

Breaking the Limits of Dimensions: Large-Scale Multi-Dimensional Databases with rasdaman

INTERGEO

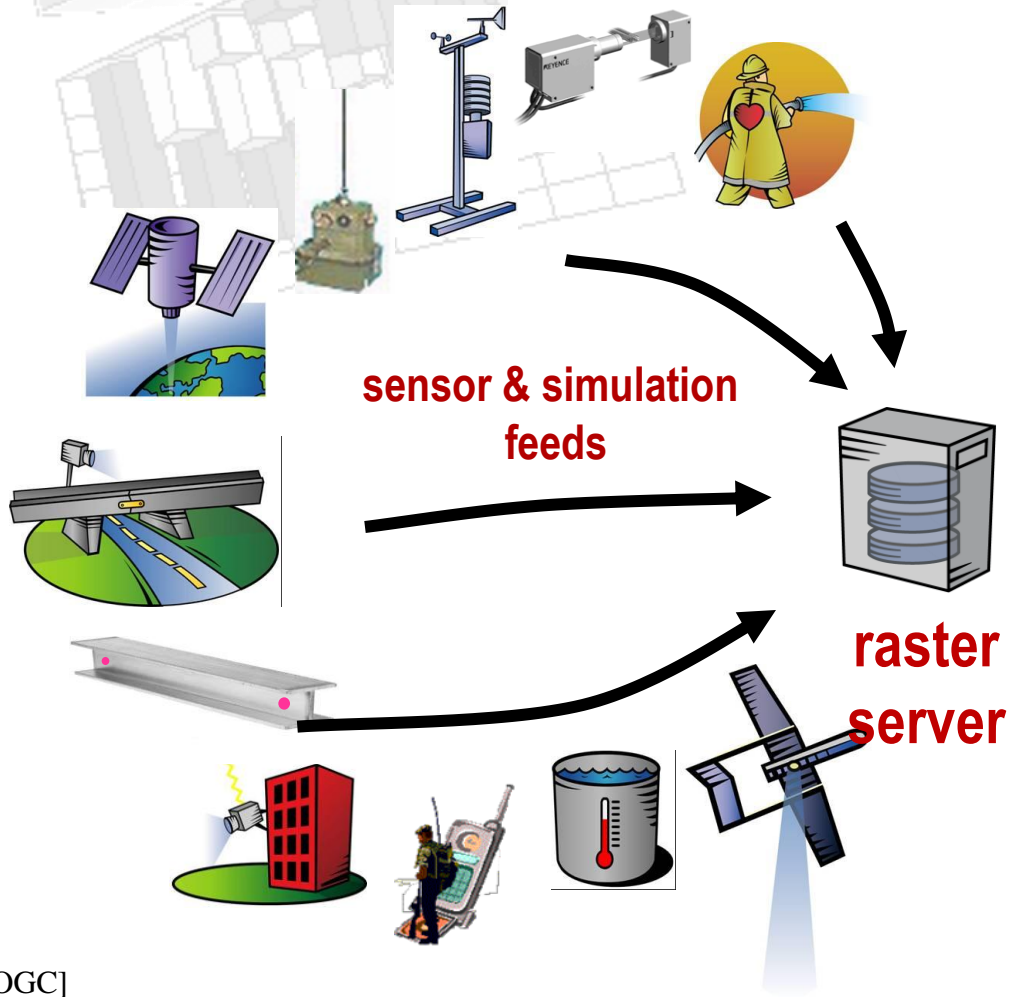
Hannover, 2012-oct-09

Peter Baumann

Jacobs University | rasdaman GmbH
Bremen, Germany

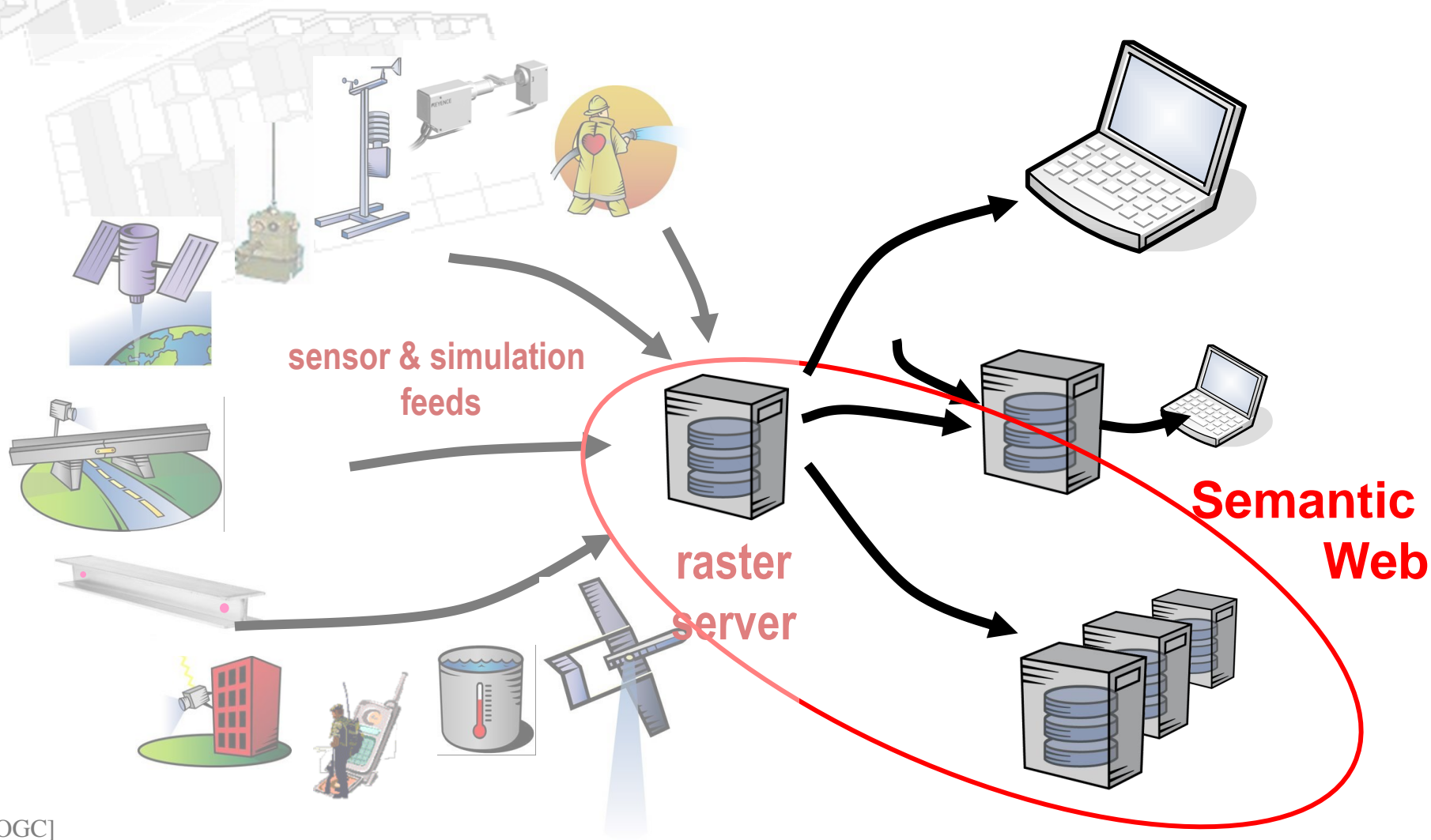
p.baumann@jacobs-university.de | baumann@rasdaman.com

Facing the Data Tsunami



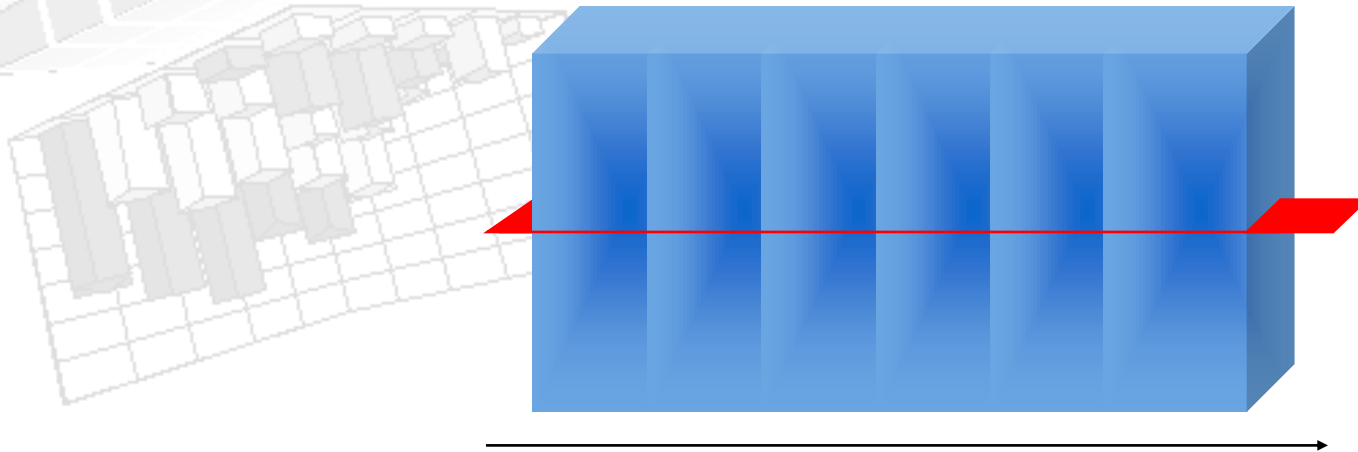
[OGC]

Taming the Data Tsunami



[OGC]

Let's Take a Closer Look...



- **Divergent access patterns** for ingest and retrieval
 - Alternative 1: simple access service, let client chisel result
 - Alternative 2: Deliver to exact needs
 - *no bandwidth waste, higher quality of service*

- Server must **mediate** between access patterns (...later more)
 - **Intelligent access interfaces** help

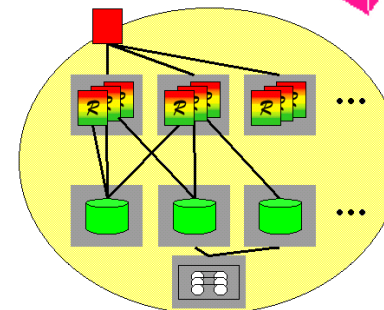
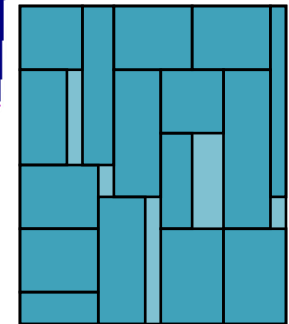
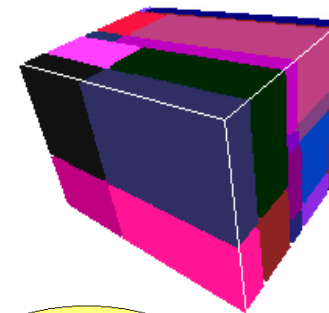
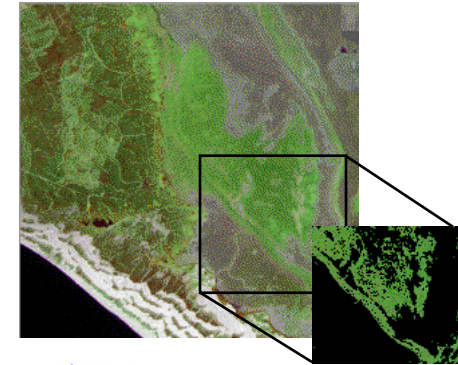
The rasdaman Raster Analytics Server

www.rasdaman.org

- **Array DBMS** for massive n-D raster data
 - new database attribute type: `array<celltype,extent>`
 - Data integration: rasters stored in standard database
- **Design goal:**
 - Fast, flexible, database integration:

```
select img.green[x0:x1,y0:y1] > 130
from LandsatArchive as img
```

- **Architecture: strictly tile-based**
 - n-D array → set of n-D **tiles**
 - extensive **optimization**, hw/sw parallelization
- **In operational use**
 - dozen-Terabyte objects
 - Analytics queries in 50 ms on laptop



Value-Added Satellite Image Archive

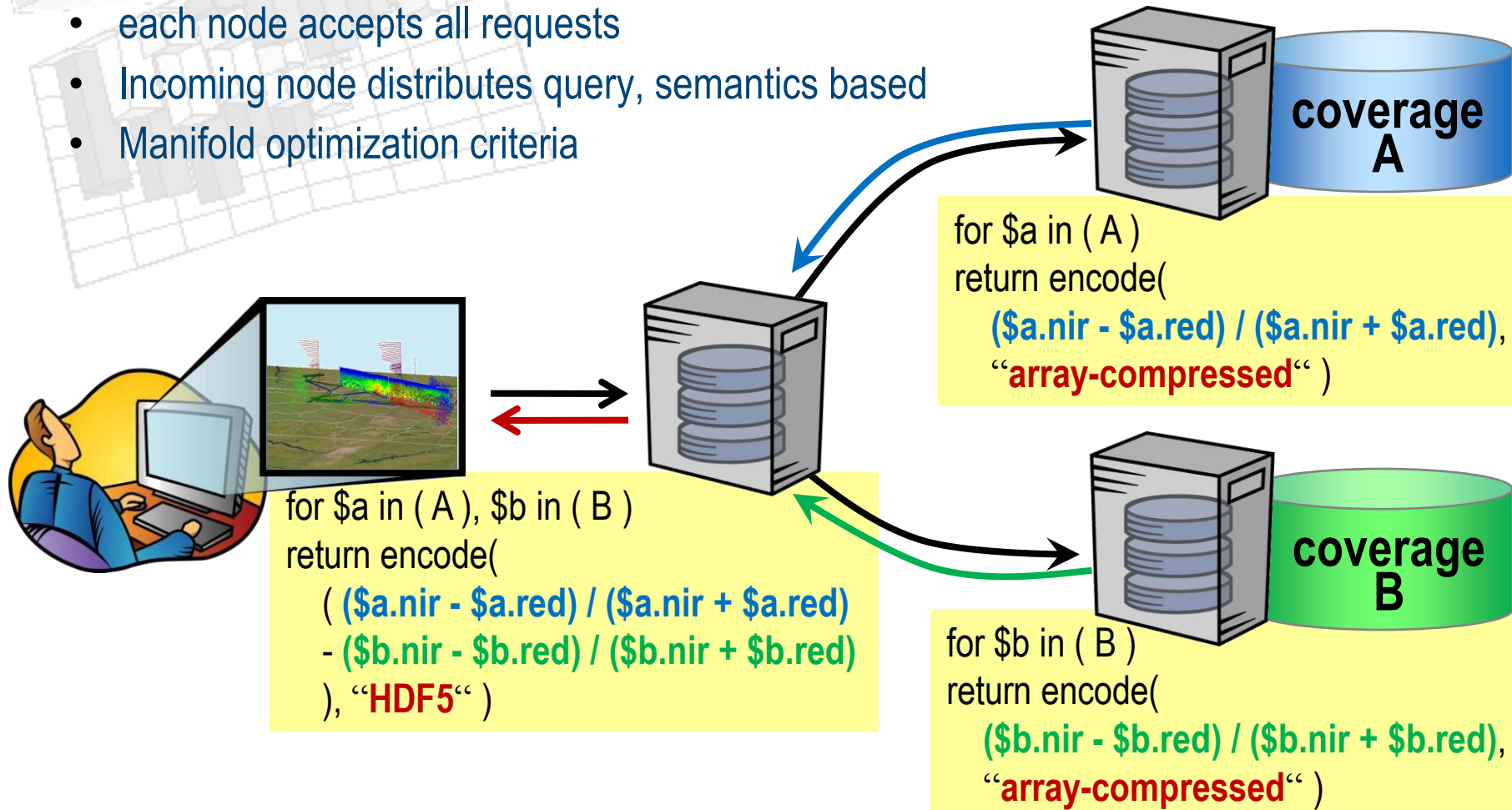
The screenshot displays the DLR EOWEB Interactive Data Service interface. At the top left is the DLR logo. The main header reads "DLR EOWEB - Interactive Data Service" and "A Service of the German Remote Sensing Data Center". Below this, a prompt says "Please, find available options below the image." The central part of the interface features a large satellite image of a coastal region, overlaid with a grid. A yellow-bordered inset on the left shows a zoomed-in view of a specific area, with vertical red and cyan lines indicating a data selection. Below the main image is a horizontal scale bar from 50 to 450. At the bottom, a date range selector is set to "1997-07-10 00:00:00" to "2000-07-24 23:59:59".

On the right side, there are several interactive elements and plots:

- A small plot showing a satellite image of a coastal area with a color scale from black to yellow.
- A button labeled "Click here to select another point".
- A note: "Note: Creating the plots needs a few moments. Click on a plot to get the numbers or to check for missing data."
- A line plot titled "SST 1999 X=Month, Y=Temperature [deg C]". The x-axis represents months (0 to 12) and the y-axis represents temperature (16 to 28). The plot shows a seasonal cycle peaking at approximately 27 degrees Celsius in month 8.
- Buttons labeled "Click here to select another point" and "Analyze dataset".
- A note: "Click on the Back button to return to date time selection. Click on the plot to get the numbers and to check for missing data."
- A bar chart titled "Jan 1996 - Dec 1998". The x-axis is labeled "X=Time[months]" and the y-axis is labeled "Y=NDVI". The plot shows monthly NDVI values with a peak of 0.709 in month 20 and a minimum of 0.261 in month 1.

Distributed Query Processing

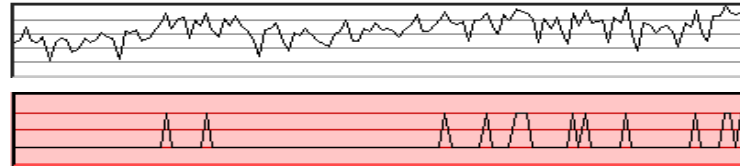
- WCPS peer-to-peer cloud
 - each node accepts all requests
 - Incoming node distributes query, semantics based
 - Manifold optimization criteria



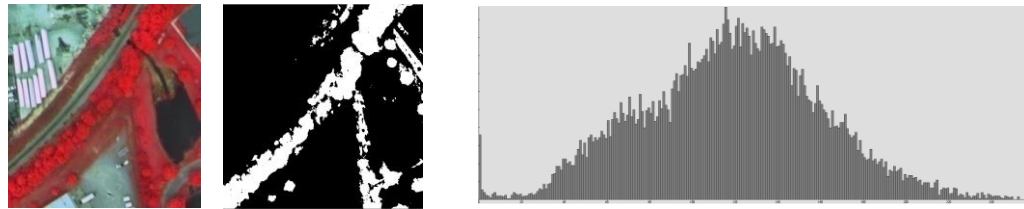
What Can You Do With rasdaman?

Raster Query Language: ad-hoc navigation, extraction, aggregation, analytics

- Time series



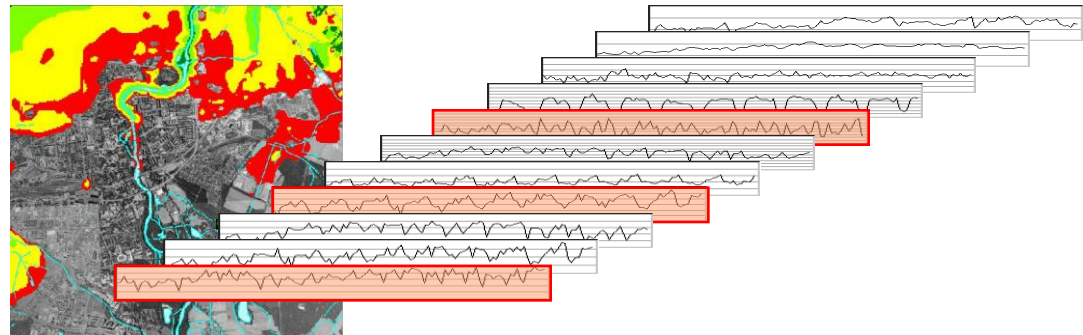
- Image processing



- Summary data

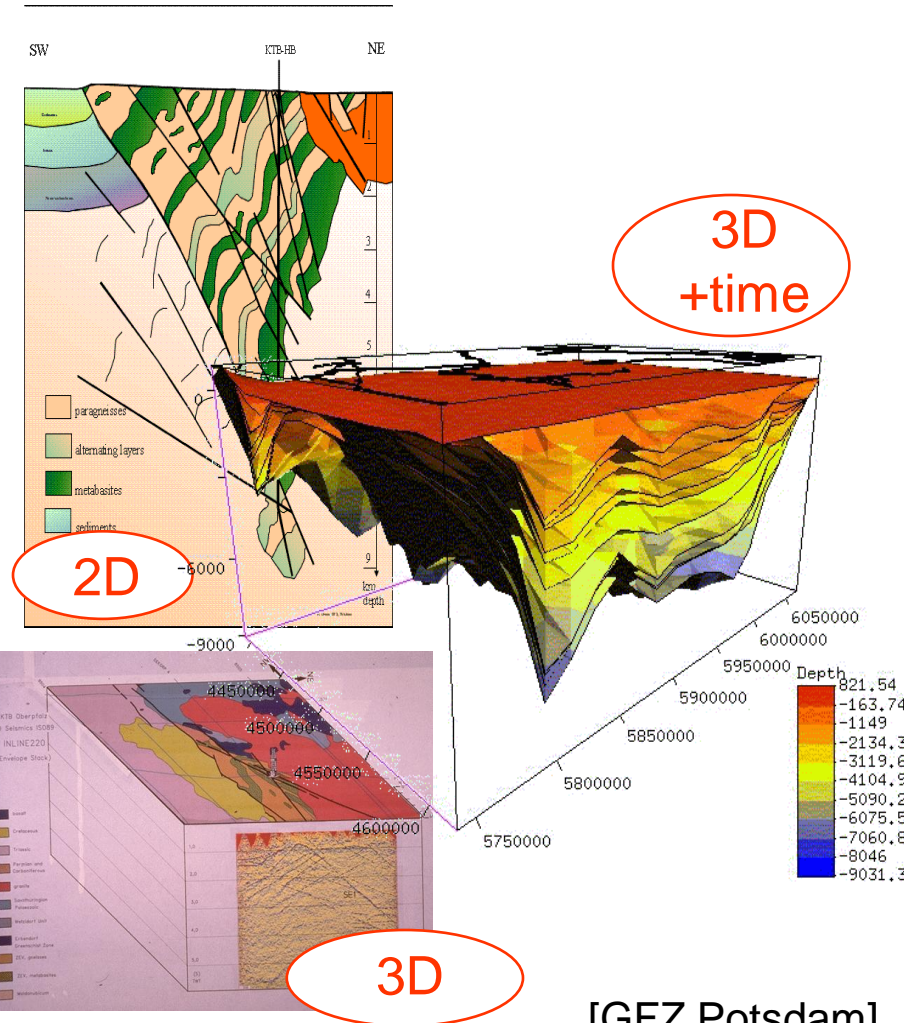
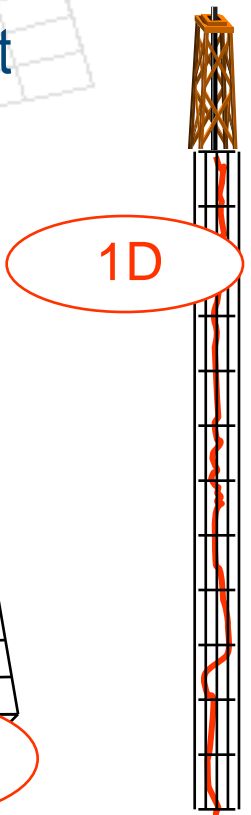
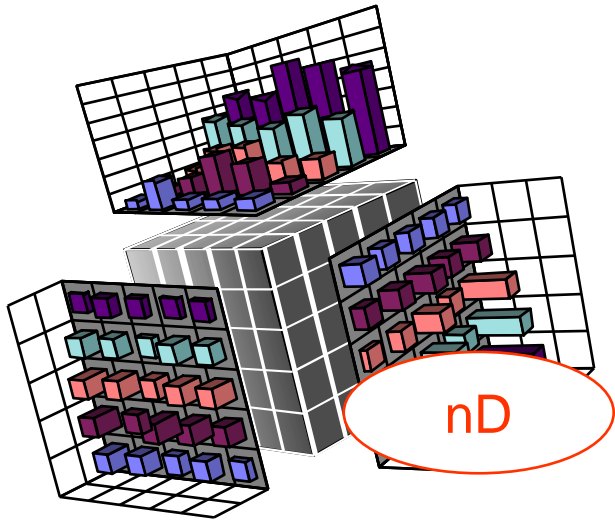
- current value is **8220.0**;
- average over all values up to now currently is **7461.7692307692305**.

- Sensor fusion & pattern mining



The Vision: Integrated Geo Warehouse

- Comprehensive geophysics data mgmt
- Integrating data + meta data



[GFZ Potsdam]



EarthServer: *Big Earth Data Analytics*

- Scalable On-Demand Processing for the Earth Sciences
 - EU funded, 3 years, 5.85 mEUR
- Platform: rasdaman
 - Distributed query processing, integrated data/metadata search, 3D clients
- Strictly open standards: OGC WMS+WCS+WPCS; W3C Xquery; X3D
- 6 * 100+ TB databases for all Earth sciences + planetary science

Cryospheric Science
landcover mapping

EOX

Airborne Science
high-altitude long-endurance drones

NASA

Atmospheric Science
climate variables

MEEO
Meteorological Environmental Earth Observation

Geology
geological models

BGS
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Oceanography
marine model runs + in-situ data

PML | PLYMOUTH MARINE LABORATORY

Planetary Science
Mars geology

JACOBS UNIVERSITY

Conclusion

- n-D rasters main contributor to Big Geo Data
= **Sensor, image, & statistics data**
 - High volumes
 - Manifold data structures, interconnected
 - complex analytics

- rasdaman transposes traditional DBMS
benefits to rasters:
 - **Flexible** processing & filtering
 - **Scalable** & **fast** data access
 - **Standards** based

- See also:
 - www.rasdaman.org
 - www.earthlook.org
 - www.ogcnetwork.net/wcs

