Proceedings of the 6th AGILE
April 24th-26th, 2003 – Lyon, France

SEMANTIC INTEROPERABILITY AMONG ITALIAN EOS INFORMATION SYSTEMS

Stefano Nativi, Paolo Mazzetti, Lorenzo Bigagli, Dino Giuli

Engineering and Telecommunications Dept.
University of Florence – sede di Prato
Piazza Ciardi, 25, I-59100 Prato, Italy

e-mail: nativi@pin.unifi.it, mazzetti@pin.unifi.it, bigagli@pin.unifi.it

1. INTRODUCTION

EOS (Earth Observation Science) data describing atmospheric or terrestrial phenomena are collected in digital form since three decades. Nowadays, in the "Internet Era", it is reasonable to expect this wealth of data to be accessible and shareable in a simple and coherent manner among the different actors of the geo-science arena, integrating the existing information systems with the ones that future missions and research activities are set to provide.

Geomatics, that is automatic management and processing of environmental and geographical information, is changing from a niche discipline to an horizontal application area, involving scientific as well as industrial and institutional bodies.

The key factor for these expectations to become reality is interoperability, that may be generically defined as "*system cooperation for the sake of information and process sharing*". Semantic interoperability is essential to implement interoperability and integration of different data sources [1]. Therefore, semantic reconciliation is required to share, and exchange heterogeneous data in a seamless way.

In order to effectively use such huge amount of data, EOS researchers and geo-scientists need a user-friendly information system which provides them with the following transparent and online services:

- to locate data;
- to evaluate data;
- to compare data;
- to access data;
- to transfer data.

Naturally, in the most general case –as well as in the most common real scenario- data is acquired and managed by heterogeneous and autonomous information systems which are distributed over an IP-enabled network.

In extreme synthesis, EOS researchers and geo-scientists need a searchable Online Catalogue System which must be "smart" enough to accomplish all the previously introduced services, in an extensible manner.

A key factor for enabling such services is the availability of complex metadata along with data. Metadata allows generic users to quickly search and evaluate heterogeneous data. Metadata plays a crucial role in data reconciliation, determining how to deal with similar or contradictory information.

This work presents a solution to implement such Online Catalogue System for the Italian EOS community. That solution is based on a Federated Information System (FIS) approach, and comprises a Domain Ontology model, an information broker, a query

software and a Web-based interface. The architecture presented as a system solution is based on mediation and wrapper component approach [2]. The introduced Domain Ontology is used as the Common User Model of the FIS.

We have developed a Web-based prototype for exploring our solution feasibility and capabilities; such prototype, called SINOTS (http://sinots.pin.unifi.it/sinots), was experimented in the framework of a research project funded by the Italian Space Agency (ASI).

2. THE ONLINE CATALOGUE SYSTEM FOR THE NATIONAL EOS COMMUNITY

The introduced solution realizes semantic interoperability and integration of heterogeneous IS on the Web. The main objectives were:

- to design a simple and extensible ontology for On-line Cataloguing of EOS Datasets and utilize such ontology as the Common User Model of a federated information system (FIS), which integrates multiple and disparate EOS information resources;
- to enable EOS information resources to join the FIS and publish their data and metadata in a secure way, without any modification to their existing resources and procedures and without any restriction to their autonomy;
- to enable Users to browse and query the FIS, receiving a combined result which incorporates relevant data and metadata from across different resources, in a transparent way and by means of standard Web technologies;
- to accommodate the growth of such FIS, either in terms of its Users or of its information resources, as well as the evolution of the underlying domain ontology.

The developed solution can be seen as a One-stop Catalogue of EOS datasets. Figure 1 depicts the main logical architecture of the system.

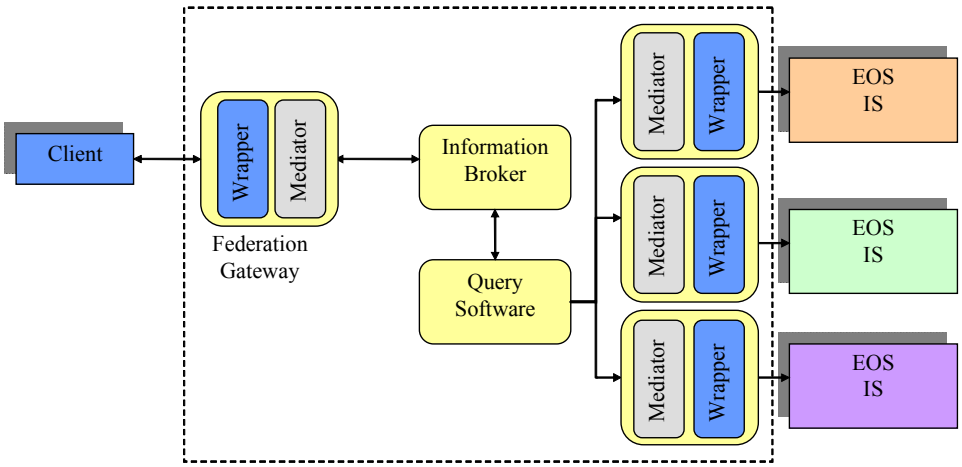

Fig. 1 The logical architecture of the system

3. THE FEDERATION COMMON ONTOLOGY

There have been several international standardisation efforts, such as the OpenGIS (http://www.opengis.org) and ISO TC211 (http://www.isotc211.org/) initiatives, which define a couple of standard semantics for geo-data and metadata. In particular, these two initiatives co-operate with each other and their specifications are very close.

On the top of a profile of such semantics (i.e. the Metadata sections) we introduced the Federation Common Ontology (FCO), which abstracts our EOS IS domain. The FCO defines a unified semantics for the federation which is recognised and shared among the federated IS.

Concerning the IS resources to be federated, we reckoned several Conceptual Models which abstract different realities: that is, possible worlds of the EOS IS domain. For examples, it was possible to distinguish a conceptual model for: meteorological-radar acquisition catalogue, AVHRR satellites data preview system, multi-source study case IS. These conceptual models can be considered as different implementations of the shared unified semantics -the FCO- which defines them [3].

Figures 2 and 3 represent two simple RDF-based diagrams about the introduced ontology, dealing with the information resource and the dataset subjects, respectively.

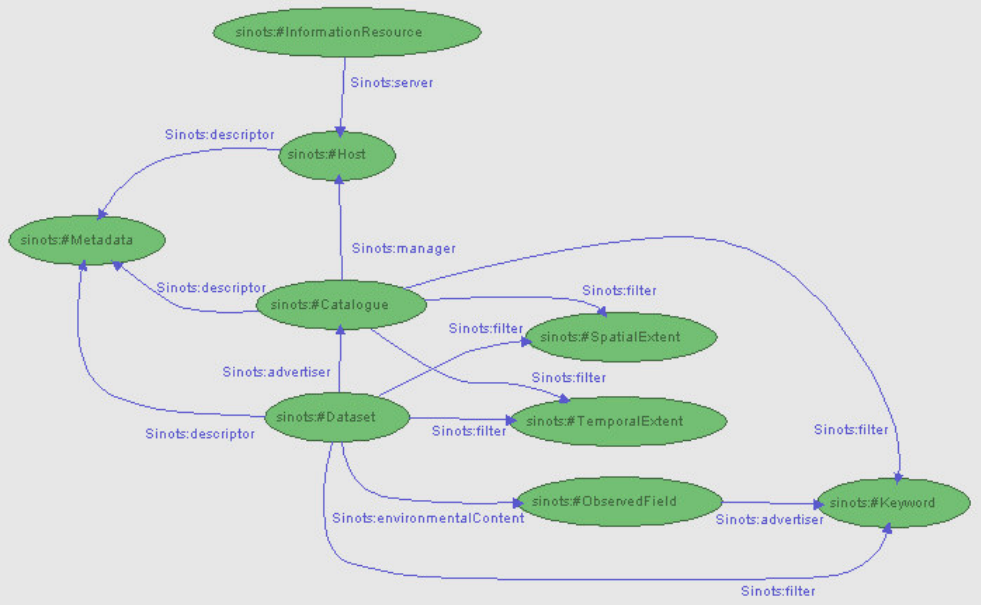

Fig. 2 FCO: information resource

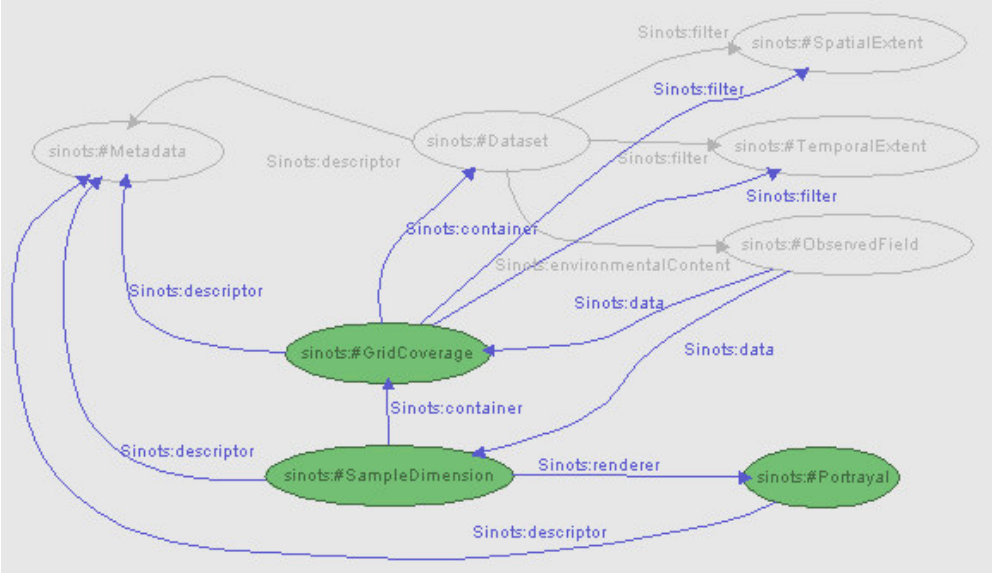

Fig. 3 FCO: dataset

4. SEMANTIC MAPPING

The introduced FCO for EOS information system domain provided enough knowledge at the intentional level for a unique interpretation of domain concepts and their relationships (Bishr and Kuhn, 2000). A Federation Common Conceptual Model (FCCM) was introduced to provide an extensional description of the FCO. The FCCM describes the useful domain conception but does not carry enough knowledge, and therefore is not semantically rich enough to provide a unique interpretation.

For the federation, we achieved the semantic reconciliation task by mapping the heterogeneous conceptual models –characterising federated EOS IS- to the FCCM. Moreover, the queries at the federation level are expressed and finalised against the FCCM.

The same approach is followed for the client application: a client application is characterised by its own conceptual model which must realise the FCO. Naturally, it is possible to support different client applications characterised by diverse conceptual model.

In order to support client applications whose conceptual models realise ontologies different from the FCO, we should conceive and utilise an ontology mapping solution.

Figure 4 depicts the adopted solution architecture.

According to our solution, each EOS IS can join the federation resorting to a Federation Gateway: made up of a couple of components: a wrapper and a mediator (see Fig. 1). The Federation Gateway performs the semantic mapping between the local IS conceptual model and the FCCM; naturally, such process is possible only because they realise the same domain ontology (i.e. the FCO) [4].

The Mediator is in charge of the conceptual models mapping task, meanwhile the wrapper achieves the generation of a hierarchical model where an entity-relationship model is present (e.g. DB models).

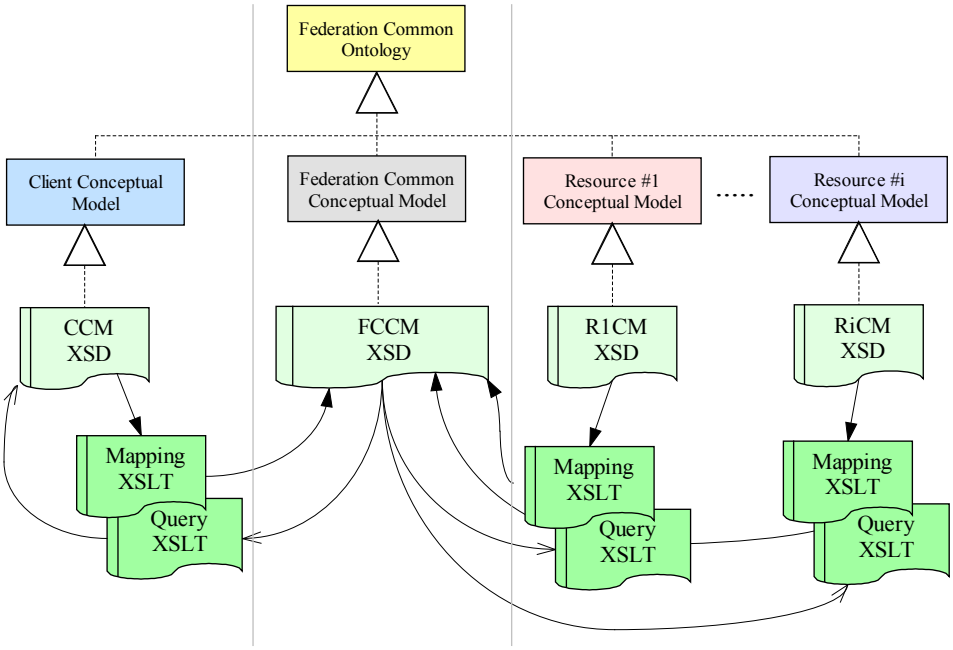

Fig. 4 Semantic mapping solution: the architecture

As far as our experimentation is concerned, Figure 5 depicts the implemented scenario.

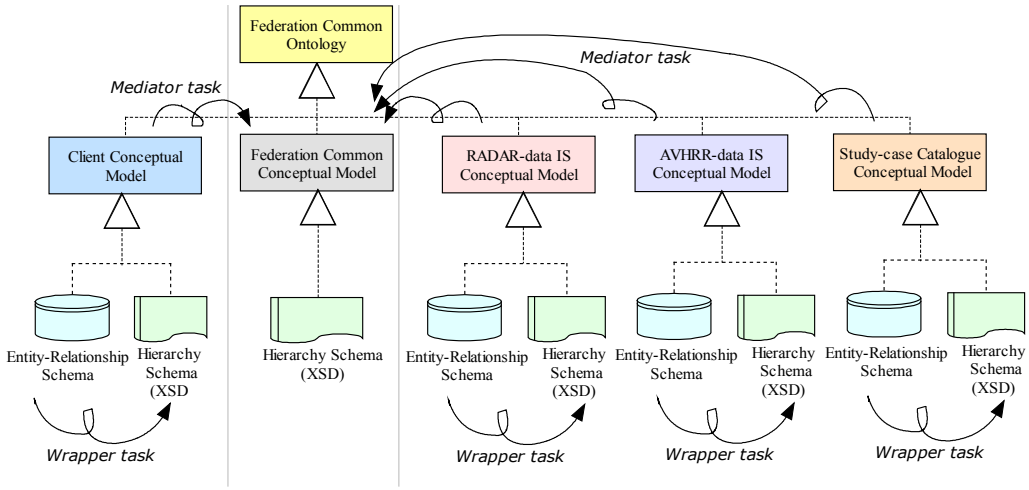

Fig. 5 Semantic mapping solution: the experimented scenario

Each ontology element is realised by a concept of the FCCM, and each concept is characterized by a set of simple and complex attributes that describe it completely. Metadata will externalises such attributes.

5. XML ENCODING AND FEDERATION QUERIES

XML is an important technology for data representation and exchange. It provides a common format for expressing both data structures and contents; it is useful in integrating structured, semistructured and unstructured data [5]. XML fulfils the syntactic interoperability requirements: an XML parser can parse any XML document and is usually a reusable component [6]. Therefore, we used XML as a common data model over the Web. We utilised wrapper components in order to perform database content publishing in XML.

Nevertheless, XML is not sufficient to solve the complex problem of data integration; in particular the issues related to the representation of semantic aspects of information resources are still open [5]. Therefore, we decided to use the XML potentialities along with the introduced FCO which abstracts the considered domain.

As far as schema reconciliation is concerned, we used the W3C XSLT (Extensible Stylesheet Language Translation) for defining mapping among heterogeneous tags. The same solution is adopted in order to "translate" a federation-level query into local schema queries; unfortunately, this mapping is cumbersome and therefore we implemented a set of pre-canned queries. Anyway, these queries provide users with fifteen selection criteria specifying a combination of as many as four intuitive information properties, defining a 4-dimensional domain, including space and time dimensions.

Heterogeneous information reconciliation is performed leveraging metadata associated to each piece of information. Each piece of information is characterized by a set of metadata that describe it completely; the following metadata categories are supported:

- *Constraint*: the restrictions placed on data
- *Content*: the content of a referring entity
- *Distribution*: the distributor of, and options for obtaining a dataset
- *Management*: the scope and the frequency of data updating
- *Data Quality*: a general assessment of the quality of the data, and the resources and production process used in producing the data entity
- *Extent*: the spatial and temporal extent of the referring entity
- *Localization*: the mechanisms used to represent spatial and temporal information in a dataset
- *Format*: the description of the computer constructs that specifies the data entity

6. THE CLIENT APPLICATION

The interaction is completely driven by the federation unified view –based on the FCCM- and integrates browsing and querying functionalities.

In order to provide a relatively simple interface and ease overall information discovery and navigation, it is necessary to reduce the overhead of working with a potentially large number of resources. We adopted two main approaches to address the this issue, introducing a precise:

1. information space segmentation approach;
2. conceptual map navigation strategy (i.e. FCCM navigation).

The overall federation information space was segmented according to a 4-W metaphor: What, When, Where, Who. Hence, user can navigate through and interact with information in a four dimensional space.

7. THE IMPLEMENTED PROTOTYPE

The main objectives of the system prototype are:

to enable EOS information resources to join a FIS and publish their data and metadata in a secure way, without any modification to their existing resources and procedures and without any restriction to their autonomy;

to enable users to browse and query the FIS, receiving a combined result which incorporates relevant data and metadata from across different resources, in a transparent way and by means of standard Internet technologies;

to accommodate the growth of such FIS, either in terms of its clients or of its information resources, as well as the evolution of the underlying data model.

We reckoned that the best solution to these requirements is a FIS where the individual participants are self-contained autonomous systems, but together form a consistent wider picture: the federation. We implemented a mid-tier integration approach which utilizes the wrapping of parts of existing systems (i.e. EOS data sources) to form a FIS. In order to guarantee data source autonomy, the developed FIS is a read-only system: the federation does not allow the updating (or insertion) of data into the participant resource systems, through the federation layer.

We developed a model-based mediating system [7], in which unified views are defined and executed at the level of conceptual models rather than at the structural level. The strategy adopted for implementing the FIS is top-down: first we introduced the FIS unified information need, i.e. the FCCM, then we plugged in the data resources, mapping their contribute to such need, in the framework of the FCO. In this process, there was no need to include –and map- resource schemas completely.

The developed FIS implements a virtual integration architecture: it materializes query results only temporarily -at the time the query is posed, implementing a mechanism to translate queries against the common schema into several semantically meaningful and executable queries against data resources.

The experimentation of the described solution has been conducted in the framework of a project funded by the ASI (Italian Space Agency). It set up a federation of EOS information resources, located in the central and southern regions of Italy: the University of Florence in Florence, the PIN research centre in Prato and the IMAA Institute of the CNR near Potenza.

The testbed has involved three heterogeneous information resources managing remotely-sensed data (i.e. satellite and ground-based radar datasets); these resources remained completely autonomous and independent.

A new extended implementation is under development in the framework of a collaboration between the CNIT (Interuniversity Consortium for Telecommunications) and the Italian Ministry of the Environment.

8. REFERENCES

[1] Shanzhen Y., Lizhu Z., Chunxiao X., Qilun L., Yong Z., “Semantic and Interoperable WebGIS”, Proceedings of the Second International Conference on Web Information Systems Engineering, Volume: 2, pp. 42-47, IEEE 2002 Page(s): 42 -47 vol.2

[2] Wiederhold G., Genesereth M., “The Conceptual Basis for Mediation Services“, IEEE Expert, Vol. 12, n. 5, pp. 38-47, 1997.

[3] Bishr Y., Kuhn W., “Ontology-based Modelling of Geospatial Information”, Proceedings of the 3° AGILE Conference on Geographic Information Science – Helsinki/Espoo, Finland, May 25th – 27th, 2000, pp. 24-27.

[4] Guarino N. Formal Ontology and Information Systems. In N. Guarino (ed.), Formal Ontology in Information Systems. Proc. of the 1st International Conference, Trento, Italy, 6-8 June 1998. IOS Press.

[5] Decker S., Melnik S., Van Hermelen F., Fensel D., Klein M., Broekstra J., Erdmann M., Horrocks I., „The semantic Web: The role of XML and RDF“, IEEE Internet Computing, September-October 2000, pp. 63-74.

[6] Bertino E., Ferrari E., “XML and Data Integration”, IEEE Internet Computing, November-December 2001, pp. 75-76.

[7] Wiederhold G., “Mediation to Deal with Heterogeneous Data Sources”, in Interoperating Geographic Information Systems, Springer LNCS 1580 (Proc. Interop'99, Zurich) March 1999, pp. 1-16.