



# The EarthServer initiative: *towards Agile Big Data Services*

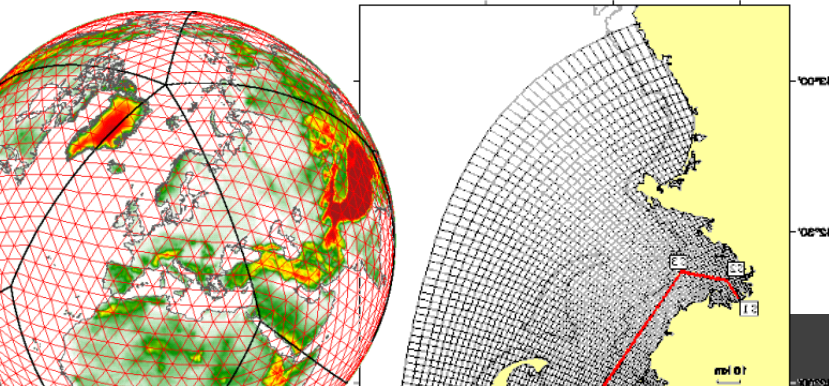
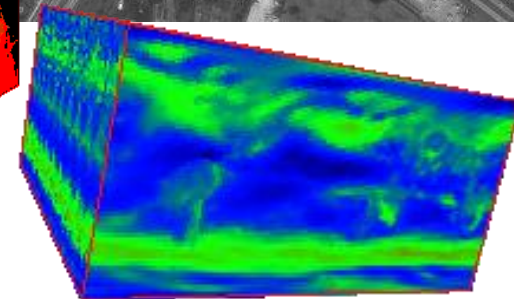
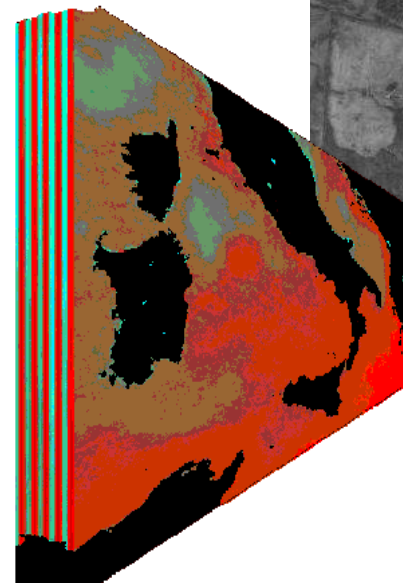
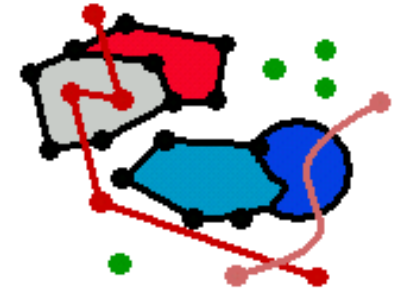
2nd GEOSS Science and Technology Stakeholder Workshop  
Bonn, Germany, 2012-aug-29

**Peter Baumann**

Jacobs University | rasdaman GmbH  
Bremen, Germany  
[p.baumann@jacobs-university.de](mailto:p.baumann@jacobs-university.de)

# Feature and Coverage Data Standards

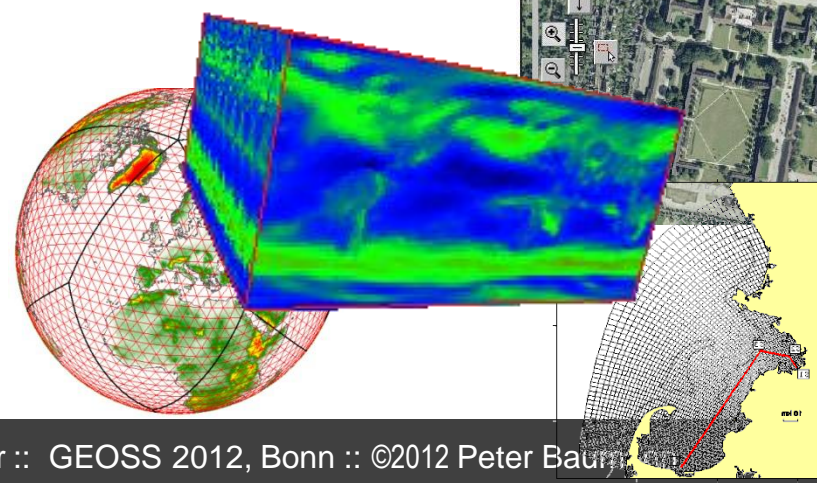
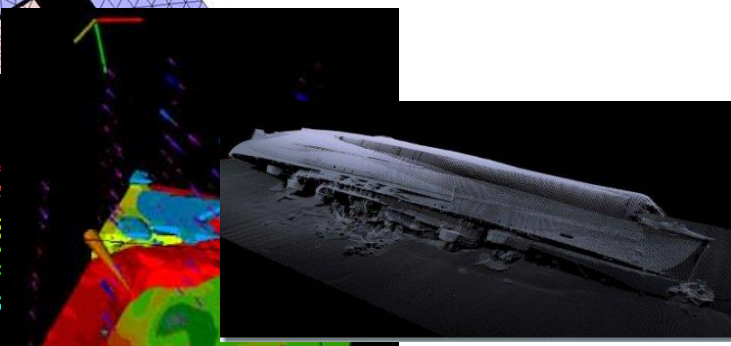
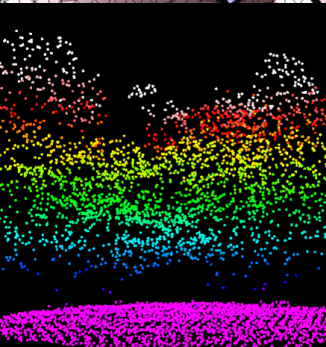
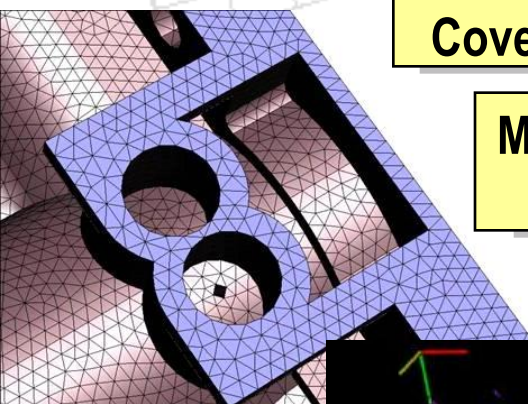
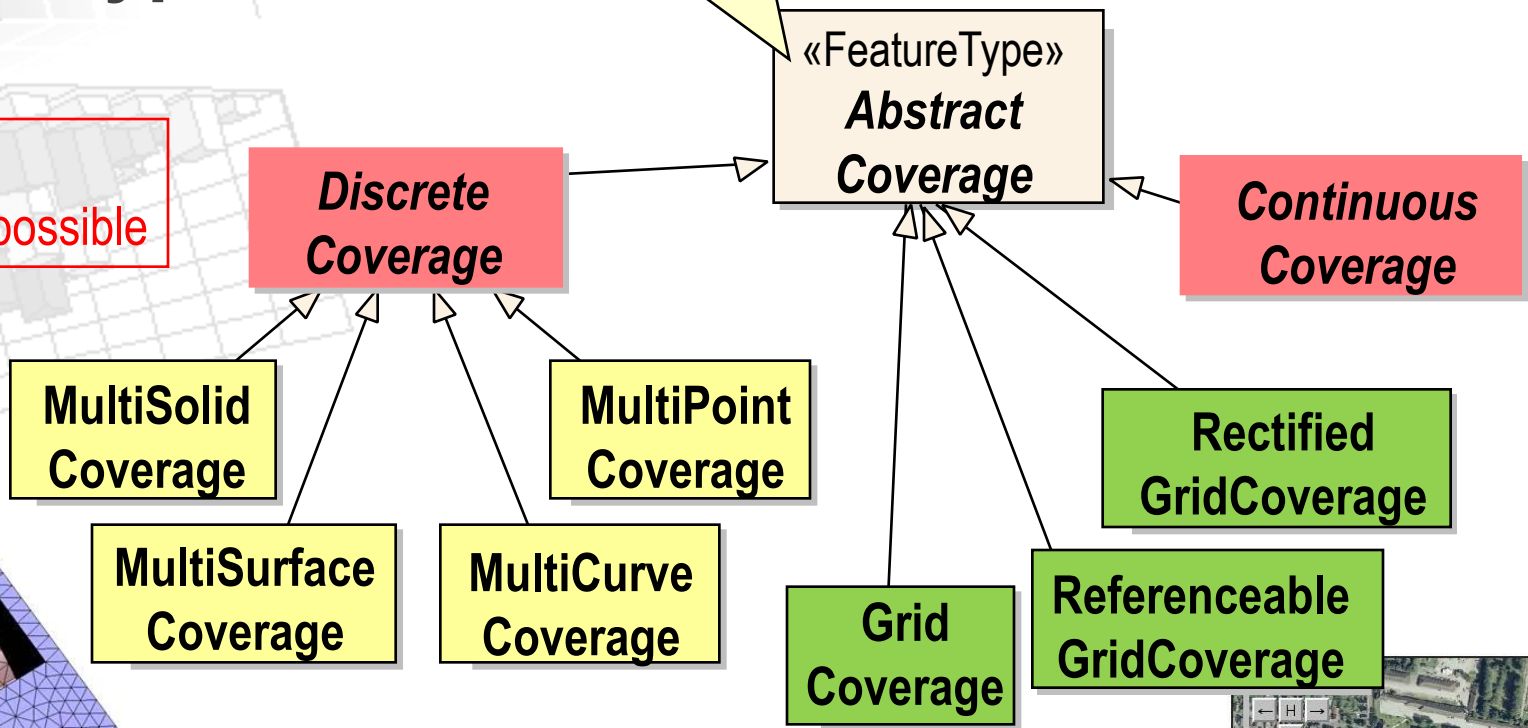
- Core element in OGC: geographic **feature**
  - = abstraction of a real world phenomenon
  - associated with a location relative to Earth
  
- Special kind of feature: **coverage**
  - = space-time varying multi-dimensional phenomenon
  - Typical representative: **raster image**
  - *...but there is more!*
  
- Typically, coverages are **Big Data**



# Coverage Types

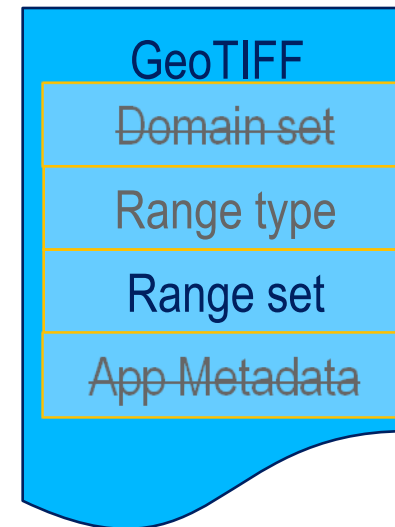
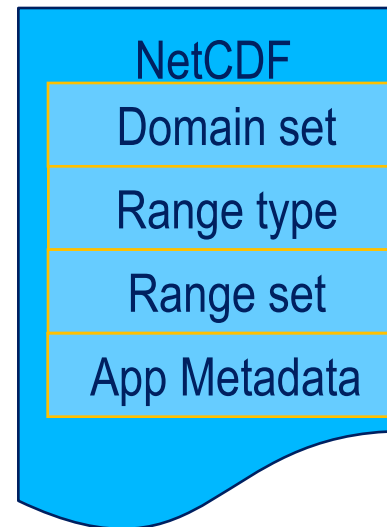
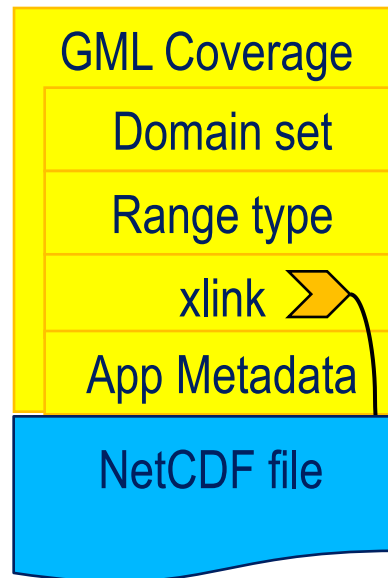
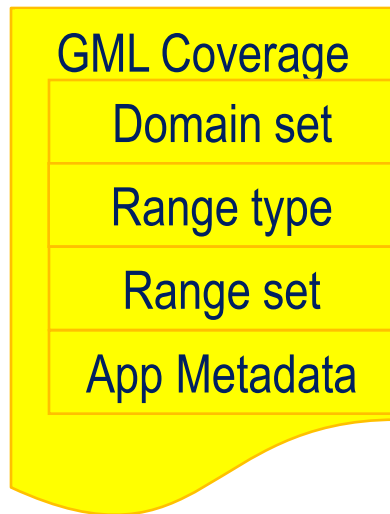
as per GML 3.2.1

all n-D  
New subtypes possible

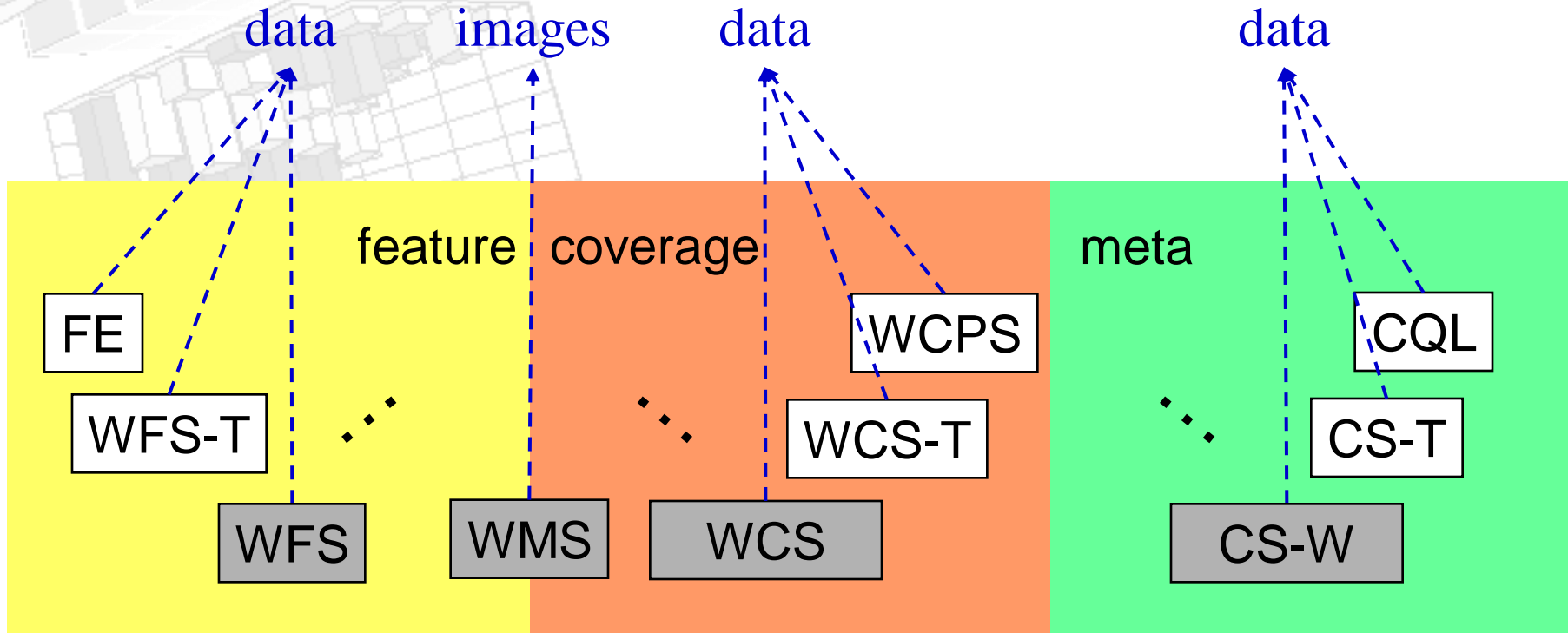


# Coverage Encoding

- **Pure GML**: complete coverage represented by GML
- **Special Format**: other suitable file format (ex: MIME type “image/tiff”)
- **Multipart-Mixed**: multipart MIME, type “multipart/mixed”



# Core OGC Service Standards



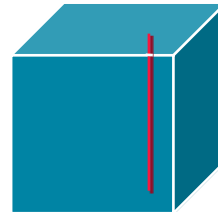
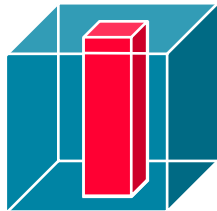
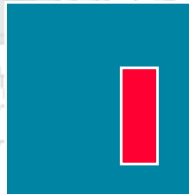
- WMS "portrays spatial data" → pictures
- WCS "provides data + descriptions; data with original semantics, may be interpreted, extrapolated, etc."

[09-110r4]

# Web Coverage Service (WCS)

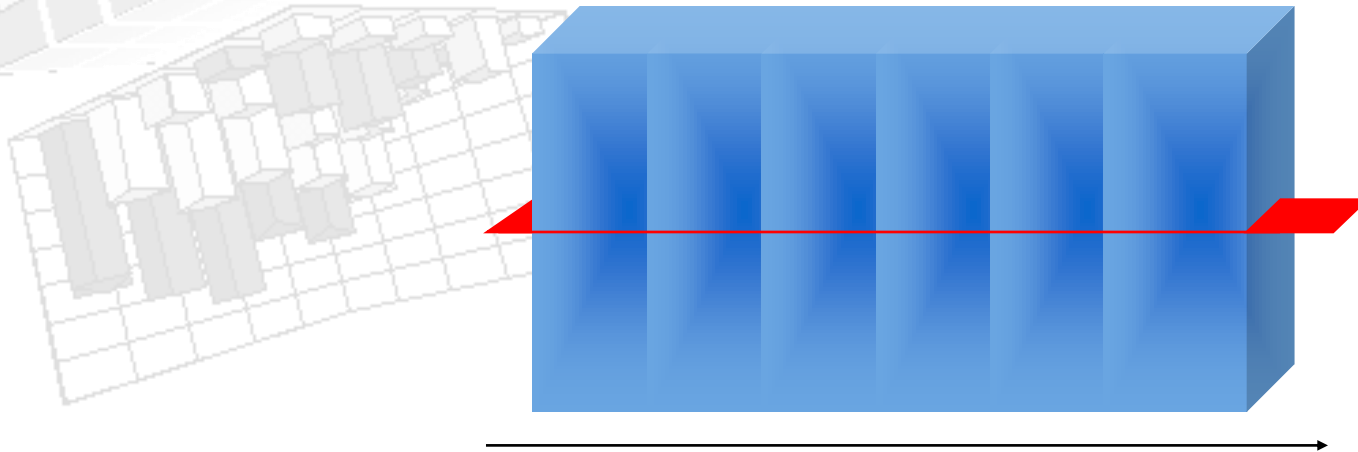
- **Core:** Simple & efficient access to multi-dimensional coverages

- subset = **trim** | **slice**



- **WCS Extensions** for additional functionality facets
  - “band extraction”, scaling, reprojection, interpolation, query language, ...
- **Application Profiles** define domain-oriented bundling

# 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



# EarthServer: *Big Earth Data Analytics*

- Scalable On-Demand Processing for the Earth Sciences
  - EU funded, 3 years, 5.85 mEUR
  - Platform: rasdaman (Array Analytics server) ↘
  - 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*

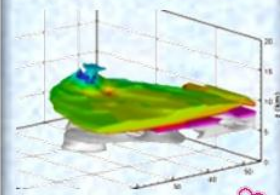



**Airborne Science**  
*high-altitude long-endurance drones*






**Atmospheric Science**  
*climate variables*



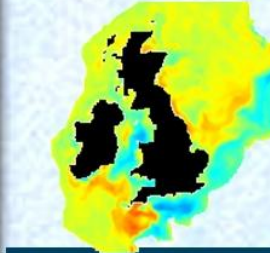
**MEEO**  
Meteorological Environmental Earth Observation

**Geology**  
*geological models*



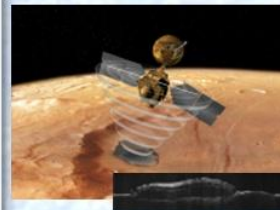

**British Geological Survey**  
NATURAL ENVIRONMENT RESEARCH COUNCIL

**Oceanography**  
*marine model runs + in-situ data*



**PML** PLYMOUTH MARINE LABORATORY

**Planetary Science**  
*Mars geology*

JACOBS UNIVERSITY



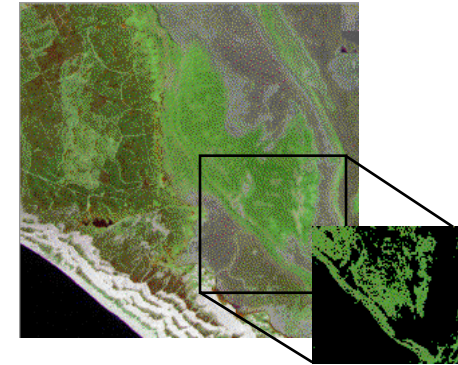
# The rasdaman Raster Analytics Server

[www.rasdaman.org](http://www.rasdaman.org)

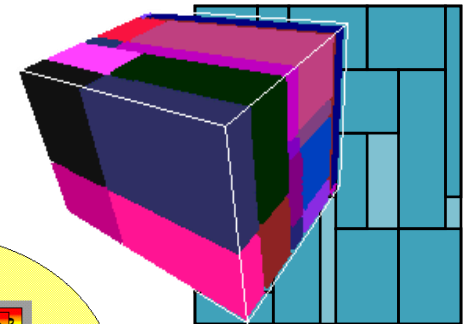
- **Array DBMS** for massive n-D raster data
  - new database attribute type: `array<celltype,extent>`
  - Data integration: rasters stored in standard database

- Extending ISO SQL with array operators:

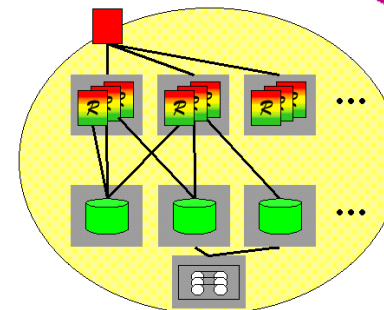
```
select img.green[x0:x1,y0:y1] > 130
from LandsatArchive as img
```



- “tile streaming” architecture
  - 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



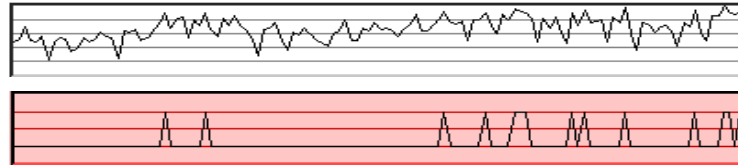
# Value-Added Satellite Image Archive

The screenshot displays the DLR EOWEB Interactive Data Service interface. At the top left is the DLR logo. The main title is "DLR EOWEB - Interactive Data Service" and "A Service of the German Remote Sensing Data Center". Below the title, it says "Please, find available options below the image." The interface features a large satellite image of a coastal region with a yellow grid overlay. A zoomed-in view of a specific area is shown on the left, with vertical red and blue lines indicating a data strip. A horizontal scale bar at the top of the main image shows values from 50 to 450. Below the main image, there is a date range selector: "Date time range from 1997-07-10 00:00:00 to 2000-07-24 23:59:59". On the right side, there are several interactive elements: a "Click here to select another point" button, a 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]" showing a seasonal cycle peaking at approximately 27 degrees Celsius in August, and a "Click here to select another point" button. Below that, there is a "Click on the Back button to return to data selection. Click on the plot to get the numbers and to check for missing data." and a bar chart titled "Jan 1996 - Dec 1998" showing monthly NDVI values with a peak of 0.709 in August 1997 and a minimum of 0.251 in February 1997.

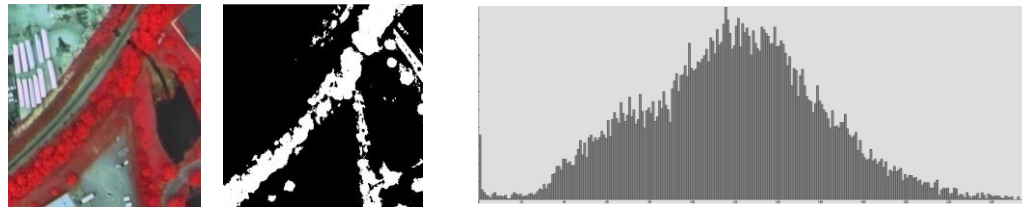
# Web Coverage Processing Service (WCPS)

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

- Time series



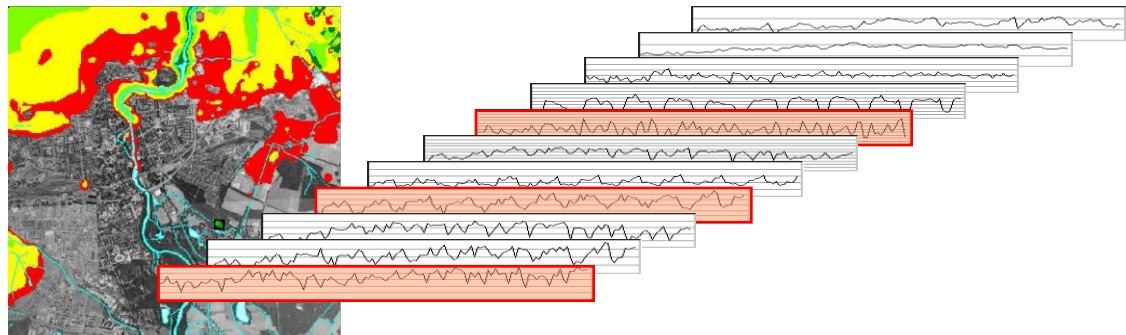
- Image processing



- Summary data

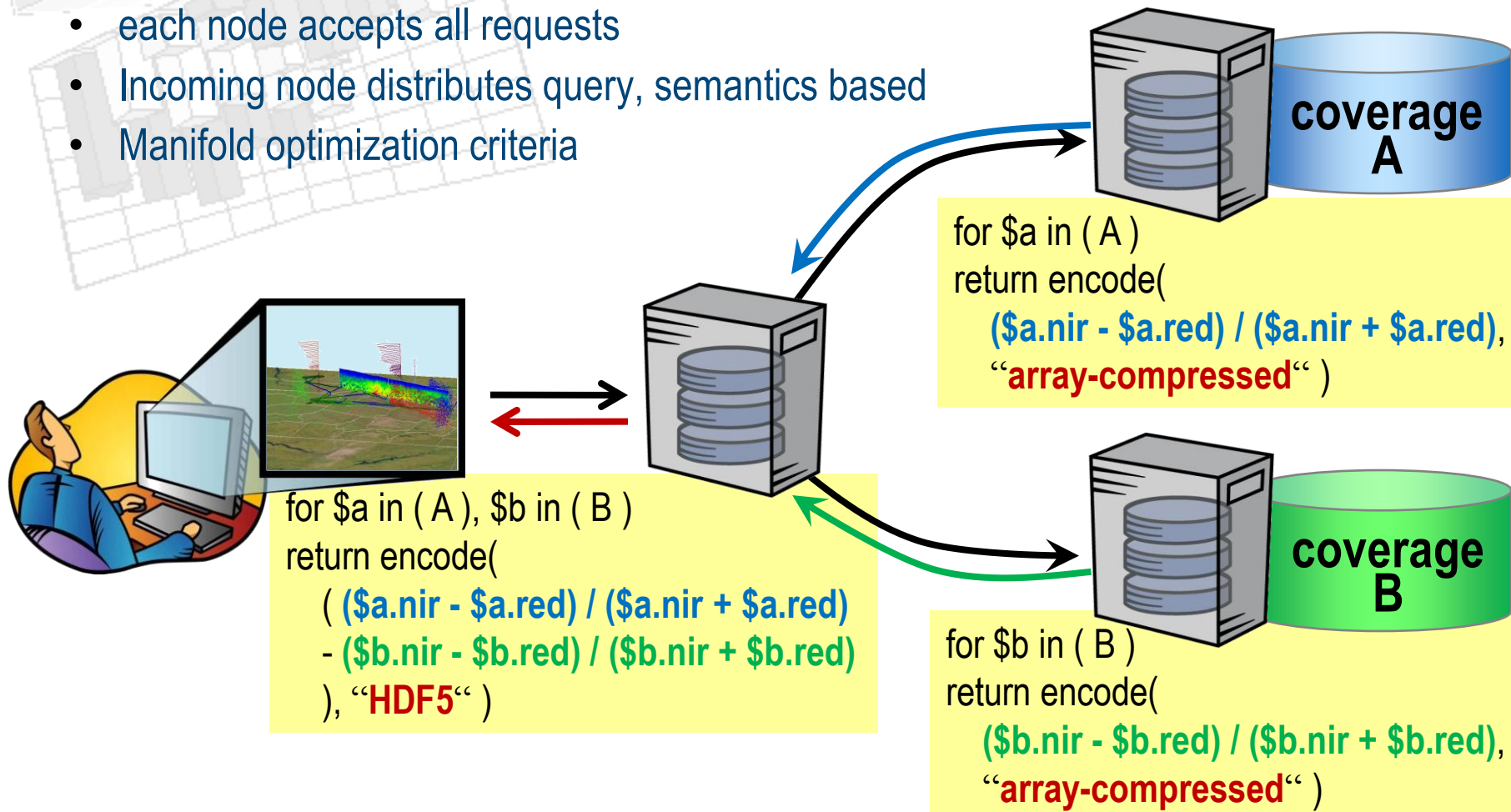
- current value is **8220.0**;
- average over all values up to now currently is **7461.7692307692305**.

- Sensor fusion & pattern mining



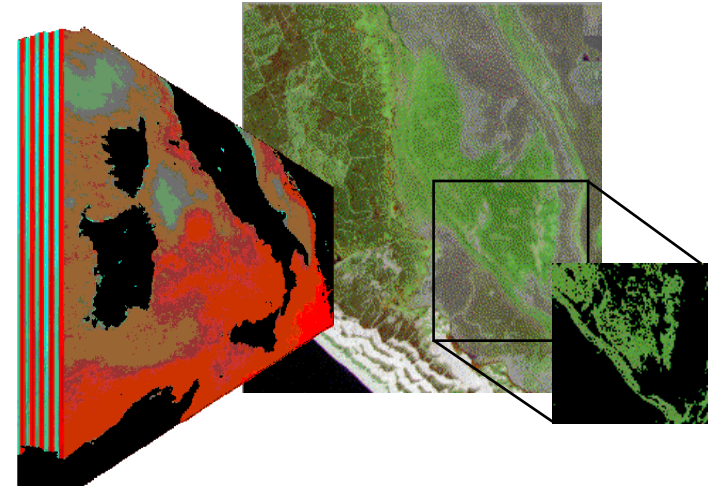
# rasdaman: Distributed Query Processing

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



# Conclusion

- **Sensor, image, & statistics data**  
= a main source of Big Data  
in Earth Sciences
  - Petrol industry has „more bytes than barrels“
- **OGC standards** offer common platform
  - spatio-temporal **coverages** – *a unified, cross-domain data model*
  - **Web Coverage Service** suite – *from simple download to flexible analytics*
  - [www.ogcnetwork.net/wcs](http://www.ogcnetwork.net/wcs)
- EarthServer can contribute **Agile Analytics**  
to GEOSS
  - OGC coverage standards
  - rasdaman technology



[www.earthserver.eu](http://www.earthserver.eu)