



The EPOS e-Infrastructure

Keith Jeffery (1) and Daniele Bailo (2)

(1) Shrivenham, United Kingdom (keith.jeffery@stfc.ac.uk), (2) INGV, Roma, Italy (daniele.bailo@ingv.it)

The EPOS e-Infrastructure

SM1.3/ESSI2.14/GD8.7/GI1.6/GMPV56/TS1.5 entitled Integrated Research Infrastructures and Services to users: supporting excellence in a science for society

Keith G Jeffery¹, Daniele Bailo²

(1) Keith G Jeffery Consultants, UK (keith.jeffery@keithgjefferyconsultants.co.uk)

(2) Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy (daniele.bailo@ingv.it)

The European Plate Observing System (EPOS) is integrating geoscientific information concerning earth movements in Europe. We are approaching the end of the PP (Preparatory Project) phase and in October 2014 expect to continue with the full project within ESFRI (European Strategic Framework for Research Infrastructures). The key aspects of EPOS concern providing services to allow homogeneous access by end-users over heterogeneous data, software, facilities, equipment and services. The e-infrastructure of EPOS is the heart of the project since it integrates the work on organisational, legal, economic and scientific aspects. Following the creation of an inventory of relevant organisations, persons, facilities, equipment, services, datasets and software (RIDE) the scale of integration required became apparent. The EPOS e-infrastructure architecture has been developed systematically based on recorded primary (user) requirements and secondary (interoperation with other systems) requirements through Strawman, Woodman and Ironman phases with the specification – and developed confirmatory prototypes – becoming more precise and progressively moving from paper to implemented system.

The EPOS architecture is based on global core services (Integrated Core Services – ICS) which access thematic nodes (domain-specific European-wide collections, called thematic Core Services - TCS), national nodes and specific institutional nodes. The key aspect is the metadata catalog. In one dimension this is described in 3 levels:

- (1) discovery metadata using well-known and commonly used standards such as DC (Dublin Core) to enable users (via an intelligent user interface) to search for objects within the EPOS environment relevant to their needs;
- (2) contextual metadata providing the context of the object described in the catalog to enable a user or the system to determine the relevance of the discovered object(s) to their requirement – the context includes projects, funding, organisations involved, persons involved, related publications, facilities, equipment and others, and utilises CERIF (Common European Research Information Format) standard (see www.eurocris.org);
- (3) detailed metadata which is specific to a domain or to a particular object and includes the schema describing the object to processing software.

The other dimension of the metadata concerns the objects described. These are classified into users, services (including software), data and resources (computing, data storage, instruments and scientific equipment).

An alternative architecture has been considered: using brokering. This technique has been used especially in North America geoscience projects to interoperate datasets. The technique involves writing software to interconvert between any two node datasets. Given n nodes this implies writing $n*(n-1)$ convertors.

EPOS Working Group 7 (e-infrastructures and virtual community) which deals with the design and implementation of a prototype of the EPOS services, chose to use an approach which endows the system with an extreme flexibility and sustainability. It is called the Metadata Catalogue approach. With the use of the catalogue the EPOS system can:

1. interoperate with software, services, users, organisations, facilities, equipment etc. as well as datasets;
2. avoid to write $n*(n-1)$ software convertors and generate as much as possible, through the information contained in the catalogue only n convertors. This is a huge saving – especially in maintenance as the datasets (or other node resources) evolve. We are working on (semi-) automation of convertor generation by metadata mapping – this is leading-edge computer science research;
3. make large use of contextual metadata which enable a user or a machine to:
 - (i) improve discovery of resources at nodes;
 - (ii) improve precision and recall in search;
 - (iii) drive the systems for identification, authentication, authorisation, security and privacy recording the relevant attributes of the node resources and of the user;

(iv) manage provenance and long-term digital preservation;

The linkage between the Integrated Services, which provide the integration of data and services, with the diverse Thematic Services Nodes is provided by means of a compatibility layer, which includes the aforementioned metadata catalogue. This layer provides 'connectors' to make local data, software and services available through the EPOS Integrated Services layer.

In conclusion, we believe the EPOS e-infrastructure architecture is fit for purpose including long-term sustainability and pan-European access to data and services.