

Vol. 54 – SUPPLEMENT

Bollettino di Geofisica

teorica ed applicata

An International Journal of Earth Sciences

IMDIS 2013

International Conference on Marine Data and Information Systems

23-25 September, 2013 - Lucca (Italy)

Organised by

ENEA, jointly with
EU SeaDataNet project, IFREMER and IOC/IODE.

Organizing committee

Giuseppe Manzella, chair - ENEA, Italy; Michele Fichaut - Ifremer, France
Dick Schaap - MARIS, The Netherlands; Lesley Rickards - NERC, UK
Vanessa Tosello - Ifremer, France

International Steering Committee

Giuseppe Manzella, chair - Italy; Steve Miller - USA
Robert Keeley - Canada; Margarita Greggs - USA
Roger Proctor - Australia; Gilbert Maudire - France
Michele Fichaut - France; Dick Schaap - The Netherlands
Lesley Rickards - UK; Peter Pissierssens, IODE
Serge Scory - Belgium; Jean-Marie Beckers - Belgium
Reiner Schlitzer - Germany; Friedrich Nast - Germany
Athanasia Iona - Greece; Alessandra Giorgetti - Italy
Simona Simoncelli - Italy; Vittorio Barale - EU



Istituto Nazionale di Oceanografia
e di Geofisica Sperimentale

ISSN 0006-6729

A system for managing metadata using XML format

Elena Partescano, OGS-Istituto Nazionale di Oceanografia e di Geofisica Sperimentale,
epartescano@ogs.trieste.it (Italy)

Alberto Brosich, OGS - Istituto Nazionale di Oceanografia e di Geofisica Sperimentale,
abrosich@ogs.trieste.it (Italy)

Alessandra Giorgetti, OGS - Istituto Nazionale di Oceanografia e di Geofisica Sperimentale,
agiorgetti@ogs.trieste.it (Italy)

By definition, metadata, is information associated with data, answering questions as where, how, when and by whom the data was acquired. Taking part in European projects such as the EU-SeaDataNet (the Pan-European infrastructure for ocean and marine data management) made it necessary to use XML (Extensible Markup Language) as a standard file format for sharing metadata.

At present, the Italian National Oceanographic Data Centre (OGS/NODC) has all its data and metadata contained in an Oracle relational database, and some metadata are managed using XML documents following standard schemata (from SeaDataNet Project).

This paper shows the system OGS/NODC is currently using to manage oceanographic XML metadata, stored as is, into a relational database and a possible future development.

The participation to EU-SeaDataNetProject has grown the need to create and share metadata in XML format. Tools for managing it (as Mikado Software) are available, therefore the next step was linking metadata XML files to data loaded into relational database and building a web service to allow the connection between the managing tool (Mikado) and the database.

We evaluated and decided that the best solution was to load into the Oracle database the whole XML metadata files, using a specific data type. Using XML and XQuery functions it is possible to store, extract and manage different kinds of information that might be exchanged at the European level. Furthermore, with a RESTful (REpresentational State Transfer) Web Service we have a simple and standard interface for quickly and easily creating, modifying and deleting records containing XML documents inside the database. Finally, through the use of that RESTful Web Service it is possible to decouple the applications from the database, so that through the use of software that manages HTTP URLs, such as the Mikado (SeaDataNet project), the XML documents can be inserted, updated and deleted inside the database without the need for a direct connection to it.

Future development

Another way to manage XML files could be to use a native XML database (Fig. 1): the documents are stored in a database designed especially for storing XML, supporting XPath and XQuery, to retrieve it. A native XML database doesn't use a relational model, used in common RDBMS databases.

In a native XML database (as for example: eXist, BaseX, Berkeley DB XML), the entire XML file can be stored in a single place. In this way, if the XML file or a part of it needs to be retrieved, only one searching index and only one reading index is required to recover the information, whereas a RDBMS database needs several indexes for searching and reading to recover the data. The main difference is that their inner model is based on XML and not something else.

The major advantage using a native XML database is that a database schema is not required to store textual or binary data and documents (as for the relational databases). XML databases allow to manages complex data relationships that are not easily managed in relational rows and columns. The database structure come from the XML schema, and can be easily adapted if necessary.

For these reasons, a native XML database fits perfectly with data-intensive uses such as archiving metadata.

In the future, we would prefer to use a native XML database to store the XML metadata file (Fig. 1), but even if it seems the more logical choice, several problems could arise, as for example how to join data and XML metadata (relational database and native XML database).

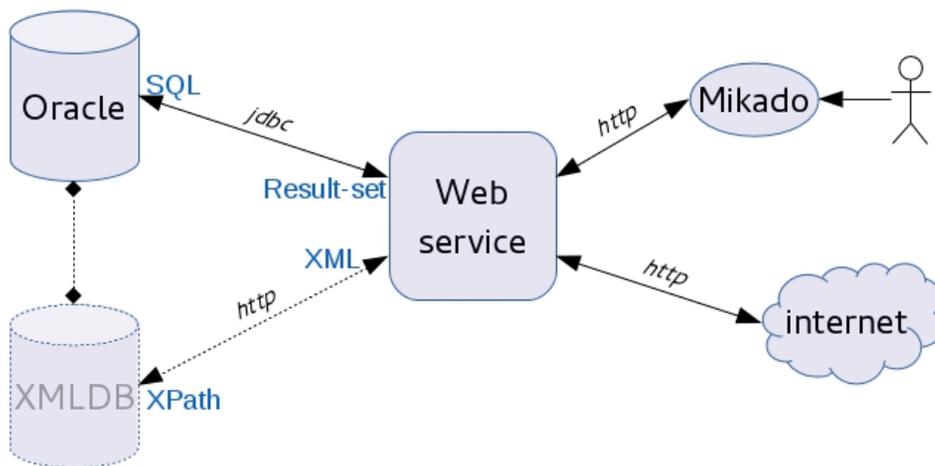


Fig. 1 - Work-flow (Dashed lines represent the future developments while continuous lines describe the present situation).

Conclusion

By using a database, relational as Oracle (currently) or Native XML database (possible future development), a RESTful Web Service and a software that manages HTTP URLs (as Mikado), we have the advantage that every single element is independent from the others. The independence between elements guarantees us more versatility. This has the benefit that changes in the database don't affect the operation of the system; at the same time, there is no obligation to use the Mikado software for managing the XML data and there is complete autonomy from the Web Service used including the implementing technology.