Made in Europe, Making a Difference:
Contributions to Open Standards and Open Source Software

EUROGI imagine Conference
Berlin, 2014-oct-08

Peter Baumann
Jacobs University | rasdaman GmbH
p.baumann@jacobs-university.de
Outreach to Science, Communities, Citizens

- **EU FP7 EarthServer, PublicaMundi**
  - earthserver.eu, publicamundi.eu

- **Citizens: 1h TV documentary** for ARTE

- **Community conferences:**
  - FOSS4G-Europe 2014 (Jacobs U)
  - NASA WorldWind Europe Challenge
  - Big Data From Space (ESA)

- **Science:**
  - pioneered **Array Databases**
  - shaping **Big Data standards**
  - **Research Data Alliance (RDA):**
    - Big Data Analytics IG, Geospatial IG
Hadoop – Not the Answer to All

- “Since it was not originally designed to leverage the structure, its performance is suboptimal” [Daniel Abadi]
- “will hit a scalability wall” [M. Stonebraker, XLDB 2012]
The rasdaman Array Database

- "raster data manager": SQL + n-D arrays

```plaintext
select ls.img.green[x0:x1,y0:y1] > 130
from  LandsatArchive as ls
where  avg_cells( ls.img.nir ) < 17
```

- Scalable parallel tile streaming architecture
  - In operational use,
    Big Data Reference Implementation
select
   encode(
       struct {
           red: (char) s.image.b7[x0:x1,x0:x1],
           green: (char) s.image.b5[x0:x1,x0:x1],
           blue: (char) s.image.b0[x0:x1,x0:x1],
           alpha: (char) scale( d.elev, 20 )
       },
       "image/png"
   )
from SatImage as s, DEM as d
Parallel / Distributed Query Processing

1 query → 1,000+ cloud nodes

```
select
  max((A.nir - A.red) / (A.nir + A.red))
- max((B.nir - B.red) / (B.nir + B.red))
- max((C.nir - C.red) / (C.nir + C.red))
- max((D.nir - D.red) / (D.nir + D.red))
from A, B, C, D
```
Secured Archive Integration

First-ever direct, ad-hoc mix from protected NASA & ESA services in OGC WCS/WCPS Web client (EarthServer + CobWeb)
Information technology — Database languages — SQL —

Part 15:
Multi-Dimensional Arrays (SQL/MDA)

Technologies de l'information — Langages de base de données — SQL —
Partie 15: Tableaux multi-dimensionnels (SQL/MDA)
Array Databases for Big Science Data

- Sensor, image, simulation, statistics data ...Big Data in sci & eng
  - n-D spatio-temporal arrays

- "Science SQL" heralding disruptive change
  - OGC, ISO, INSPIRE

- "1 datacube says more than 1,000,000 images"