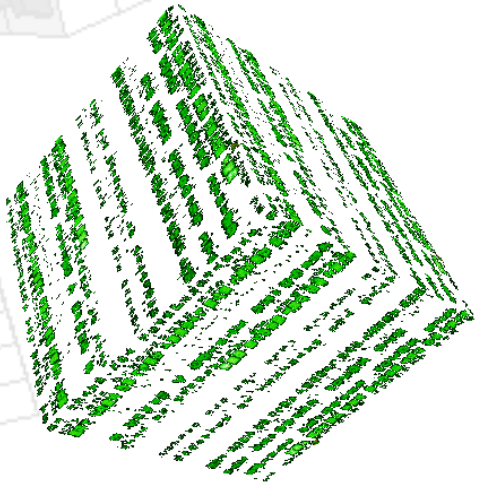


Managing Spatio-Temporal Big Data through Scalable OGC Web Services

FOSSDEM, Brussels, 2016-jan-31

Alex Dumitru

Jacobs University | rasdaman GmbH

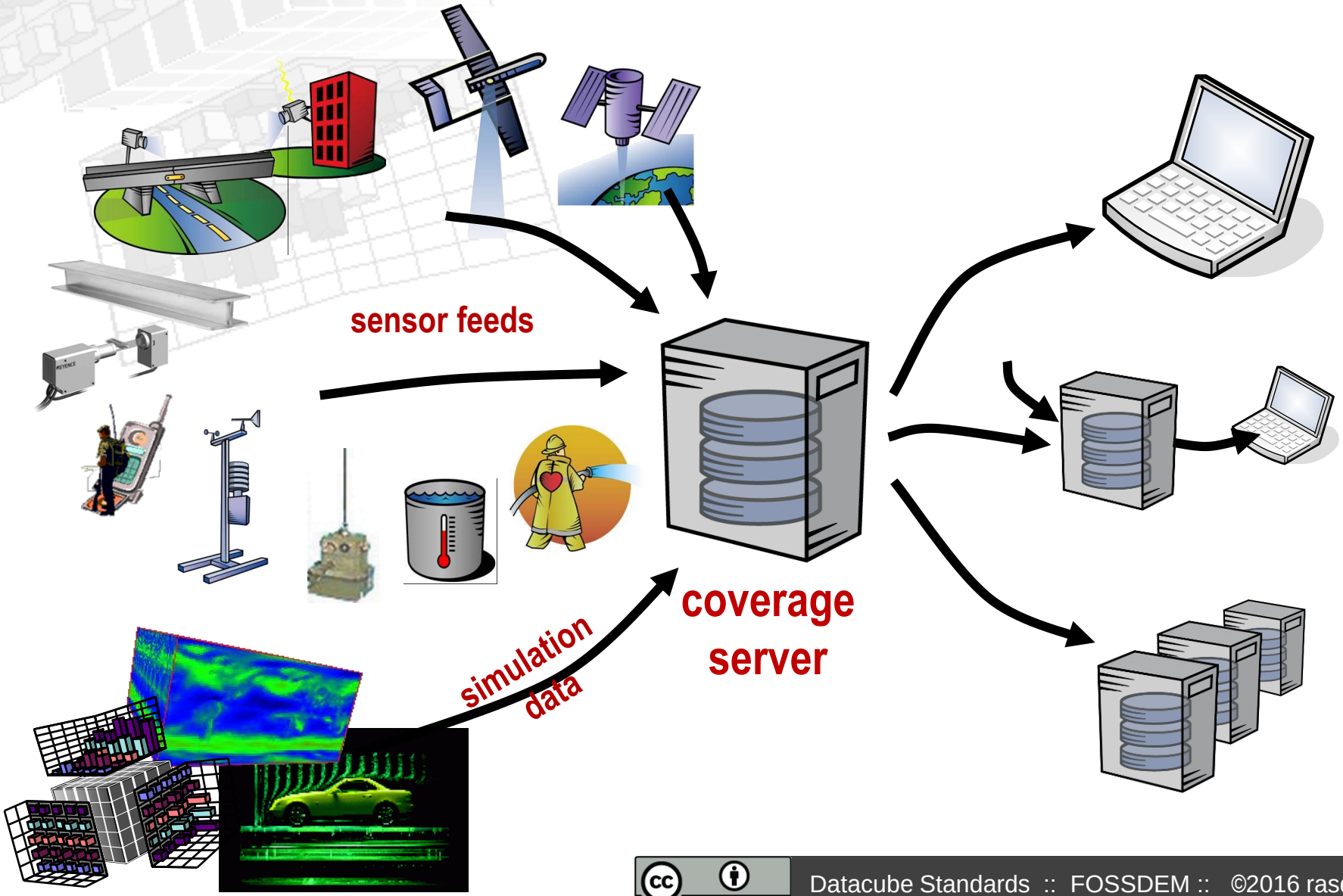
dumitru@rasdaman.com

[gamingfeeds.com]

[co-funded by EU through EarthServer 1/2, PublicaMundi]

CC BY

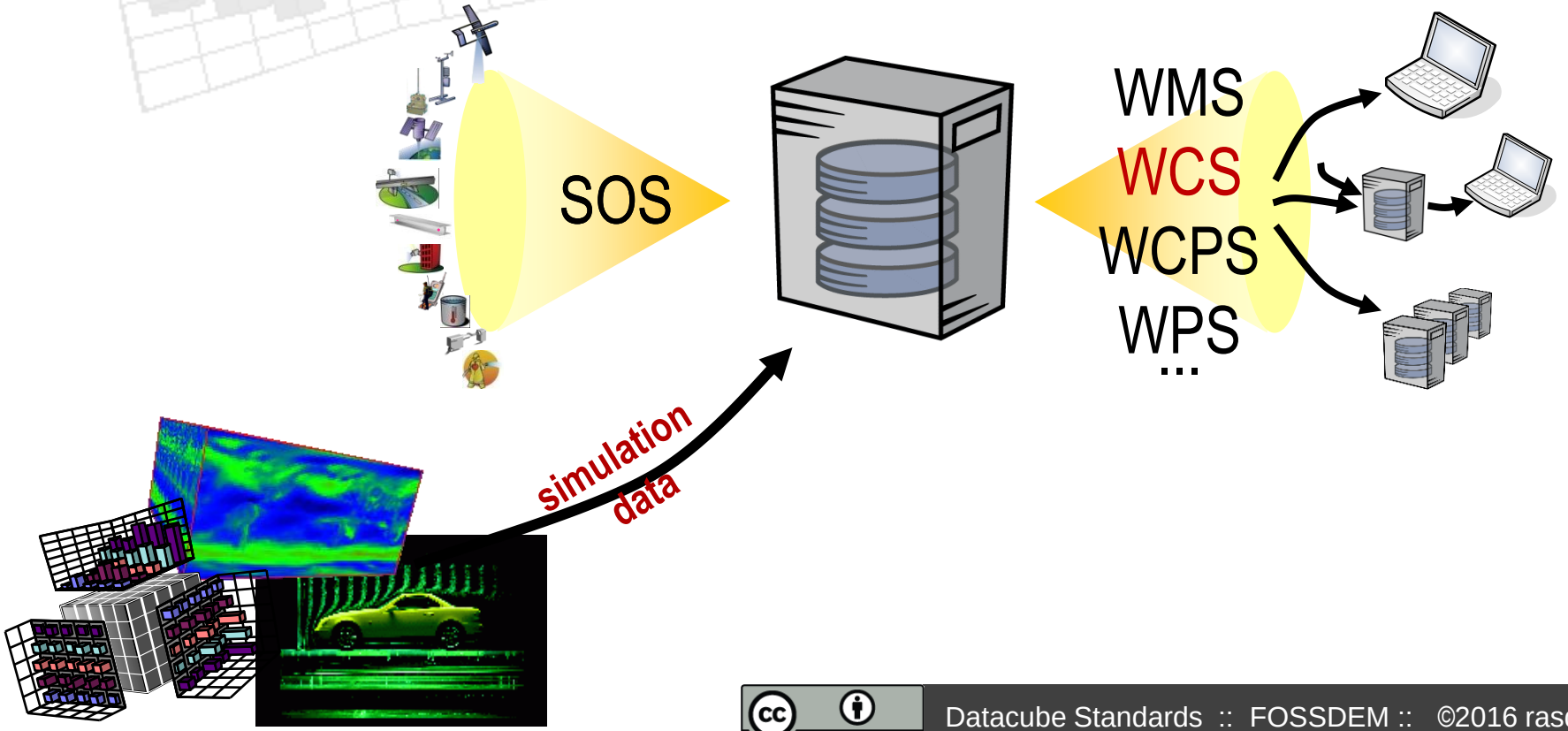
Data Homogenization With OGC Standards



Data Homogenization With OGC Standards

SOS: upstream
sensor data capturing

W*S: downstream
download, processing, visualization



Mini workshop

- Retrieval
  - Using WCS
- Ingestion
  - Using WCSTImport
- Processing
  - Using WCPS
- Display
  - Using WMS

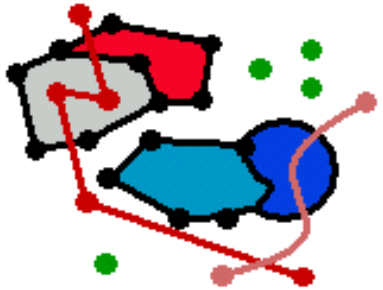
 ©2016 rasdaman

Coverages

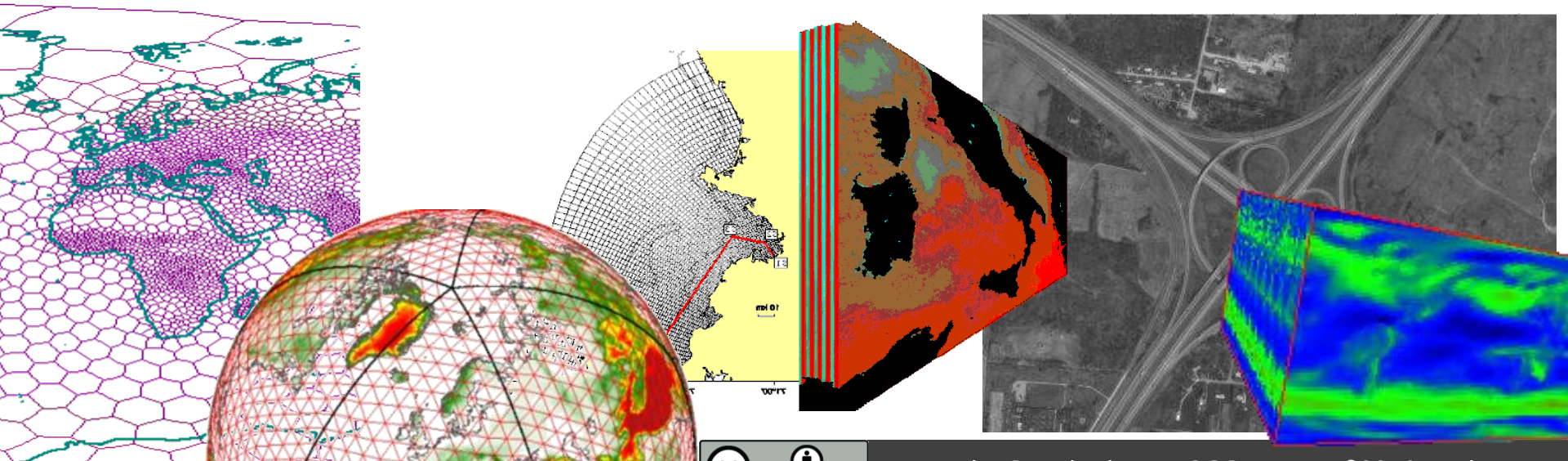
 ©2016 rasdaman

Coverages & WCS in OGC®

Making location count.



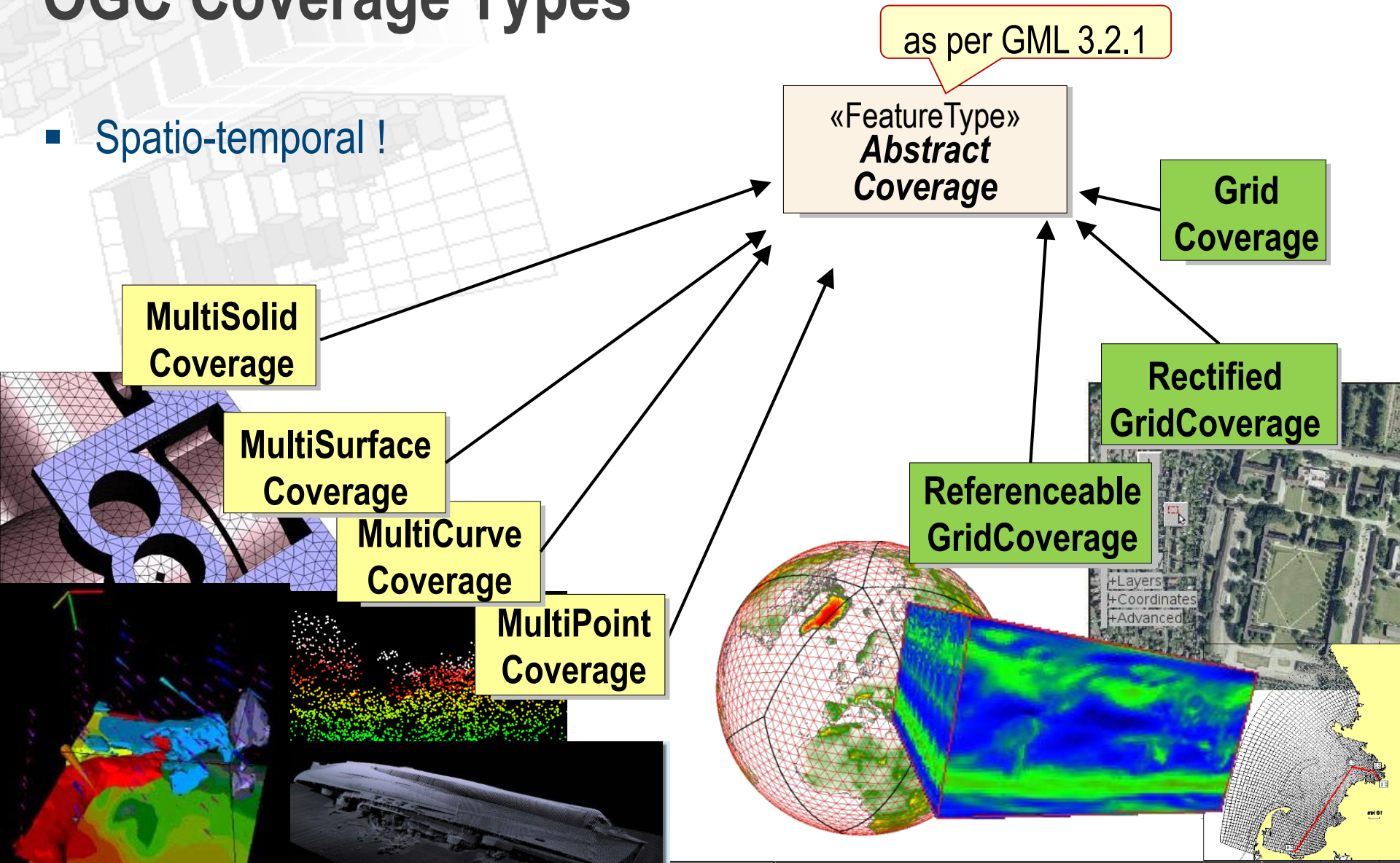
- The basis of all: geographic feature
- coverage: special type of feature
  - regular & irregular grids, point clouds, meshes
- Usually, Big Geo Data are coverages



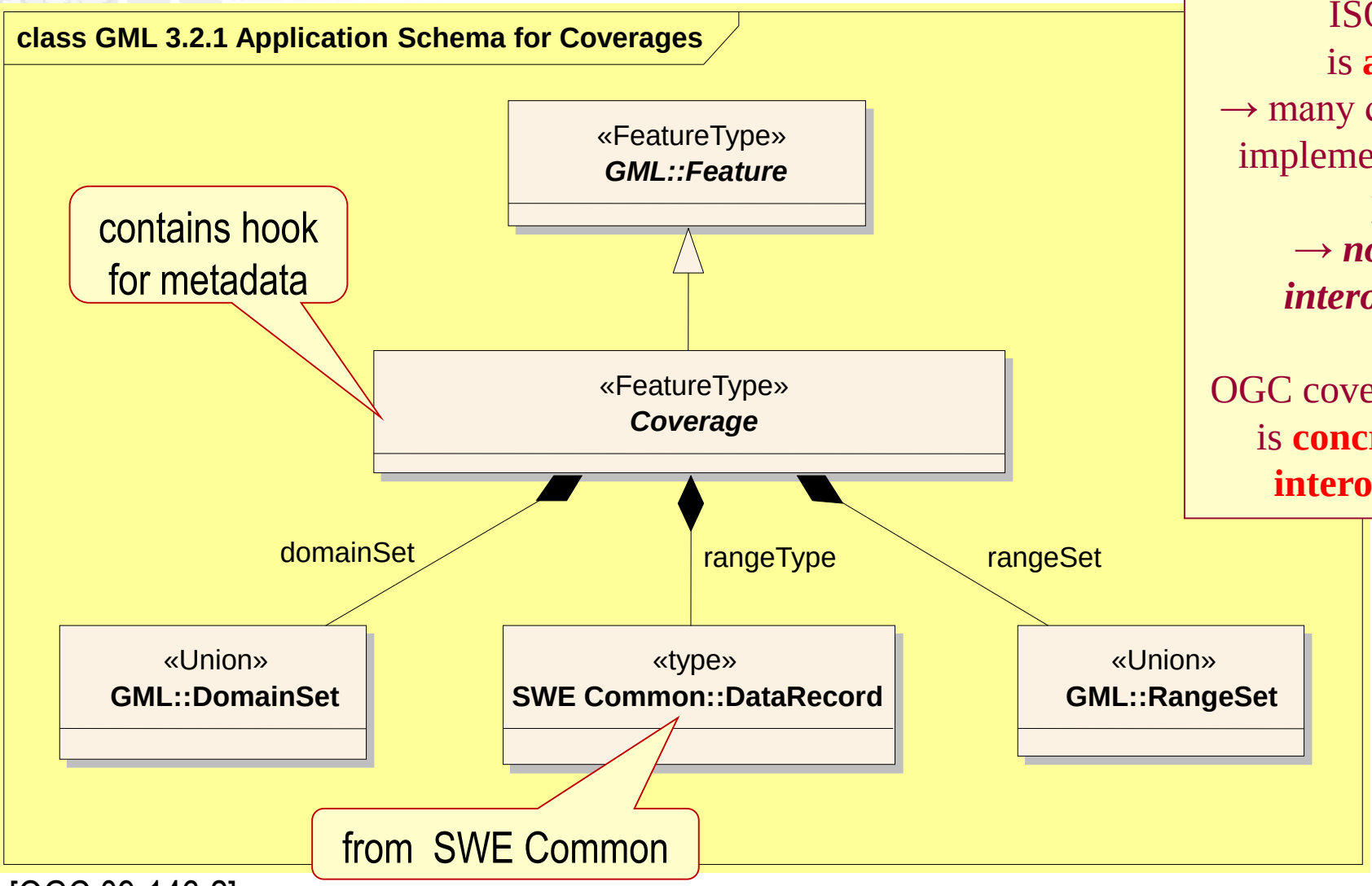
CC BY

OGC Coverage Types

- Spatio-temporal !



Coverage Definition



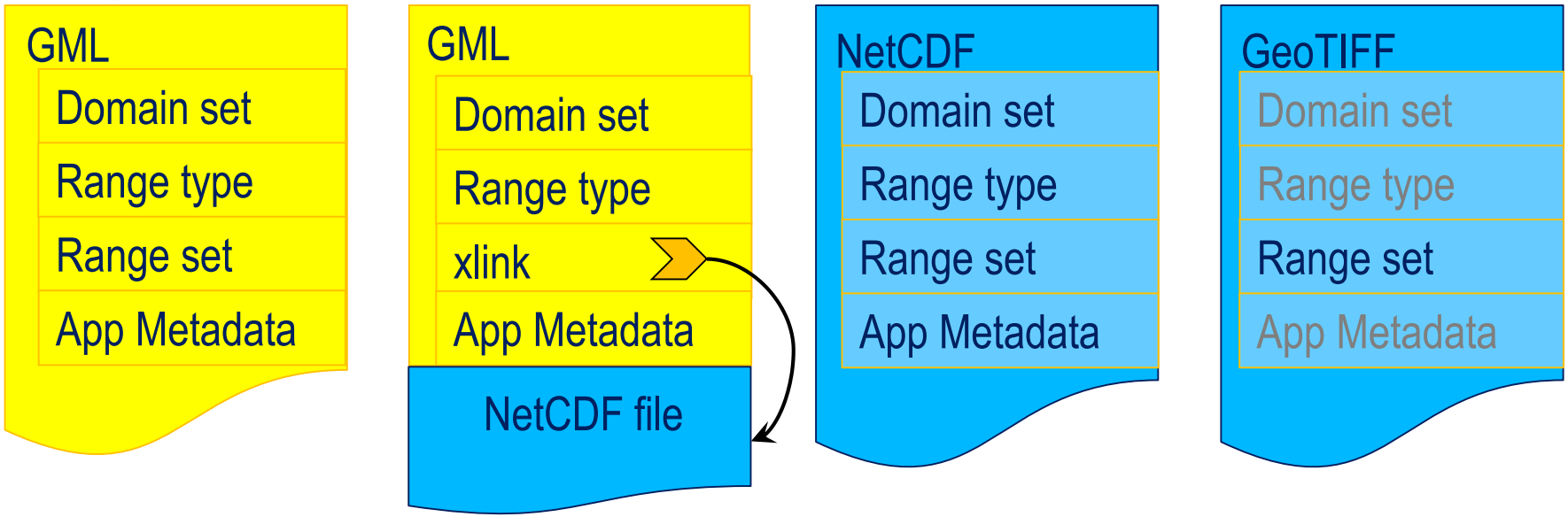
ISO 19123 is **abstract** → many different implementations possible → ***not per se interoperable***

OGC coverage std is **concrete** and **interoperable**

[OGC 09-146r2]

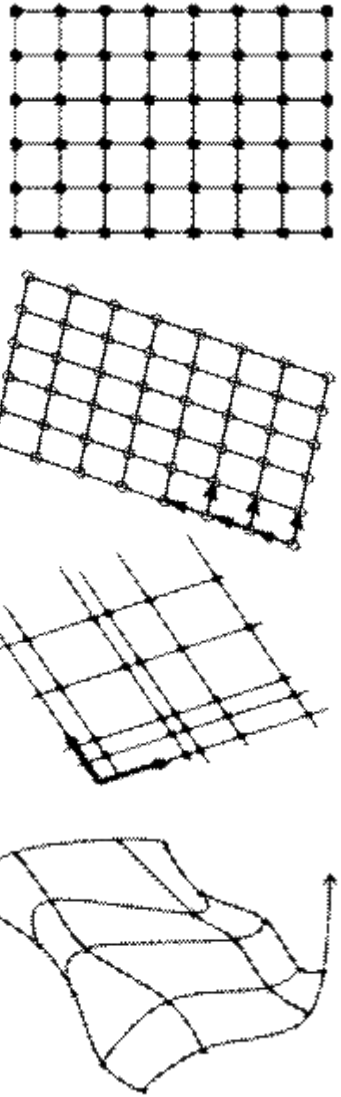
Coverage Encoding

- Any ASCII or binary format
  - GML , JSON (under work), GeoTIFF, NetCDF, …
- Multipart-Mixed: part 1 “header”, part 2 binary “pixels”/”voxels”



Outlook: OGC Coverages v1.1

- Renaming to resolve misunderstandings:
  - „GML 3.2.1 Application Schema – Coverages“ (GMLCOV) → „Coverage Implementation Schema“ (CIS)
  - GMLCOV 1.0 → CIS 1.1, backwards compatible
- Regular & irregular grids
  - Generalizes GML 3.3, SensorML 2.0, etc.
- Coverage carries list of applicable interpolation methods
- Partitioned representation of coverages
  - mosaicking, timeslicing as per WaterML, ...
- Status: RFC
  - Also to become ISO 19123-2

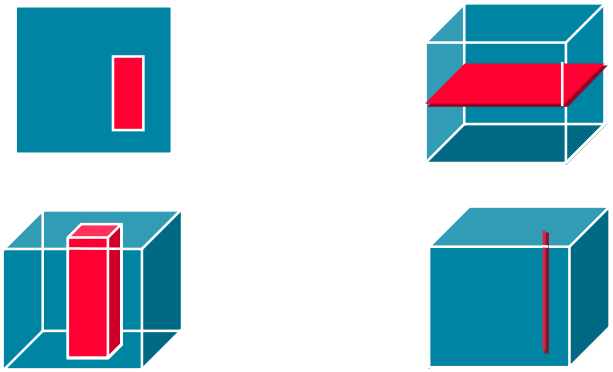


Web Coverage Service (WCS)

 ©2016 rasdaman

Web Coverage Service (WCS)

- Coverages designed to work with WFS, WMS, WCS, WCPS, WPS, SOS, …
  - WCS specialized → most functionality specific to coverages
- WCS Core: access to spatio-temporal coverages & subsets
  - format encoding on the fly
  - subset = trim | slice



- WCS Extensions: optional functionality facets
  - Scaling, CRS transformation, …

Large, growing implementation basis: rasdaman, GDAL, QGIS, OpenLayers, OPeNDAP, MapServer, GeoServer, NASA WorldWind, EOx-Server; Pyxis, ERDAS, ArcGIS, ...

WCS Core *GetCoverage*

- Download a coverage (or a subset thereof), values guaranteed unchanged
- Ex: „*download coverage c001*“

```
http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0
    & REQUEST=GetCoverage & COVERAGEID=c001
```

- Ex: „*coverage c001, lat/long cutout, time slice t=2009-11-06T23:20:52*”

```
http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0
    & REQUEST=GetCoverage & COVERAGEID=c001
    & SUBSET=Long(100,120) & SUBSET=Lat(50,60)
    & SUBSET=time("2009-11-06T23:20:52")
```

- Ex: “*coverage c002, in GeoTIFF*”

```
http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0
    & REQUEST=GetCoverage & COVERAGEID=c002 & FORMAT=“image/tiff“
```

 ©2016 rasdaman

WCS Extension – Range Subsetting [OGC 12-039]

- Extract range components
  - „bands“, „variables“
  - Extension to *GetCoverage* request
- Ex:

```
http://www.acme.com/wcs ? SERVICE=WCS & VERSION=2.0
    & REQUEST=GetCoverage & COVERAGEID=c001
    & RANGESUBSET=red
```

  - or: …& RANGESUBSET=nir,red,green &…
  - or: …& RANGESUBSET=green,red,blue &…
  - or: …& RANGESUBSET=nir:green &…
  - or: …& RANGESUBSET=band01,band03:band05,band19:band21 &…

WCS Extension – Transaction [OGC 13-057]

- = WCS-T: Modify coverage offerings on a server via Web
  - New requests:
    InsertCoverage + *DeleteCoverage* + *UpdateCoverage* (partial replacement)
- Core design goal: *GetCoverage* → *InsertCoverage*
- Ex:

```
http://www.acme.com/wcs
    ? SERVICE=WCS & VERSION = 2.0
    & REQUEST=InsertCoverage
    & COVERAGEREF=http://bcme.com/archive/hurricane.nc
    & USEID=new
```

 ©2016 rasdaman

OGC Web Coverage Processing Service (WCPS)

- = high-level spatio-temporal geo analytics language

[JacobsU, Fraunhofer; data courtesy BGS, ESA]

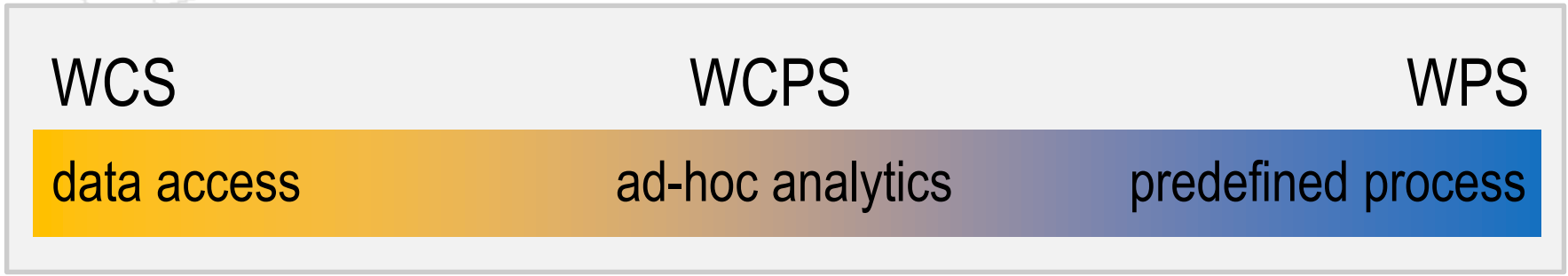
- "From MODIS scenes M1, M2, M3: difference between red & nir, as TIFF"
  - …but only those where nir exceeds 127 somewhere

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

→ $(\text{tiff}_A, \text{tiff}_C)$

Q: W...whatever...S ?

- from data-intensive to processing-intensive services



Coverages & WCS in

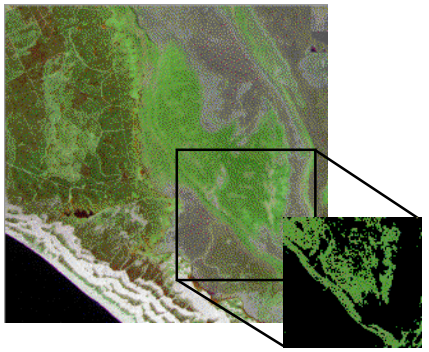
- Coverages in several Annex II & III themes
  - Ortho image, elevation, geology, ...
- INSPIRE Coverage model based on OGC
  - Caveat: modified; eg, using xs:any in domainSet → interoperability issues
  - Discussion started
- INSPIRE WCS under work as “Download Service”
  - first results expected early 2016

Implementation: rasdaman

 ©2016 rasdaman

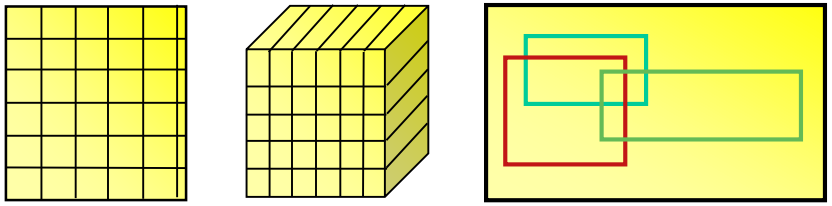
WCS Core Reference Implementation: rasdaman

- „raster data manager“: SQL + n-D arrays
- Scalable parallel “tile streaming” architecture
- Supports R, QGIS, OpenLayers, MapServer, GDAL, EOxServer, Pyxis, ERDAS, ArcGIS, ...
- Blueprint for ISO Array SQL standard

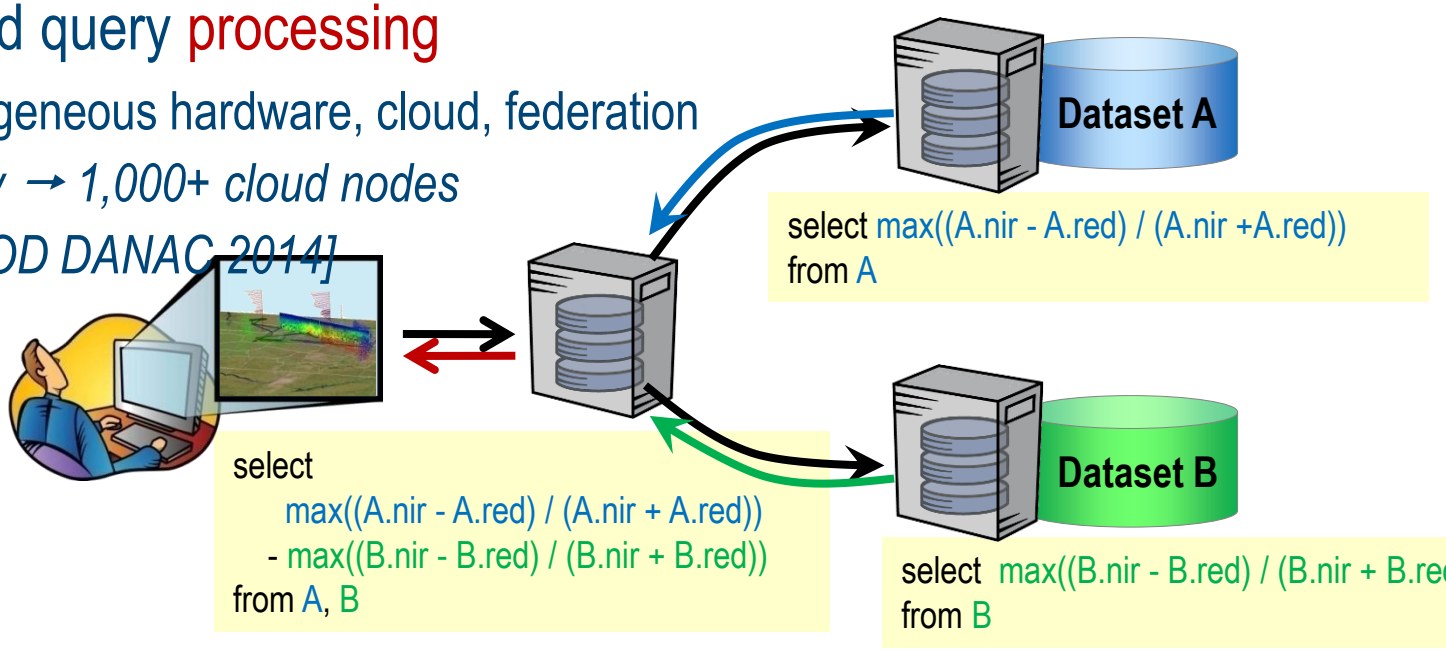


rasdaman Scalability

- Adaptive data partitioning & distribution
  - *130+ TB datacubes*

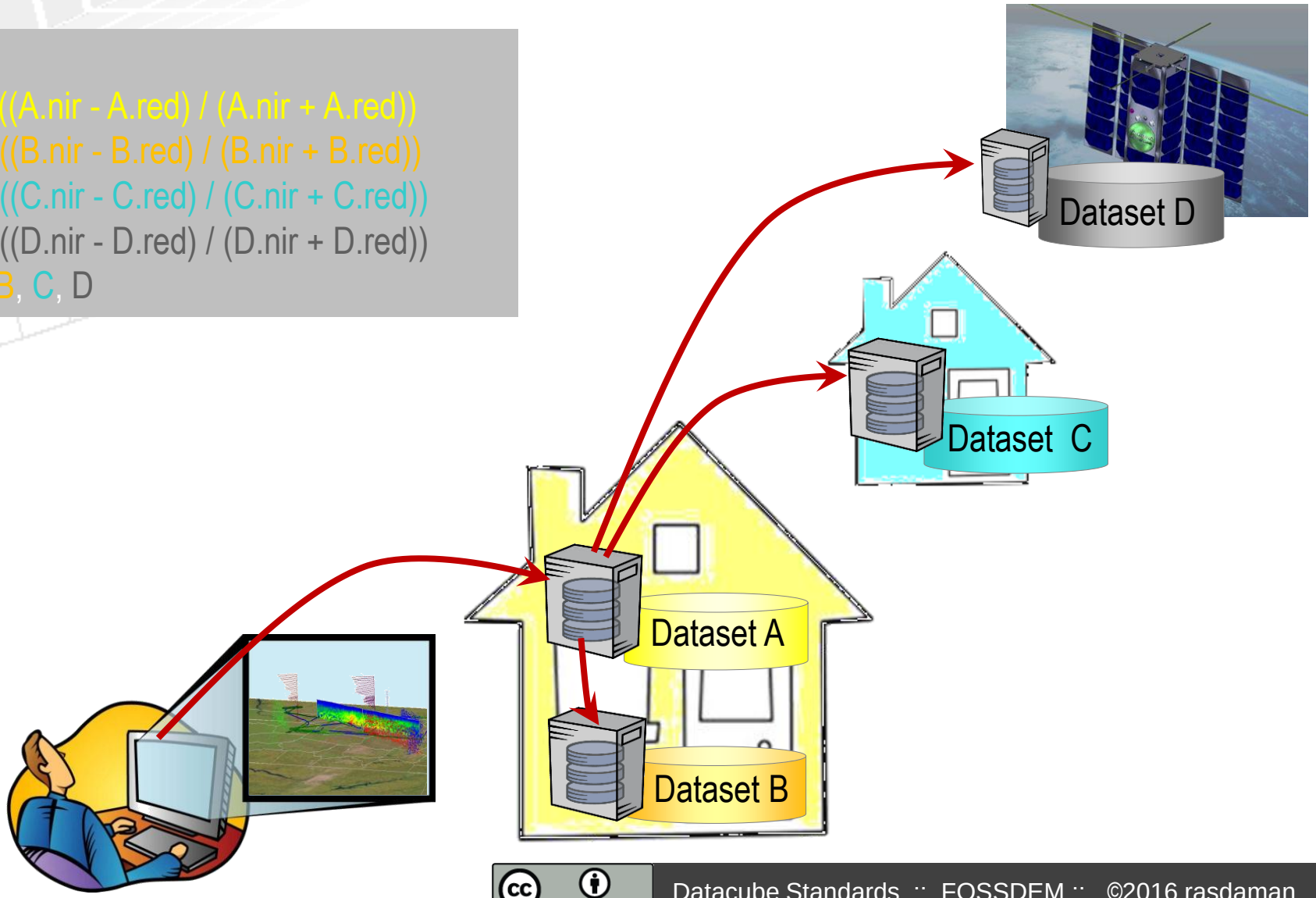


- Distributed query processing
  - Heterogeneous hardware, cloud, federation
  - *1 query → 1,000+ cloud nodes*
  - *[SIGMOD DANAC 2014]*



Parallel / Distributed Query Processing

```
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
```



 ©2016 rasdaman

Parallel / Distributed Query Processing

> Use case: Scientists want to retrieve readily processed data, rather than doing large downloads and programming extraction and analytics with tedious own programming.

> The service: Derivation of the Normalized Difference Vegetation Index NDVI on the fly. The NDVI is a measure for the probability of vegetation in remote sensing: closer to +1 a pixel is, the more likely it is plants. Pull the slider to the desired value. The original image has 6 MB - compare download speed!

Loaded in **3680** ms.

Loaded in **1461** ms.

Big Datacubes: The Big Picture

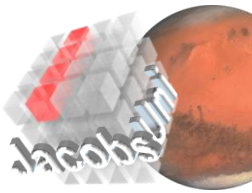
 ©2016 rasdaman

rasdaman raster data management

JACOBS UNIVERSITY

EarthServer: Datacubes at Your Fingertips

- Operational Agile Analytics on 1+ Petabyte space/time datacubes
  - Earth & Planetary Science
- Based on & extending rasdaman
- Intercontinental initiative:
  EU + US + AUS



Phase 1 reviewers:
"proven evidence" that rasdaman will "*significantly transform […] access and use data*" …and *"with no doubt has been shaping the Big Earth Data landscape" …*

Co-funded by the European Union

www.earthserver.eu

CC BY

Conclusion

- OGC coverages as unifying concept for space/time grids & beyond
- OGC WCS: Functionality & flexibility
  - from simple extraction (WCS Core) to complex analytics (WCPS)
- Consensus across OGC, ISO, INSPIRE
- rasdaman:
  mature, scalability-proven, in use

[rasdaman screenshots]