Towards Global Data Interoperability

Costantino Thanos, Leonardo Candela, Donatella Castelli

Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, via G. Moruzzi 1, 56124 Pisa, Italy, e-mail: costantino.thanos@isti.cnr.it

Data Interoperability is a multi-level and context-specific concept that requires management of the full data life cycle. We can distinguish diverse interoperability levels: technical, syntactic, semantic, operational, secure, temporal, and language interoperability. They form dependent layers: operational interoperability is only possible if semantic interoperability is ensured; semantic interoperability is only possible if syntactic interoperability is ensured; and so on.

There are two key problems hindering the achievement of data interoperability: (i) the heterogeneity of the exchanged data which covers all types of syntactic, structural, and semantic diversities among systems used to modeling data and (ii) inconsistencies between the use of data as intended by the originator and its exploitation by the recipient.

The main concept enabling interoperability is mediation. This concept has been used to cope with many heterogeneity dimensions spanning terminology, representation format, transfer protocols, semantics, etc. The mediation concept is implemented by a mediator, which is a software device that supports a mediation schema capturing user requirements, and an intermediation function between this schema and the distributed data sources.

A key feature which characterizes a mediation process is the kind of intermediation function implemented by a mediator. There are two main functions: Mapping and Matching.
Mapping refers to how data structures, properties, relationships are mapped from one representation scheme to another one, equivalent from the semantic point of view.
Matching refers to the action of verifying whether two strings/patterns match, or whether semantically heterogeneous data match.
There are several approaches to implementing the intermediation function, of particular relevance the approaches based on standard data modeling formalisms and ontologies.
In order to effectively implement the intermediation function exchanged data should be complemented with contextual, provenance/lineage, and quality information.

This paper identifies relevant interoperability levels for a number of important application domains and presents efficient approaches to implement data mediation.