

## **Datacubes as a Novel Paradigm for Massive Earth Data Analysis: Concepts and Implementation**

**Peter Baumann**

### **Abstract (1)**

Formerly we focused on data as they are delivered from satellites or other sensors. By providing access to hyperspectral scenes, however, services were tied to the instrument characteristics, rather than offering users a comprehensive, unified data view. Datacubes change this radically: like seamless maps caused a revolution in map access for non-experts, in datacubes images are combined along time, or height, or both. The result is a single homogeneous data object - the datacube - which represents Analytics-Ready Data (ARD) along space and time. On server side, datacube operations can be parallelized and distributed across clusters, clouds, and even form cross data center federations allowing on-demand extraction, analysis, aggregation, and fusion on datacubes.

In our talk, we introduce the datacube concept as a service paradigm and discuss the datacube standards available in OGC, ISO, and INSPIRE. Real-life examples, demonstrated live, show how to operate on Petascale datacubes while exploiting cloud parallelization and use of heterogeneous hardware. Participants can reproduce and vary most of the examples on their Internet-connected laptop.

## **Big Data Standards in OGC, ISO, and INSPIRE: Overview, Concepts, and Use**

**Peter Baumann**

### **Abstract (2)**

In OGC standards terminology, coverages - aka digital representations of space-time varying phenomena - represent the data category contributing most to "Big Data". In terms of practical application, coverages resemble spatio-temporal sensor, image, image time series, simulation, and statistics data. OGC offers both an abstract definition (with Abstract Topic 6) and a concrete definition (with the Coverage Implementation Schema, CIS) which establishes interoperability down to the level of single pixels, manifest through open, free conformance tests. ISO already has the mirror standard 19123 and currently is in the process of adopting OGC CIS and the corresponding service model, Web Coverage Service (WCS). The European SDI initiative, INSPIRE, likewise has adopted coverages and WCS for spatio-temporal data. In particular OGC Web Coverage Processing Service (WCPS) - as part of WCS - defines a spatio-temporal datacube query language suitable for operating on Analysis-Ready Data (ARD). Hence, a fruitful convergence of standards can be observed which benefits implementers, service providers, and users alike.

In our seminar we introduce the coverage data and service model, starting from simple extraction requests and moving up to on-demand analytics. We put the WCS suite in context with related standards and illustrate the state of standardization as well as future trends. The seminar relies on real-life multi-dimensional data sets where participants can run most of the examples on their Internet-connected laptop, as well as modify them for exploration.

**Peter Baumann**  
**Professor of Computer Science**  
**Jacobs University, Bremen Germany**

Dr. Peter Baumann is Professor of Computer Science, inventor, and entrepreneur. At Jacobs University, Bremen, Germany he researches on scalable multi-dimensional array databases ("datacubes") and their application in science and engineering. With his work on algebra, query languages, and efficient architectures culminating in the rasdaman array DBMS he has coined the research field of array databases. He has published 130+ book chapters and journal and conference articles, holds international patents on array database technology, and has received numerous international innovation awards for his work. The rasdaman technology is in operational use on Petabyte-scale spatio-temporal databases.

Peter Baumann is active, often leading contributor to standardization in the Open Geospatial Consortium (OGC) and ISO bodies, being editor of a series of adopted international standards. In OGC he is chairing "Big Geo Data" working groups. In OGC and ISO he is editor of the "Big Datacube" standards CIS, WCS, and ISO SQL/MDA ("Multi-Dimensional Arrays"). In the Research Data Alliance (RDA), he has co-founded and co-chairs the Big Data and Geospatial Interest Groups, as well as the Array Database Assessment Working Group. In 2014, OGC has honored his contribution to Big Data standardization with the prestigious Kenneth Gardels Award.

See [www.peter-baumann.org](http://www.peter-baumann.org) for more information.