Geophysical Research Abstracts, Vol. 7, 06509, 2005
SRef-ID: 1607-7962/gra/EGU05-A-06509


# Mediation to deal with information heterogeneity

L. Bigagli (1,2), S. Nativi (1,2), P. Mazzetti (1,2), G. Villoresi (1)

(1) Institute of Methodologies for Environmental Analysis of the National Research Council (IMAA-CNR), Italy, (2) University of Florence, Italy, (bigagli@imaa.cnr.it, nativi@imaa.cnr.it)

## 1 Introduction

In computer science, interoperability is defined as *the ability to exchange and use information, usually in a large heterogeneous network made up of several local area networks* [1]

On a broader basis, it may be seen as the capacity to move information across the boundaries between the source and the destination of such information. Architectural variants are possible, for example, in the endpoints cardinalities (distribution/integration of information), in the communication paradigm (client-server, master-slave, P2P, etc.), in the messaging pattern (request-response, one-way, publish-subscribe, etc.)

The above mentioned boundaries may be of several types, e.g. technical, geographical, temporal, administrative, linguistic, cultural, social, so that interoperability may be hard to achieve. A most relevant example of this multi-faceted problem is e-learning, that implies the exchange of information between subjects that are usually quite distant from each other under several points of view (e.g. their level of instruction).

A very clear distinction is usually made between information and data: the first is defined as *knowledge acquired through study or experience or instruction*, where knowledge is defined as *the psychological result of perception and learning and reasoning* [1].

According to these definitions, information is always a subjective concept, whose meaning depends on its context, and whose value is apparent to its human destina-

tion (actually part of the context itself).

Different representations may be derived from a given information, for example, for the sake of compactness, transmission, storage, etc. The encoding of information in an automatically-managed form is referred to as data.

Data-management systems provide the capability to restore the original representation of information, that may be seen as the "meaning" of that data. Likewise, a data-management system can be seen as the context in which that data is meaningful.

It is usually distinguished among data interoperability and (information) interoperability.

In the above terminology, data interoperability refers to the capacity to move data from a source machine to a destination machine, in such a way that the latter is able to restore the original representation of information.

In the end, we may say that the difference between information and data resides in the interpreting context and resembles the difference between a human and a machine: basically, just a richer context. The larger part of our context is formalized and adapted to deterministic management, the more we become able to use machines to cross the unavoidable boundaries of the real world.

# 2 Scope

In this work, we address the problem of data and information interoperability, focusing on the Earth System Science information domain.

We argue that, nowadays, well-established architectures and standard technologies are available to address and implement data interoperability.

In particular, mediation proves a valuable and flexible approach to harmonize data. The mediated approach relies on the identification of *articulation points* around a particular boundary and the encapsulation of the adaptation logic into a specialized component: the mediator [2].

For example, given an administrative boundary protected by a firewall, an HTTP proxy is in charge of handling HTTP traffic between internal clients and external servers.

On a conceptual basis, data integration (which is an aspect of data interoperability) is easily achieved by mediation, by adapting the source data model to the destination one before data exchange. To avoid excessive proliferation of mediation compo-

nents (given n different system to interconnect, nˆ2-n mediators would be necessary), federate system solutions are usually implemented, choosing a particular model (the federated model) as a common Esperanto [3].

Since a federal model tends to be more general than federated ones (being a sort of least common denominator), loss of information may occur in translation, hence mediation is particularly well-suited for read-only systems (e.g. typical client-server architectures).

When a federated model proves successful, it may be natively adopted by federation members, hence avoiding the need for mediation. This happened for example with TCP/IP spreading in internal networks (Intranet).

Mediation has been successfully applied to the task of interconnecting diverse networks and transporting data on it. Other data interoperability issues are basically *language* issues, in the usual meaning of language as a persistent representation of knowledge, and pertain to the three different levels of semantic, syntax and lexicon. Recently, a federal model for lexical and syntactic data interoperability has emerged: XML and related technologies [4].

The impact of XML has been crucial, because of its simple yet powerful data model (hierarchical semi-structured data), that can be easily mediated to legacy relational models, as well as to file systems, and generally to web data (intrinsically semi-structured). Besides, XML has proven actually eXtensible and powerful enough to allow description of its dialects in XML itself (i.e. XML Schema Language, XQuery, eXstensible Stylesheet Language Tranformations [5]).

Several XML dialects are being developed by the diverse Information Communities to provide a means to exchange data between the community members. Science Digital Libraries must support interdisciplinary exchange of information and provide a framework for markup languages to be extended even further as they are tested and applied in science education settings [6]. In this context, mediation tools and services may provide translating interfaces between data representations in different markup languages, or may support access to data in a given markup language by a wide variety of users [6].

We have successfully experimented a federated system mediating disparate sources of Earth System Science Data (ESSD). Briefly, such system accommodates geographic boundaries (i.e. distance) by means of Internet; administrative boundaries by means of SSL; execution environments by means of Java; data models by means of XML technologies.

The chosen federal data model is based on the ISO19115 standard series and may not

be implemented by participating system (or supported at higher levels of abstraction only). The XSL Transformation language is used to convert flowing XML resources from source to federal schema.

Concerning information interoperability, efforts are ongoing to formalize semantic and context so as to enable automatic management of “meaningful” data. This could support, for example, the automated mapping of different conceptual models.

With this regard, we present our experience in mapping concepts of different ESSD management systems: basically, an ad-hoc crafting of content structure and syntactic mapping.

We also introduce some ideas for a more formal approach to the problem of supervised metadata profiling, in the context of information overloading mitigation.

# 3 Conclusions and future work

Main conclusions are:

- presently, technologies are available to address and implement data interoperability; particularly, mediation approach and XML technologies prove suitable to the task;
- although still at an early investigation stage, information interoperability can be facilitated by carefully identifying articulation points;
- geospatial data community has achieved a certain maturity as far as data model and interoperability interfaces are concerned; it should continue to leverage such achievements, extending and enhancing them to accommodate its valuable peculiarities;
- the logic that mediates from model A to model B is not easily applicable to mediating from B to A: mediation inversion could be investigated to clarify its possible extent, with or without human supervision (e.g. with respect to syntactic mediation as performed by XSLT).

# 4 References

[1] http://www.hyperdictionary.com/search.aspx?define=interoperability

[2] Gio Wiederhold, “Mediators in the Architecture of Future Information Systems”, IEEE Computer March 1992, pp.38-49.

[3] S. Busse, R.-D. Kutsche, U. Leser, H. Weber, Federated Information Systems: concepts, terminology and architectures, Technical Report Nr. 99-9, TU Berlin, 1999.

[4] Tim Bray, et al., “Extensible Markup Language (XML) 1.0 (Third Edition)”, W3C Recommendation 04 February 2004

[5] James Clark, “XSL Transformations (XSLT) Version 1.0”, W3C Recommendation 16 November 1999.

[6] Laura M. Bartolo et. Al, “NSF / NSDL Workshop: Scientific Markup Languages”, the National Science Foundation, Arlington, Virginia, June 14-15, 2004.