W3C Linking Geospatial Data Conference March 5-6 London

Position paper

EPOS: An approach to Linked Open Spatial Data

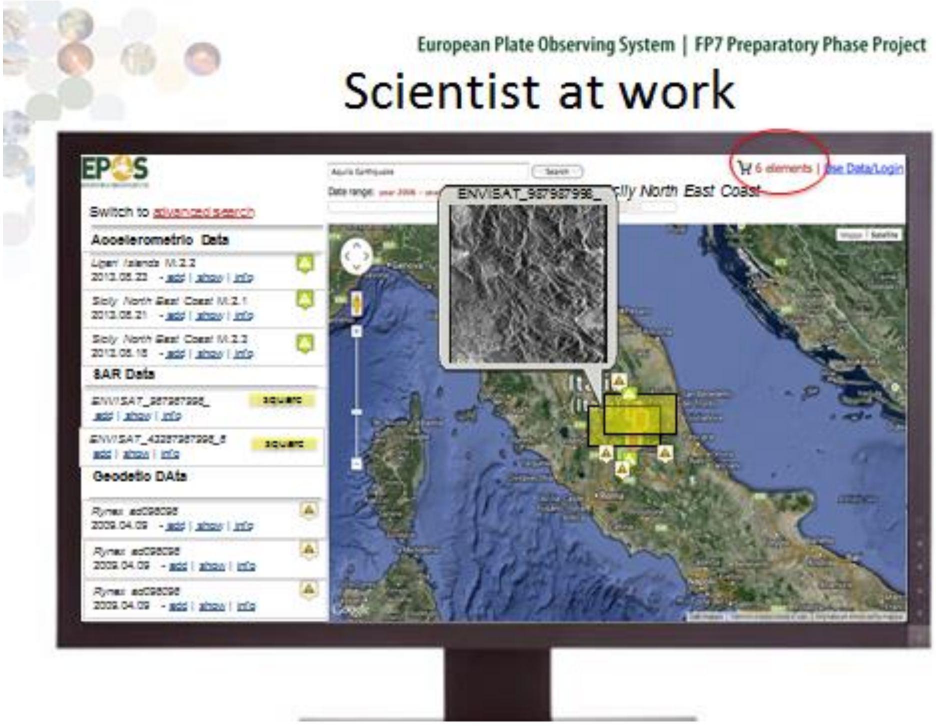
Keith G Jeffery (keith.jeffery@keithgjefferyconsultants.co.uk)

Abstract

An approach to providing linked open data in a geoscience environment is presented. The requirement is for homogeneous access to services, software, resources, users and datasets over heterogeneous provider nodes. The major part of the system is based on advanced heterogeneous distributed database technology using the catalog approach. However, a LOD/semantic web 'image' of the system is provided to allow access, browsing and integration in that environment.

Position Paper

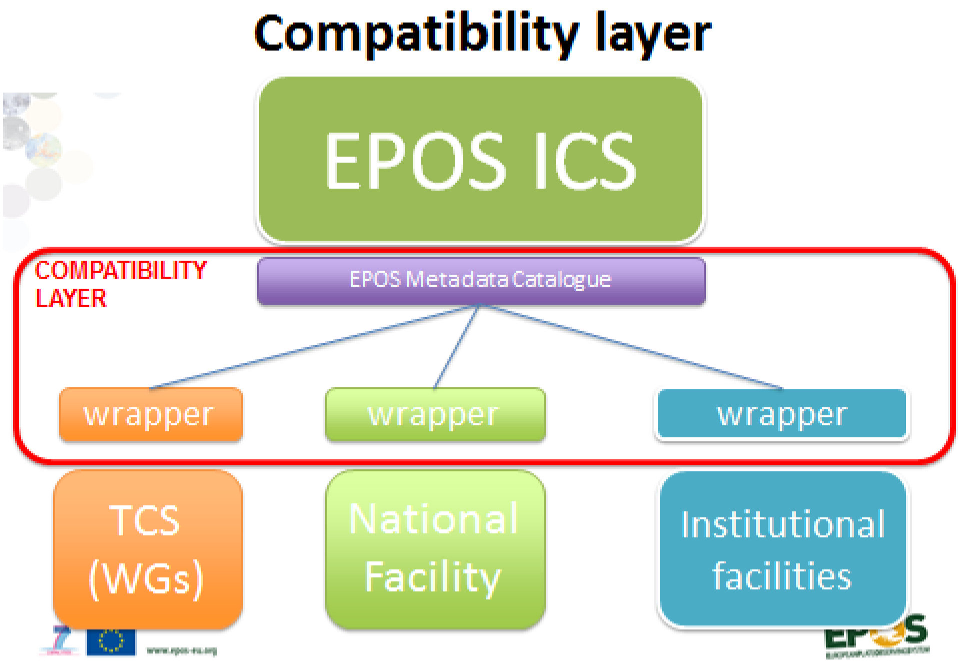
The European Plate Observing System (EPOS) is integrating geoscientific information concerning earth movements in Europe. We are approaching the end of the PP (Preparatory Project) phase and in October 2014 expect to continue with the full project within ESFRI (European Strategic Framework for Research Infrastructures). The key aspects of EPOS concern providing services to allow homogeneous access by end-users over heterogeneous data, software, facilities, equipment and services. Simple access from a portal to datasets – even preceded by discovery metadata – is insufficient and too human-intensive. The end-user requirement is for integration of datasets with appropriate software and integration of several datasets to provide a richer understanding within a workflow in an e-Research (a.k.a. e-Science) environment. This implies 'mash-ups' but in as automated a way as possible for scalability and sustainability. The key 'mash-up' integrators identified in user requirements analysis are geospatial and temporal coordinates.


European Plate Observing System | FP7 Preparatory Phase Project

Scientist at work

EPOS
Switch to advanced search
Accelerometric Data
Lipari Islands M.2.2
2013.05.23 - add | show | info
Sicily North East Coast M.2.1
2013.05.21 - add | show | info
Sicily North East Coast M.2.3
2013.05.16 - add | show | info
SAR Data
ENVISAT_987987998_
add | show | info
square
ENVISAT_43287987998_8
add | show | info
square
Geodetic DAta
Rynex ad098098
2009.04.09 - add | show | info
Rynex ad098098
2009.04.09 - add | show | info
Rynex ad098098
2009.04.09 - add | show | info
Aquila Earthquake
Search
Date range:
ENVISAT_987987998_
Sicily North East Coast
6 elements | Use Data/Login
Italia

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.

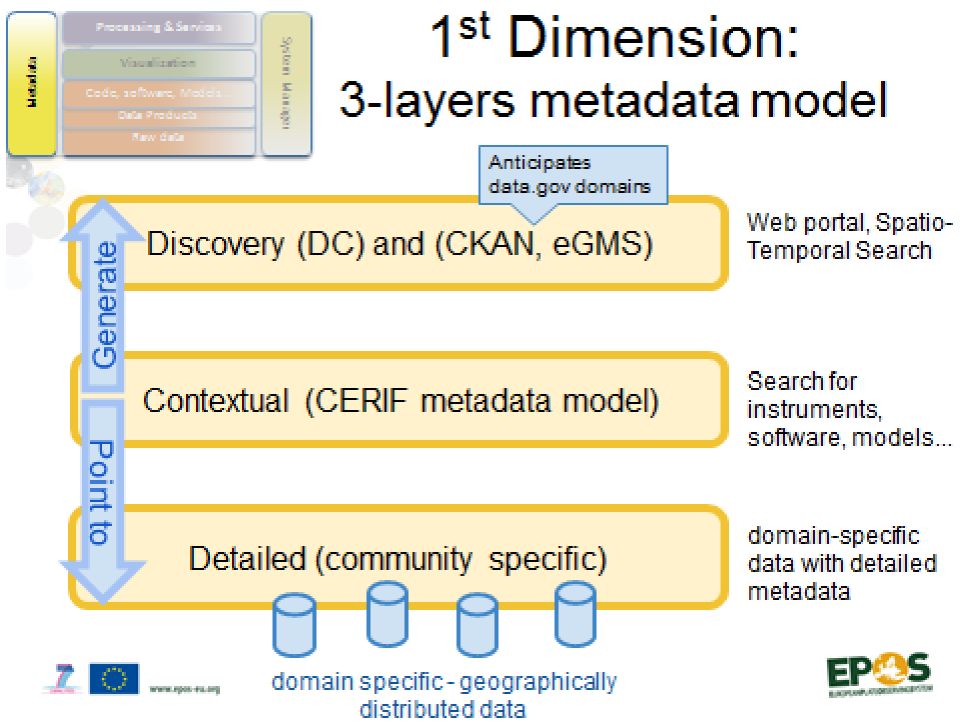


The EPOS architecture is based on global core services (Integrated Core Services – ICS) which access thematic nodes (domain-specific European-wide collections, called thematic Core Services - TCS), national nodes and specific institutional nodes. The key aspect is the metadata catalog. In one dimension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



1st Dimension:
3-layers metadata model

Anticipates data.gov domains

Discovery (DC) and (CKAN, eGMS)

Web portal, Spatio-Temporal Search

Contextual (CERIF metadata model)

Search for instruments, software, models...

Detailed (community specific)

domain-specific data with detailed metadata

domain specific - geographically distributed data

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;

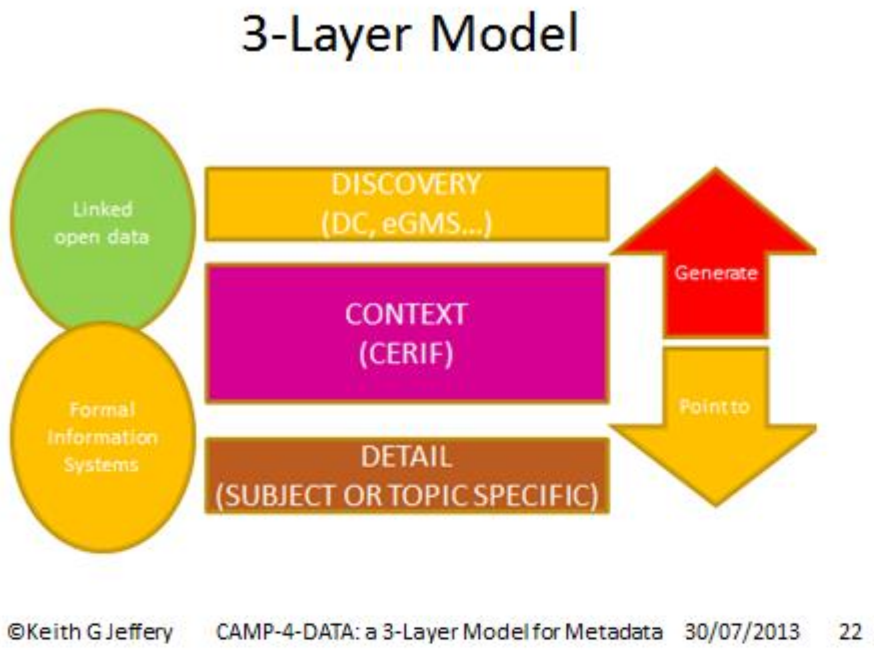
   (ii) improve precision and recall in search;

   (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' or 'convertors' as mentioned above to make local data, software and services available through the EPOS Integrated Services layer.

Thus far the architecture described is following the evolutionary line of research and development in heterogeneous distributed databases. However, because of work in the EC-funded ENGAGE project (which concerns open government data linked to research datasets and a social network approach to problem solving using the linked open data) this architecture provides also a linked open data / semantic web option. The metadata catalog is centred on CERIF. CERIF represents base entities/objects and linking entities/objects with role and temporal validity; basically a 'triple' structure as understood in logic processing or RDF. From CERIF we can generate and interoperate with other implementations of DC, DCAT, CKAN, eGMS, INSPIRE etc. The intercommunication is through RDF encoded as XML or native XML. However, to represent a relationship instance (tuple) between two base entities in CERIF requires several RDF statements. Thus a LOD 'image' is produced (either for permanent storage or generated 'on the fly') which can be used for browsing across the linkages between datasets. Better yet, CERIF has formal syntax and declared semantics. The semantic layer of CERIF (which is also multilingual) provides facilities for crosswalking and can be converted to / represented in OWL. Thus the architecture ensures there will be an 'image' of EPOS available in the LOD/semantic web (OWL) environment.



3-Layer Model

Linked open data

Formal Information Systems

DISCOVERY (DC, eGMS...)

CONTEXT (CERIF)

DETAIL (SUBJECT OR TOPIC SPECIFIC)

Generate

Point to

©Keith G Jeffery CAMP-4-DATA: a 3-Layer Model for Metadata 30/07/2013 22

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 via conventional ICT interfaces and also using LOD/semantic web technology.

Acknowledgement: the author acknowledges gratefully the work of colleagues on both EPOS and ENGAGE.