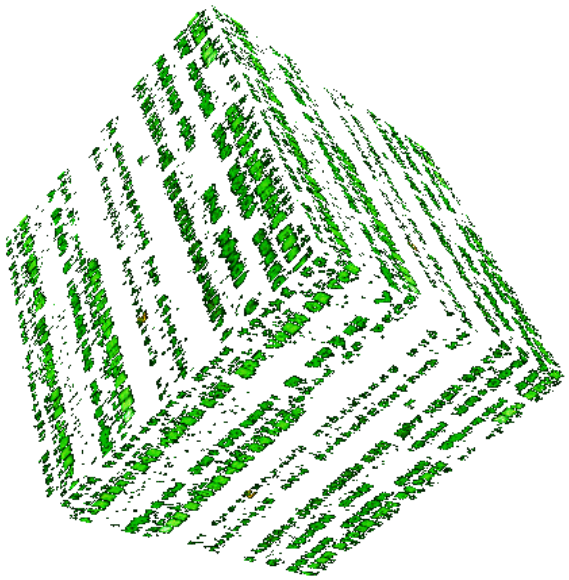
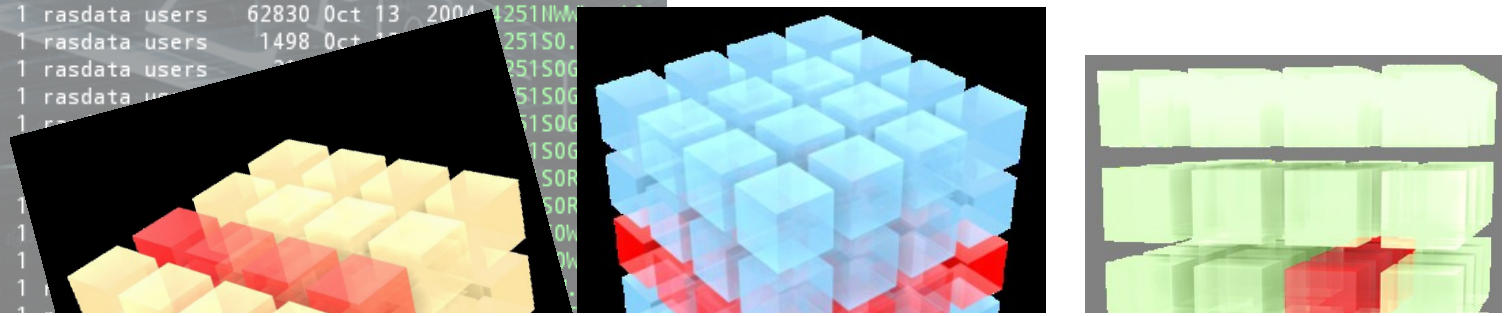


Scalable Datacube Analytics with rasdaman

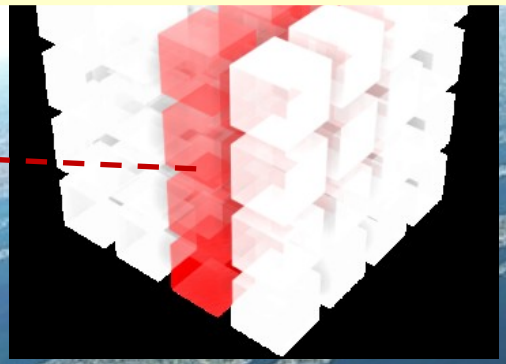
Peter Baumann
rasdaman GmbH, Germany

[gamingfeeds.com]

SERVICE QUALITY



SERVICE QUALITY



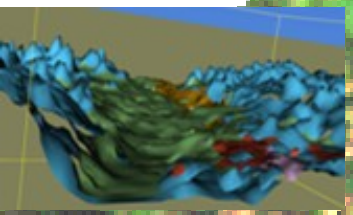
OGC WCPS: Analyzing Datacubes

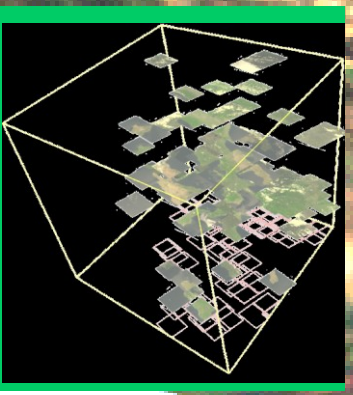
- Web Coverage Processing Service
  - spatio-temporal datacube analytics language
  - Part of the OGC WCS “Big Earth Datacube” standards suite

- "From MODIS scenes M1, M2, M3: difference 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" )
```

OGC WCPS: Elevation & Image Fusion



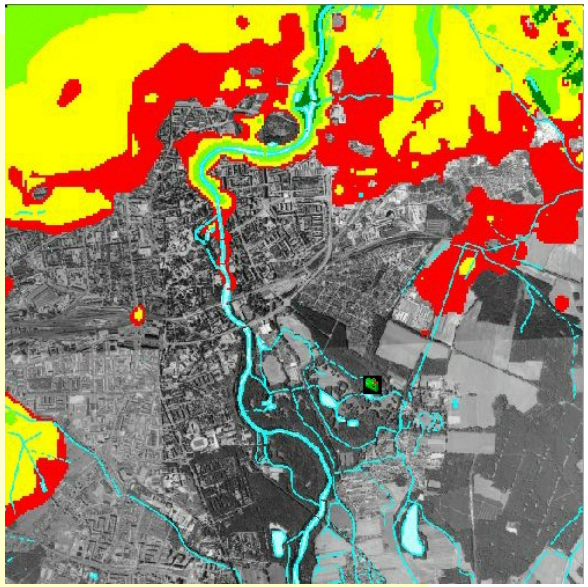


```
for $s in (SatImage), $d in (DEM)
where $s/metadata/@region = "Glasgow"
return
  encode(
    struct {
      red:   (char) $s.b7[x0:x1,x0:x1],
      green: (char) $s.b5[x0:x1,x0:x1],
      blue:  (char) $s.b0[x0:x1,x0:x1],
      alpha: (char) scale( $d, 20 )
    },
    "image/png"
  )
```

[JacobsU, Fraunhofer; data courtesy BGS, ESA]

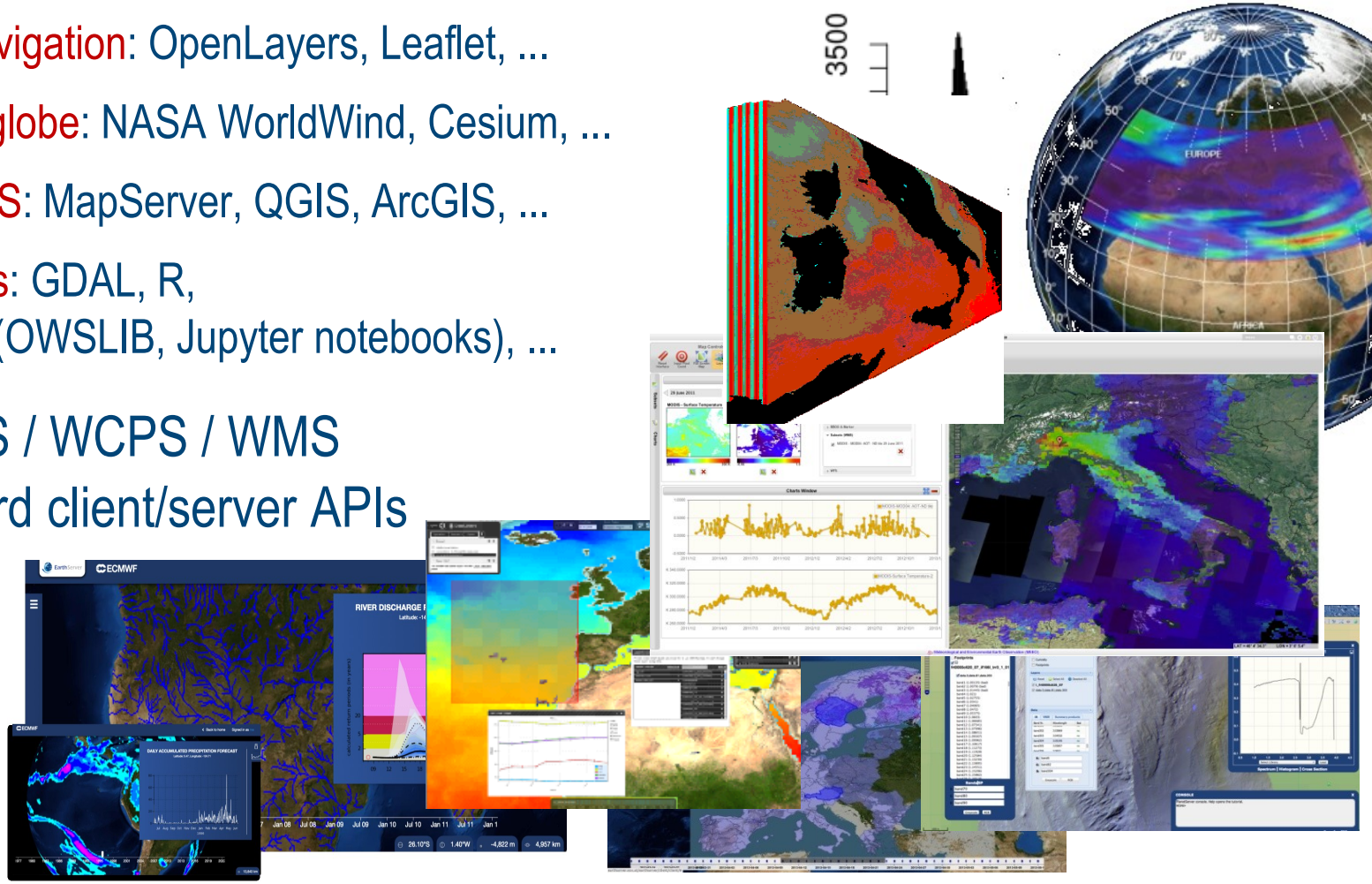
OGC WCPS: Emulating WMS

```
for $p in (OrthoPhoto),
    $wl in (WaterLines), $wa in (WaterAreas),
    $d in (DEM)
return
    encode(  (unsigned char) (
          $p * { 1, 1, 1 }
        overlay
          $wl * { 0, 128, 255 }
        overlay
          $wa * { 191, 255, 255 }
        overlay
          switch $d
          case $d > 260 return { red:255, green:0, blue:0 }
          case $d > 262 return { red:0, green:255, blue:0 }
          case $d > 264 return { red:0, green:0,  blue:255 }
          default       return { red:0, green:0, blue:0 }
          end
       ),
      "image/png" )
```



...But That's Not What You Want to See

- Let users remain in comfort zone of well-known tools
  - Map navigation: OpenLayers, Leaflet, ...
  - Virtual globe: NASA WorldWind, Cesium, ...
  - Web GIS: MapServer, QGIS, ArcGIS, ...
  - Analysis: GDAL, R,
    python (OWSLIB, Jupyter notebooks), ...
- ...via WCS / WCPS / WMS
  as standard client/server APIs

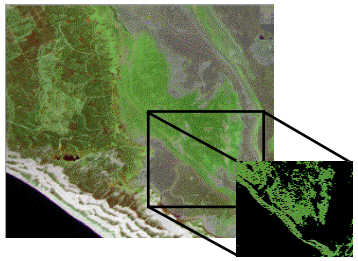


[screenshots: rasdaman-based portals]

Sample Client on 4D Datacube

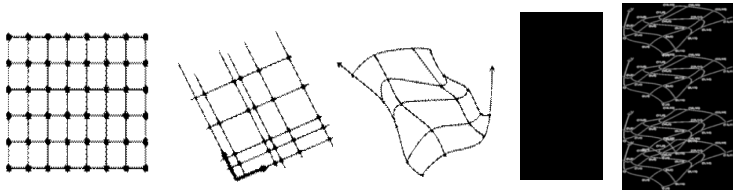
Technology

rasdaman: Agile Array Analytics



= „raster data manager“: declarative QL on n-D arrays

- Scalable parallel “tile streaming” architecture
- Any spatio-temporal regular & irregular grid
- OGC WMS + WCS + WCPS



- Mature, in operational use worldwide
  - Blueprint for standards
  - OGC & INSPIRE Reference Implementation

GEOSPATIAL WORLD FORUM™

CIO Review 100 MOST PROMISING BIG DATA SOLUTION PROVIDERS 2016

The Earth Monitoring Competition

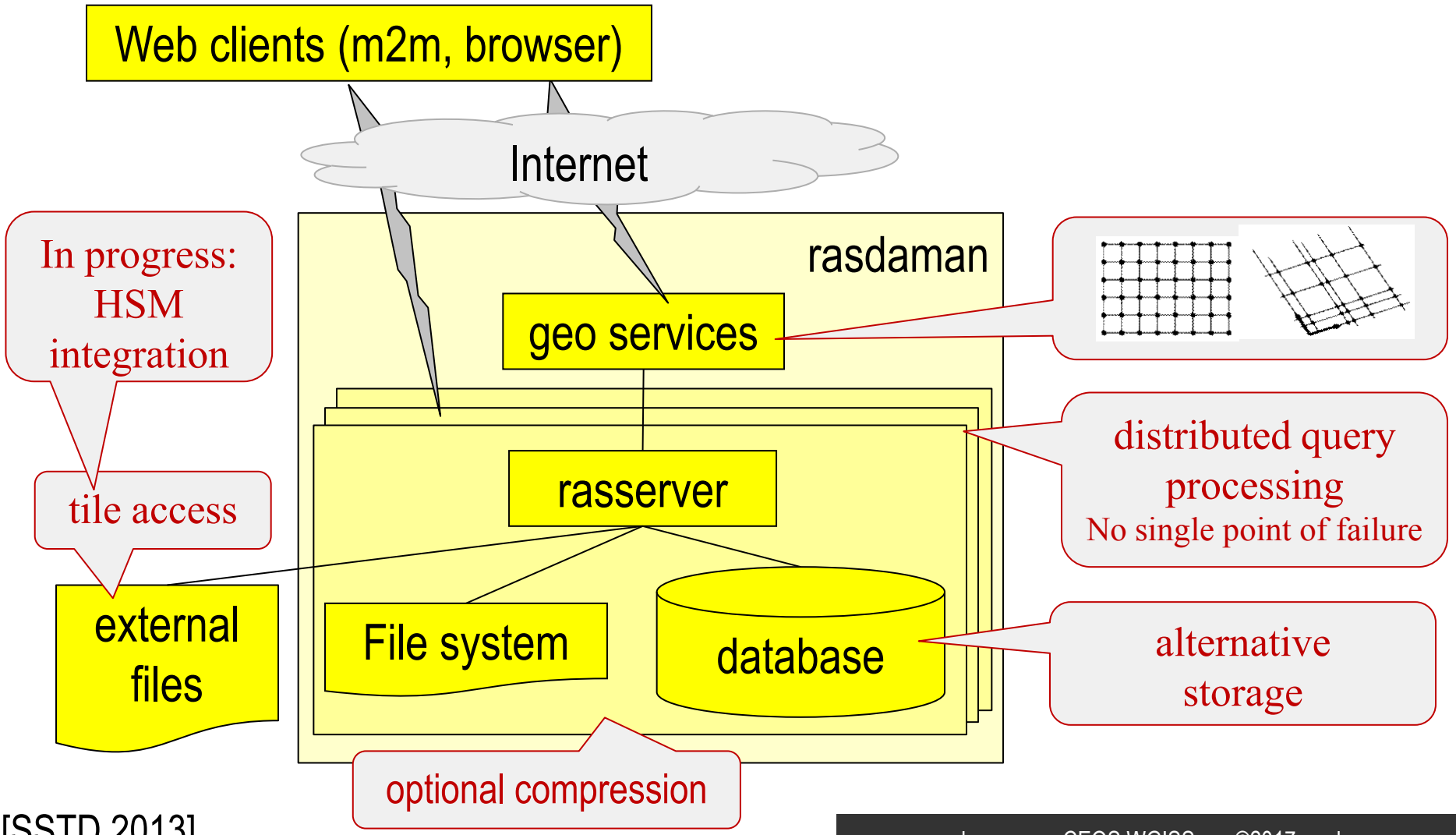
Copernicus masters

WINNER

T-SYSTEMS BIG DATA CHALLENGE 2014

T · · Systems ·

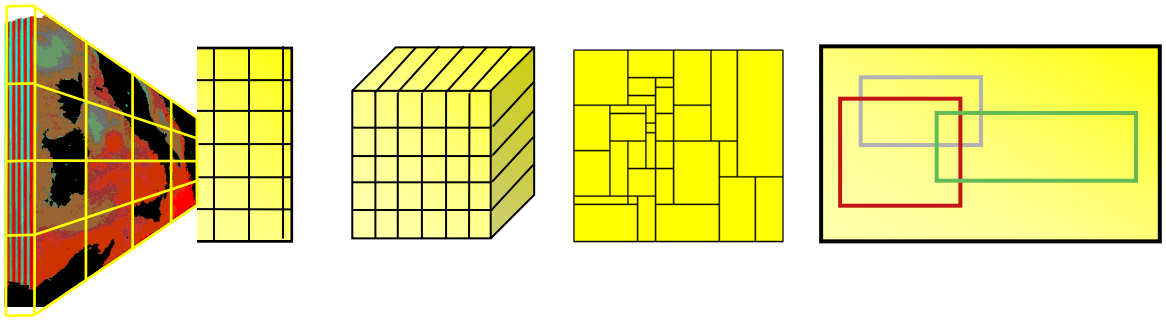
Architecture



[SSTD 2013]

Adaptive Partitioning („Tiling“)

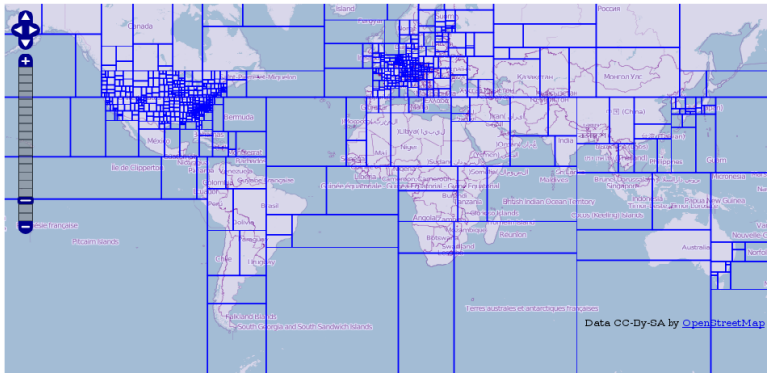
- Any tiling [Furtado 1999]
  - Cast into strategies



  - 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
```

- Why irregular tiling?



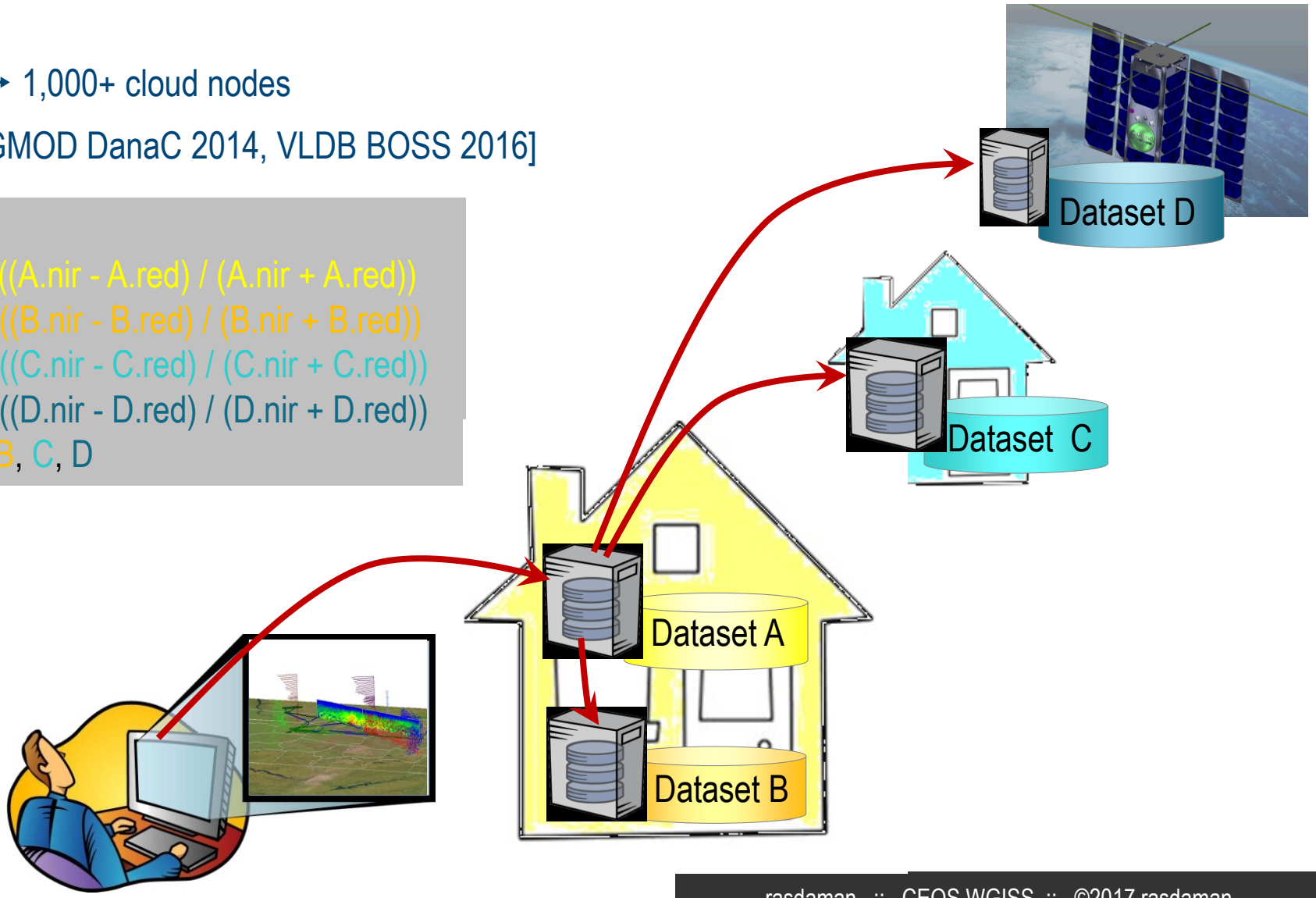
[OpenStreetMap]

Parallel, Distributed Processing

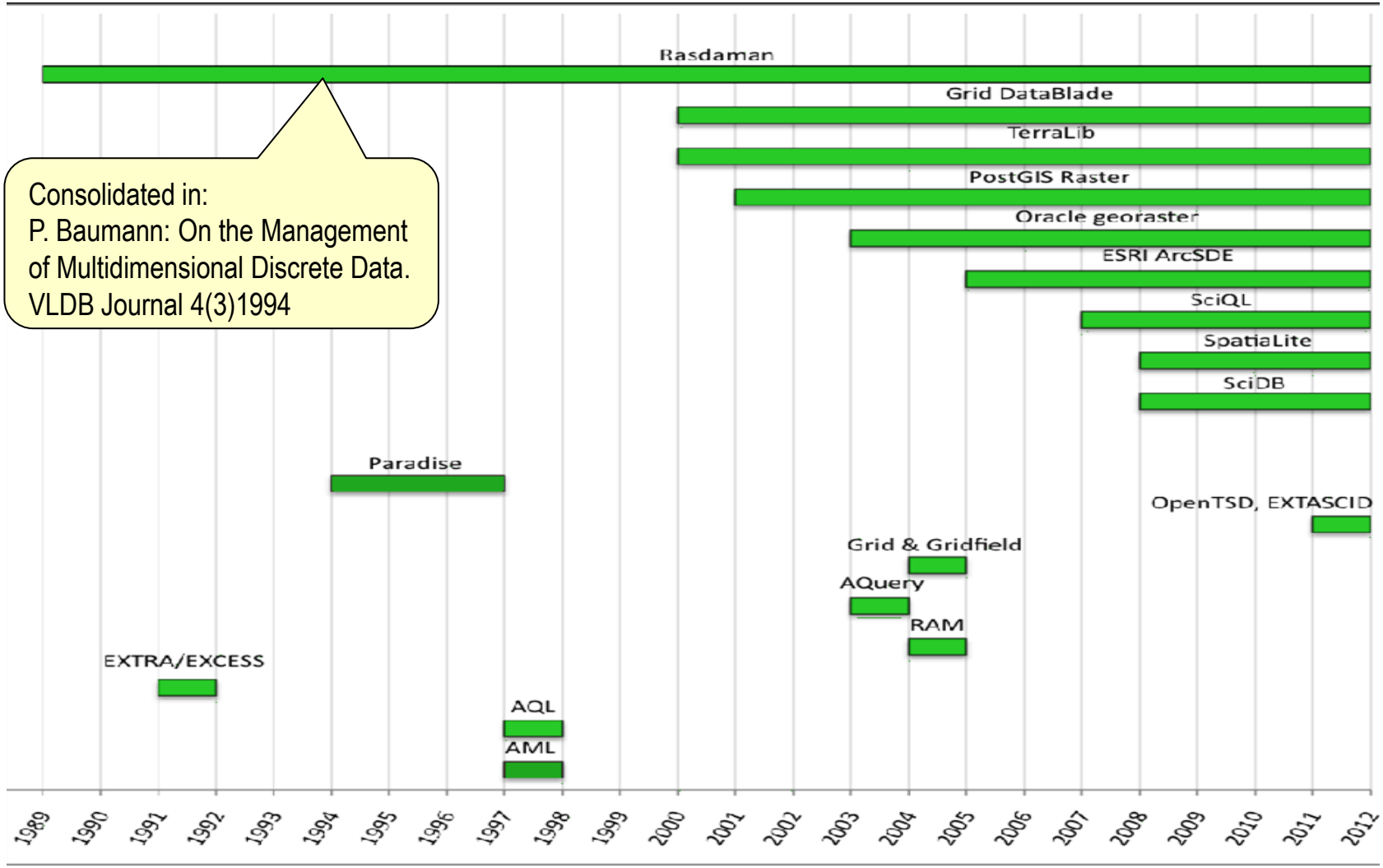
1 query → 1,000+ cloud nodes

[ACM SIGMOD DanaC 2014, VLDB BOSS 2016]

```
select
      max((A.nir - A.red) / (A.nir + A.red))
    - max((B.nir - B.red) / (B.nir + B.red))
    - max((C.nir - C.red) / (C.nir + C.red))
    - max((D.nir - D.red) / (D.nir + D.red))
from A, B, C, D
```



The Beginnings of Array Databases

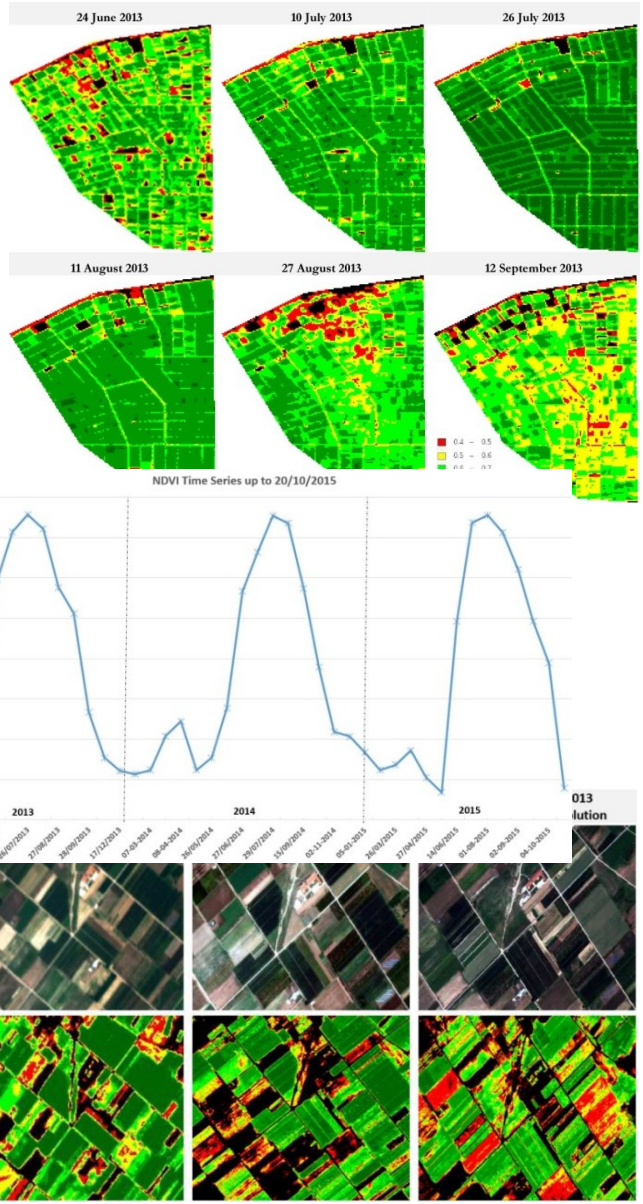


Consolidated in:
P. Baumann: On the Management of Multidimensional Discrete Data. VLDB Journal 4(3)1994

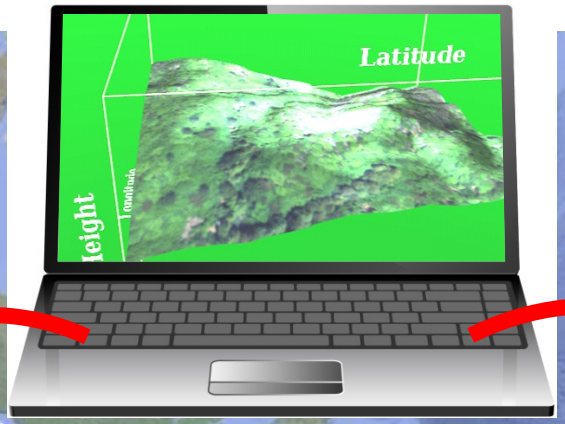
Applications

EOfarm Startup (Greece)

- Big Data Analytics for farmers
  - rasdaman via OGC WCS & WCPS
  - similar framework deployed for water quality monitoring
- Data: Landsat8, Sentinels, RapidEye
- Functionality:
  - Color Composites, Band Ratios and Indices
  - Vegetation Detection
  - Canopy Greenness Estimation
  - Land Surface Temperature
  - Time series over AOI



First-ever direct, ad-hoc mix from protected NASA & ESA services in OGC WCS/WCPS Web client



WCPS

WCPS

rasdaman raster data management

NASA

esa

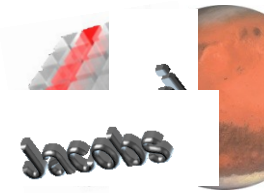
rasdaman raster data management

- Hosting:
  - Drone image server: Ames Research Center
  - DEM server: esa
  - Installation: MEEO
- Technology:
  - ESA, NASA image server: rasdaman raster data manager
  - Security: Secure Dimensions
- Sponsored by: EarthServer, COBWEB Citizen Observatory Web

rasdaman
raster data management

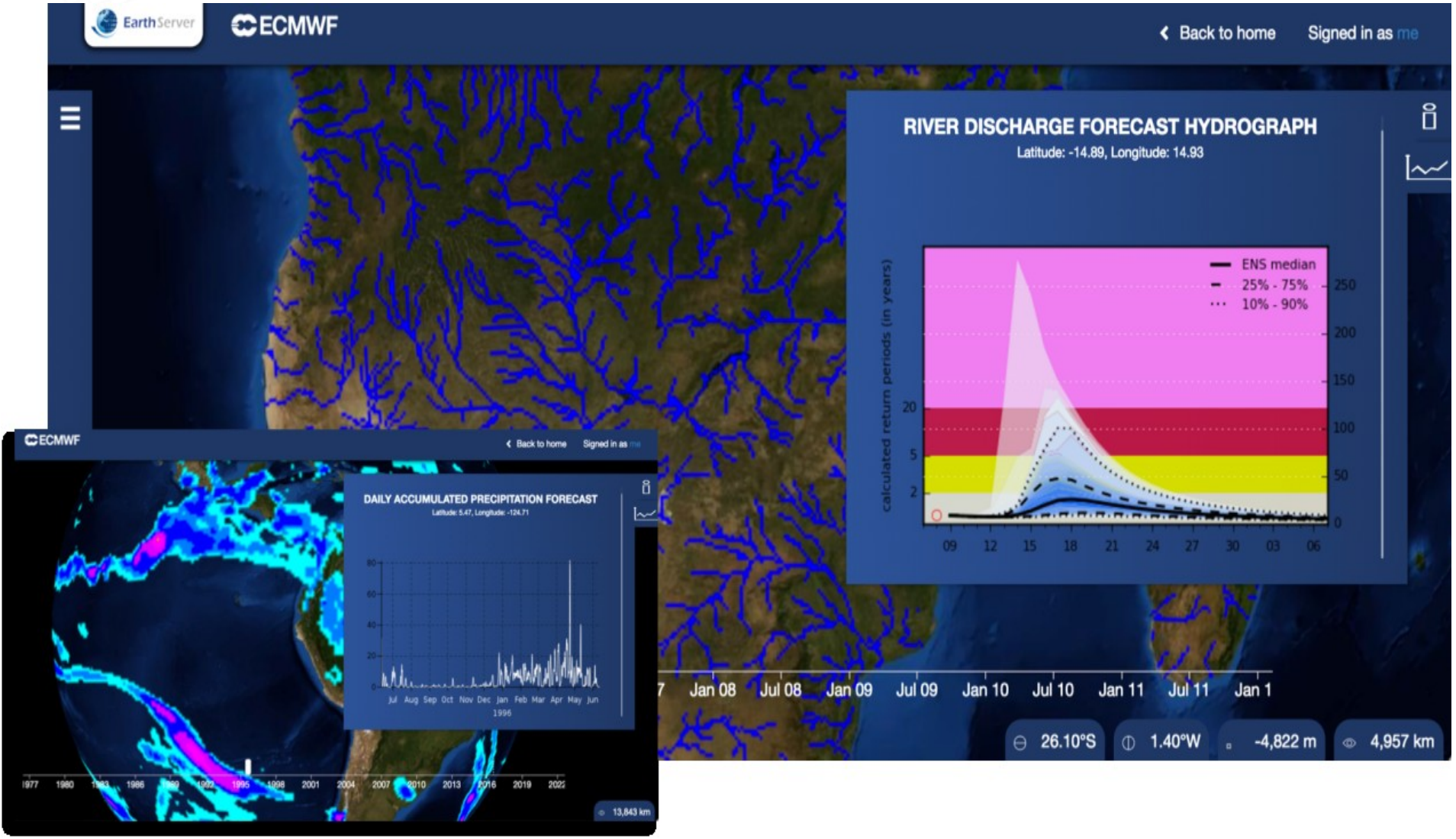
EarthServer: Datacubes At Your Fingertips

- Agile Analytics on x/y/t + x/y/z/t Earth & Planetary datacubes
  - Rigorously standards: OGC WMS + WCS + WCPS
  - EU rasdaman + NASA WorldWind
  - 500+ TB → 1+ PB
- Intercontinental initiative,
  3+3 years: EU + US + AUS
- www.earthserver.eu,
  www.planetserver.eu

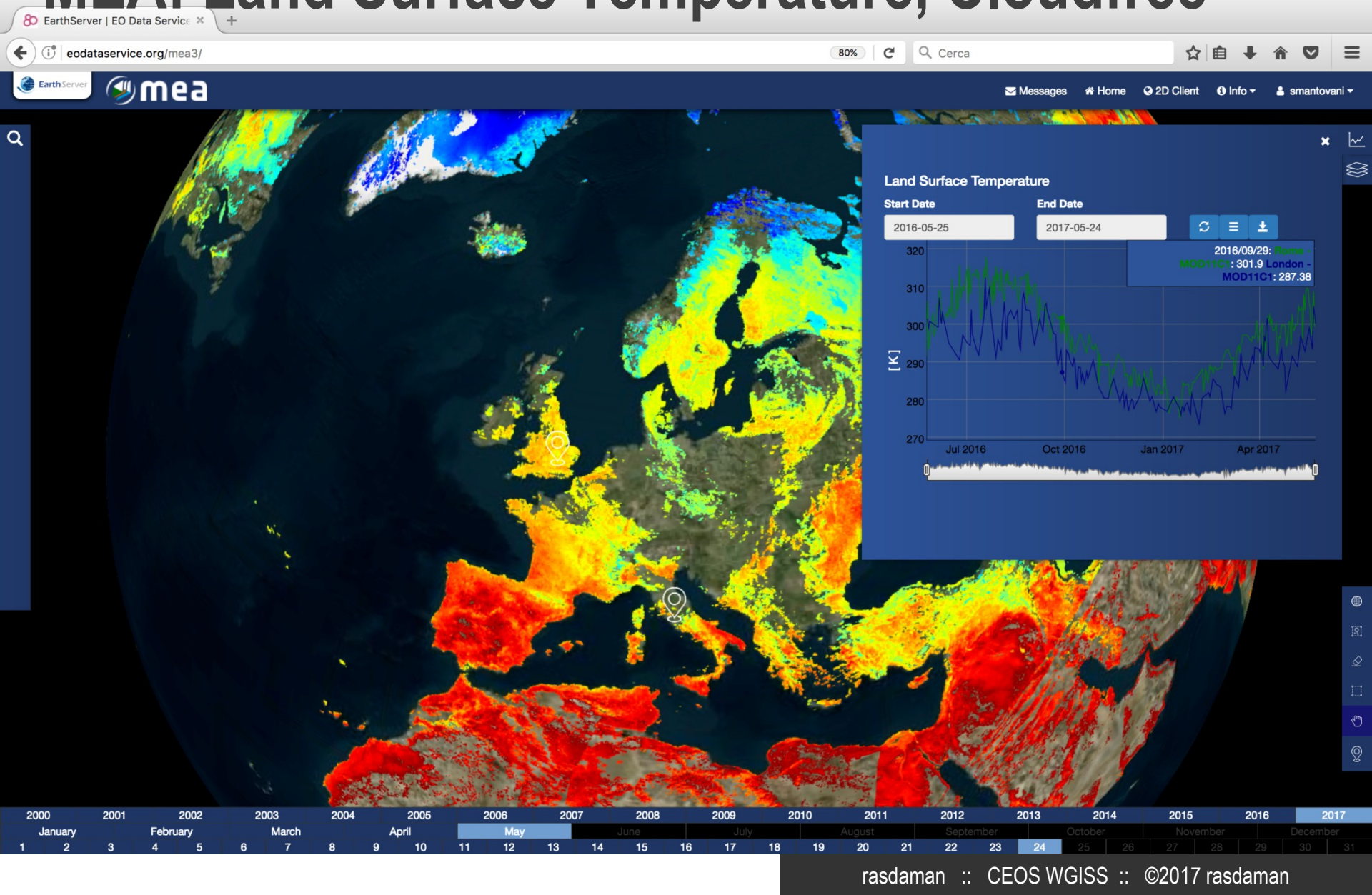


Co-funded by
the European Union

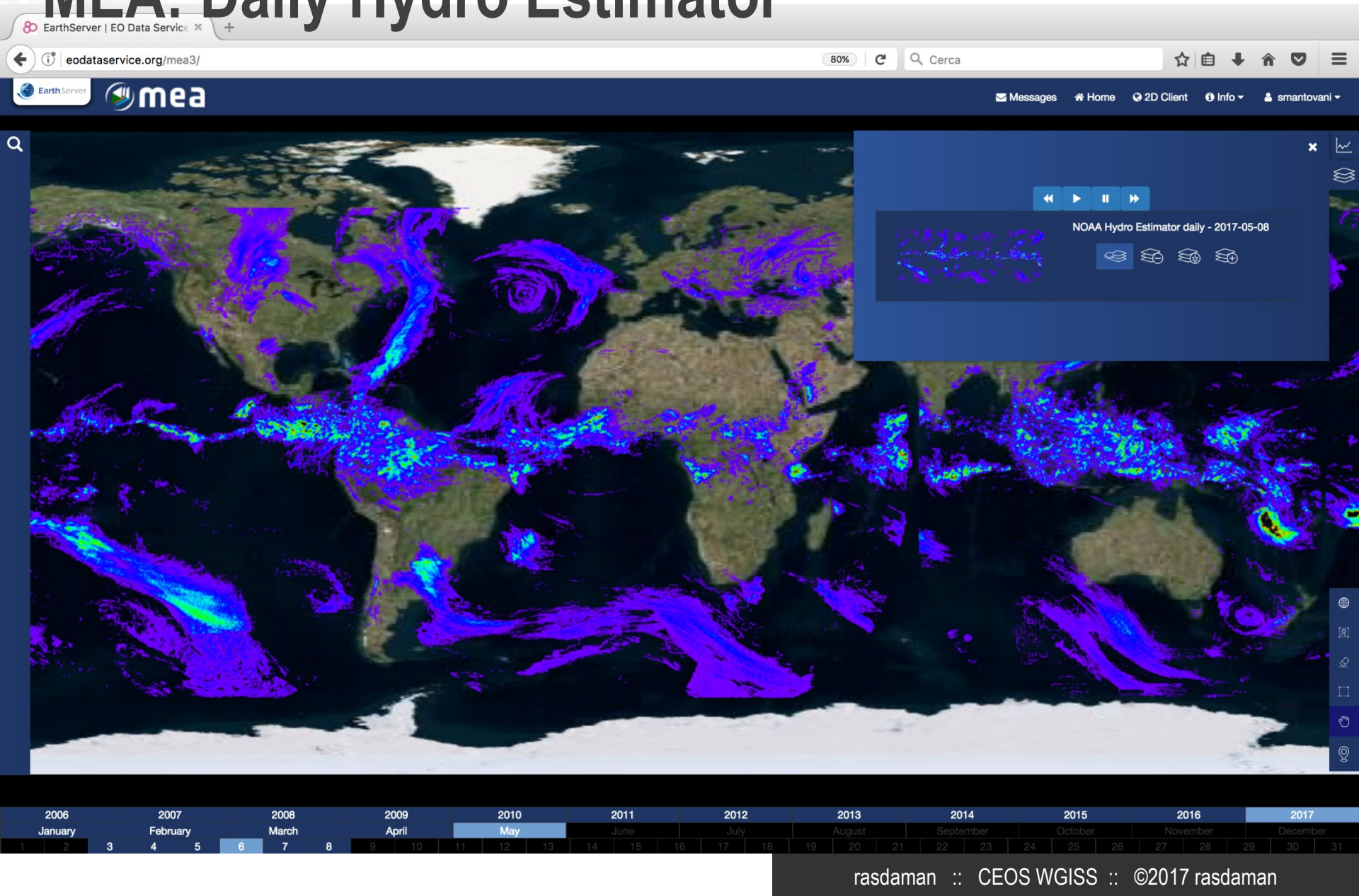
ECMWF: River Discharge



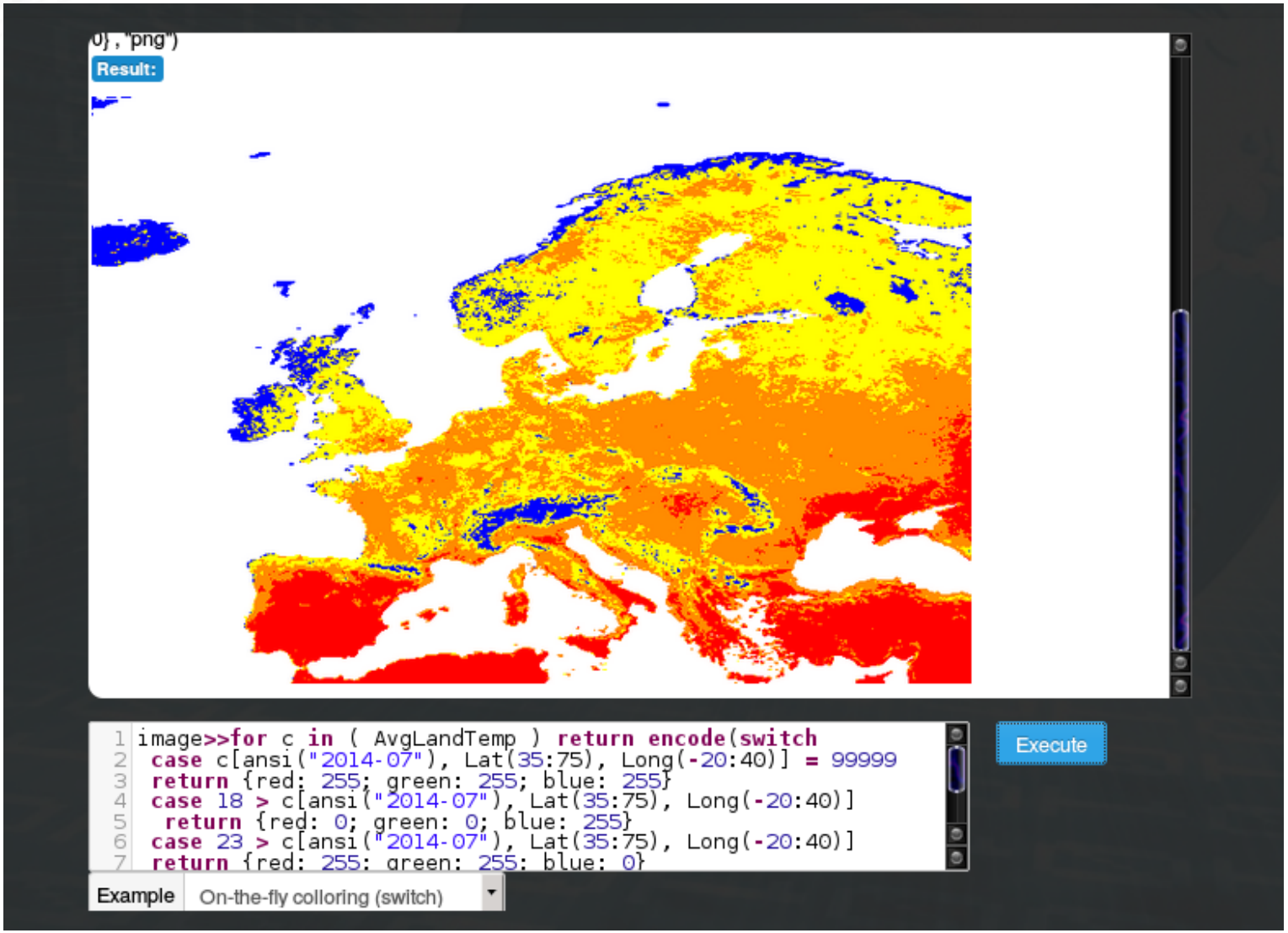
MEA: Land Surface Temperature, Cloudfree



MEA: Daily Hydro Estimator



Server-Side Processing: Flexibility



Server-Side Processing: Federation

```
SELECT ENCODE(CASE
WHEN (CONDENSE + over i_i in [42364:42368] using
d[0:3600, 0:1800, i_i[0]] / 1423 + 1.47) > ((CONDENSE +
over i_i in [42364:42368] using (c) [*:*,*:*,i_i[0]])*(1000))
THEN ((255) * {1c,0c,0c,0c} + (255) * {0c,1c,0c,0c} + (255)
* {0c,0c,1c,0c} + (0) * {0c,0c,0c,1c})

WHEN (CONDENSE + over i_i in [42364:42368] using
d[0:3600, 0:1800, i_i[0]] / 1423 + 4) > ((CONDENSE + over
i_j in [42364:42368] using (c) [*:*,*:*,i_j[0]])*(1000))
THEN ((0) * {1c,0c,0c,0c} + (128) * {0c,1c,0c,0
```

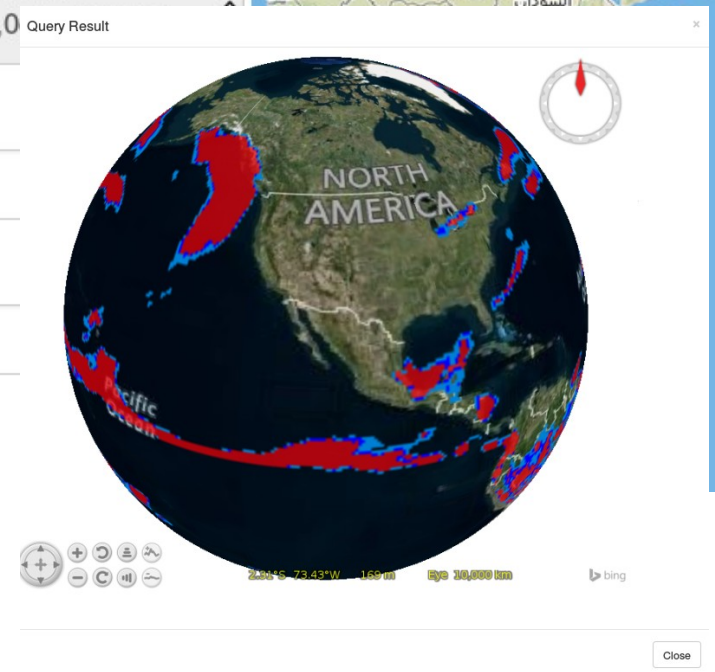
Query:

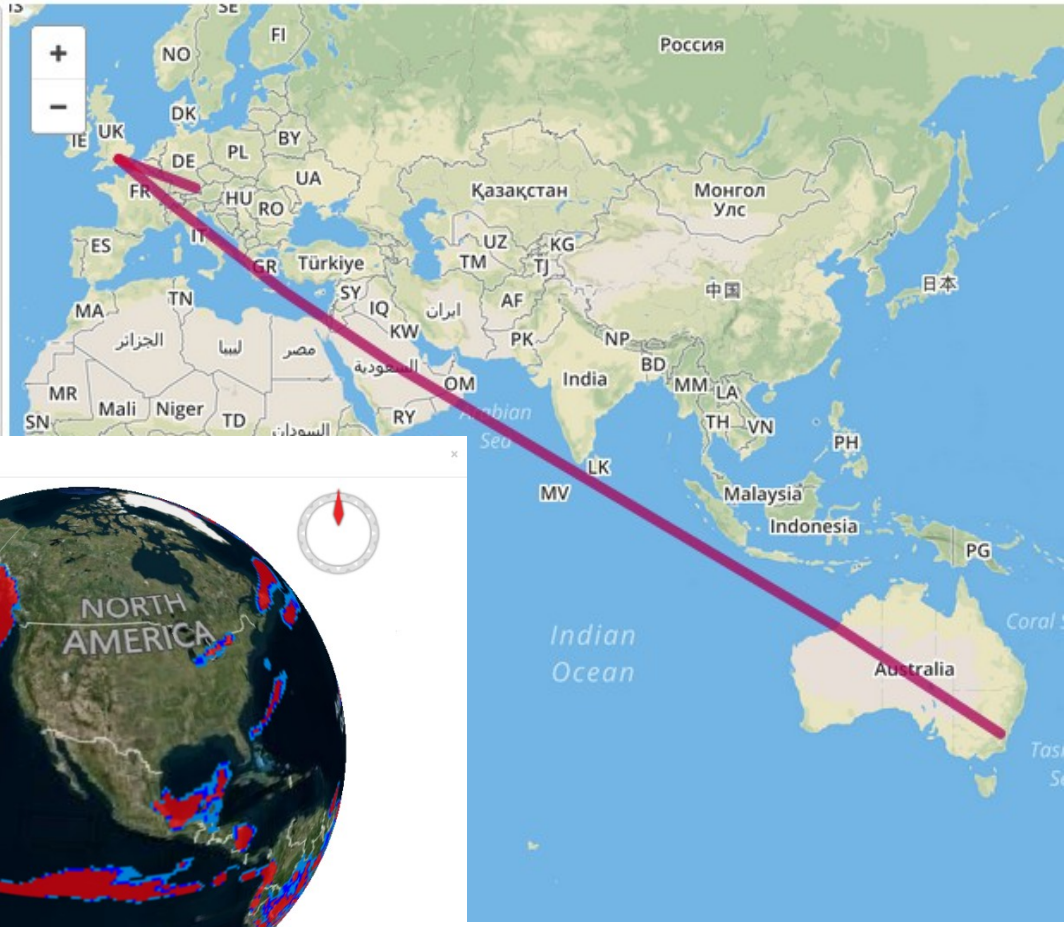
Heavy rainfall risk areas

Server:

ECMWF

Run Query

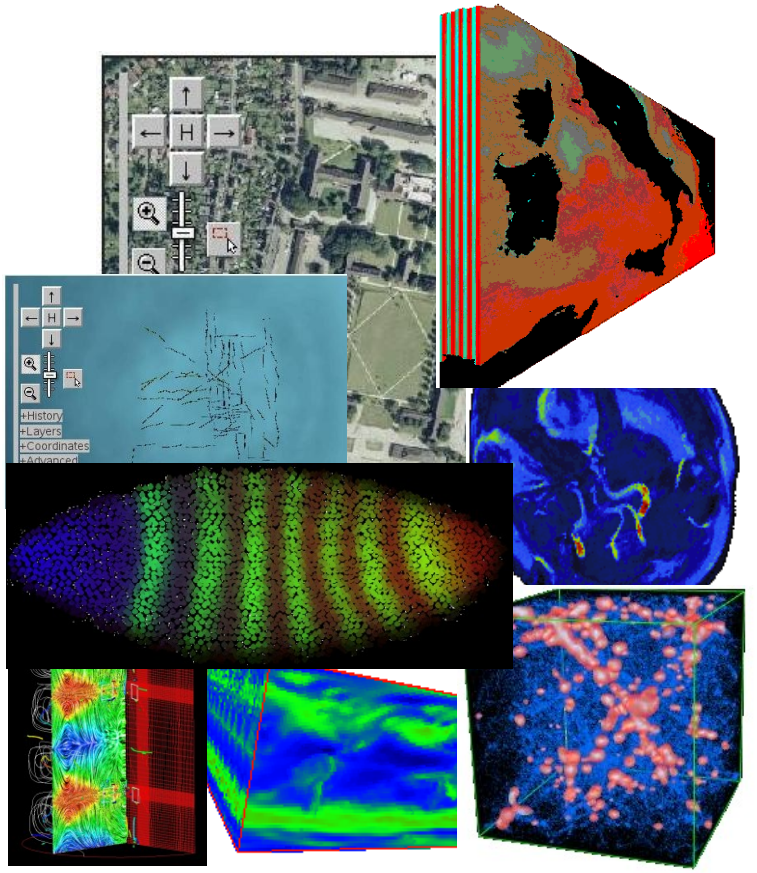




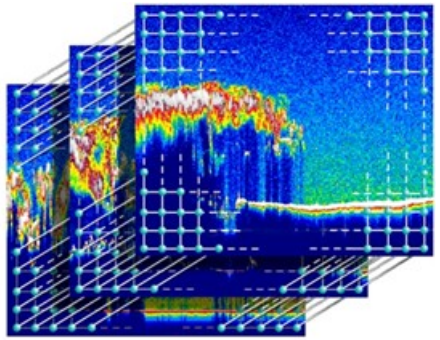
Wrap-Up

Rasdaman: Cross-Domain Datacubes

- Geo
  - Environmental sensor data, 1-D
  - Satellite / seafloor maps, 2-D [VLDB 1999]
  - Geophysics (3-D x/y/z)
  - Climate modelling (4-D, x/y/z/t)
- Life science
  - Gene expression simulation (3-D) [InfSys 2003]
  - Human brain imaging (3-D / 4-D) [TiNS 2001]
- Other
  - Computational Fluid Dynamics (3-D)
  - Astrophysics (4-D)
  - Statistics (n-D)



Summary



- rasdaman = scalable datacube engine
  - Pioneering Array Databases since 1992
  - www.rasdaman.org
- Flexibility: any query, any time, any dimension, any size
- Scalability: CPU/GPU, distributed processing, query rewriting, caching, ...
- Datacube federations
  - Handheld - laptop - cloud - data centers - moving devices
- Mature, large-scale deployments
  - Petascale databases; working on ECMWF's 220 PB
- Open datacube standards: blueprint, reference implementation