Approach & Technology

FOSS4G-Europe, Bremen, 2014-07-14

Peter Baumann
Jacobs University | rasdaman GmbH
p.baumann@jacobs-university.de

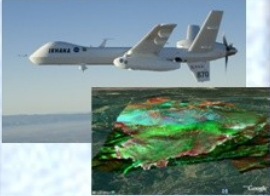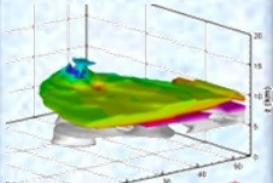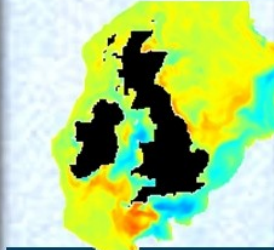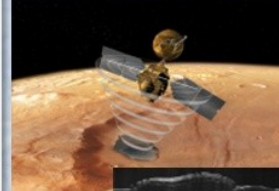# EarthServer Goals

- **Scalable On-Demand Processing** for the Earth Sciences
  - Agile geo analytics – query language approach
  - Platform: pioneer Array Database technology, rasdaman
    - *Extend with: Integrated filtering & processing on metadata, regular/irregular grids, point clouds, ...*

- Large-scale deployment – 6 Lighthouse Applications covering Earth & Planetary Sciences
  - 4x 20 TB, 2x 100 TB

- Open standards – client/server interfaces are strictly relying on WMS, WCS, WCPS
  - Strong impact on standards development

Earth Science Lighthouse Applications

- 6 Lighthouse Applications covering all Earth Sciences

- Established data centers adding EarthServer technology to service portfolio

- Summer 2014: ~260 TB operational Earth & Planetary data

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
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Oceanography
marine model runs + in-situ data
PML | PLYMOUTH MARINE LABORATORY

Planetary Science
Mars geology
JACOBS UNIVERSITY

EarthServer Technology Advances

- Big Geo Data engine development, based on rasdaman Array Database

- Geo service interfaces in rasdaman: OGC WMS, WCS, WCPS

- Extending data types: Regular & irregular grids, point clouds, meshes
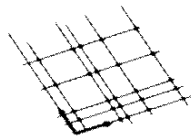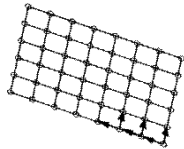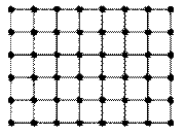
- Coupling: Hadoop, R, MatLab, MapServer, ...

- Data/metadata search integration

- Scalability: distributed processing

- Visual 1D/2D/3D client toolkit, mobile clients

rasdaman: Agile Array Analytics

- „raster data manager": SQL + n-D raster objects

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

- Scalable parallel "tile streaming" architecture

- In operational use since many years

- OGC WCS Core Reference Implementation
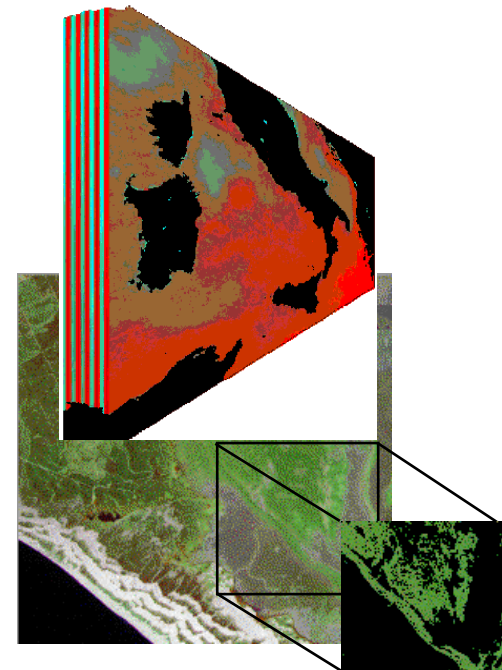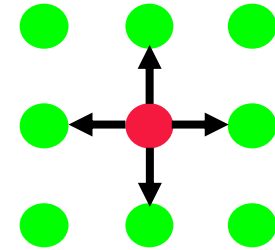
rasdaman website visitors

The rasdaman Array Database

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

```
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, several innovation awards
  - OGC WCS Core Reference Implementation

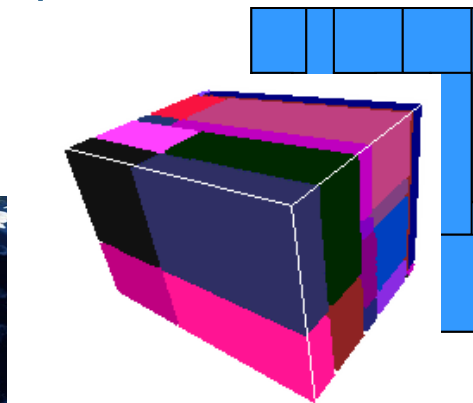**rasdaman Web visitors**
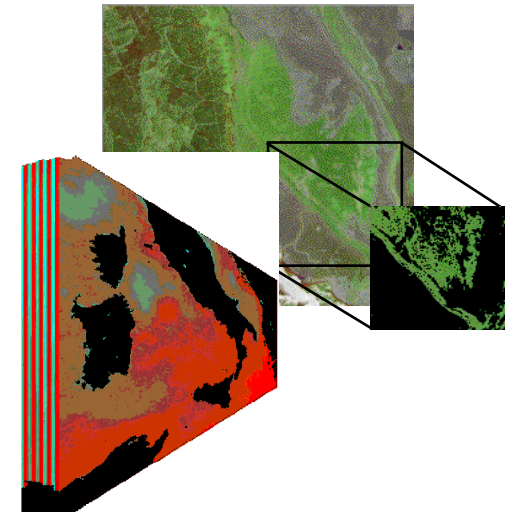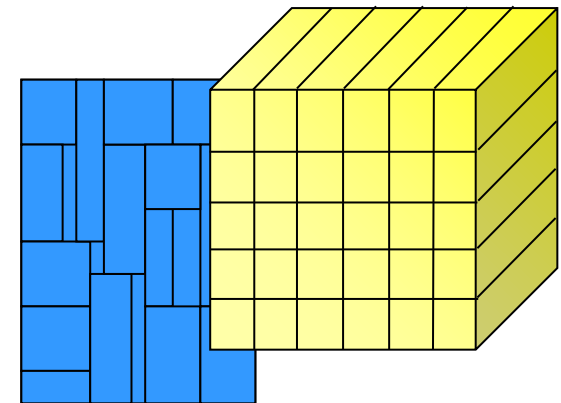
The rasdaman Modular Server Toolkit

- modular & powerful, but easy handling through meaningful defaults
  - UNIX philosophy

Web GUIs		WMS client	raster Web client toolkit	direct queries	server admin
Web services		OGC WMS	OGC WCS	OGC WCPS	OGC WPS	Server control
programming, queries		data definition	data manipulation	C++ API	Java API
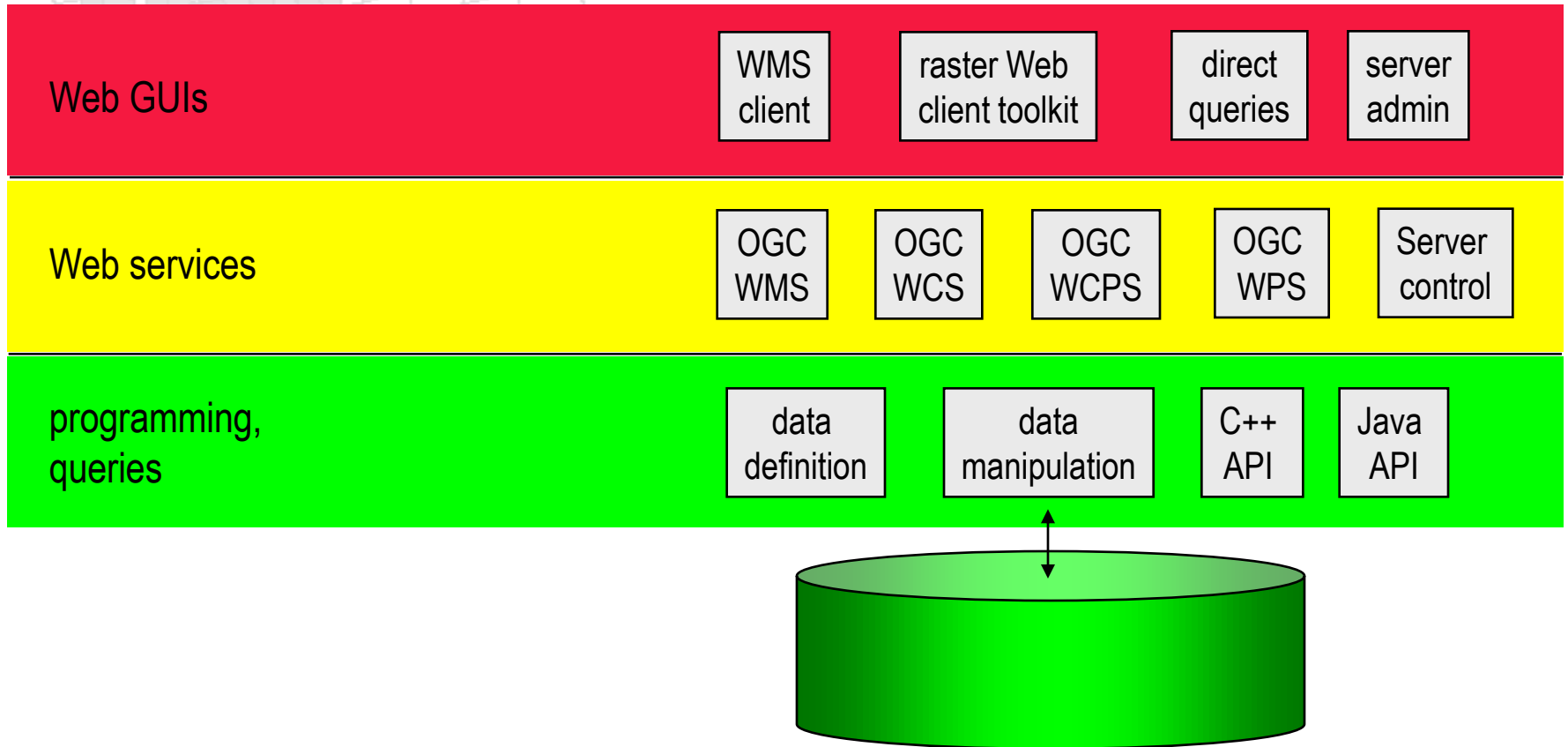
Tiling: Tuning Data for Applications

- tiling strategies as service tuning [Furtado]:

  - regular   directional   area of interest
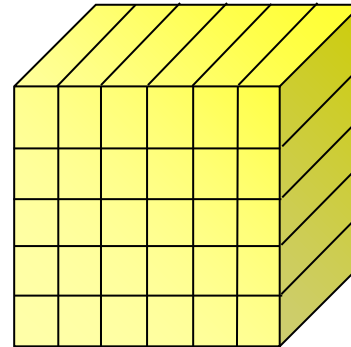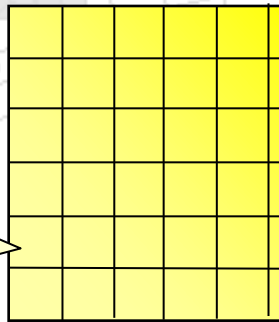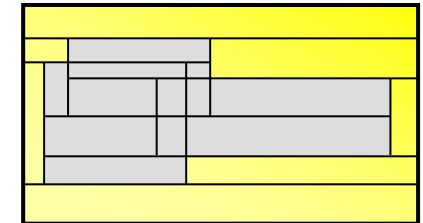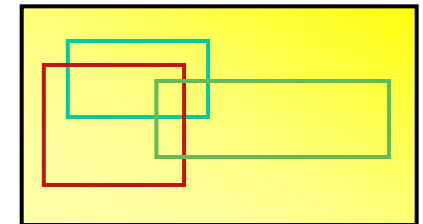
„chunks"
[Sarawagi, DeWitt, ...]

- rasdaman storage layout language

```
insert into MyCollection
   values ...
   tiling area of interest [0:20,0:40], [45:80,80:85]
   tile size 1000000
   index d_index storage array compression zlib
```

EarthServer Integration

- „no one size fits all" [M. Stonebraker]

- Intergration with MapServer, EOxServer

- Via database UDFs: C++, R, Hadoop, ...
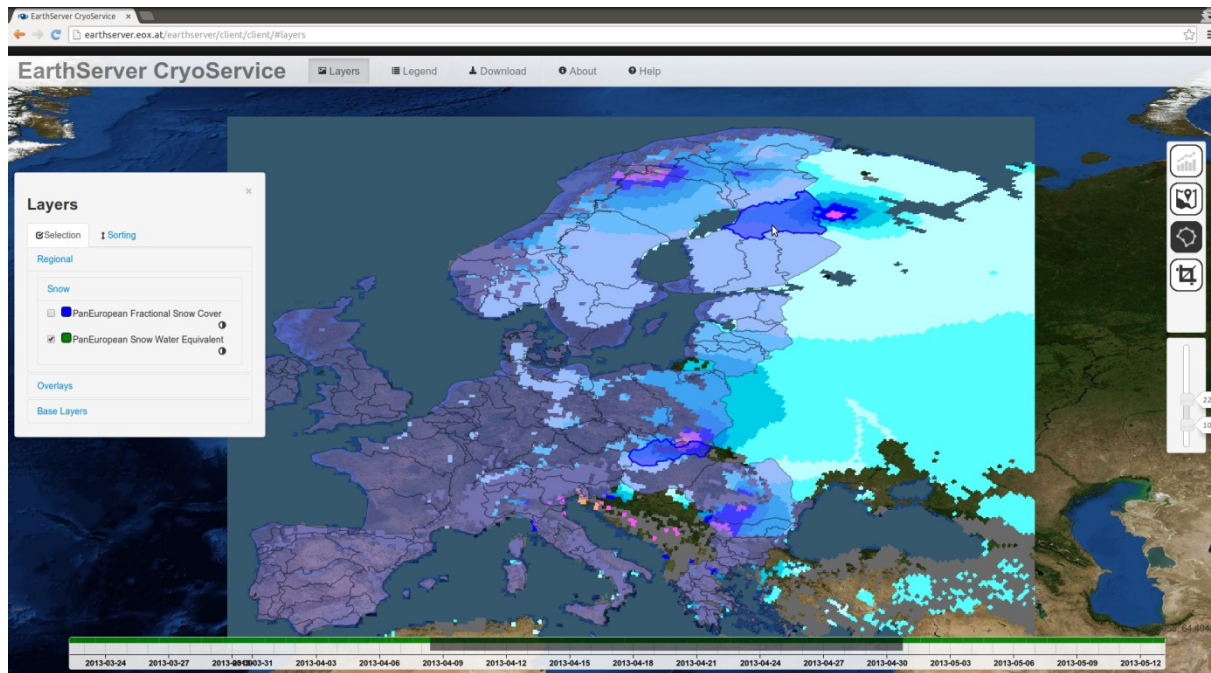
Data / Metadata Integration

- Remember: OGC WCPS
  = high-level grid coverage filtering & processing language
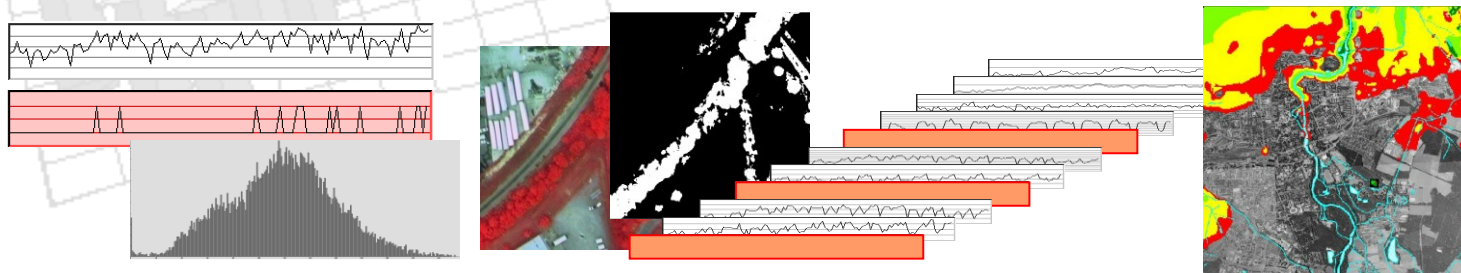
- "From MODIS scenes M1, M2, M3: difference between red & nir, as TIFF"
  - …but only those where nir exceeds 127 somewhere

```
for $c in ( M1, M2, M3 )
where
    some( $c.nir > 127 )
return
    encode(
        $c.red - $c.nir,
        "image/tiff"
    )
```

$\rightarrow$ (tiff$_A$, tiff$_C$)

Data / Metadata Integration

- Idea: merge WCPS with Xquery
  - Ex1: „difference of red, nir bands for all coverages on Austria"

```
for $c in doc("http://acme.com")//coverage
where
    some( $c.nir > 127 ) and metadata/@region = "Austria"
return
    encode( $c.red - $c.nir, "image/tiff" )
```

  - Ex2: „name & location of coverages showing some phenomenon"
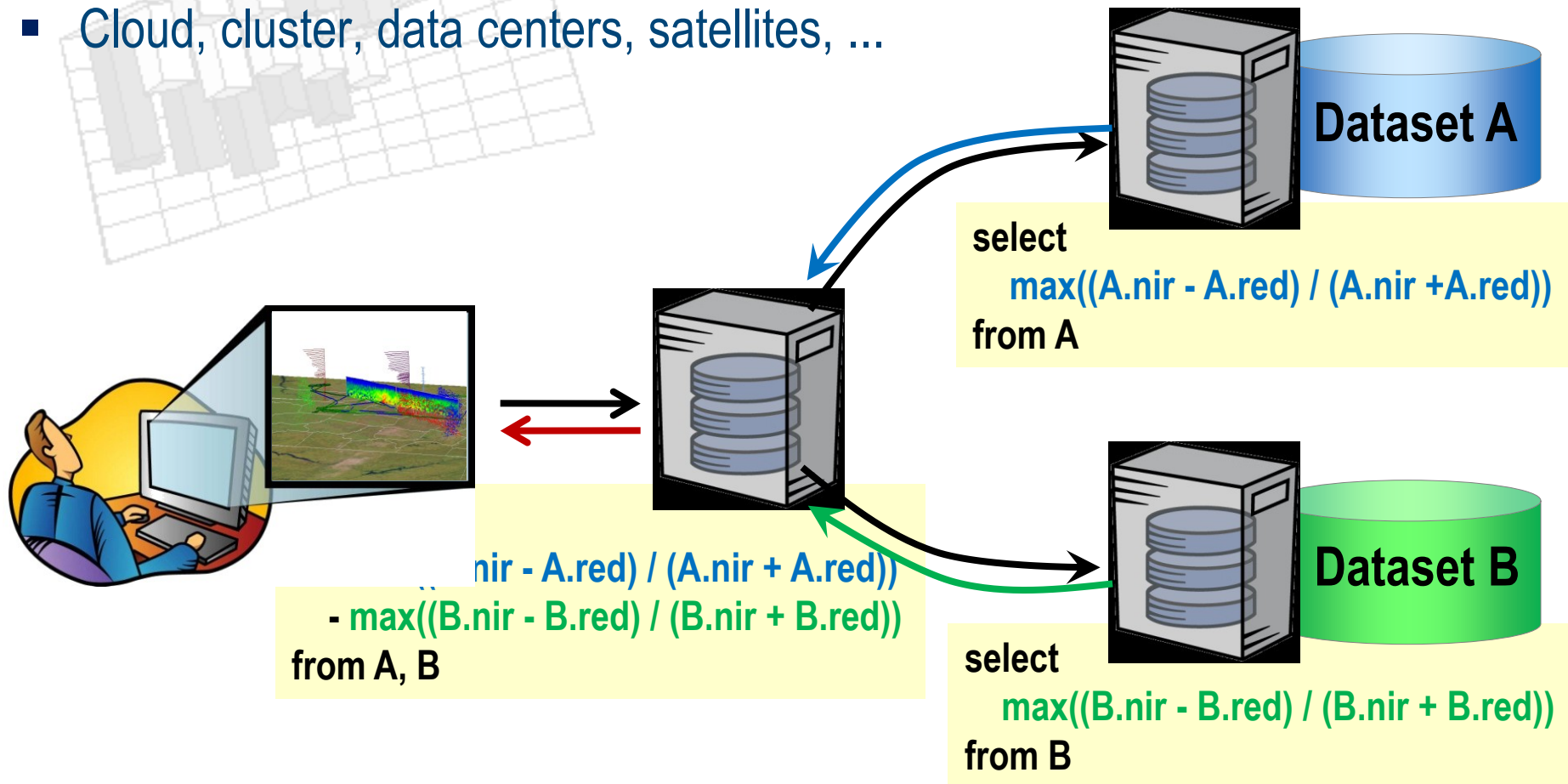
```
for $c in doc("WCPS")//coverage/[ some( $c.nir > $c.red )]
return
    <id> { $c/@id } </id>
    <area> { $c/boundedBy } </area>
```

- WCPS 2.0, in progress [JacobsU]
  - Implementation: federation of eXist + rasdaman [Athena]

# Federated Query Processing

- Heterogeneous rasdaman peer networks

- Cloud, cluster, data centers, satellites, ...

Dataset A

```
select
    max((A.nir - A.red) / (A.nir +A.red))
from A
```

Dataset B

```
select
    max((B.nir - B.red) / (B.nir + B.red))
from B
```

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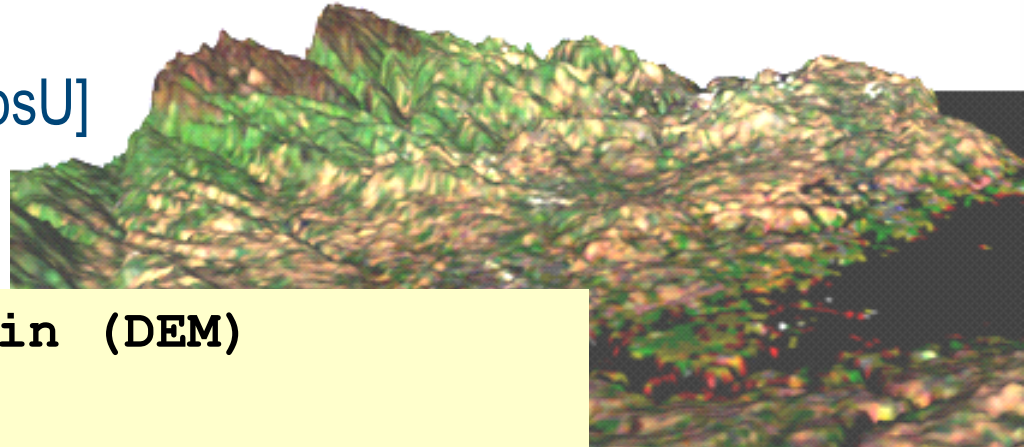
Secured Archive Integration

First-ever direct, ad-hoc mix from protected NASA & ESA services
in OGC WCS/WCPS Web client (EarthServer + CobWeb)
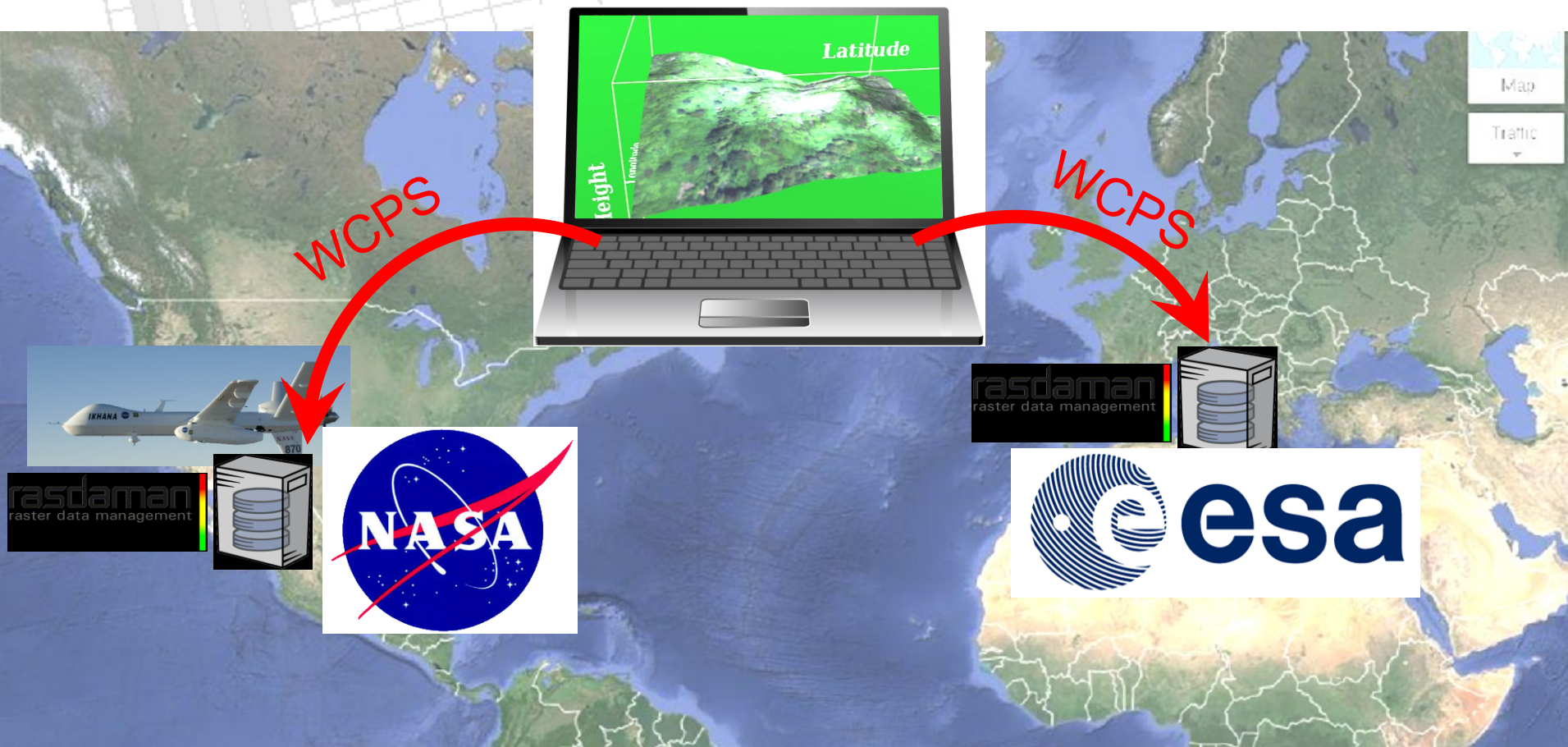
Visual Frontends

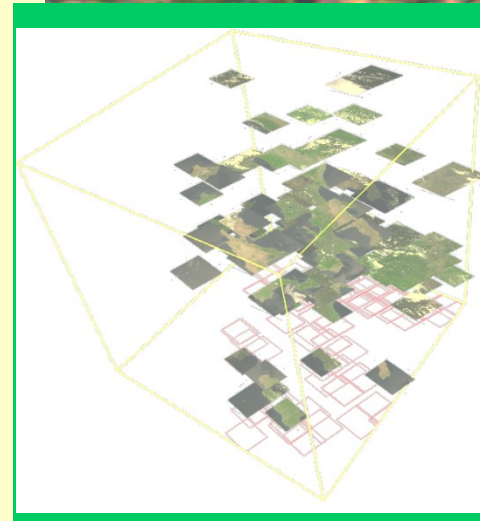- Mobile clients [COMETA, SE.IT]

- Rasdaman browser toolkit [JacobsU]

- 3D browser clients [FhG]

```
for $s in (SatImage), $d in (DEM)
return
  encode(
    struct {
      red:    (char) $s.b7[x0:x1,x0:x1],
      green:  (char) $s.b5[x0:x1,x0:x1],
      blue:   (char) $s.b0[x0:x1,x0:x1],
      alpha:  (char) scale( $d, 20 )
    },
    "image/png"
  )
```

[data courtesy BGS, ESA]

Take Home Messages

www.earthserver.eu

- EarthServer: agile analytics on spatio-temporal Big Geo Data
  Platform: Scalable Array Database, rasdaman

  - images → actionable data

  - Operational Earth science services on 100s of TB

- Future services need query languages!

  - Flexibility + scalability + integration

- Impact on science, industry, business

  - Next-gen service standards : OGC, ISO, RDA