



Contribution ID: 2

Type: **not specified**

Science SQL: Advancing from Data to Service Stewardship

Thursday 1 October 2015 14:30 (30 minutes)

In today's science archives, data typically are managed separately from the metadata, and with different, restricted retrieval capabilities. While databases are good at metadata modelled in tables, XML hierarchies, and RDF graphs, they traditionally do not support "the data", in particular: multi-dimensional arrays. Consequently, file-based solutions let users "drown in data files" rather than presenting just a few datacubes for dissection and rejoining with other cubes. In the quest for improved service quality the new paradigm is to allow users to "ask any question, any time" thereby enabling them to "build their own product on the go".

This requires a new generation of services with new quality parameters, such as flexibility, ease of access, embedding into well-known user tools, and scalability mechanisms that remain completely transparent to users. In the field of massive spatio-temporal arrays this gap is being closed by Array Databases, pioneered by the scalable rasdaman ("raster data manager") array engine. Its declarative query language, rasql, extends SQL with array operators which are optimized and parallelized on server side, including dynamic mash-up configuration. As of today, rasdaman is in operational use on hundreds of Terabytes of satellite image time-series datacubes, with transparent query distribution across more than 1,000 nodes. Its concepts have shaped international Big Data standards in the field, including the forthcoming array extension to ISO SQL and the Open Geospatial Consortium (OGC) Big Geo Data standards, with manifold take-up by both open-source and commercial systems.

We show how array queries enable flexible data access and describe the rasdaman architecture with its optimization and parallelization techniques.

Presenter: BAUMANN, Peter (Jacobs University Bremen)