

# Cataloguing in EPOS

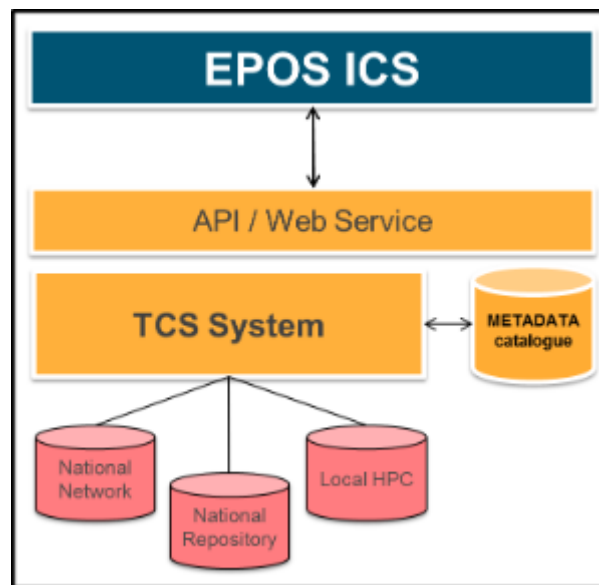
## Context of cataloguing in EPOS

Complete report on Cataloguing in EPOS available at: <https://envriplus.manageprojects.com/projects/requirements/notebooks/470/pages/42/comments/257/attachments/230/download>

## Summary of EPOS requirements for cataloguing

### Detailed requirements

EPOS requires to the different communities that belongs to TCS to use a catalogue (database) where each data object is referenced and described by the metadata. The data is different depending on the community (e.g. seismic waveforms, GPS time-series, geological maps). In general terms, EPOS distinguishes between two types of community data: the communities that provide files (or data objects) and the others that provide data-streams.



For ICS, EPOS uses a metadata catalogue (called CERIF) to keep tracks of the Research Information. The Research Information includes information about research entities such as people, projects, organisations, publications, patents, products, funding, or equipment, etc. and the relationships between them. Information

CERIF (developed in the Librarian domain and exported later in the enterprise domain) is a standard recommended by EU. Basically is an entity relation metadata model database, which can be defined as:

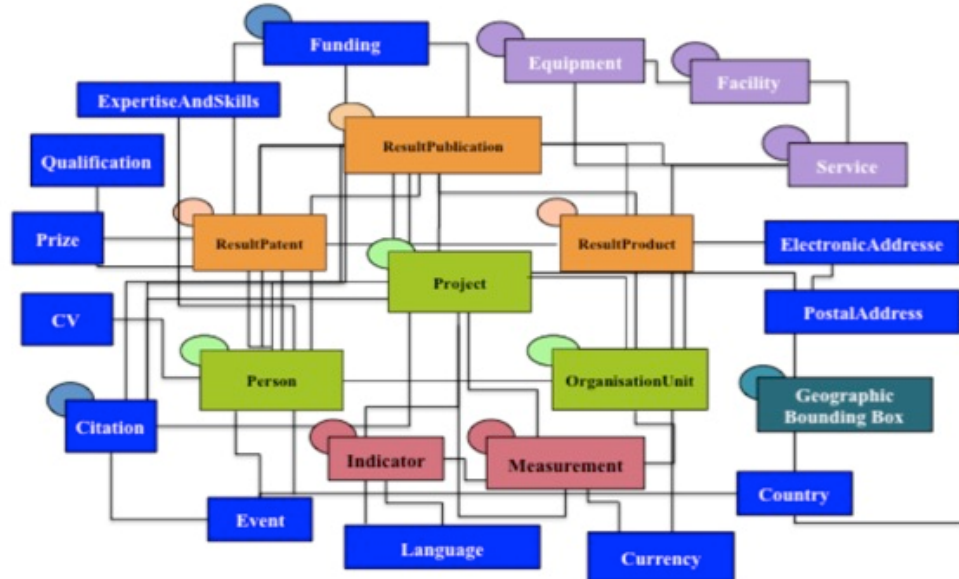
- A concept about research entities and their relationships – Specification (Conceptual Level) .
- A description of research entities and their relationships – Model (Logical Level).
- A formalization of research entities and their relationships – Database Scripts (Physical Level) .

See bellows figure, where different entities and their relationships (abstract view) are represented.

## Contextual (CERIF metadata model)

(<http://www.eurocris.org/>)

## Common European Research Information Format













### Item descriptions

CERIF (the Common European Research Information Format) is a formal conceptual model to support the management of Research Information, including the set up of and the interoperation between Research Information Systems. Its main features are described in [1]. The full data model (FDM) is introduced and specified here [2], and additional details can be found here [3].

It is very difficult to summarize the fields used, since CERIF is an ER database, which has many attributes.

CERIF supports semantics, since each term can be linked with a vocabulary. The structure and strength of the Semantic Layer as part of the CERIF model has been presented in the following document [4].

CERIF Entity Types		CERIF Features	
	Base Entities [base]		Multiple Language [lang]
	Result Entities [result]		Semantics [class]**
	2 <sup>nd</sup> Level Entities [2nd]		Geographics [geo]
	Link Entities [link]		Additional [add]
	Infrastructure Entities [infra]		Measurement [mm]

\*\* The CERIF Semantics additionally provides a vocabulary that is maintained outside of the CERIF data model and available from the public euroCRIS Website.

EPOS maintains cross-links and inter-links between catalogue items, and fields for item description and actual items, because it is an ER database.

Currently, EPOS uses CERIF data model and *Postgresql* to manage their metadata.

### Inputs

#### Human Inputs

At ICS level, EPOS requires different human inputs, like cvs, workflows validations. At TCS level, EPOS requires that each community provide: *data + metadata + API access the data*. However, different communities could have different requirements, different infrastructures and different maturity levels.

EPOS has defined two strategies to deal with the TCS metadata, one that requires human input and the other that requires machine input:

1. Metadata dump: the metadata from TCS is fully 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 to access the data. However, it requires periodic (e.g. daily) polling/copying procedures to ensure data dump. Also, in case metadata from one digital object (e.g. file) changes, synchronization mechanisms must be put in place to guarantee consistency. -à **HUMAN INPUT**
2. Metadata Runtime Access: The access to metadata is done runtime by querying the web services with the defined APIs. The APIs specification must be stored into 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 by the TCS. -à **MACHINE INPUT**

The reason behind these two strategies is that some communities want that EPOS copies their metadata, and other communities they want that EPOS accesses to their metadata at run-time through the web services.

### **Machine Inputs**

For populating catalogues, EPOS has a metadata runtime access strategy, which deals with TCS metadata (described just above).

Regarding with duplication, EPOS is currently thinking how to deal with. It is still a challenge.

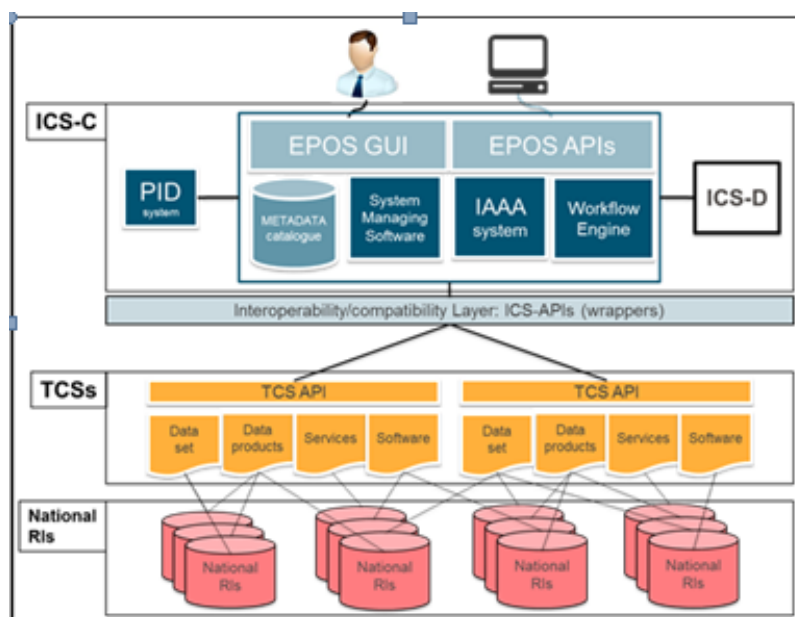
### **Outputs**

#### **Human Outputs**

EPOS has still to set up a feature for web discovery. It can be a multi-grid area.

EPOS uses a key performance indicator to keep track of the usability.

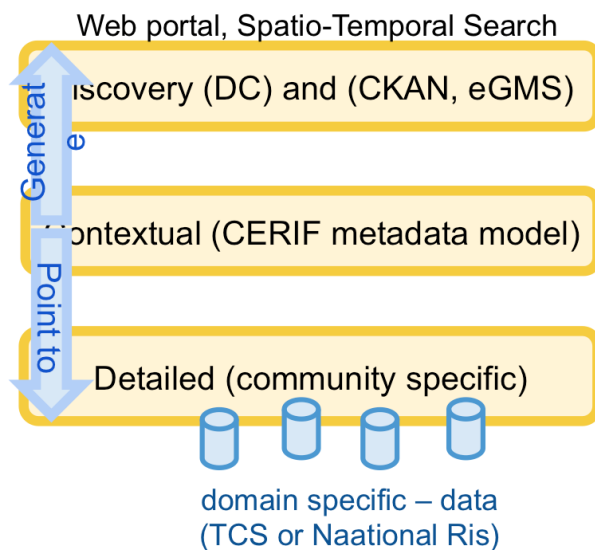
For reading the catalogue, users need to access to portal, and the portal make use of the catalogue. And users need authenticate to the portal in order to use it. Therefore, indirectly EPOS requires login and password to use the catalogue. But we have to remember that the catalogue is only one of the components of the portal. See bellows Figure to see the other components of the portal:



#### **Machine Outputs**

EPOS provides machines access to the catalogue through the portal. EPOS wants to give access to a number of APIs: e.g. OpenSearch, CERIF-API.

Regarding to fit in Applicable Regulations, EPOS will be complied to INSPIRE and GEOSS, because we can extract GEOSS and INSPIRE metadata from the catalogue itself. See bellows figure for further information.



## Metadata model

- 3-layers
- Manage the complexity (see previous slide)
- CERIF (formal syntax, declared semantics)

### Functions

- Discovery, selection
- Impact assessment
- Manage interoperability

### Issues (selection of)

- heterogeneity
- Many standards to be mapped from TCS
- Ontologies
- Complex work

#### References:

1. <http://www.eurocris.org/cerif/main-features-cerif>
2. [http://www.eurocris.org/Uploads/Web%20pages/CERIF1.3/Specifications/CERIF1.3\\_FDM.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF1.3/Specifications/CERIF1.3_FDM.pdf)
3. <http://www.eurocris.org/cerif/feature-tour/cerif-15>
4. [http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3\\_Semantics.pdf](http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf)

## Formalities (who & when)

Go-between	Rosa Filgueira
RI representative	Daniele Bailo
Period of requirements collection	From September to November 2015
Status	Finished