On-board Analytics of Earth Observation Satellite Sensor Data

14th SPACOMM, Barcelona, Spain, 2022-apr-25

Peter Baumann, Dimitar Misev

Jacobs University | rasdaman GmbH

baumann@rasdaman.com

Supported by:

Federal Ministry of Economics and Technology

Technically supported by:

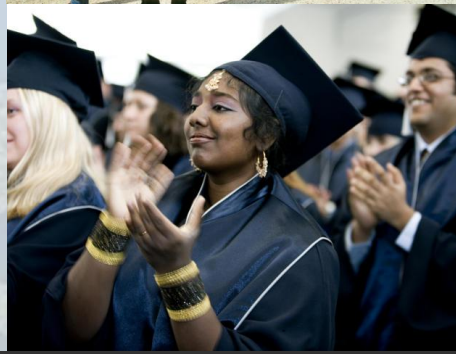
About the Presenter

Dr. Peter Baumann is Professor of Computer Science and entrepreneur. At Jacobs University he researches on flexible, scalable datacube services and their application in science and engineering. With the rasdaman engine he and his team have pioneered datacubes and Array Databases, and have set the de-facto standard for datacube services, documented by 160+ scientific publications, international patents and numerous high-ranking innovation awards. For his continued effort in IT education he was elected Teacher of the Year 2020. As founder and CEO he leads the successful international commercialization of rasdaman. Since many years, Baumann is leading datacube standardization in ISO, OGC, and EU INSPIRE. Baumann is chair, IEEE GRSS Earth Science Informatics Technical Committee; co-chair, OGC Coverages.SWG and Coverages.DWG and BigData.DWG; German delegate, INSPIRE; editor, ISO 19123-x suite.

See details on https://peter-baumann.org.

Datacube Research @ Jacobs U

- Large-Scale Scientific Information Systems research group
  - large-scale n-D raster services & beyond
  - https://l-sis.org
- Spin-off company: rasdaman GmbH
- Main visible results:
  - pioneer Array DBMS, rasdaman
  - Datacube standards



Roadmap

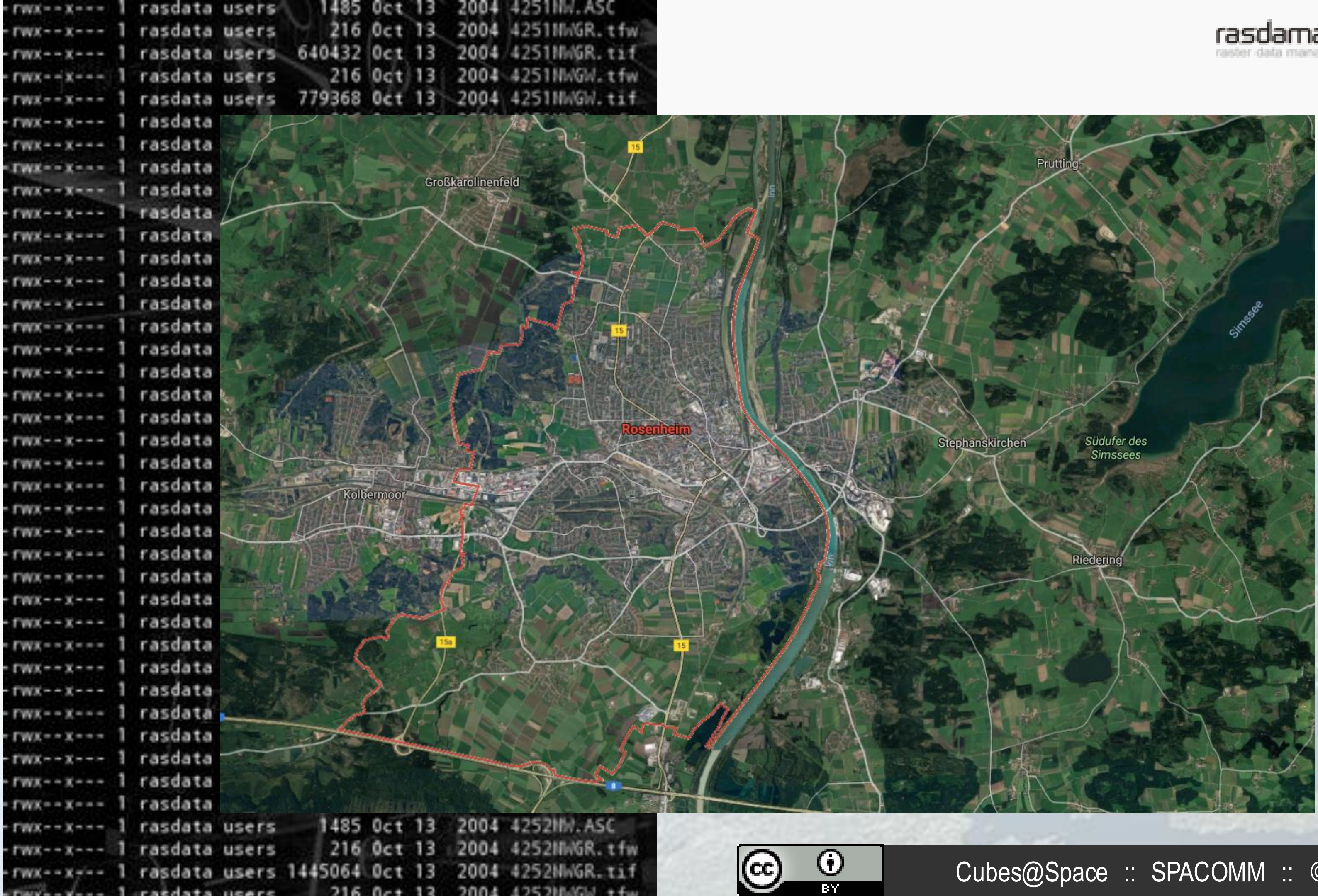
- Introduction
- Datacubes
- Array Databases
- ORBiDANSe
- Further Projects
- Conclusion & Outlook

```
-rwx--x--- 1 rasdata users    1485 Oct 13  2004 4251NW.ASC
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251NWGR.tfw
-rwx--x--- 1 rasdata users  640432 Oct 13  2004 4251NWGR.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251NWGW.tfw
-rwx--x--- 1 rasdata users  779368 Oct 13  2004 4251NWGW.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251NWRL.tfw
-rwx--x--- 1 rasdata users  712492 Oct 13  2004 4251NWRL.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251NWWL.tfw
-rwx--x--- 1 rasdata users   62830 Oct 13  2004 4251NWWL.tif
-rwx--x--- 1 rasdata users    1498 Oct 13  2004 4251SO.ASC
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SOGR.tfw
-rwx--x--- 1 rasdata users 1076750 Oct 13  2004 4251SOGR.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SOGW.tfw
-rwx--x--- 1 rasdata users  197142 Oct 13  2004 4251SOGW.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SORL.tfw
-rwx--x--- 1 rasdata users  936348 Oct 13  2004 4251SORL.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SOWL.tfw
-rwx--x--- 1 rasdata users  119990 Oct 13  2004 4251SOWL.tif
-rwx--x--- 1 rasdata users    1485 Oct 13  2004 4251SW.ASC
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SWGR.tfw
-rwx--x--- 1 rasdata users  577868 Oct 13  2004 4251SWGR.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SWGW.tfw
-rwx--x--- 1 rasdata users  352188 Oct 13  2004 4251SWGW.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SWRL.tfw
-rwx--x--- 1 rasdata users  913032 Oct 13  2004 4251SWRL.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4251SWWL.tfw
-rwx--x--- 1 rasdata users   74152 Oct 13  2004 4251SWWL.tif
-rwx--x--- 1 rasdata users    1485 Oct 13  2004 4252NO.ASC
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NOGR.tfw
-rwx--x--- 1 rasdata users  355774 Oct 13  2004 4252NOGR.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NOGW.tfw
-rwx--x--- 1 rasdata users   49046 Oct 13  2004 4252NOGW.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NORL.tfw
-rwx--x--- 1 rasdata users  600964 Oct 13  2004 4252NORL.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NOWL.tfw
-rwx--x--- 1 rasdata users   46714 Oct 13  2004 4252NOWL.tif
-rwx--x--- 1 rasdata users    1485 Oct 13  2004 4252NW.ASC
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NWGR.tfw
-rwx--x--- 1 rasdata users 1445064 Oct 13  2004 4252NWGR.tif
-rwx--x--- 1 rasdata users     216 Oct 13  2004 4252NWGW.tfw
```

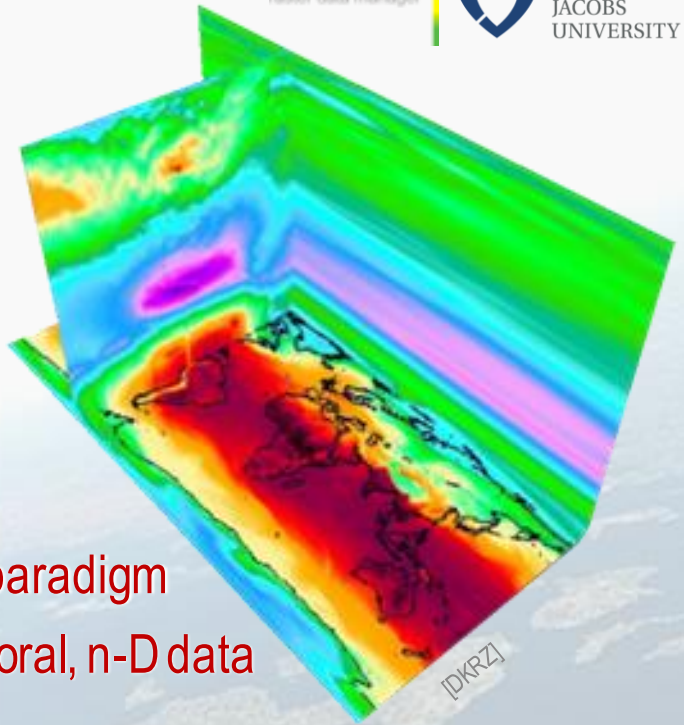
rasdaman raster data manager

JACOBS UNIVERSITY

rasdaman
raster data manager
JACOBS UNIVERSITY

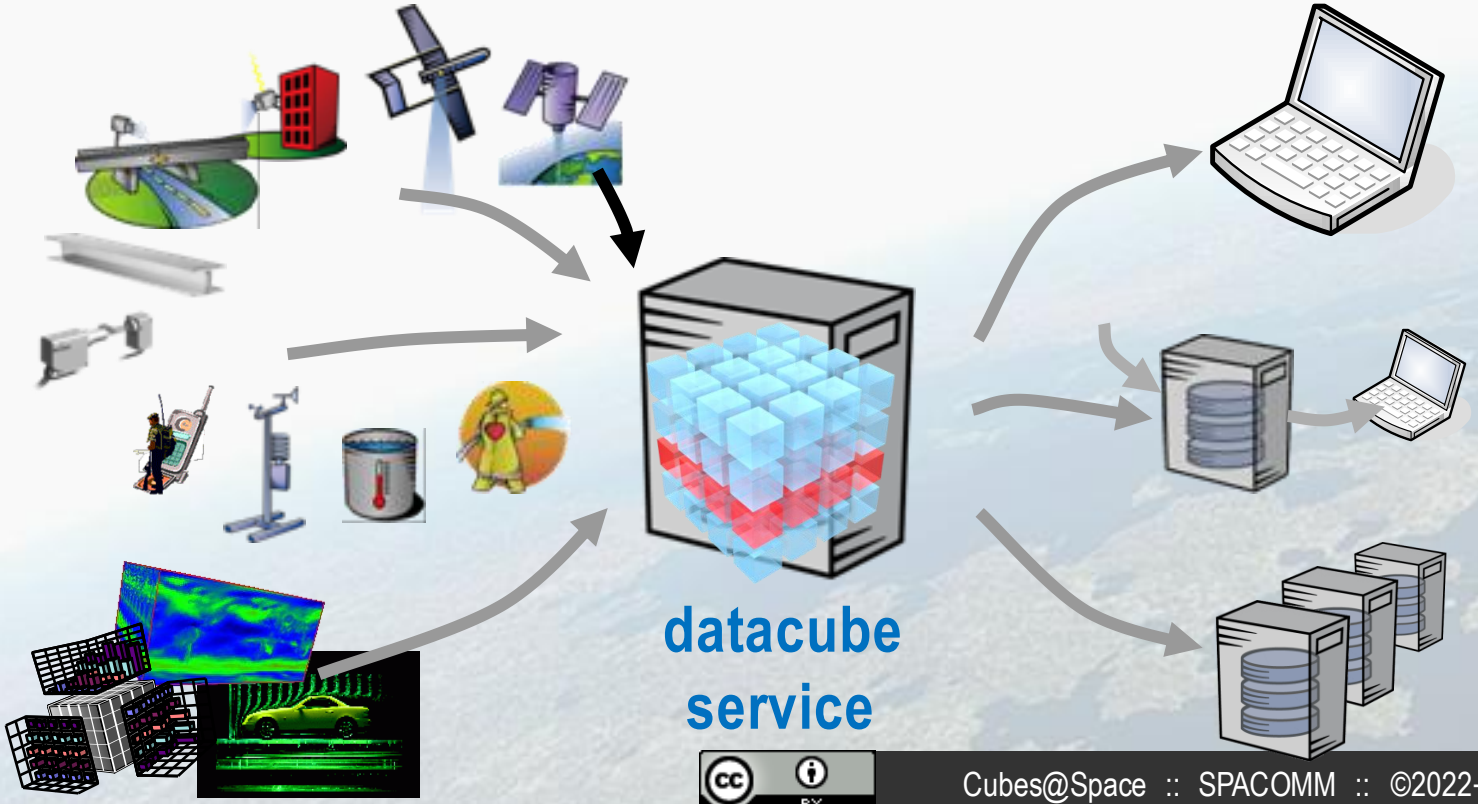


Datacubes?



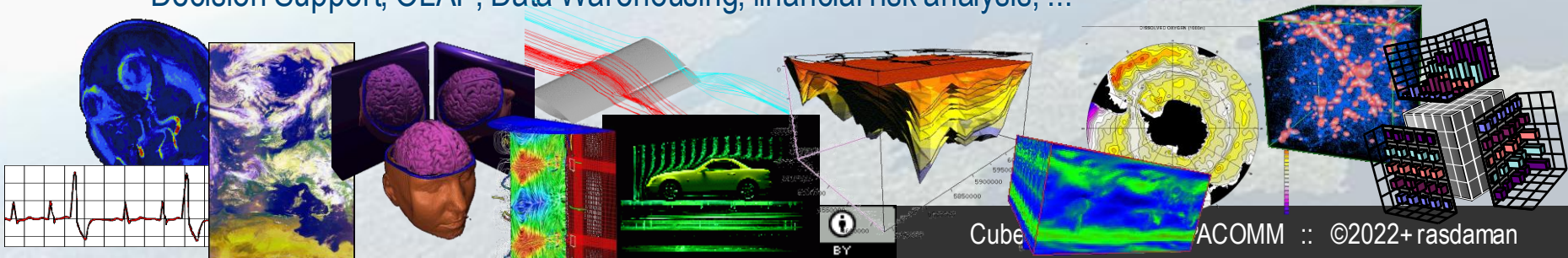
- Sensors & simulations
  -> gridded („raster“) data
- Datacubes are the natural paradigm
  to interact with spatio-temporal, n-D data
- Avoid undue complexity
  -> data + service on high semantic level

Homogenized, Analysis-Ready Datacubes



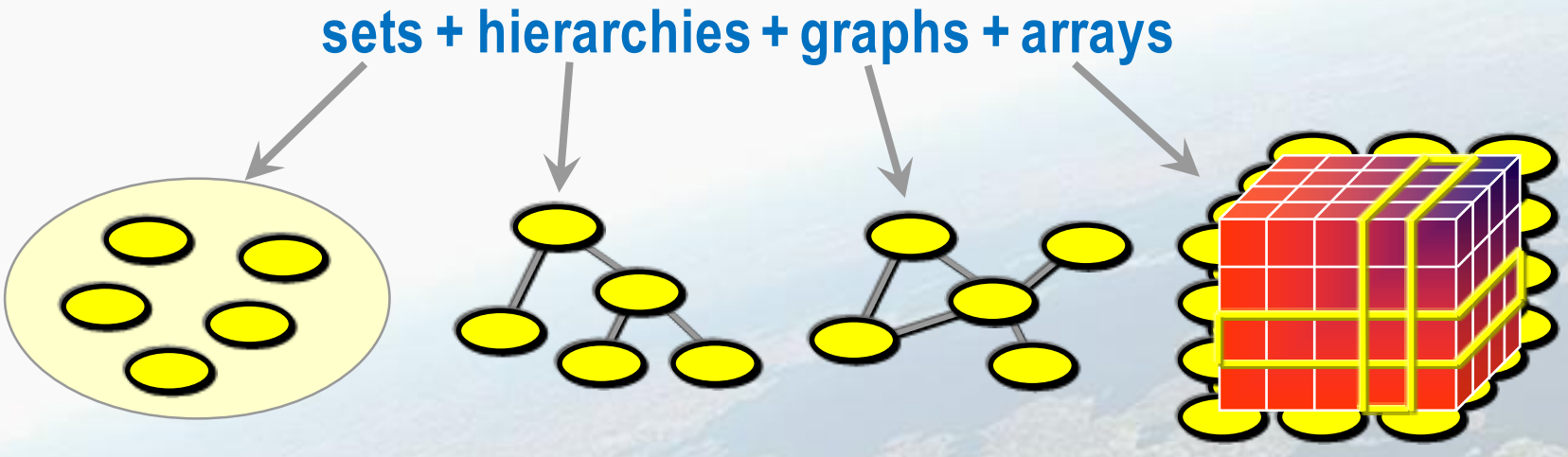
Arrays: *Sensor, Image, Model, Statistics Data*

- **Life Sciences**
  - Pharma/chem, healthcare / bio research, bio statistics, genetics, ...
- **Earth Sciences**
  - Geodesy, geology, hydrology, oceanography, meteorology, remote sensing, ...
- **Space Sciences**
  - Astronomy, planetary science, astrobiology, ...
- **Engineering**
  - Simulation & experimental data in automotive / shipbuilding / aerospace industry, turbines, process industry, ...
- **Management/Controlling**
  - Decision Support, OLAP, Data Warehousing, financial risk analysis, ...



An Array Database Perspective

Fundamental Information Categories (at Least in Databases)

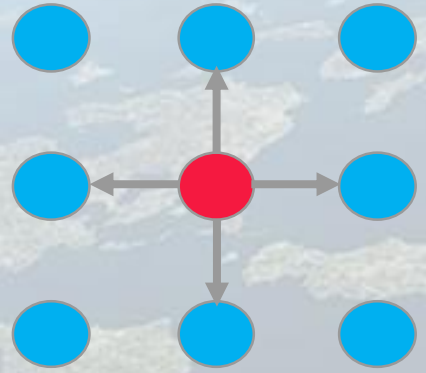


Array Analytics

- Array Analytics :=

  Efficient analysis on multi-dimensional arrays of a size orders of magnitude above evaluation engine's main memory

- Essential data property: n-D Cartesian neighborhood
  - Secondary: #dimensions, density, ...
- Operations: Linear Algebra++



[EDBT/ICDT Array Databases Workshop, 2011]

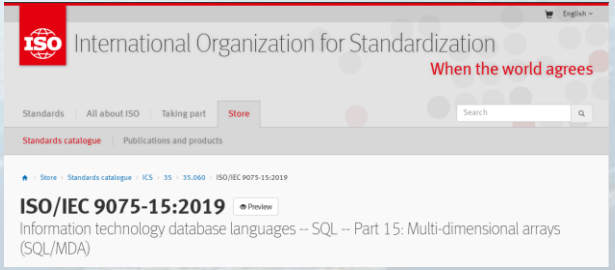
State of the Art

- Actionable datacubes: Baumann 1991, 1994
- OLAP: Business & statistics datacubes: Inmon 1992
- Today:
  - Array Databases
    - *rasdaman, SciQL, SciDB, Oracle GeoRaster, PostGIS Raster, EXTASCID, Teradata Arrays, NorthLight, ...*
  - Libraries & programming tools
    - *xarray:,OpenDataCube, CubeX, …*
    - *other: Wendelin.core, boost::geometry, Ophidia, Google Earth Engine, ...*
  - MapReduce based: SciHadoop, GeoTrellis, ClimateSpark, SciSpark, ...
- Research Data Alliance report: 19 tools deep-comparison, 4 benchmarked
  - https://rd-alliance.org/system/files/Array-Databases_final-report.pdf

rasdaman

„raster data manager“ - pioneered actionable datacubes: 160+ publications, patents

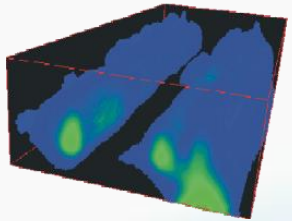
- High-level datacube query language
- massively scalable Datacube Management & Analytics engine
  - 30+ PB; 1000x parallelization, planetary-scale federation
  - Intelligent ETL: automated datacube maintenance & optimization
- Reference implementation, multi-award winning



Array Query Operators: rasql

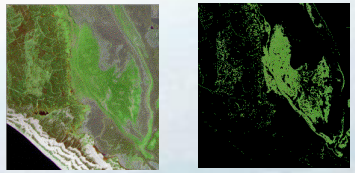
- selection & section

```
select c[ *:*, 100:200, *:*, 42 ]
from   ClimateSimulations as c
```



- result processing

```
select img * (img.green > 130)
from   LandsatArchive as img
```



- search & aggregation

```
select mri
from   MRI as mri, masks as am
where  some_cells( mri > 250 and m )
```

- data format conversion

```
select encode( c[ *:*, *:*, 100, 42 ], „png“ )
from   ClimateSimulations as c
```

rasql Basis: 3 Algebraic Operators

- Array constructor: generate array, initialize with values
  - Ex:

```
marray p in [ 0:255, 0:511 ]
values A[p] + B[p]
```

shorthand: `A + B`

- Array condenser: aggregate array into summary value
  - Ex:

```
condense +
over p in sdom(A)
using A[p] * B[p]
```

shorthand: `add( A * B )`

- Array sorter: slice, then resort according to some ranking expression

Linear Algebra: Beginnings

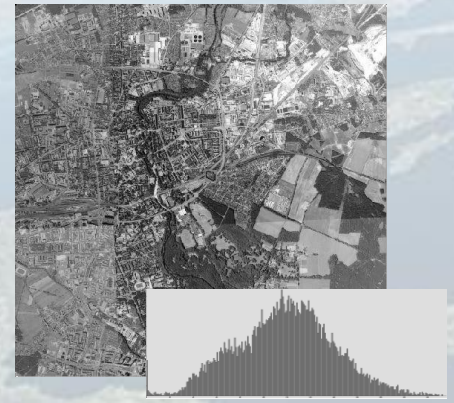
- Matrix multiplication

```
select marray i in [0:m], j in [0:p]
       values condense +
              over     k in [0:n]
              using    a [ i, k ] * b [ k, j ]
from   matrix as a, matrix as b
```

$$(\mathbf{AB})_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}$$

- Histogram

```
select marray bucket in [0:255]
       values count_cells( img = bucket )
from   img
```



- AI-Cube, CENTURION: add ML (Tensor Algebra)

Server-Side Extensibility

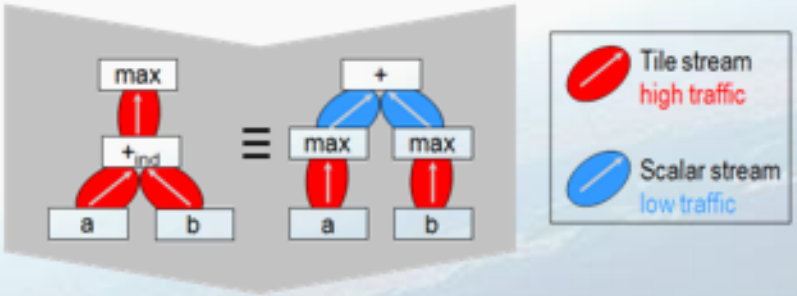
- External code dynamically linked into server, aka User-Defined Function (UDF)
  - Ex: "NDVI from raw Sentinel subset, orthorectified with Orfeo Toolbox"

```
select
  encode(
    otb.orthoRectifFilter(
      ((img.red-img.nir)/(img.red+img.nir))[x0:x1,y0:y1],
      outputSpacing, deformationFieldSpacing
    ),
    "png"
  )
from   LandsatRawArchive as img
```

Query Optimization

- Incoming queries get analyzed: better evaluation possible?
- Query rewriting:
  - Ex:

```
select max_cells( a + b )
from   a, b
```

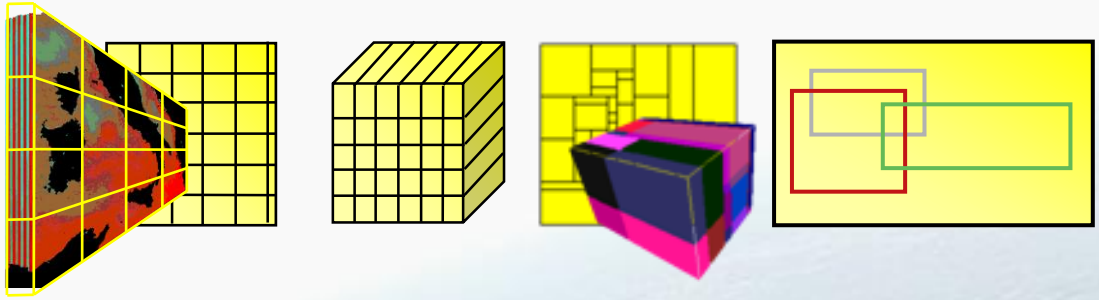



```
select max_cells( a ) + max_cells( b )
from   a, b
```

- Physical optimization: exploit hw, distributed processing, ...

Adaptive Partitioning („Tiling“)

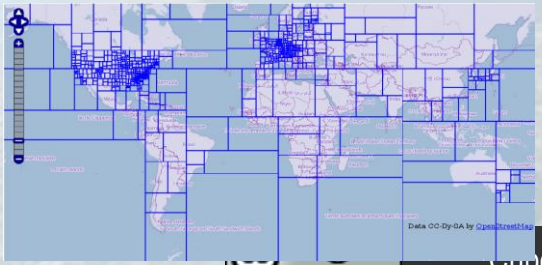
- Any tiling [Furtado 1999]
  - strategies



  - rasdaman storage layout language

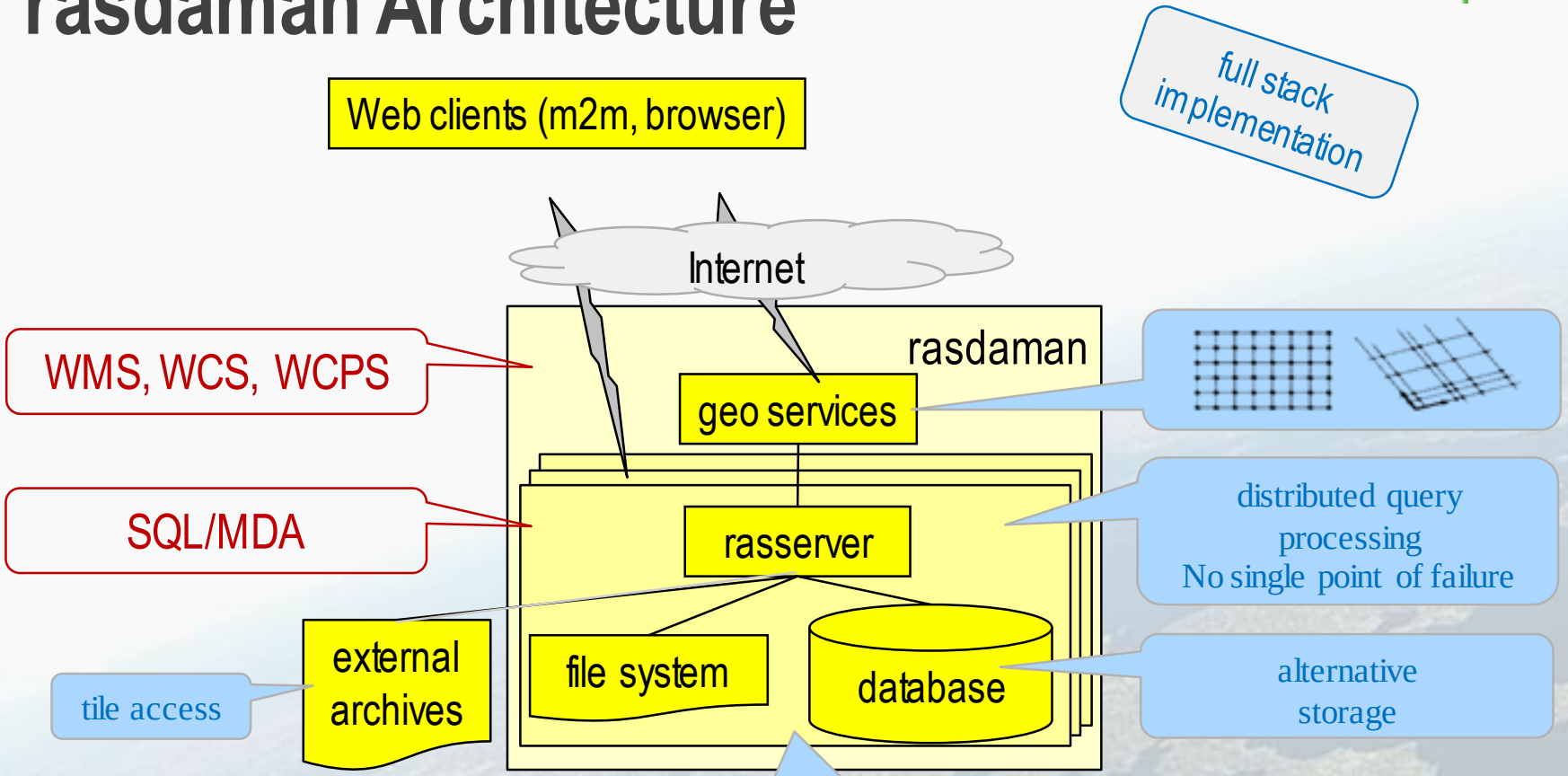
```
insert into MyCollection
   values ...
   tiling
     area of interest [0:20,0:40], [45:80,80:85]
     tile size 1000000
     index d_index storage array compression zlib
```

- Why irregular tiling?



[OpenStreetMap]

rasdaman Architecture



Optimized Distributed Processing

```
max( (A.nir - A.red) / (A.nir + A.red) )
+ avg( B.green )
+ max( (C.red + C.green + C.blue) / 3 )
```

1 query → 1,000+ cloud nodes

[ACM SIGMOD, VLDB]

rasdaman raster data manager

JACOBS UNIVERSITY

from 1D to 4D:

https://standards.rasdaman.com

EarthServer Datacube Federation

- spatio-temporal analytics & fusion, dozens of PB
  - location-transparent
  - open standards, coding-free
  - Open for code & data
- Open, free, transparent, democratic
  - Open & private; free & commercial
  - quickly growing community
- Powered by rasdaman

EarthServer:
https://earthserver.xyz

ORBiDANSe

Downsizing

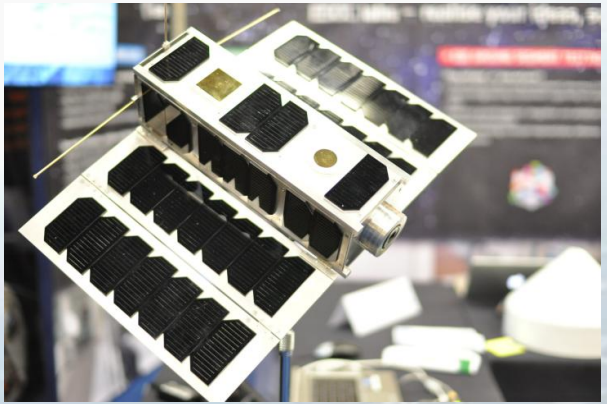
- Oracle (recommended): AMD Opteron etc, 4 GB RAM + swap, 10 GB free diskspace [doc.rasdaman.com]
  - Otherwise, „may not run”

ORBiDANSe

- Orbital Big Data Analytics Service
- Big Data: „Process close to source“
  → datacube engine on satellite
- Deliver insight, not pixels
- Answer analytics questions = avoid full download
- why?
  - https://standards.rasdaman.com/demo_power.html

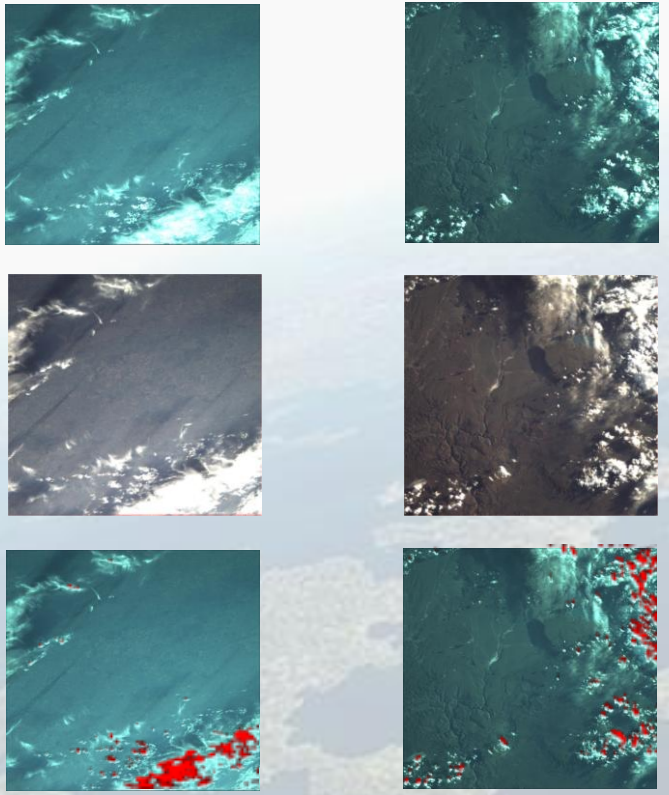
ORBiDANSe

- rasdaman @ OPS-SAT cubesat
  - ARM, 1 GB RAM, 16 GB flash memory
  - 2048x1944 RGB camera
- Prep step: port to Raspberry Pi
- December 2021: sccessful experiments in space



ORBiDANSe: Examples

- Picture taken with onboard camera
  - Goal: show onboard query processing
  - Non-goal: high-quality image analysis
- Ex 1: (naive) radiometry correction
- Ex 2: (naive) cloud mask



Performance: OPS-SAT vs PC

- Sample query:

```
select encode(
         marray i in [0:255]
         values count_cells( c[1:300,1:300].red = i[0] ),
         "csv" )
from Images_20211027 as c
```

- Comparison: OPS-SAT vs COTS office PC
- Result:
  - Onboard queries slower 8x – 23x

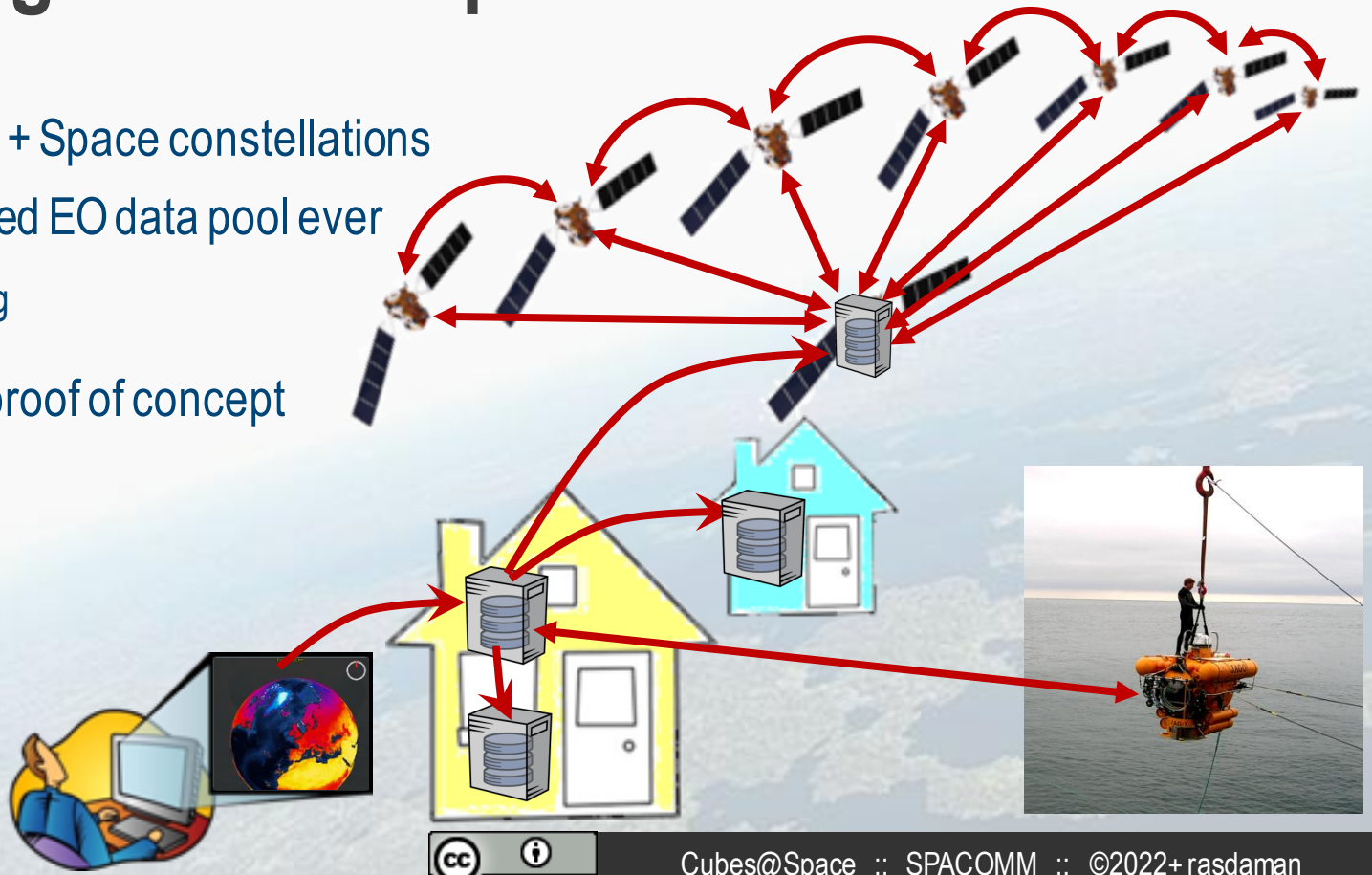
query#	6	7	8	9	10
satellite	2.34	2.61	2.61	28.31	14.9
desktop	0.1	0.32	0.32	3.45	1.43
query#	**11**	**12**	**13**	**14**	
satellite	9.28	20.90	17.37	11.6	
desktop	1.05	1.79	0.94	1.48	

Lessons Learnt

- Datacubes feasible on minimalistic onboard architecture
- Massive bandwidth reduction possible (↳demo), more satellite power
- Onboard processing not replacing, but augmenting basic complete data supply
- Onboard IT still far from industry standards, porting laborious and cumbersome
- Vision:
  - Nanosats as edge devices, seamlessly integrated in federations etc
  - Nanosats building their own federation in space
  - Democratized data access and analysis, via standard clients

Vision: Integrated Dataspaces

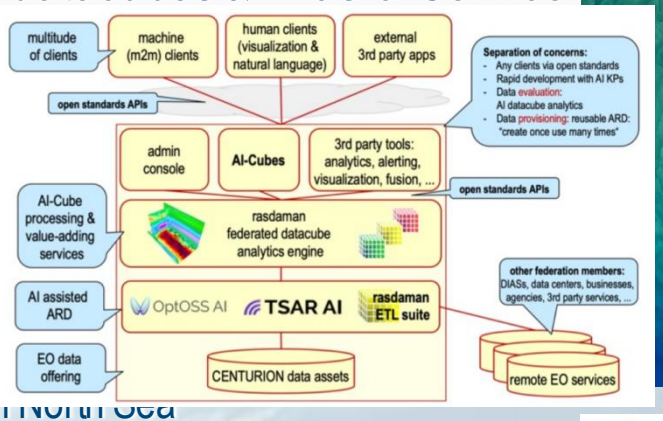
- Earth federations + Space constellations
  = largest integrated EO data pool ever
  - Query → tasking
- ORBiDANSe as proof of concept
- Submarines?

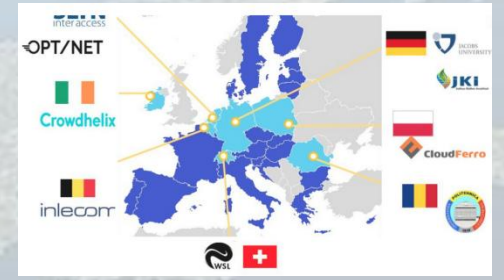


Projects

CENTURION

- combine & advance spatio-temporal datacubes & AI-as-a-Service
  - Copernicus for EO & non-EO markets
- 5 use cases:
  - Analysis Ready Data (ARD) generation
  - Agricultural weather index configurator
  - Characterization of shipping dynamics in North Sea
  - Near Real Time monitoring of the forest dynamics
  - Near Real Time Flood zone mapping
- EU H2020, May 2021 – April 2024

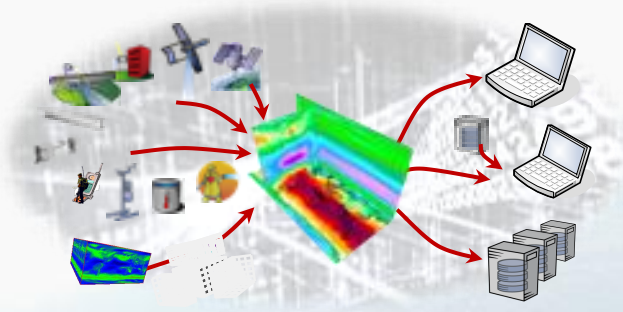




centurion-project.eu

coordination), OPT/NET (BE tech mgmt) + 5 more partners

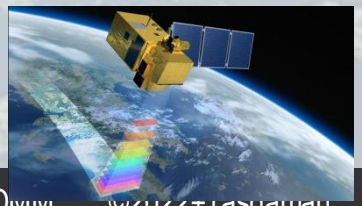
Cube4EnvSec



- NATO need: joint forces operating "off the same map"
  - from anywhere to anywhere, with realtime updates
- Cube4EnvSec: ad-hoc realtime mix&match
- Approach: analysis-ready spatio-temporal datacubes: sat, drone, metoc, …
  - Multi-source federation + moving sensors realtime integration
- Strong end user involvement; open for further collaboration (resources permitting)
- NATO SPS, start 2022-03-01, 18 months
  - Partners: Jacobs U, DE; Tel Aviv U, Israel; Greenland Institute of Natural Resources, GL/DK

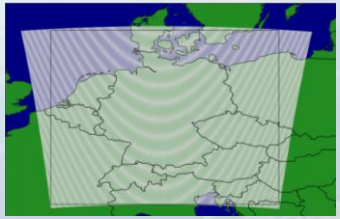


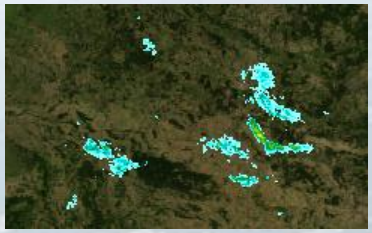




DeepRain

- improved rain prediction for mountaneous areas through ML + datacubes
- COSMO models, precipitation radar, topography, etc
  - Challenge: rotated grids
    → enhanced rasdaman, GDAL, PROJ

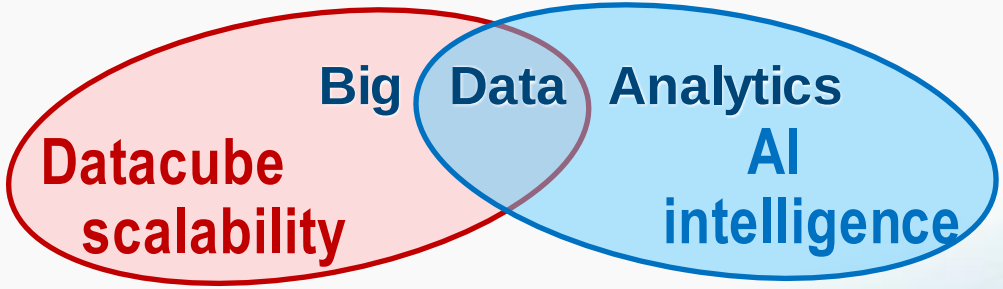




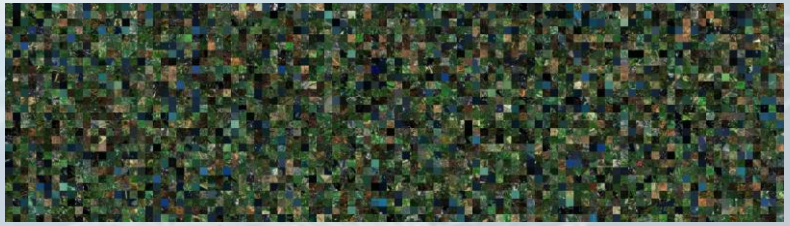
- EarthServer member: climate datacubes
- Jülich Supercomputing Centre, Osnabrück U, Bonn U, Jacobs U, DWD

deeprain-project.de

AI-Cube



- federated EO datacubes + advanced EO AI + natural language processing
- BigEarthNet Large-Scale Benchmark Archive
  - 590,326 tagged Sentinel-1/2 pairs



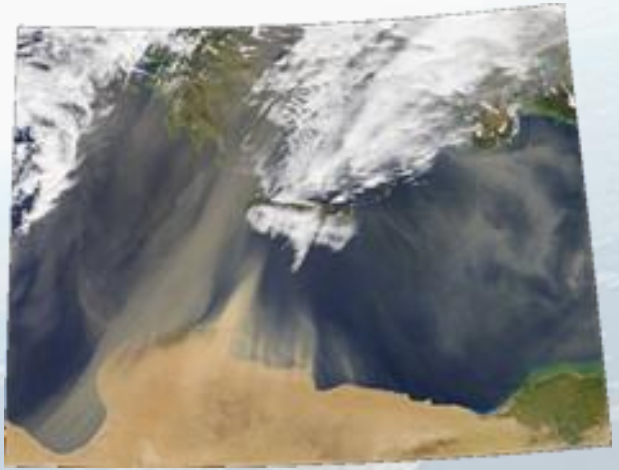
- Jacobs University, TU Berlin, rasdaman GmbH; Sep 2021, 18 months

Bundesministerium für Ernährung und Landwirtschaft

ai-cu.be

ML-Cube

- NASA remote sensing EO + rasdaman datacubes
  - rasdaman federation NASA / ESA DIAS
- Research question: transfer learning MODIS – Sentinel-2
- Use case: atmospheric dust transport
- Supported by IEEE
  Geoscience and Remote Sensing Society (GRSS)

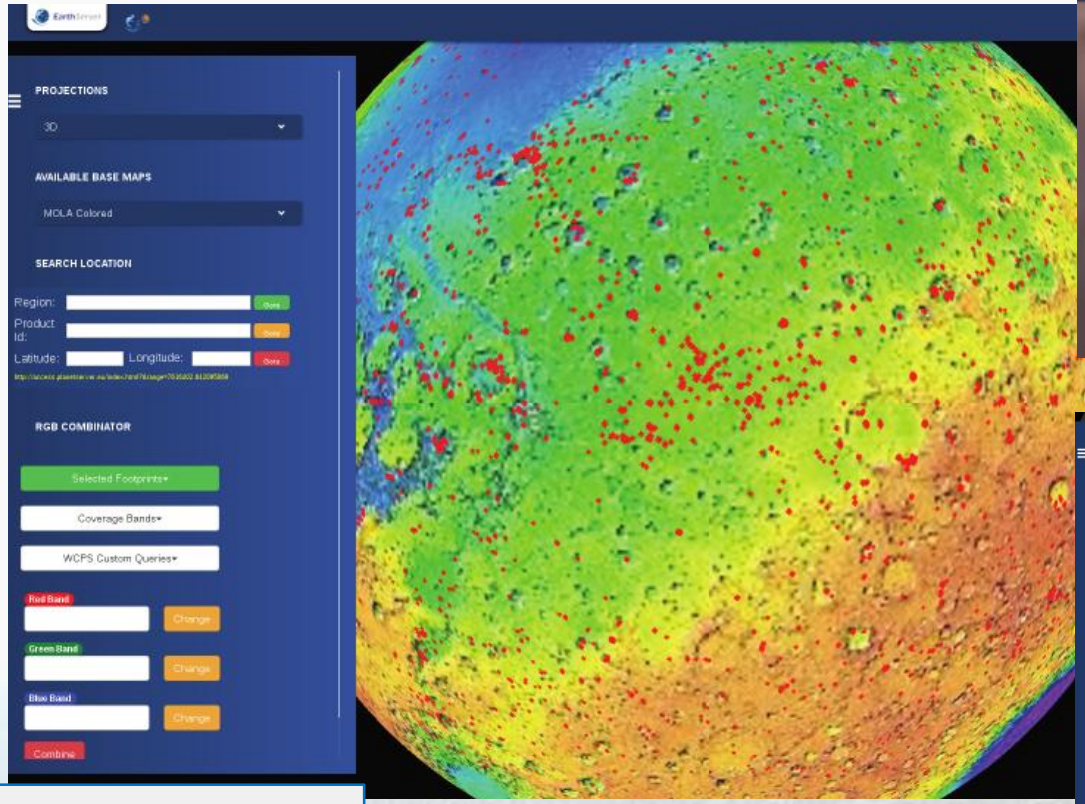


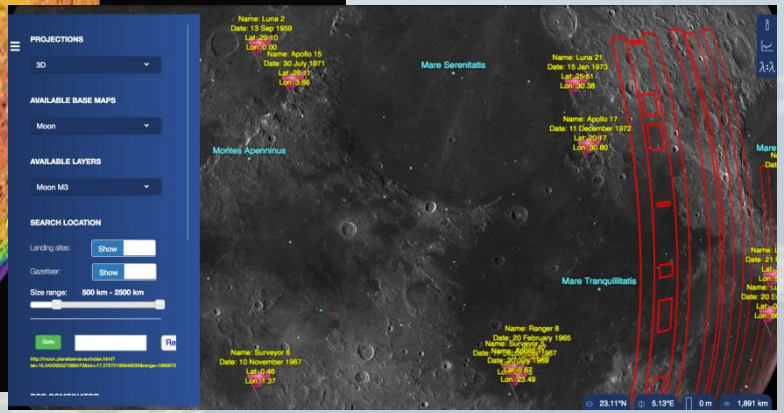
[earthobservatory.nasa.gov]

PlanetServer

rasdaman raster data manager

JACOBS UNIVERSITY

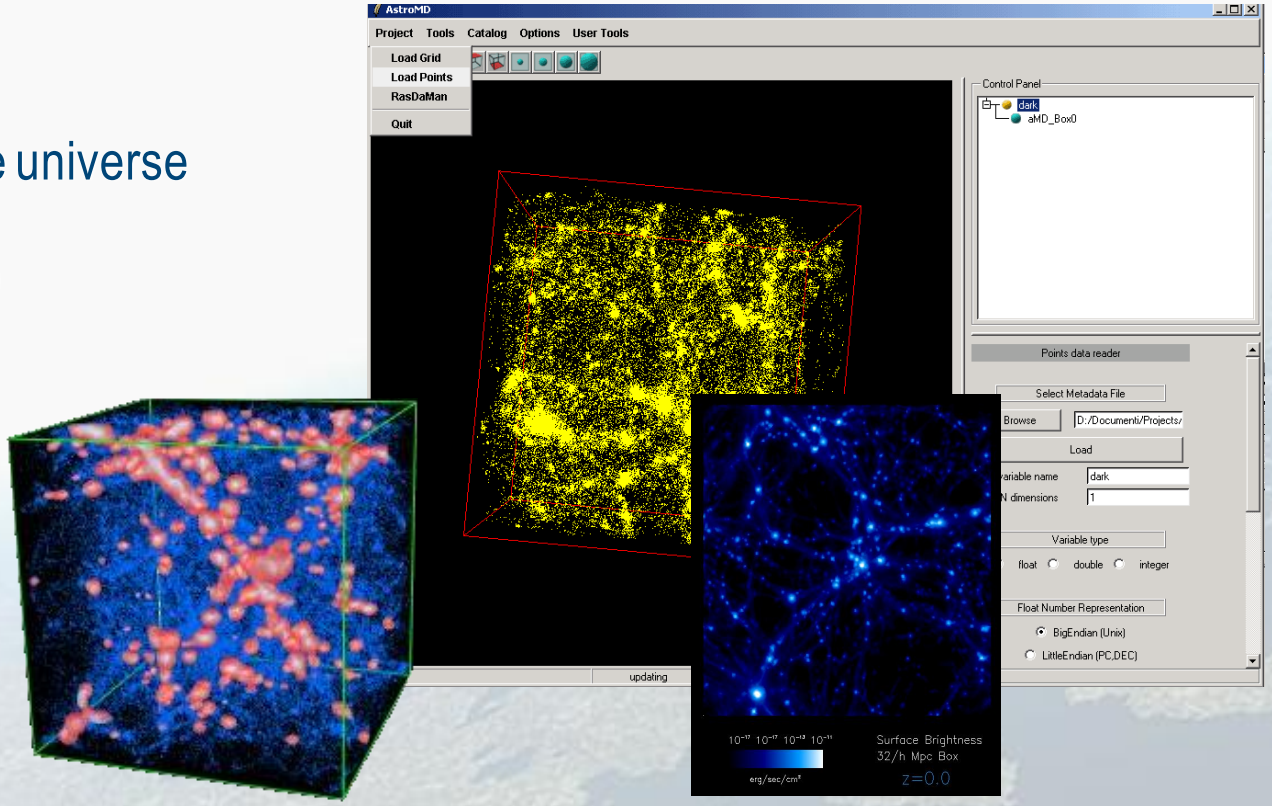




planetserver.eu

Cosmological Simulation

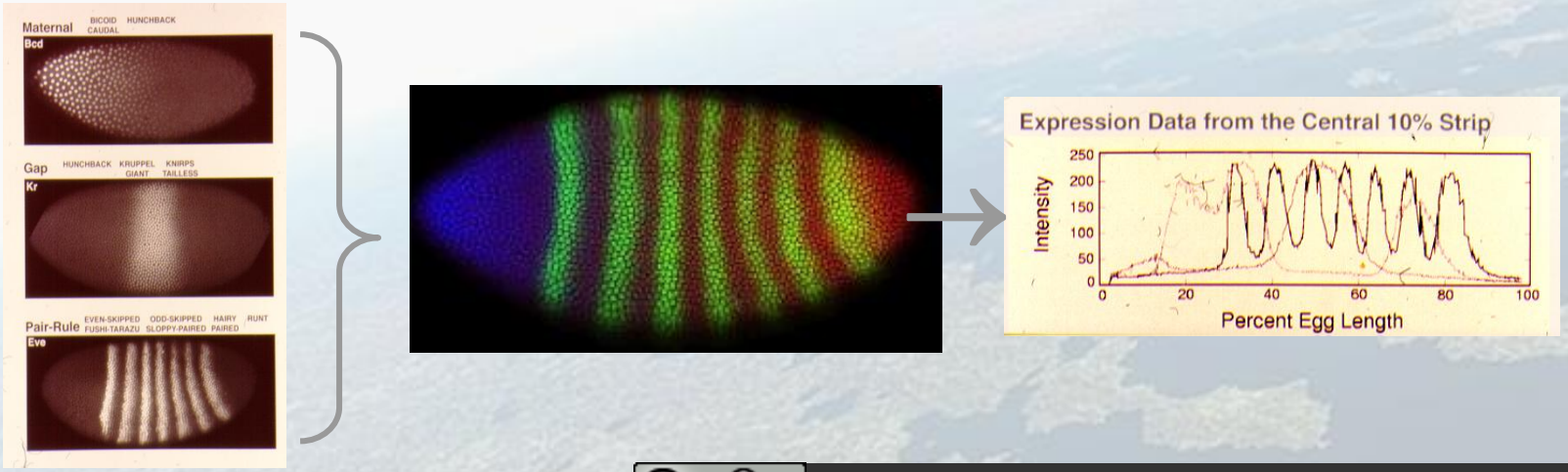
- 3D/4D regions from the universe
- Screenshots: AstroMD
  [Gheller, Rossi 2001]



Gene Expression Analysis

[Samsonova et al]

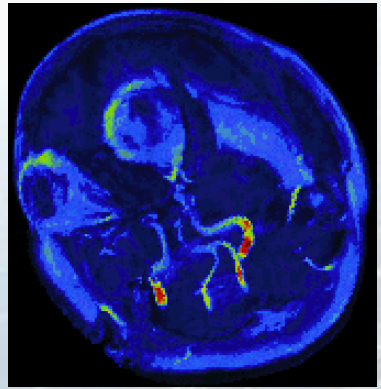
- Gene expression = reading out genes for reproduction
- Goal: capture spatio-temporal expression patterns in Drosophila genes

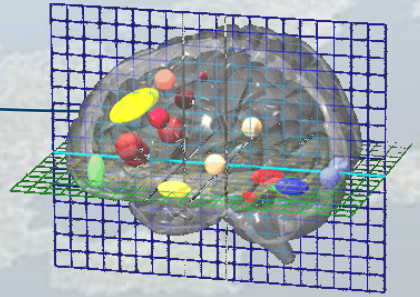


Human Brain Imaging

- Goal: structural-functional relations in human brain
- Experiments (PET, fMRI) → activation maps
- Ex: "*a parasagittal view of all scans containing critical Hippocampus activations, TIFF-coded*."

```
select tiff( ht[ $1, *:*, *:* ] )
from   HeadTomograms as ht,
       Hippocampus as mask
where  count_cells( ht > $2 and mask )
       / count_cells( mask )
       > $3
```



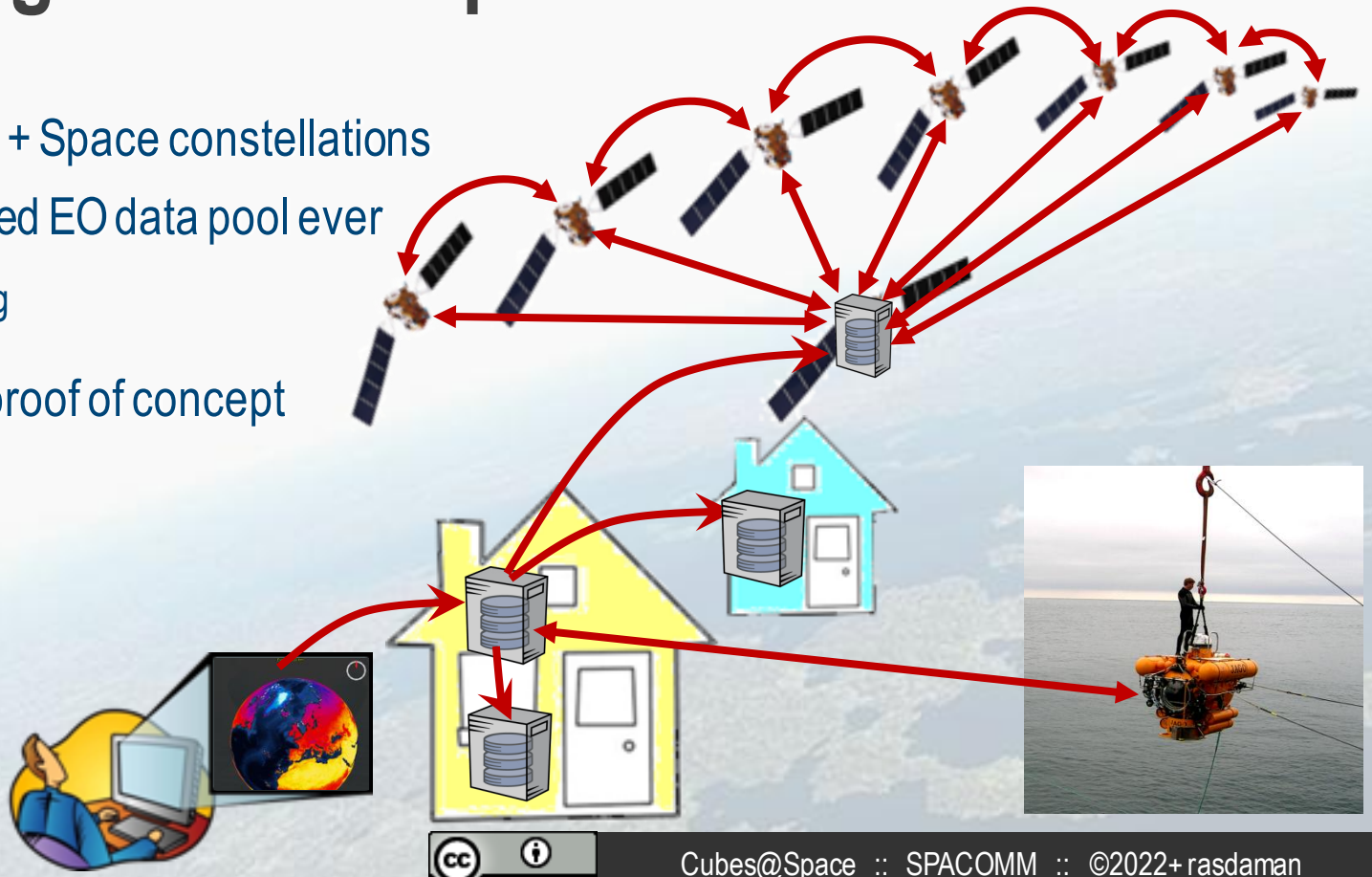


INSPIRE Coverages

- INSPIRE = EU legal framework for common SDI
  - Geo raster data modelled as coverages
  - Web Coverage Service (WCS)
- rasdaman official EU / JRC INSPIRE Good Practice & first full implementation
  - Established in-kind by rasdaman GmbH, DataCove E.U.
  - Data from ES, FI, DE
  - See https://inspire-wcs.eu

Vision: Integrated Dataspaces

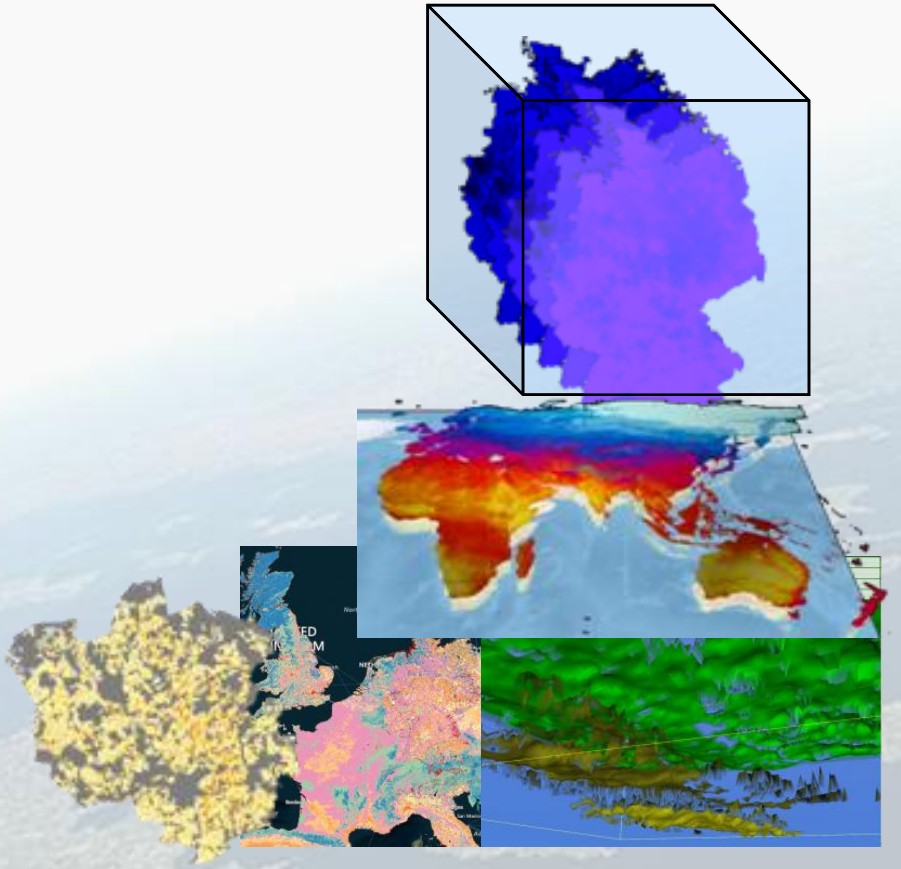
- Earth federations + Space constellations
  = largest integrated EO data pool ever
  - Query → tasking
- ORBiDANSe as proof of concept
- Submarines?



Conclusion

Wrap-Up

- ORBiDANSe = rasdaman @ cubesat
- On-board ad-hoc processing
  - satellite → Actionable Datacube® engine
- Current research on rasdaman: distributed datacubes, AI-Cubes
- Vision: seamless Earth + Space federation



Python?

Wake up –

We‘re in the age

of Web services!