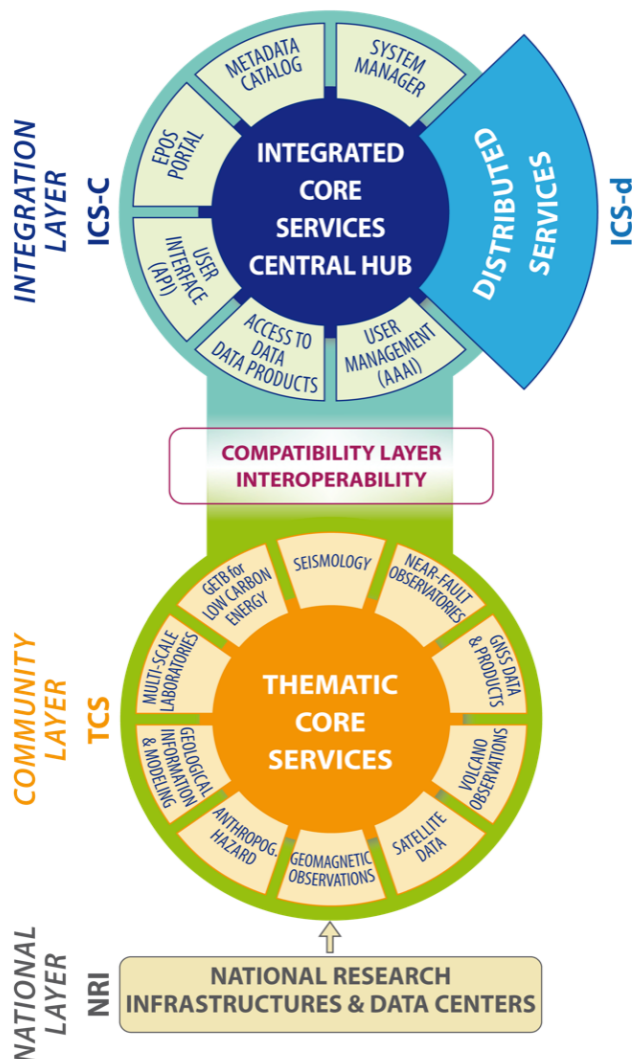


General requirements for EPOS

Context of general requirements in EPOS

Complete report at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/223/attachments/186/download>

EPOS is a long-term plan for the integration of RI for Solid Earth Science in Europe. Its main aim is to integrate communities to make scientific discovery in the domain of solid earth science. EPOS integrates the existing (and future) advanced European facilities into a single, distributed, sustainable infrastructure taking full advantage of new e-science opportunities.



The EPOS community is organized in

1) National Layer, made up of Research Infrastructures (RIs) providing data and services. 244 RI, 128 Institutions, 22 countries, 2272 GPS services, 4939 seismic stations, 464 TB Seismic data, 828 instruments in 118 Laboratories.

2) Community Layer, made up of pan European e-Infrastructures which disseminate data and services of a single discipline (e.g. seismology with ORFEUS/EIDA), These are the Thematic Core Services (TCS) of EPOS.

3) Integration Layer, (Integrated Core Services - ICS) the e-Infrastructure designed and operated by EPOS; this is the place where the integration of data and services provided by the Thematic Core Services (TCS, Community Layer) occurs. This layer is what EPOS is building. It has several layers because it has several communities.

In this complex framework the major challenge is to enable scientists to make use of multidisciplinary data, which are usually heterogeneous in terms of formats, metadata and accessibility.

A general use-case of EPOS could be seen when an earth-scientist wants to study the earth dynamics of a natural hazard event, like an earthquake or a volcanic eruption, by using different types of real (recorded) data and modeled data. Some examples of real data are seismology data (e.g. seismic waveforms); geological maps; GPS data (GPS velocity and displacement) or satellite data (e.g. SAR interferograms, integrated satellite products). Scientists can compare the real data that has been recorded for a particular hazard event with the modeled data.

Summary of EPOS general requirements

Now, more specifically, EPOS can identify two basic uses cases:

- A basic multidisciplinary use case, dealing with the discovery of heterogeneous data by a user who connects to the ICS-C (Integrated Core Services Central Hub) portal to discover and access (e.g. download) such data.
- An extended, single discipline, computational oriented use case, dealing with the usage, from user's side, of a computational seismology or geodesy tool which orchestrates the access to data and to computational resources on behalf of the user

Scientists by using EPOS could:

- Make integrated use of SAR, GPS, Accelerometric Data, etc.
- Use different codes and languages (python, fortran, any other...)
- Perform heavy processing online (use of HPC resources)
- Compare results (e.g. focal mechanism catalogues)
- Compare different data
- Save data in personal area
- download the data

Detailed requirements

How data is acquired, curated and made available to users depends on the community. The real recorded data is acquired by using different sensors (e.g. seismic stations, gps receivers, satellites, chemical sensor, geomagnetic sensors). The curation and availability depends on the specific domain community. For example the GPS community stores the data in a single side server and used to share it by FTP. Seismologists have a more mature system, where each institution stores the data in their local servers. The data is backed up regularly and those repositories are federated. The data is made available for the users by the ICS interface, which will be a GUI (website or portal). And the metadata will be available in different formats, like RDF export (ENVRI), OAI-PMH, CKAN, opensearch (EUDAT) and other standards. For registering and citing data or publications, EPOS will use PID system because DOI can be uniquely referenced, and we could assign a PID at data creation times. The data from TCS services has to be available in a reasonable amount of time. A user normally connects with the ICS. The ICS on behalf of the user fetches data and metadata from the TCS. Therefore, the TCS has to react in a reasonable time.

In EPOS there are different software that scientists made use, like community libraries (e.g Obspy), workflow systems (e.g disep4py, ER-FLOW), and HPC/Cloud resources (e.g. SuperMuc and CINECA). The users have full responsibility of the results produced by EPOS platform. EPOS does not guaranty the *trustability* of the results.

EPOS might have interactions with other RIs to access some computational services, like SuperMUC or CINECA, but always staying at the scope of the environment science.

EPOS follows the open access policy. Therefore, most of the data is available for any registered users. However, it is required a login for measuring the impact of the data used. Small portion of the data might be available under special conditions, for example, after the embargo period (6 months for writing papers) or paid data.

EPOS has different software libraries for building their own systems. EPOS also provides software libraries for analyzing data. For example workflow software, called dispel4py and its provenance recording system, is an open-source library and users in ENVIRplus can use it.

EPOS sets up policies for regulating the transnational access (TNA), which is the access to the data from the laboratories instruments. Basically, this means that scientists could go to another laboratory for doing an experiment with a certain type of instrument.

EPOS makes the technical reports public and available through different project websites.

Regarding data management and its exploitation, EPOS is using CERIF metadata model (Common European Research Information Format [1], an integrated metadata model to build the integrated core-services. At community level (TCS), users are free to use any standards as long as the data is accessible and discoverable by the ICS. However, EPOS is currently deciding in to change the currently standards, and it is open to EUDAT solutions.

EPOS needs to improve the Interoperable AAI system (federated & distributed), taking already existing software and make it available and scalable across communities. It purposes is to authenticate and authorize users, and provide a transparent access to TCS and ICS-D data and services.

EPOS does not have the data management plan available yet.

There are different non-functional constraints depending on the ICS or TCS layer, like maintaining cost, capital, and operational. At national layer the funding comes from the government, and at the TCS layer they come part from EPOS and part from the government. At ICS, the funding comes from EU and from fees of the member states (remember that EPOS is becoming an ERIC, and all the member states have to pay a fee).

Regarding to the security and access approach, users access to the ICS in a secure way, which means login and password with all the existed credentials, which could be a certificate, an OpenID google account or a typical registration of the user. The ICS satisfies the user request accessing to appropriate resources (thematic cores services and/or computational services).

EPOS has 85 % of the data open. Only for a small amount of data is not open, which is subject to an embargo period (6 months) or paid data.

References:

1. <http://www.eurocris.org/>

Formalities (who & when)

Go-between	Rosa Filgueira
RI representative	Daniele Bailo
Period of requirements collection	From September to November
Status	Finished