

Asterix & Obelix: How Standards Reunite Data and Metadata

P. Baumann

Jacobs University, Large-Scale Scientific Information Systems, 28759 Bremen, Germany – p.baumann@jacobs-university.de

THEME: DATA – Data and information systems and spatial data infrastructure

Invited contribution to the session on “Geospatial Information Analysis in Digital Earth”.

KEY WORDS: standards, OGC, WCS, WCPS, ISO, SQL/MDA, rasdaman

ABSTRACT:

There is a traditional saying that metadata are understandable, semantic-rich, and searchable whereas data are big, with no accessible semantics, and just downloadable – a little bit like smart little Asterix and fat, unintelligent Obelix in the well-known comics. Not only has this led to an imbalance of search support from a user perspective, but also underneath to a deep technology divide often using relational databases for metadata and bespoke archive solutions for data. Looking at the latter, arguably a major part of today’s “Big Data” is constituted by multi-dimensional arrays, such as 1-D sensor timeseries, 2-D remote sensing imagery, 3-D x/y/t image timeseries and x/y/z geophysical voxel models, as well as 4-D x/y/z/t climate and ocean data.

In our work, we attempt to bridge the data/metadata barrier by extending SQL with declarative model and query support for massive multi-dimensional arrays. This effectively allows filtering and processing of both data and metadata in a single query, like Asterix and Obelix fighting together. Integrated retrieval gives a new degree of freedom to users for expressing their needs, and more efficiency as only the actually intended result get transmitted. On server side, the high-level language enables highly effective optimization, parallelization, and distribution methods. With rasdaman, this has been shown on 130+ TB databases and splitting of incoming queries over 1,000+ cloud nodes.

The OGC Web Coverage Processing Service (WCPS) language standard follows this principle, embedding itself into XQuery to accommodate the prevailing XML metadata sets. In ISO, SQL is being extended with arrays as new attribute types.

In our talk we outline the current state of array query languages in concepts, implementation, use, and standardization. We report on pertinent OGC and ISO activities and real-life use cases where rasdaman is being used on services covering the Earth Sciences.