

Example 5: Using the Reference Model to explain the technology details of common services (WP4 practices)

Descriptions of the Example

ENVRI working package 4 responses to deliver common services to support the constructions of ESFRI ENV RIs. Initially, the implementations focus on a [data access subsystem](#) that supports integrated data discovery and access. In order to help ESFRI project managers, architects, and developers understand the design and implementation of these services, this example uses the terms and concepts from the Reference Model to explain the technology details of these services.

How to Use the Reference Model

We start with the semantic harmonisation service developed by the team in Task 4.2 [39]. The development is conducted to support the [use case "Iceland Volcano Ash"](#). The goal is to support scientists to analyse Iceland behaviour using data provided by different research infrastructures during a specific time period.

Science Viewpoint

Defined by the Reference Model Science Viewpoint, the [semantic harmonization](#) is a [behaviour](#) belong to the [data publication community](#), which captures the business requirements of unifying similar data (knowledge) models based on the consensus of collaborative domain experts to achieve better data (knowledge) reuse and semantic interoperability.

Computational Viewpoint

A data publication community interacts with a [data access subsystem](#) to conduct user roles. The computational specification of the data access subsystem is given in Figure 1. The model specifies a **data access subsystem** which provides [data broker](#) that act as intermediaries for access to data held within the data curation subsystem, as well as [semantic brokers](#) for performing semantic interpretation. These brokers are responsible for verifying the agents making access requests and for validating those requests prior to sending them on to the relevant data curation service. These brokers can be interacted with directly via [virtual laboratories](#) such as [experiment laboratories](#) (for general interaction with data and processing services) and [semantic laboratories](#) (by which the community can update semantic models associated with the research infrastructure).

Figure 1: Computational specification of data access subsystem



Definitions

A [data broker](#) object intercedes between the data access subsystem and the data curation subsystem, collecting the computational functions required to negotiate data transfer and query requests directed at data curation services on behalf of some user. It is the responsibility of the data broker to validate all requests and to verify the identity and access privileges of agents making requests. It is not permitted for an outside agency or service to access the data stores within a research infrastructure by any means other than via a data broker.

An [experiment laboratory](#) is created by a science gateway in order to allow researchers to interact with data held by a research infrastructure in order to achieve some scientific output.

A [semantic broker](#) intercedes where queries within one semantic domain need to be translated into another to be able to interact with curated data. It also collects the functionality required to update the semantic models used by an infrastructure to describe data held within.

A [semantic laboratory](#) is created by a science gateway in order to allow researchers to provide input on the interpretation of data gathered by a research infrastructure.

Please click the links to find out the specification details of these computational objects and the interactions between them.

The implementation conducted by WP4 T4.2 is an **instantiation** of the above computational objects specified in the Reference Model, that uses existing software components and developed approaches to enable integration and harmonization of data resources from cluster's infrastructures and publication according unifying views.

Figure 2 depicts the computational components deployed in the prototype implementation. The service receives users' requests via the [SPARQL-endpoint](#). Then, it can automatically retrieve and integrate real measurement data collections from distributed data sources. The current prototype focuses on datasets from two different ESFRI projects:

- ICOS, which is organized by atmospheric stations which perform measurements of the CO2 concentration in the air and
- EURO-Argo observations that were provided in separate collections grouped according to the float that performed measurements of the ocean temperature.

The prototyped service uses two semantic models to provide mapping between representations: the [RDF Data Cube vocabulary](#) and the [ENVRI vocabulary](#). The ENVRI vocabulary is derived from the OGC and ISO “Observations & Measurements” standard ([O&M](#)), [SWEET](#) and [GeoSparql Vocabulary](#).

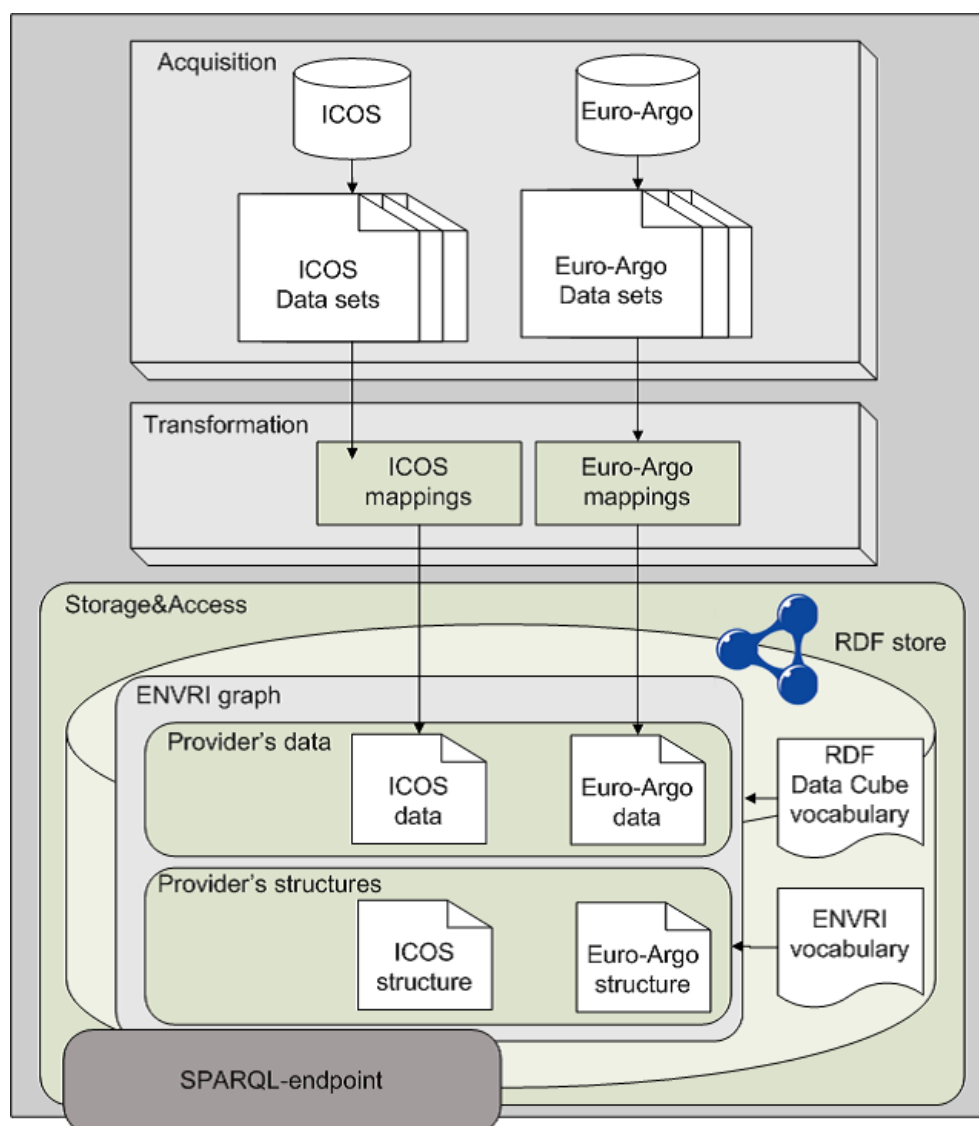


Figure 2: The Deployed service components for semantic harmonization [39]

Table 1 provides the mapping between Reference Model computational objects and the deployed service components. Among them, the *Transformation* component serves as a **data broker** to negotiate data access with data stores within heterogeneous research infrastructures. An (instance of the) **semantic broker** is implemented using the RDF store technology which provides the semantic mappings and translations.

Table 1: Mapping of the deployed service components to the Reference Model computational objects

RM Computational Objects	Deployed Service Components
💡 Data Broker	Transformation (ICOS mappings, EuroArgo Mappings)
💡 Experiment Laboratory	SPARQL-endpoint
💡 Semantic Broker	Provider's data (ICOS data, EuroArgo data) Provider's structures (ICOS structure, EurArgo structure)
💡 Semantic Laboratory	RDF Data Cube Vocabulary, ENVRI Vocabulary

In the following, we explain the design of the information model of the semantic harmonisation service.

Information Viewpoint

Analysing the environmental data schema results in identifying the common structural concepts, the ENVRI vocabulary, which include the terms such as “metadata attributes”, “observation”, “dataset”. Data retrieved from the different sources are firstly mapped to this uniform semantic model. Figure 3 gives two examples, and shows how datasets of ICOS and EuroArgo can be mapped to the ENVRI vocabulary, respectively.

Metadata Attributes							Observations		Dataset
Site	Year	Month	Day	Hour	Date	CO2	SD	Flag	
mhdall	2011	1	1	0	2011.000000	400.135			0.526
mhdall	2011	1	1	1	2011.000114	399.893			0.670
mhdall	2011	1	1	2	2011.000228	401.878			1.073
mhdall	2011	1	1	3	2011.000342	400.474			0.499
mhdall	2011	1	1	4	2011.000457	402.787			2.611
mhdall	2011	1	1	5	2011.000571	406.205			3.125

Metadata Attributes		Observations					Dataset
PLATFORM	ARGOS_ID	DATE	LAT	LONG	PRES (decibar)	TEMP (degree_C)	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	4.4	11.595	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	8.8	11.617	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	19.1	11.629	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	29.3	11.656	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	38.4	11.566	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	48.8	11.649	
4900679	37751	2010-03-01T03:24:00Z	48.817	-31.648	58.3	11.638	

Figure 3: Datasets as provided by ICOS (above) with CO2 concentrations and by EURO-Argo (below) with ocean temperature measurements

Semantic mappings are based on observation statements. For example, the following observation statement declares the measurements about “air”:

“Observation of the CO2 concentration in samples of air at the Mace Head atmospheric station which is located at (53_20'N, 9_54'W): CO2 concentration of the air 25m above the sea level on Jan 1st, 2010 at 00:00 was 391.318 parts per million”.

“Air” is represented as the concept of air in GEneral Multi-lingual Environmental Thesaurus (GEMET) by assigning the [URI](#) to it (entity naming). The GEMET concept of air is then defined as an instance of envri:FeatureOfInterest (entity typing).

The mapping rules are specified by using the Data cube plug-in for Google Refine. The mappings are executed to obtain RDF representations of the source data files. As such they are uploaded to the [Virtuoso OSE RDF store](#) and are ready to be queried at a SPARQL-endpoint.

The data harmonization process described above is captured by the Reference Model. As shown in Figure 4, the Information Viewpoint models the mapping of data according to [mapping rules](#) which are defined by the use of [local](#) and [global conceptual model](#). Ontologies and thesauri are defined as **conceptual models**, and those widely accepted models such as, GEMET, O&M, Data Cube, are declared [global conceptual models](#) whereas the ENVRI vocabulary is specified as a [local](#) one, because it has been developed within the current project without being yet accepted by a broad community.

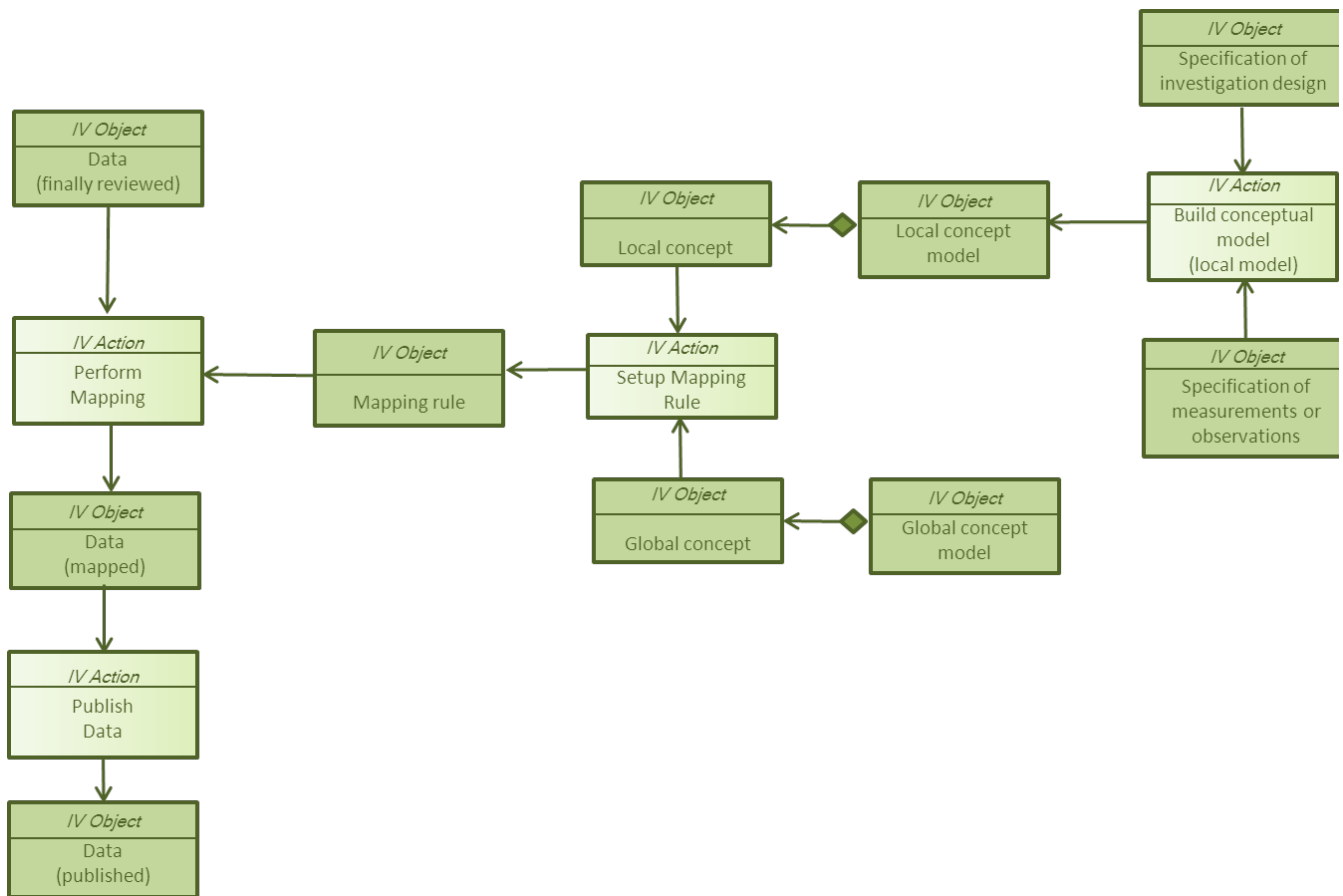


Figure 4: The RM Information specification related to the semantic harmonisation

Describing a process using the ENVRI Reference Model concepts is to instantiate the concepts that can be mapped to the process. Figure 5 illustrates the instantiation (all boxes with a dashed line) of the ENVRI Reference Model concepts focusing at the harmonization process described above. The same could be demonstrated for the EuroArgo dataset with the feature of interest being ocean. For each part of the observation mapping rules have to be defined to be able to query both datasets at a certain time period.

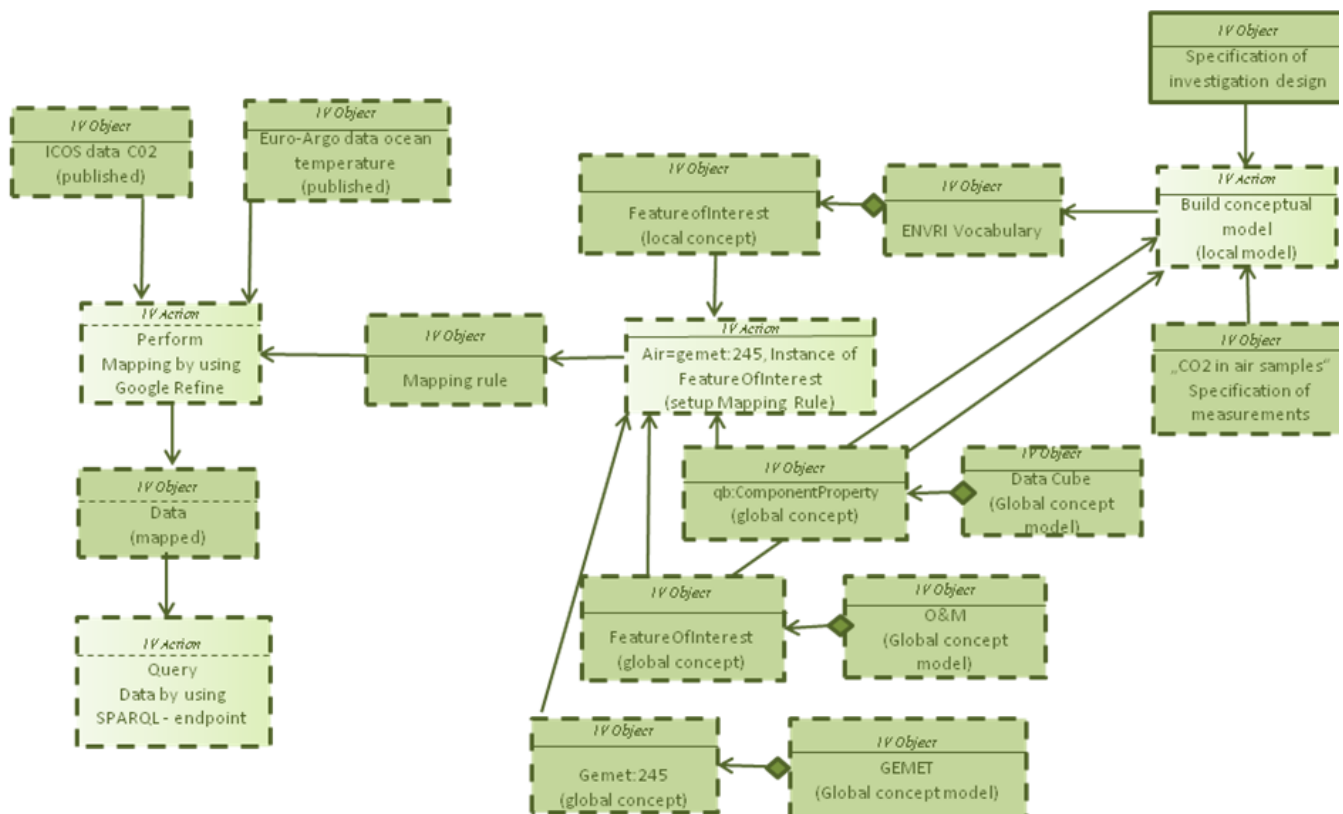


Figure 5: Mapping of the deployed information model with that of the the Reference Model

The tables below show the mapping between the harmonisation process and the concepts in the ENVRI RM information viewpoint. The example shows that both bottom up (from the applied operation to the model description) and top down approaches (from the model definitions back to the applied solution) can lead to a better understanding of the Reference Model itself and of how components should work properly in a complex infrastructure.

Table 2: Mapping between the Reference Model  **Information objects** and those in the deployed service














Information Object in RM	Component/Object in Task 4.2
 specification of measurements or observations	Observation of the CO2 concentration in samples of air at the Mace Head atmospheric station which is located at (53_20'N, 9_54'W): CO2 concentration of the air 25m above the sea level on Jan 1st, 2010 at 00:00 was 391.318 parts per million
 mapped	GEMET:245 is instance of FeatureOfInterest class
 conceptual model	GEMET, O&M, DataCube
 conceptual model	ENVRI vocabulary
 local concept	FeatureOfInterest (ENVRI vocabulary)
 global concept	Component Property, GEMET:245, FeatureOfInterest (O&M)
 mapping rule	GEMET:245 create as instance of FeatureOfInterest class
 published	ICOS data CO2 of air, EuroArgo data ocean temperature

Table 3: Mapping between the Reference Model  **Action Types** and those in the deployed service

Information Action Types in RM	Operation in Task 4.2
 build conceptual models	Build ENVRI vocabulary as extension of DataCube and on basis of O&M concepts
 setup mapping rules	Define rule: GEMET:245 create as instance of FeatureOfInterest class
 perform mapping	Perform Mapping using Google Refine
 query data	SPARQL query: http://staff.science.uva.nl/~ttaraso1/html/queries/Q1.rq

Summary

This example demonstrate the feasibility of the design specifications of the reference model. Instances of selected model components can be developed into common services, in this case, a  **subsystem** that supports integrated data discovery and access. Data products from different environmental research infrastructures including, measurements of deep sea, upper space, volcano and seismology, open sea, atmosphere, and biodiversity, can now be pulled out through a [single data access interface](#). Scientists are using this newly-available data resource to study environmental problems previously unachievable including, the study of the climate impact caused by the eruptions of the Eyjafjallajökull volcano in 2010.