



United in Variety: The EarthServer Datacube Federation

Big data and machine learning in geosciences

EGU, 2020-may-07

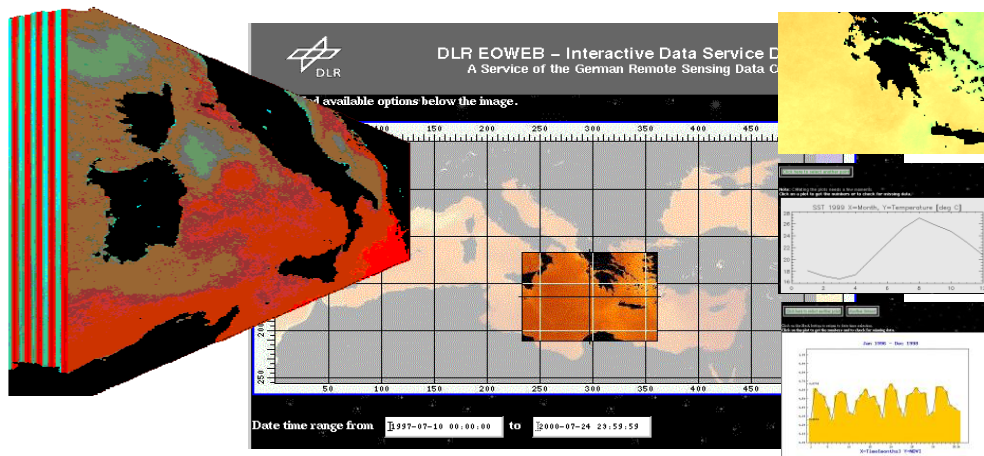
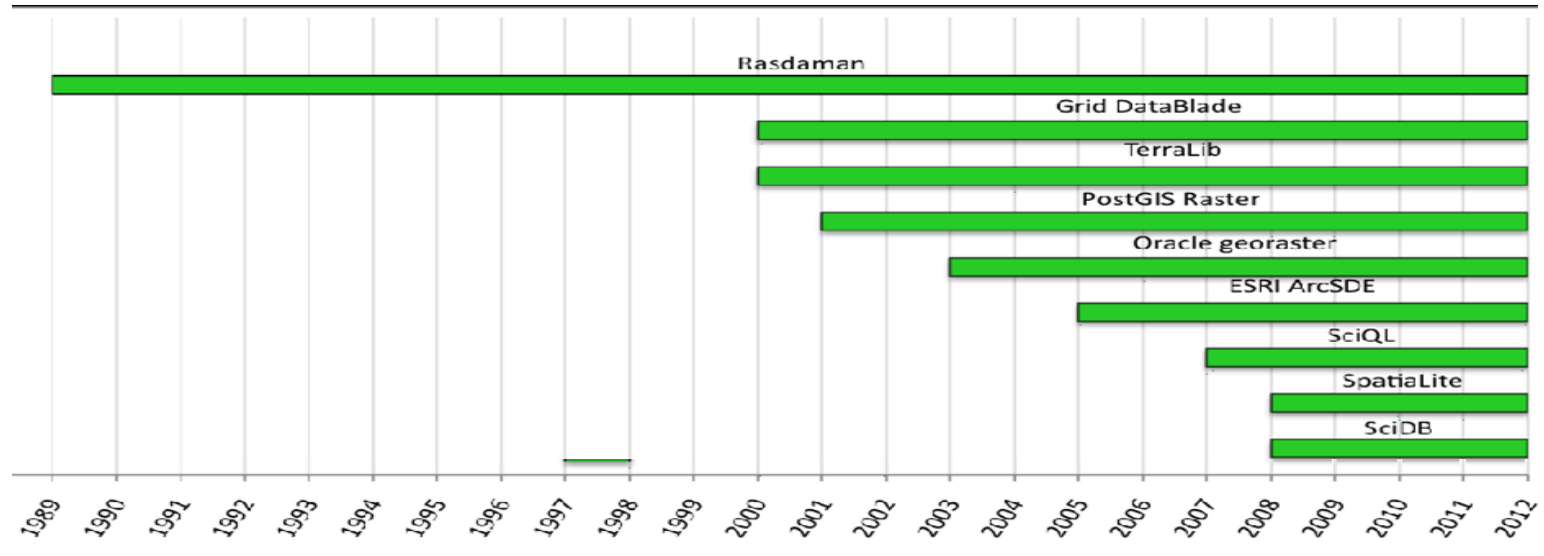
Peter Baumann

Jacobs University | rasdaman GmbH

Motivation

- Databricks accepted as cornerstone for analysis-ready data
 - and visualization, and fusion, ...
- Pioneered with [Baumann 1992] and rasdaman [www.rasdaman.org]
 - rasdaman = full-stack implementation, array queries, distributed processing
 - Recently many epigons: SciQL, SciDB, PostgreSQL Raster, ODC, TileDB, ...

Datacubes: Experience Background



[Diedrich et al 2001]

Motivation

- Databricks accepted as cornerstone for analysis-ready data
 - and visualization, and fusion, ...
- Pioneered with [Baumann 1992] and rasdaman [www.rasdaman.org]
 - rasdaman = full-stack implementation, array queries, distributed processing
 - Recently many epigons: SciQL, SciDB, PostgreSQL Raster, ODC, TileDB, ...
 - Research Data Alliance 2018, leading tools benchmarked:
„rasdaman can be 304x faster than other tools“
- **EarthServer** = Earth datacube federation
 - location-transparent access, analysis, fusion
 - Open data provider community, open standards, freedom in client choice

Datacube Federation: Beginnings

German recommendations for Roadmap

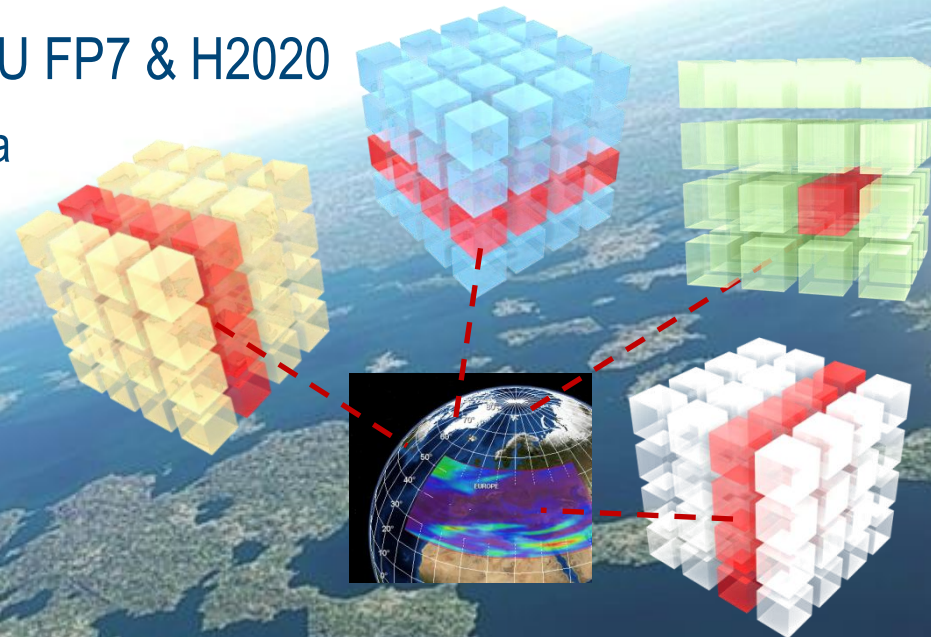
EO Big Data

- Establish „European Datacube Federation“ as a strategic goal
 - Advancing user service quality, specifically: timeseries analytics
 - Enable flexible framework for value-adding services, through open standards
 - Maintain European technology lead
 - EarthServer as reference example to be enlarged
- Foster and Organize joint activities with platform developers and operators in member states
 - Analyse existing EO Exploitation Platforms with respect to common functions and interoperable interfaces
 - Define federation requirements including IT security aspects on appropriate levels
 - Support harmonization and standardization on interfaces of federated platforms
 - Develop components and workflows supporting harmonized scenarios and interfaces





- Agile, location-transparent **analysis + fusion + visualization** ready datacubes
- **Open federation** of large-scale data providers
 - DIASs, research institutes, agencies, universities, companies, ...
 - 20+ PB and growing: Sentinel SAR & hyperspectral, thematic, products, ...
 - open standards, community governance
- Intercontinental initiative, started with EU FP7 & H2020
 - free of charge; no need to publish all data
 - Now accepting membership requests



Reviewers & EC:

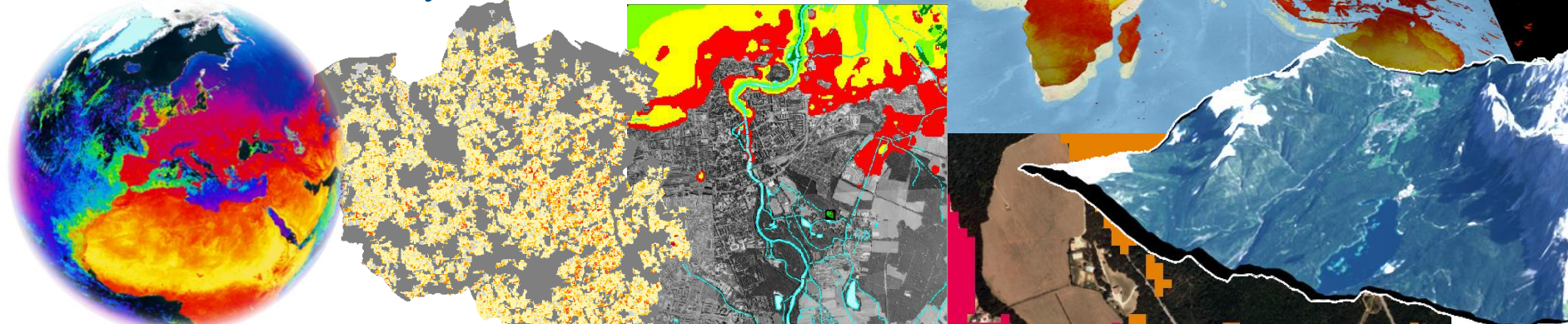
"proven evidence", will "significantly transform [how to] access and use data" ...and "with no doubt has been shaping the Big Earth Data landscape"



Back to the User

- OGC W*S → 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), ...
- Server-side polygon clipping, visualization, analytics, fusion, ...

```
In [12]: import requests
         query = ""
         for a in (CCI_V2_monthly_chlor_a) return encode (switch case 0.05 > a[Lat(30:70)],Long
         ""
         resp = requests.post('http://earthserver.pml.ac.uk/rasdaman/...')
         from IPython.display import Image
         Image(data=resp.content)
```



[rasdaman-based portals]

