



e-infrastructure

rasdaman
raster data management

JACOBS
UNIVERSITY

Bringing Ad-Hoc Analytics to Big Earth Data:

the  **Earth Server Experience**

EGU 2014, Vienna

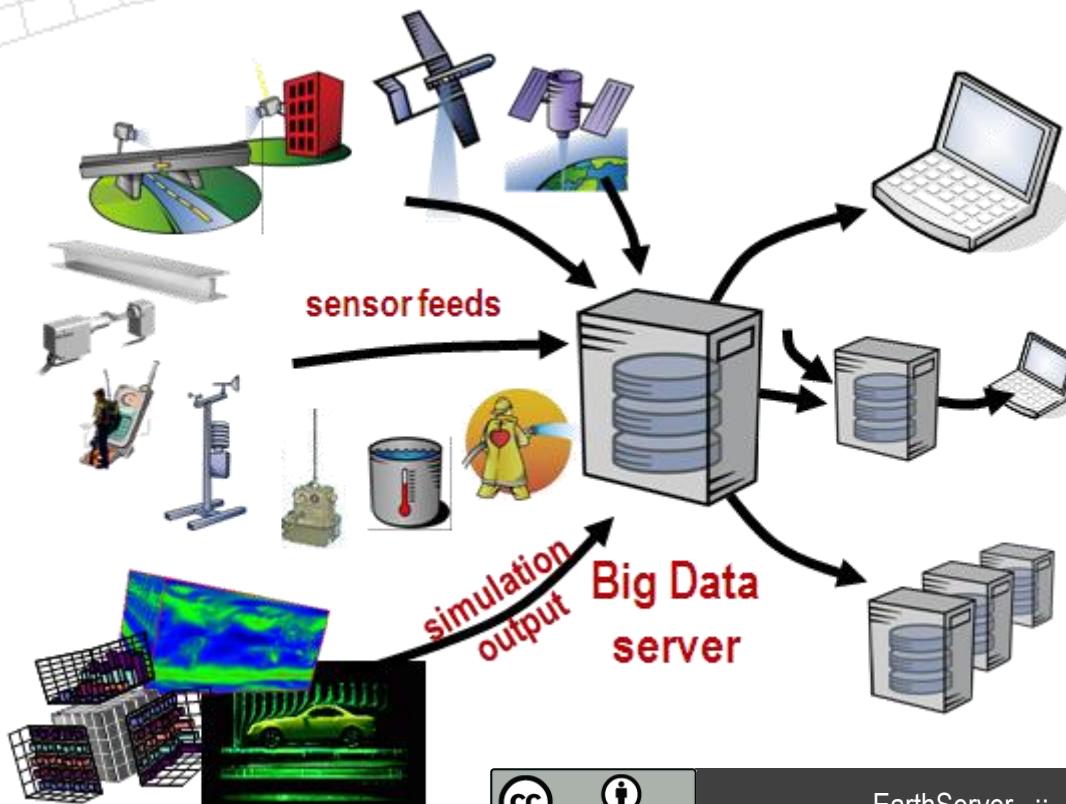
Peter Baumann

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



What are the Big Data in Geo?

- Roughly: Spatio-temporal **sensor**, **image**, **simulation**, **statistics** data
 - Often n-D raster data = arrays
- „**massive arrays & graphs next great challenges**“ [major DBMS vendors]





EarthServer: *Big Earth Data Analytics*

- Scalable On-Demand Processing for the Earth Sciences
 - EU FP7-INFRA, Sep 2011 – Aug 2914, ~6 mEUR
- Platform: pioneer Array Database technology, rasdaman
 - Integrated filtering & processing on metadata, regular/irregular grids, point clouds, ...
- 11 partners (3 SMEs):



JACOBS
UNIVERSITY

rasdaman
raster data manager



Consiglio Nazionale delle Ricerche



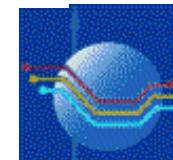
British Geological Survey
NATIONAL ENVIRONMENT RESEARCH COUNCIL

PML

Plymouth Marine
Laboratory



EOX
Fraunhofer

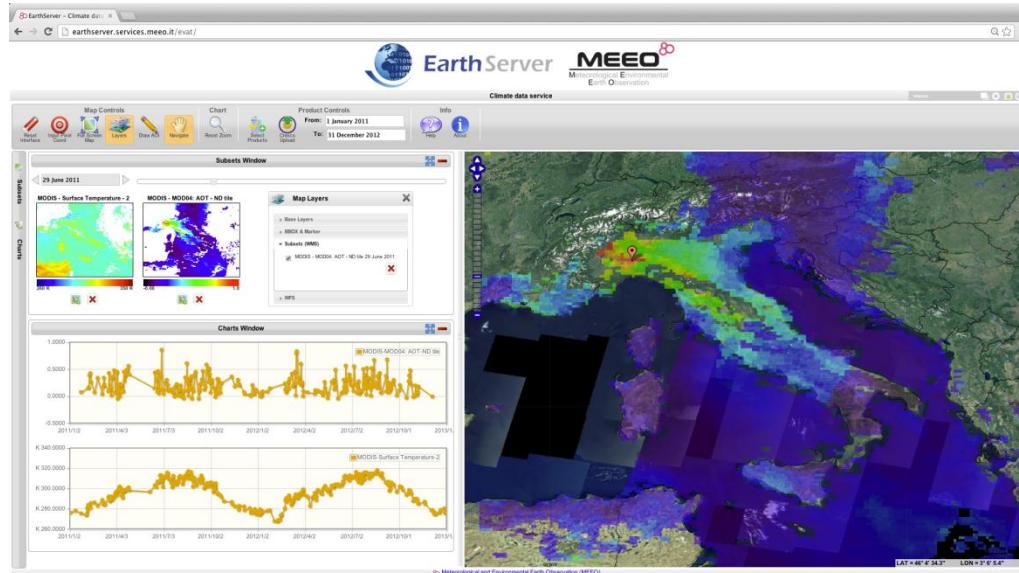


MEEO
Meteorological Environmental
Earth Observation

- CP-CSA → 3 activity pillars: RTD, Services, Networking

Service Activity

- 6 Lighthouse Applications covering all Earth Sciences
- Summer 2014:
 - all 20+ TB
 - 2x 100+ TB

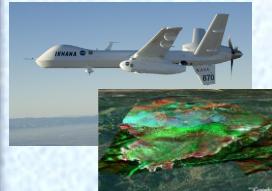


Cryospheric Science
landcover mapping



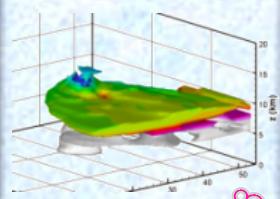
EOX

Airborne Science
high-altitude long-endurance drones



NASA

Atmospheric Science
climate variables



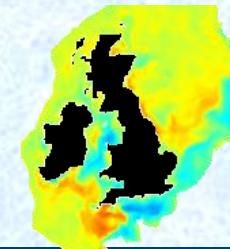
MEEO
Metereological Environmental Earth Observation

Geology
geological models



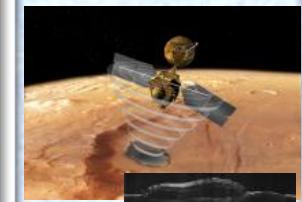
British Geological Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL

Oceanography
marine model runs + in-situ data



PML PLYMOUTH MARINE LABORATORY

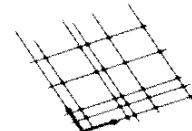
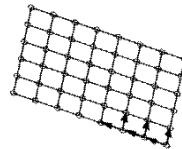
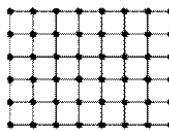
Planetary Science
Mars geology



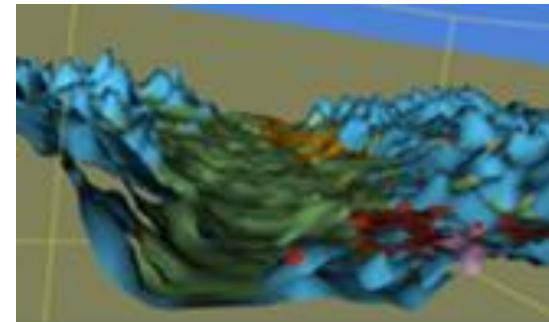
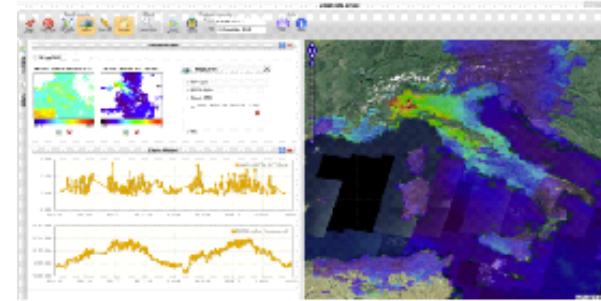
JACOBS
UNIVERSITY

RTD Activity: Overview

- Big Geo Data **engine** development
 - Based on rasdaman Array Database
 - strictly open standards (OGC WCS, WMS, WCPS)
- Regular & irregular grids, point clouds, meshes



- **Coupling:** Hadoop, R, MatLab, MapServer, ...
- Data/metadata **search integration**
- **Scalability:** distributed processing
- **Visual** 1D/2D/3D client toolkit

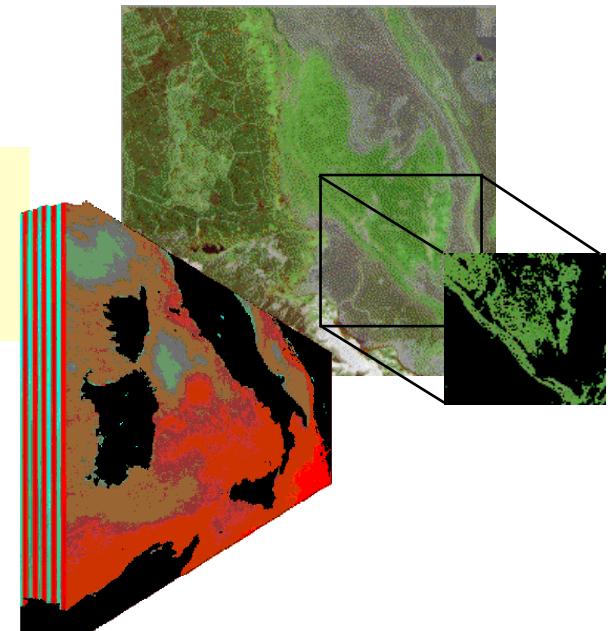
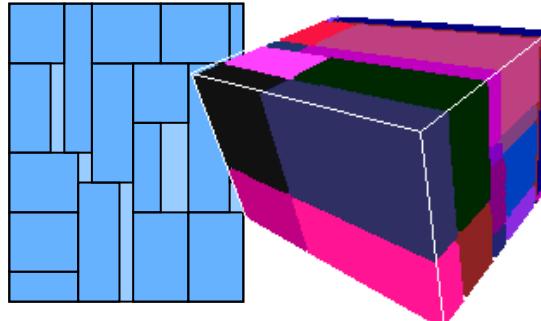


rasdaman: Agile Array Analytics

- „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



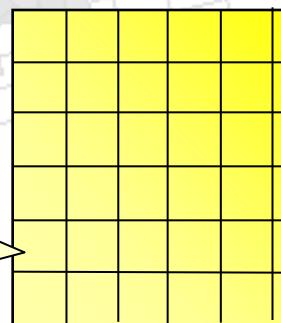
- In operational use since 10+ years
- OGC WCS Core Reference Implementation



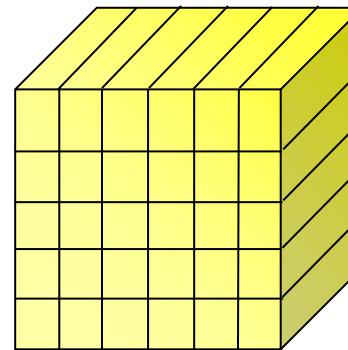
Tiling: Tuning Data for Applications

- tiling strategies as service tuning [Furtado]:

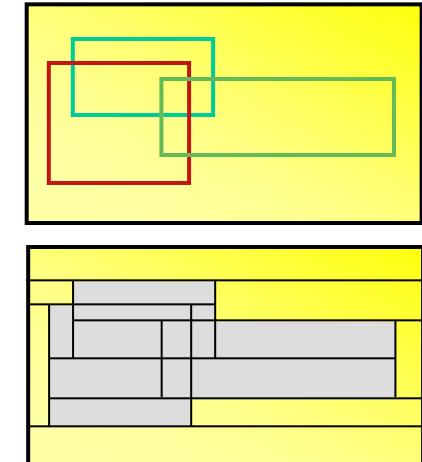
- regular



- directional



- area of interest



- rasdaman storage layout language

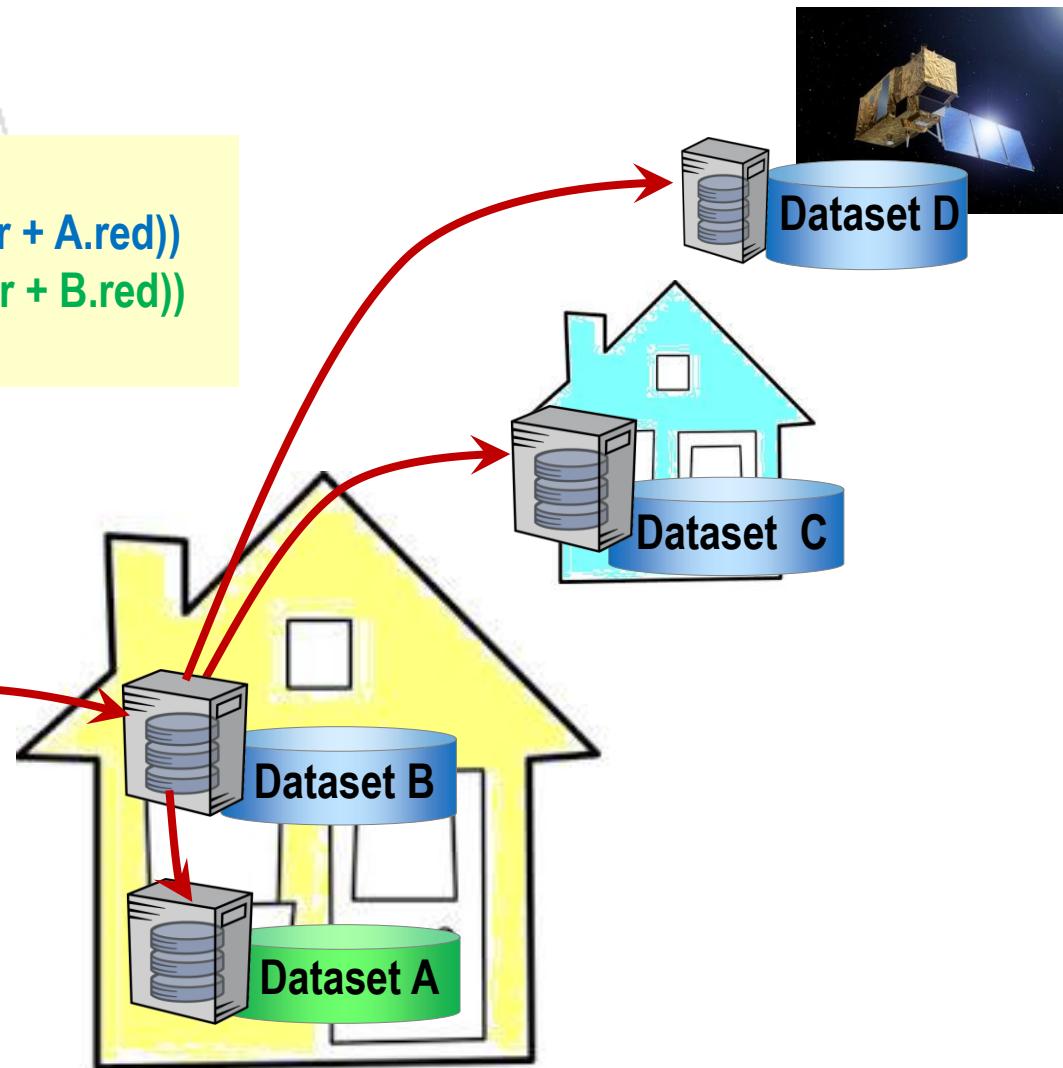
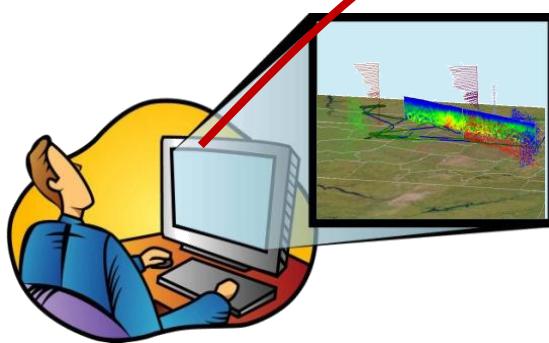
```
insert into MyCollection
  values ...
  tiling area of interest [0:20,0:40], [45:80,80:85]
  tile size 1000000
  index d_index storage array compression zlib
```

rasdaman Federation Scenarios

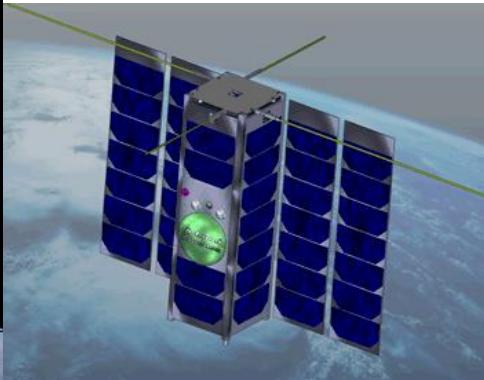
select

```
max((A.nir - A.red) / (A.nir + A.red))  
- max((B.nir - B.red) / (B.nir + B.red))
```

from A, B



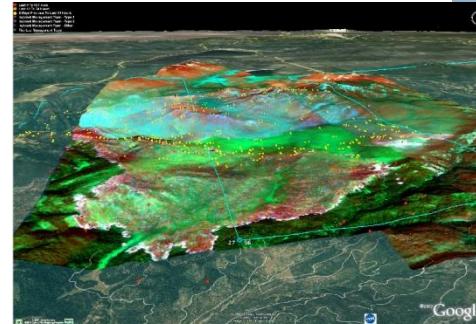
Next: On-Board Query Intelligence



ESA

Democratize direct data access

NASA



Inset: OGC Big Data Coverage Service Portfolio

- OGC standards cover **full range** from **data-intensive** to **processing-intensive** „Big Data“ coverage services

WCS

data access

WCPS

ad-hoc analytics

WPS

predefined process

- **Web Coverage Service (WCS):**
easy data access & extraction
- **Web Coverage Processing Service (WCPS):**
agile analytics, enabling automatic parallelization
- **Web Processing Service (WPS):**
predefined processes of arbitrary complexity as „black boxes“

Networking Activity: Outreach (select)

- Making the **standards**
 - Open Geospatial Consortium (OGC)
 - *chairing 4 WGs, editor of WCS standards suite*
 - ISO: new work item „Array SQL“
 - Research Data Alliance: Big Data Analytics IG
- Conferences
 - Initiated ESA conference series „Big Data From Space“
 - FOSS4G-Europe („free & open-source 4 geospatial“), July 2014
 - Sessions at EGU
- TV documentaries for German ZDF & arte
- Innovation awards
 - Ex: Geospatial World Forum



Conference Announcements

www.foss4g-e.org

Home Organization Program



OSGeo's European Conference on
Free and Open Source Software for
Geospatial

FOSS4G-Europe 2014

Independent Innovation
for INSPIRE, Big Data and Citizen Participation

Public viewing of
the Soccer World
Cup 2014 Endgame

GIn. "Big Geo Data &
INSPIRE" Forum on
July 14, 2014 (in German)

NASA World Wind Europe
2014 Challenge
for Academia

→ CONFERENCE ON BIG DATA FROM SPACE
Research, Technology and Innovation (RT&I)

12-14 November 2014
ESA-ESRIN Frascati, Italy
[congrexprojects.com/2014-
events/BigDatafromSpace](http://congrexprojects.com/2014-events/BigDatafromSpace)

Take Home Messages

- Big Data Analytics requires **high-level languages**
 - Problem-adjusted → Array Databases like rasdaman
- **EarthServer: agile analytics** on spatio-temporal Big Geo Data
 - **Flexibility + scalability + integration**
 - pictures → actionable data
- **Impact** on science, industry, business
 - Next-gen **service standards** : OGC, ISO, RDA

www.earthserver.eu

