to following formats/standards:
  - QuakeML
  - NRML
  - Restful web services
- Private-services:
  - http://www.seismo.ethz.ch/eq/latest/index_EN
  - Example: http://quake.ethz.ch/rdfdemo?r=smi:ch.ethz.sed/sc3rg/origin/NLL.[SMI (Seismological Metadata Identifier) e.g. 20140116172827.628545.50083]
  - Metadata Service for different formats (base URL: http://quake.ethz.ch/metadata?r=[SMI]):
    - http://quake.ethz.ch/metadata?r=smi:ch.ethz.sed/sc3rg/origin/NLL. 20140116153824.100521.34854
    - http://quake.ethz.ch/metadata?r=smi:ch.ethz.sed/sc3rg/origin/NLL. 20140116153824.100521.34854&f=rdf-xml
    - http://quake.ethz.ch/metadata?r=smi:ch.ethz.sed/sc3rg/origin/NLL. 20140116153824.100521.34854&f=rdf-ttl
    - http://quake.ethz.ch/metadata?r=smi:ch.ethz.sed/sc3rg/origin/NLL. 20140116153824.100521.34854&f=rdf-json
- Demonstrator Resources
  - Hazard Map http://gemmsrvr.ethz.ch/cgi-bin/mapserv?MAP=/var/www/mapfile/sharehazard.01.map&

    ○   Borders   http://gemmsrvr.ethz.ch/cgi-bin/mapserv?map=/var/www/mapfile/worldvector.map&

http://gemmsrvr.ethz.ch/cgi-bin/mapserv?map=/var/www/mapfile/worldvector.map&

3.2.3 Geology
EPOS Working Group 3
OneGeology Europe

- Contact Person: Lucie Kondrova, CGS, lucie.kondrova@geology.cz
- http://www.onegeology-europe.org/
- Services: Images (WMS) , GeoSciML (WFS), other (W*S)
- Demonstrator Resources
    - Geology http://www.bgr.de/Service/OneGeology/BGR_Geological_Units_IGME5000/
    - Faults http://mapdmzrec.brgm.fr/cgi-bin/mapserv?map=/carto/ogg/mapFiles/GISEurope_Bedrock_and_Structural_Geology.map&

http://mapdmzrec.brgm.fr/cgi-bin/mapserv?map=/carto/ogg/mapFiles/GISEurope_Bedrock_and_Structural_Geology.map&

3.2.4 Geodesy
GSAC repositories

Repositories to different providers, all are using same infrastrucutre GSAC:

- Demonstrator Resources
    - INGV http://bancadati2.gm.ingv.it:8081/gsacring/
    - Geo Azur http://epos.unice.fr:8080/renagbgsac
    - INOGS (FredNet) http://www.crs.inogs.it/frednet/ItalianSite/XFReDNetHome.htm
    - Univ Athens http://dionysos.survey.ntua.gr:8080/ntuagsac/gsacapi/
    - NOA Greece http://194.177.194.238:8080/noanetgsac/gsacapi/

http://194.177.194.238:8080/noanetgsac/gsacapi/

3.2.5 Supersites
Futurevolc

Still not accessed, but in future close cooperation planned:

- GSAC repository for storing geodetic information of Iceland
    - will be provided as soon as possible
    - contact person: Benedikt Gunnar Ofeigsson, IMO, <bgo@vedur.is>
    - own master database, and fill the gsac with it
    - active directory LDAP service to authenticate
- WIND DATA:
    - a portal showing calculated points from a model with nice interface
    - generate reports and download data

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

18

- WATER CONSUMPTION:
  - map with points representing users consuming water (users, utiliy company)
  - faceted searches
- Developments based on Microsoft Windows Server (Active Directory)

MAR-SITE

Still not connected

- OGC compliant Webservices (WMS, WFS, SOS)
- ISO Metadata catalogue
  - registering in MOIST database for use in GEO-GEOSS

3.2.6 General Resources
VERCE

- Computational
- Contact person: Alessandro Spinuso, KNMI, alessandro.spinuso@knmi.nl
- http://www.verce.eu/
- Demonstrator Resources
  - base source code

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

19

4. From prototype to ICS-C and EPOS-IP

As already mentioned, the EPOS Integrated Core Services (ICS) represent the novel e-infrastructure which will provide an integrated environment enabling scientists, citizens, policy makers and other stakeholders to access to multidisciplinary data and data products, synthetic data from simulations, and most importantly to use processing and visualization tools. The ICS are composed by the ICS-Central Hub (ICS-C) and distributed computational resources (ICS-d). ICS-d may be (a) additional computing/storage/detector array facilities outside the scope of EPOS; (b) nodes providing general software services used across all TCS such as input/validation, data management, analytics, simulation, mining, visualisation; (c) replicates / mirrors of ICS-C in distributed locations for resilience and performance.

In this new phase of EPOS, the Implementation Phase (IP), the ICS play a key role and are going to demonstrate their maturity and sustainability. Indeed, while in EPOS PP the final twofold goal was the prototype and the design of the e-infrastructure, in EPOS IP the priority is the implementation of the ICS-Central Hub. However, in parallel to this, the design control of the ICS-d and the management of implementation of input/validation, data management, analytics, simulation, mining, visualisation requirements (derived from use cases of the TCS and others) will be done. Similarly, management of access to other computing facilities (e.g. HPC), to services for digital preservation, etc will be managed by ICS and implemented jointly by ICS and TCS.

In EPOS IP, Thematic Core Services are supposed to build e-Infrastructures which are 1) internally coherent; 2) agreed upon by the respective community; 3) ICS-interoperable using the standards defined by ICS and agreed with TCS. ICS and TCS should be of course strictly connected and highly interoperable.

However, in the implementation and construction phase, a clear vision and definition of competencies and tasks is of primary importance. In this framework, the ICS governmental framework will be structured to have a triple function:

1. **Provide resources to TCS in order to help them building the services they need**. This action will be carried on with different means according to TCS maturity. Some TCS may indeed require starting a service from scratch, thus needing suggestions about technologies and process workflow to be used. Other TCSs may have an higher level of maturity, thus needing only support to build the interfaces to interoperate with ICS-C.
2. **Harmonize and scale TCS services**. Services built by one TCS may in many case be reused by other TCSs, but it require some effort in the phase of design and implementation in order to build scalable and re-usable services. ICS will provide knowledge and resources to scale and harmonize TCS services, both existing and envisaged.
3. **Provide processing, visualization and data management tools and initiatives, interfacing ICS-d with TCS.** One of the most challenging goals of EPOS is trying to harmonize existing initiatives and big analytics/visualization infrastructures so that they can be used by several TCS in an integrated way. ICS WPs propose, therefore, to interface all such tools to all the TCS. It means, for instance, the building of proper software interfaces that can enable

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

20

seismology or geodetic scientists to distribute their workflow over a number of distributed resources.

Further, the construction of the ICS, and in particular of the ICS-C, will follow two main concepts:

a) **TCS support**, which implies the support that the ICS will provide to the development of the TCS. Such support will be provided in terms of software developments, guidelines, provision of computational and other technological resources, knowledge transfer, etc.

b) **Coherency,** this ensures that the whole development of the Services is coherent a) internally (within the TCS), b) horizontally (with other TCS), c) vertically (with the ICS) and d) externally (with other projects or initiatives). We call it "ICS Coherency".

ICS Coherency and TCS support will be the base for the construction of solid, reliable, sustainable and technological leading edge Integrated Core Services Central-Hub (ICS-C

5. Long-term technical sustainability

The architecture is designed for sustainability. As indicated above the NRIs and TCSs are funded independently and have their own sustainability plans. The ICS architecture consists of ICS-C a central hub, with portal and metadata catalog, and ICS-d multiple nodes providing centrally-approved services. ICS-d are distributed because they include existing data centres, supercomputers and systems or software services. NRIs or TCSs can provide an ICS-d but equally centres of expertise can offer to be an ICS-d for EPOS.

The architecture utilises current, leading edge concepts (but not bleeding edge – they must be reliable) gathered from EC and national ICT R&D especially in the areas of e-Research and VREs (Virtual Research Environments) linked with Big Data nd Open Data. ICS-C is based on the metadata catalog. This insulates the portal from the fast-changing landscape of NRIS, TCSs, ICS-d and their offerings and characteristics. By placing all the relevant metadata information (for discovery, contextualisation and utilisation) in the catalog it is exposed to services rather than being hidden inside software such as used in classical brokering. Furthermore, the catalog can be updated either from ICS-C (central control) or from the NRI, TCS or ICS-d acting on the metadata records describing their offerings and characteristics (distributed control with responsibility lying with the service provider). This ensures (under appropriate security measures) great resilience. It also provides great flexibility in ensuring the ICS-C 'view of the world of interest' is current and complete.

ICS-C will be constructed during EPOS-IP utilising the experience gained from EPOS-PP. The validation of the architecture in EPOS-PP provides confidence for scaling up to full-scale production. Following EPOS-PP, the architecture uses 'standards' whenever possible particularly those of W3C and of the GRIDs/CLOUDs community. The metadata standard used is an EU Recommendation to Member States. Thi ensures maximal interoperability. After the implementation phase (EPOS-IP) the maintenance and further development of ICS will be the responsibility of the ERIC.

APPENDIX I: CERIF IMPLEMENTATION

To model the TCS webservice information with CERIF, the new entity "ServiceInterfaceDescrip-tion" and the associated linking entity "ServiceInterfaceDescription_Srv" was developed to store EPOS product types, product levels, and research domains.

The ServiceInterfaceDescription holds the service endpoint (url), geographic projection, geographic coordinate reference system, the access method and the return format. The access method could be a standard like FDSN-Station, WFS, WMS, or a custom method. A return format might be quakeML, FDSN-StationXML, octet-stream, or a custom format. The ServiceInterfaceDescription_Srv linking entity is modeled in line with the CERIF linking entities. It stores the attribute SrvIntId of Ser-viceInterfaceDescription and the attribute cfSrvId of the CERIF entity cfSrv. In addition, it stores cfClassId and cfClassSchemeId to characterize the service for a given time frame with the attributes cfStartDate and cfEndDate. EPOS vocabulary was added to build a thesaurus of research domains, a thesaurus of return formats, and to specify EPOS data product levels of webservices.

The definition of the database tables for the new entities and the new linking entities are as follows:

ServiceInterfaceDescription (
char (128) SrvIntId,
text url,
char (128) projection,
char (128) ref_coord,
char (128) standard,
text returntype) PK (SrvintId);

ServiceInterfaceDescription_Srv (
char (128) cfSrvId,
char (128) SrvIntId,
char (128) cfClassId,
char (128) cfClassSchemeId,
timestamp cfStartDate,
timestamp cfEndDate)
PK (cfSrvId,SrvIntId,cfClassId,cfClassSchemeId,cfStartDate,cfEndDate);

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

22

APPENDIX II: COMMON METADATA ELEMENTS

The following metadata elements were identified to be common throughout all connected services. Naturally there is a set of mandatory elements - without them storing data in an catalogue system makes no sense or is impossible. A set of optional metadata allows better presentation or navigation of datasets and is thus recommended. Elements in bold font we already selected - elements in normal font are candidates. In addition the services require to be separated into two groups. The "Research access" group offers direct access to research data, i.e. a PDF download or a seismic data stream. The "Discovery" group is used to narrow the search process down, i.e. a service providing information about seismic stations or gnss stations.

- Mandatory
  - ID
  - Title
  - Producttype: Research access or Discovery
  - Keywords
  - Rights
- Optional
  - Class (internal EPOS classification of dataproducts)
  - Description
  - Domain
  - Format/Standard (e.g. SEED, RINEX, XLS, PDF ..)
  - ResourceType (Datacite Vocabulary: Event, Image, Model, Service, Other, Text, .)
  - Geographic Projection / Coordinate Reference System (EPSG Code)
  - Geographic BBox
  - Start Time
  - End Time
  - Size (maybe category KB/MB/GB/TB)
- Access Services
  - FDSN dataselect
  - RINEX FTP link
- Discovery Services
  - FDSN station
  - FDSN event
  - GSAC station
  - GSAC file

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

23

Table 1: The new cfClassSchemes needed by EPOS - red indicates new EPOS vocabulary.

cfClassScheme
EPOS_Product_Type
EPOS_Product_Level
SITE_OF_OPERATION

Table 2: The new cfClass - cfClassSchemes groupings needed by EPOS to store webservices - red indicates new EPOS vocabulary

cfClass	cfClassScheme
Seismology	Research_Output
Geodesy	Research_Output
MSEED	EPOS_Product_Type
JSON	EPOS_Product_Type
XML	EPOS_Product_Type
CSV	EPOS_Product_Type
0	EPOS_Product_Level
1	EPOS_Product_Level
EPOS_TechnicalDescription	IDENTIFIER_SERVICE_ROLES

Table 3: The cfClass - cfClassSchemes groupings needed to migrate RIDE metadata into the CERIF catalog - red indicates new EPOS vocabulary.

cfClass	cfClassScheme
FIXEDPHONE	ELECTRONIC_ADDRESS_TYPES
EMPLOYEE	PERSON_ORGANISATION_ROLES
CONTACT	PERSON_PROJECT_ENGAGEMENTS
EMAIL	PERSON_CONTACT_DETAILS
PHONE	PERSON_CONTACT_DETAILS
PRESENTED_NAME	PERSON_NAMES
COORDINATOR	ORGANISATION_PROJECT_ENGAGEMENT
PART	ACTIVITY_STRUCTURE

RESPONSIBLE	PERSON_RESEARCH_INFRASTRUCTURE_ROLES
ELECTRONIC_ADDRESS	CERIF_ENTITIES
ORGANISATION_LEGAL_POSTAL_ADDRESS	ORGANISATION_CONTACT_DETAILS
ORGANISATION_FINANCIAL_POSTAL_ADDRESS	ORGANISATION_CONTACT_DETAILS
PART	INTER_FACILITY_RELATIONS
PROVISION	RESEARCH_INFRASTRUCTURE_RELATIONS
POSTAL_ADDRESS	SITE_OF_OPERATION
MONITORING_AREA	SITE_OF_OPERATION
FUNDER	RESEARCH_INFRASTRUCTURE_FUNDING_ROLES
OPERATING_COSTS	RESEARCH_INFRASTRUCTURE_COSTINGS
OWNER	ORGANISATION_RESEARCH_INFRASTRUCTURE_ROLES

EUROPEAN PLATE OBSERVING SYSTEM   phone +39 06 51860.401 / 577 / 652 / 636 | fax +39 06 51860565 | epos@ingv.it | epos.secretariat@ingv.it

25