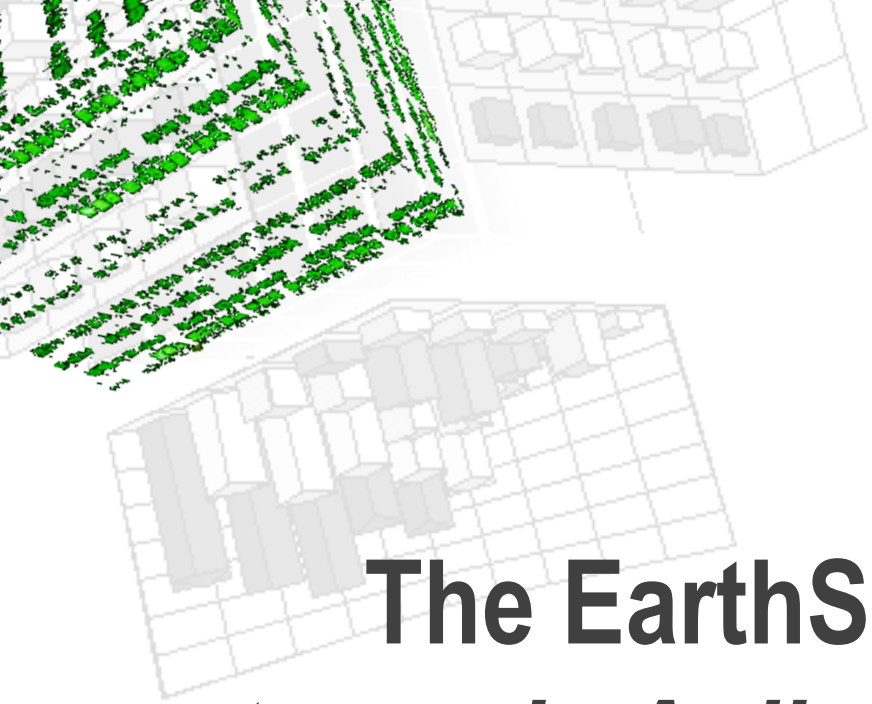


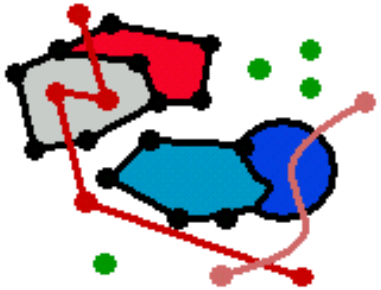
The EarthServer initiative: *towards Agile Big Data Services*

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

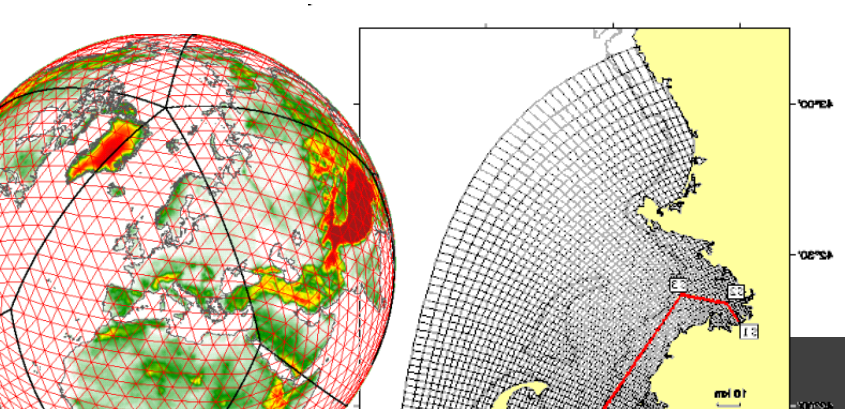
Peter Baumann

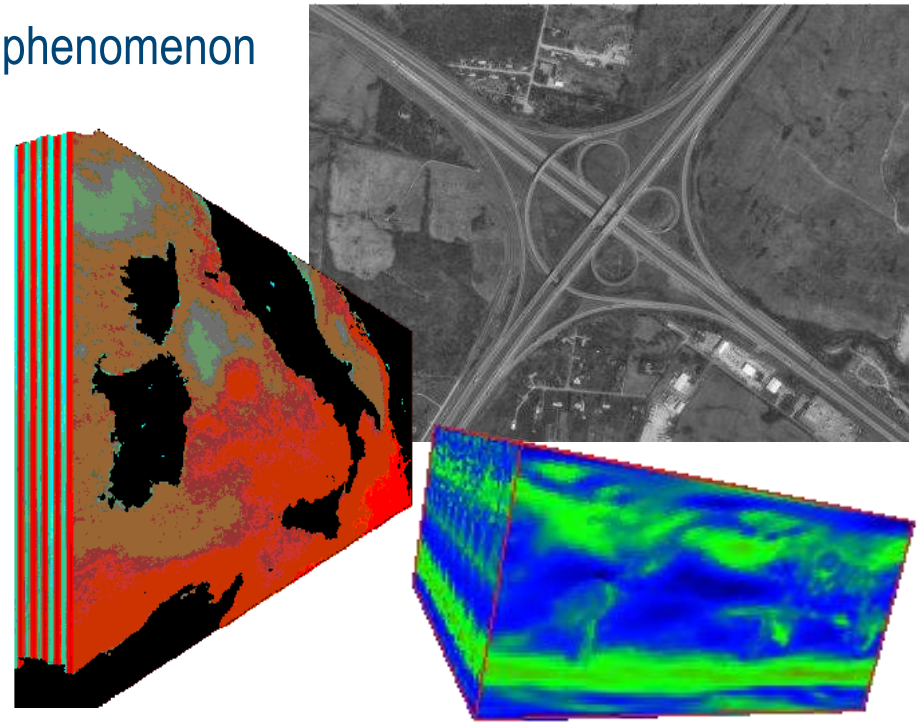
Jacobs University | rasdaman GmbH
Bremen, Germany
p.baumann@jacobs-university.de

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

«FeatureType»
Abstract Coverage

all n-D
New subtypes possible

Discrete Coverage

Continuous Coverage

MultiSolid Coverage

MultiPoint Coverage

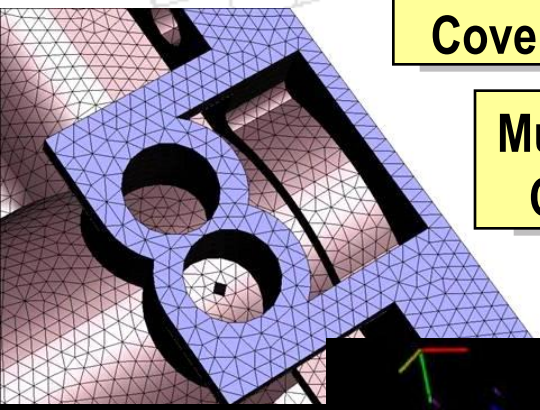
MultiSurface Coverage

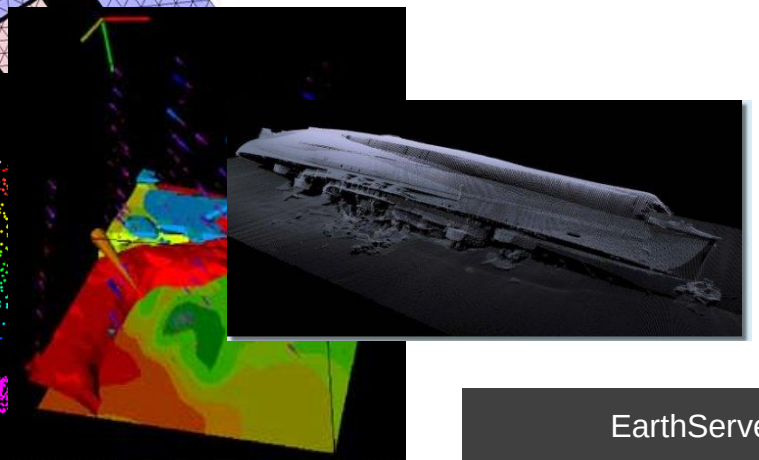
MultiCurve Coverage

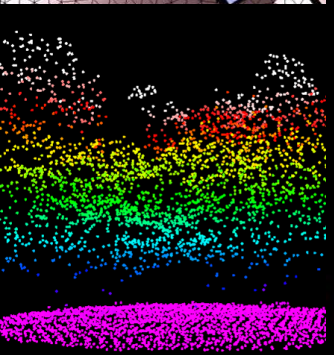
Grid Coverage

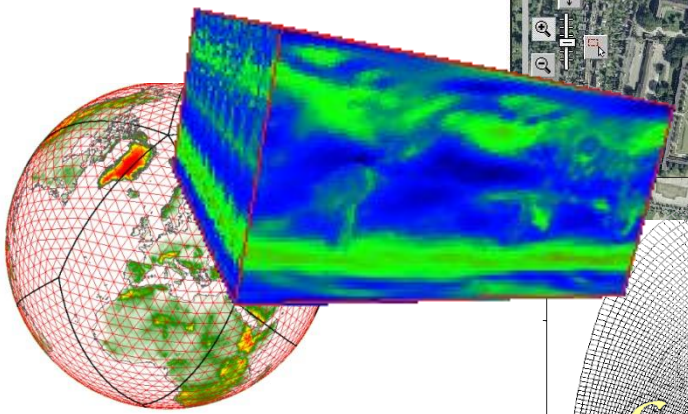
Referenceable GridCoverage

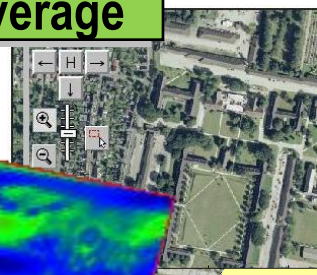
Rectified GridCoverage

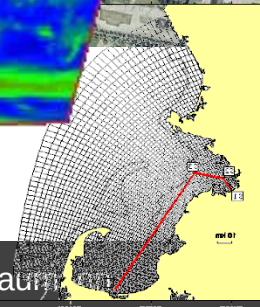






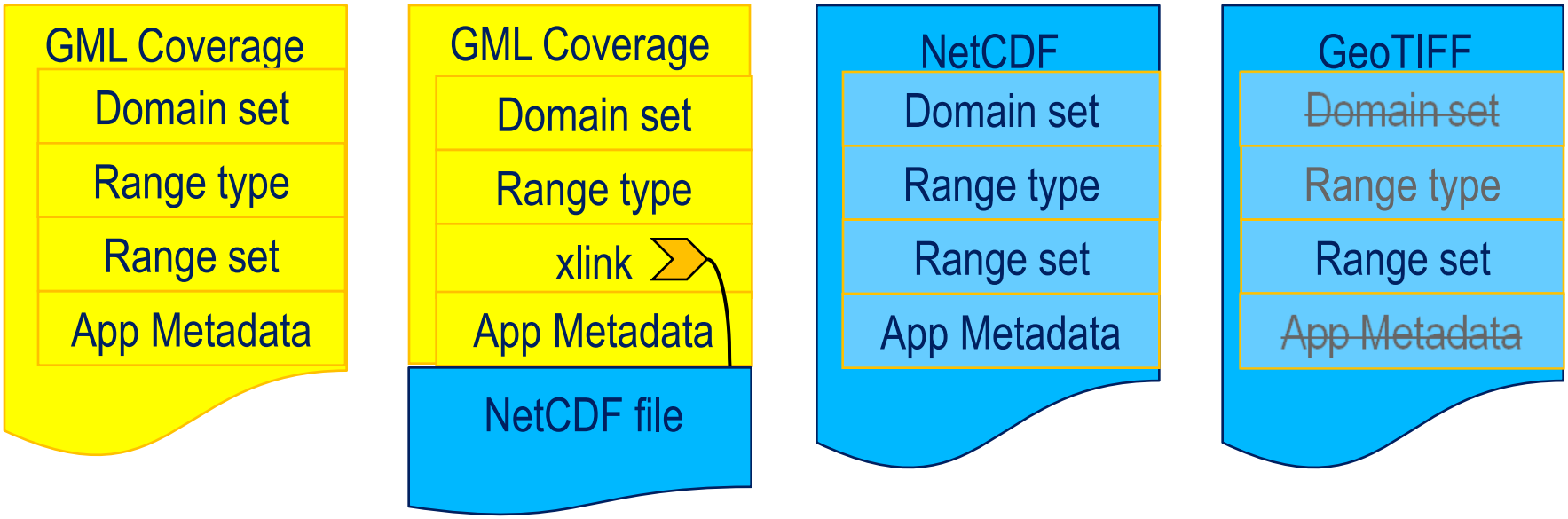




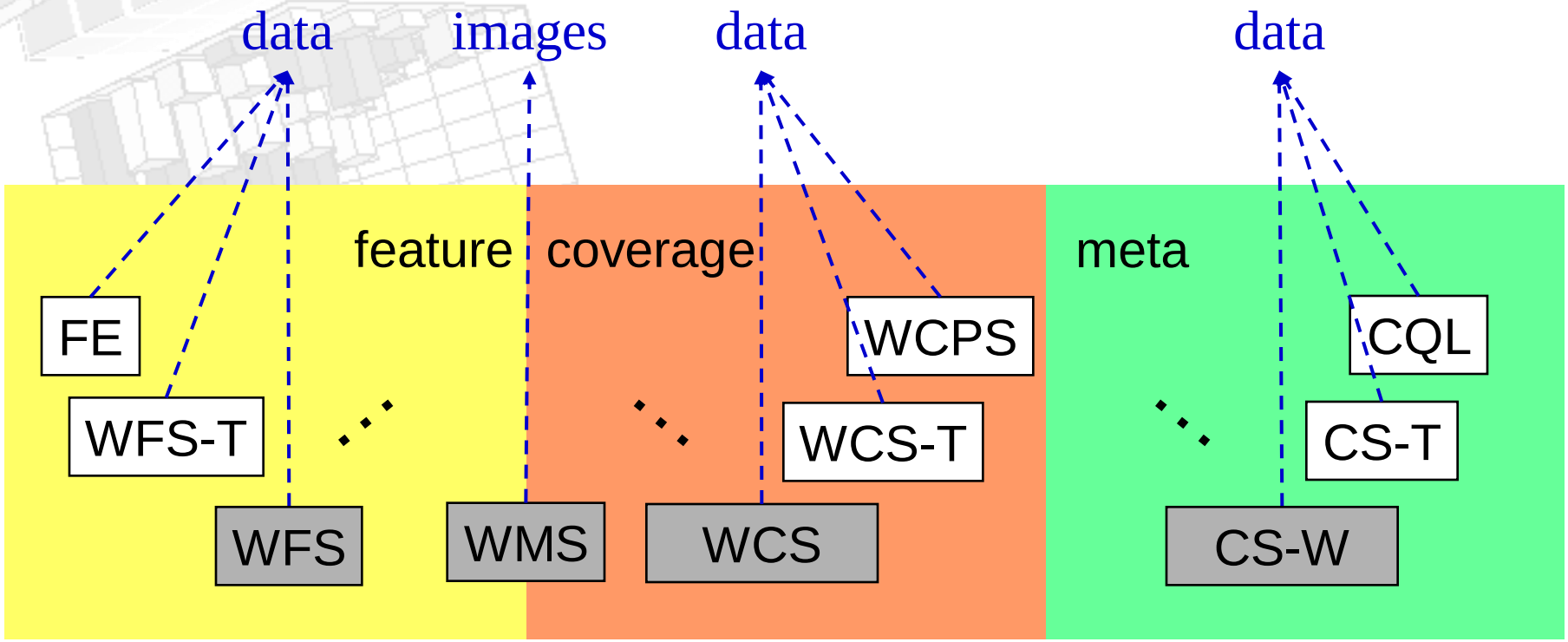


Coverage Encoding

- 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

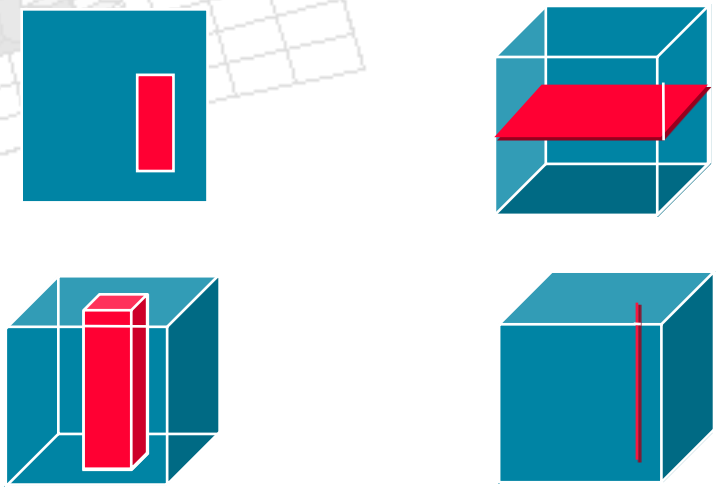


- WMS "portrays spatial data" → pictures
- WCS "provides data + descriptions; data with original semantics, may be interpreted, extrapolated, etc."

[09-110r4]

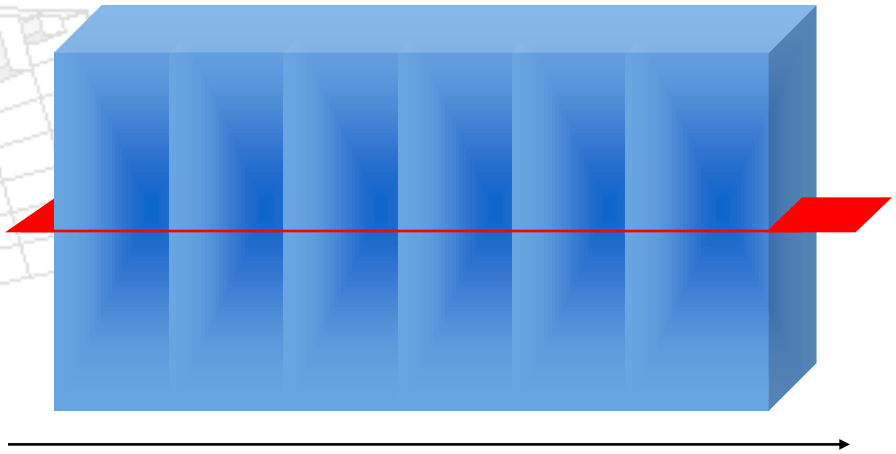
Web Coverage Service (WCS)

- 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

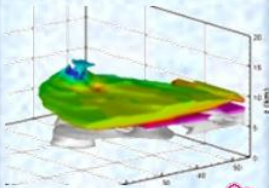
EarthServer: *Big Earth Data Analytics*

- 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+WCPS; W3C Xquery; X3D
- 6 * 100+ TB databases for all Earth sciences + planetary science

Cryospheric Science
landcover mapping

Airborne Science
high-altitude long-endurance drones

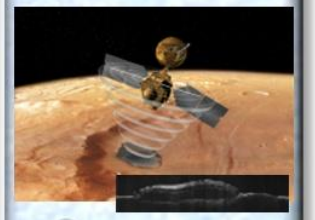
Atmospheric Science
climate variables



Geology
geological models

Oceanography
marine model runs + in-situ data

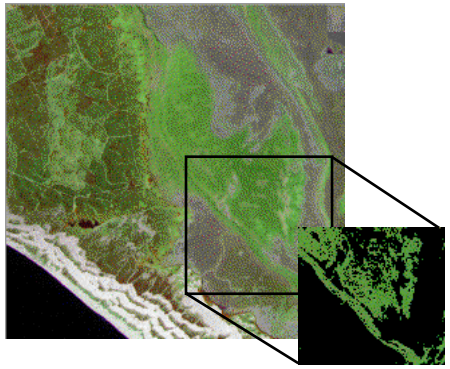
Planetary Science
Mars geology



The rasdaman Raster Analytics Server

www.rasdaman.org

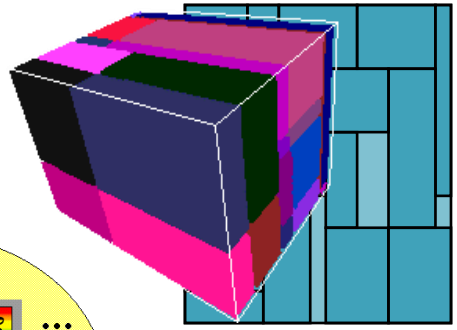
- Array DBMS for massive n-D raster data
  - new database attribute type: array<celltype,extent>
  - Data integration: rasters stored in standard database



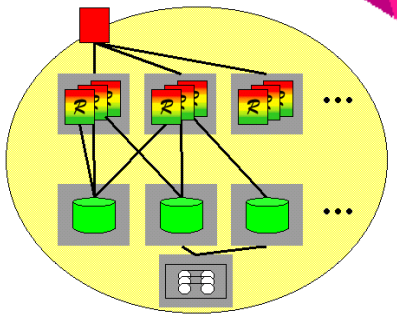
- Extending ISO SQL with array operators:

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

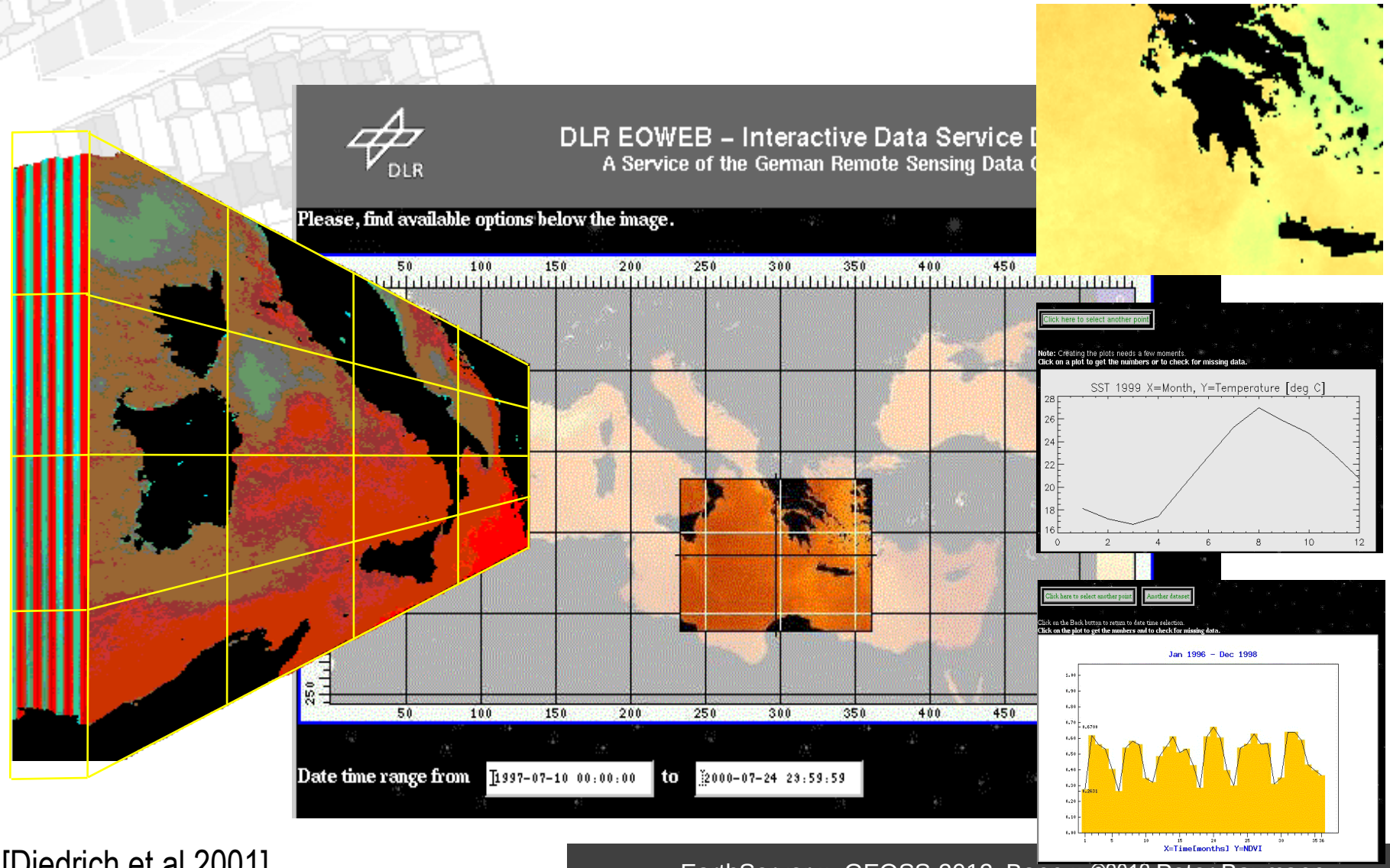
- "tile streaming" architecture
  - n-D array → set of n-D tiles
  - extensive optimization, hw/sw parallelization



- In operational use
  - dozen-Terabyte objects
  - Analytics queries in 50 ms on laptop



Value-Added Satellite Image Archive

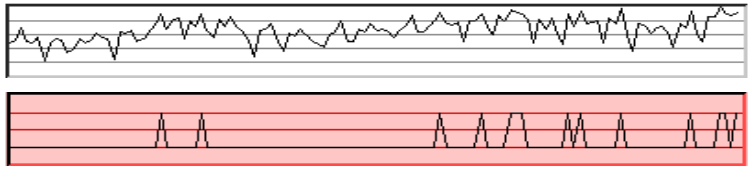


[Diedrich et al 2001]

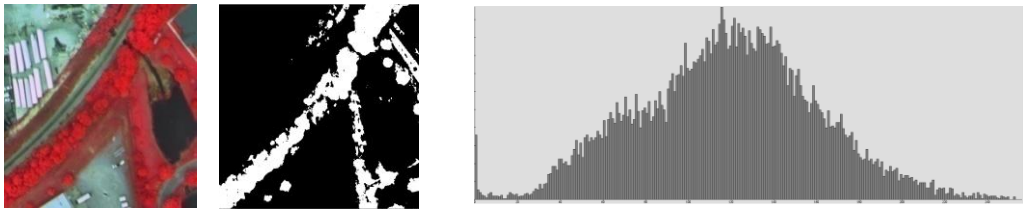
Web Coverage Processing Service (WCPS)

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

- Time series



- Image processing

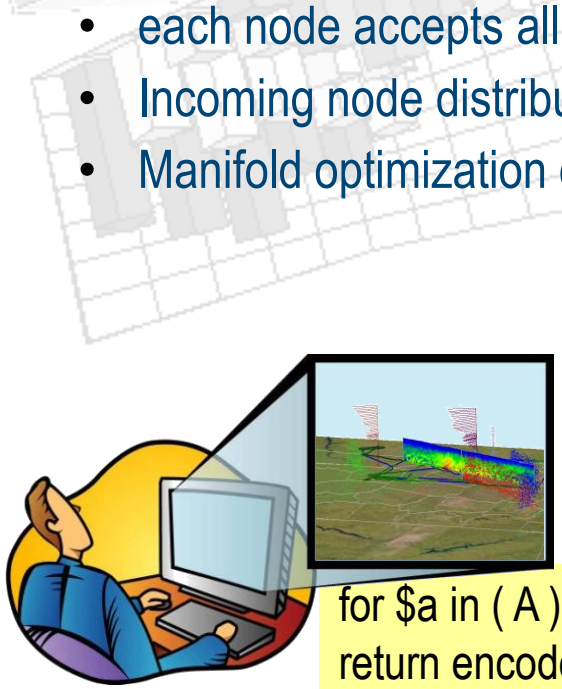


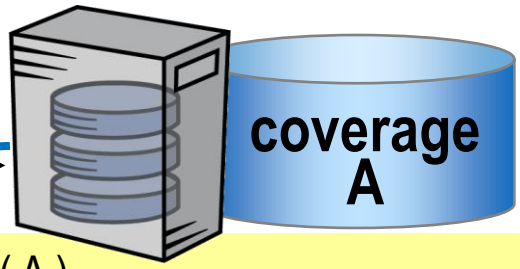
- Summary data

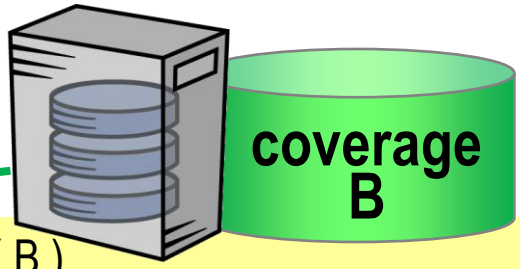
- Sensor fusion & pattern mining

rasdaman: Distributed Query Processing

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







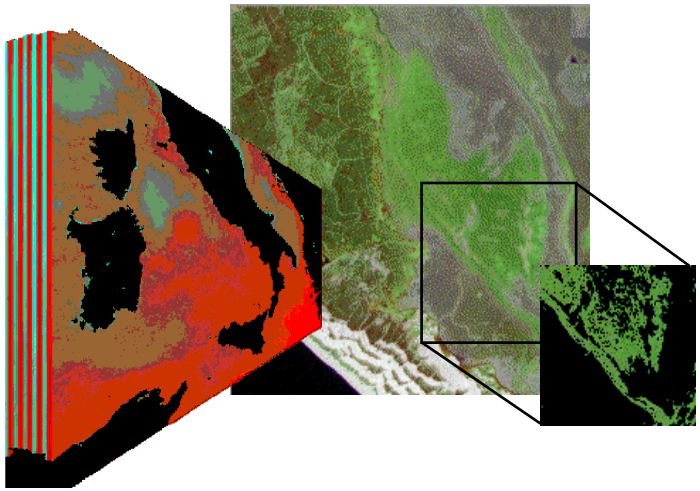

```
for $a in ( A )
return encode(
    ($a.nir - $a.red) / ($a.nir + $a.red),
    "array-compressed" )
```

```
for $a in ( A ), $b in ( B )
return encode(
    ( ($a.nir - $a.red) / ($a.nir + $a.red)
    - ($b.nir - $b.red) / ($b.nir + $b.red)
    ), "HDF5" )
```

```
for $b in ( B )
return encode(
    ($b.nir - $b.red) / ($b.nir + $b.red),
    "array-compressed" )
```

[Owonibi 2012]

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
- EarthServer can contribute Agile Analytics
  to GEOSS
  - OGC coverage standards
  - rasdaman technology

www.earthserver.eu