Big Data Cubes in Agriculture

INSPIRE Conference, Antwerp, 2018-sep-18

P. Baumann¹, Katrin Kohler²
¹ rasdaman GmbH, ² Spatial Business Integration GmbH

Supported by German Ministry of Food and Agriculture (BMEL)
Satellite Image Based Biomass Maps

a) winter wheat on 2015-04-24

b) winter oilseed rape on 2016-06-05

c) multi-year Field Potential Map (source: Landwirtschaftliche Betriebsgemeinschaft GbR Groß-Germersleben; includes material © (2018) Planet Labs Germany GmbH, Google Earth, Spatial Business Integration GmbH)
BigPicture Project

- Diagnosis in the field
  - Big-Data-based determination of causes for satellite image derived and site-specific variations
  - Goal: Recommendations for targeted measures
    - fertilizer placement, application of plant protection products, choice of species to grow, etc.
  - Ground truthing: 500 farmers

- rasdaman via OGC WCS & WCPS

- Supported by German Federal Ministry of Food and Agriculture
rasdaman: Agile Array Analytics

- "raster data manager": SQL + n-D arrays
  - Scalable parallel “tile streaming” architecture
  - Spatio-temporal regular & irregular grids

- Mature, in operational use, on OSGeo Live
  - blueprint for Big Datacube standards: ISO, OGC, INSPIRE
  - reference implementation
Parallel, Distributed Processing

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

1 query → 1,000+ cloud nodes
ECMWF: River Discharge
NCI Australia: Landsat8

WCS service endpoint: http://rasdaman.nci.org.au/rasdaman/ows

Available coverages

Footprint of geo-referenced coverages

Coverage: L58_CU_NBART_SE_29S
Coverage: L57_CU_NBART_SE_29S

Coverage Extent: lon_min=-164.7963, lon_max=-141.4494, lat_min=-25.51715, lat_max=-25.51725, lon_max=142.07660

25.57°S 141.53°E  151 m  Eye 10.000 km
Intercontinental Datacube Federation

ECMWF

NCI Australia
Startup Example: EOfarm/GR

- 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
Wrap-Up

- **Datacubes** = analysis-ready spatio-temporal „Big Data“
  - sensor, image (timeseries), simulation, statistics datacubes

- **rasdaman**: European Datacube Engine
  - only if you ask: ODC vs rasdaman

- **emerging**: European Datacube Federation
WCS Core

- Download a coverage (or a subset thereof), values guaranteed unaltered
  
  Ex: „download coverage c001“
  
  ![Image](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“
  
  ![Image](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 c001, in GeoTIFF"
  
  ![Image](http://www.acme.com/wcs?SERVICE=WCS&VERSION=2.0&REQUEST=GetCoverage&COVERAGEID=c001&FORMAT="image/tiff"

http://standards.rasdaman.com
create table LandsatScenes(
    id: integer not null, acquired: date,
    scene: row( band1: integer, ..., band7: integer ) mdarray [ 0:4999,0:4999 ]
)

select id, encode(scene.band1-scene.band2)/(scene.nband1+scene.band2), "image/tiff"
from LandsatScenes
where acquired between "1990-06-01" and "1990-06-30" and
    avg( scene.band3-scene.band4)/(scene.band3+scene.band4)) > 0
Comfort Zone of Well-known Tools

- **Map navigation**: OpenLayers, Leaflet, ...
- **Virtual globe**: NASA WorldWind, Cesium, ...
- **Web GIS**: MapServer, GeoServer, QGIS, ArcGIS, ...
- **Analysis**: GDAL, R, python, ...

...via standards

[screenshots: diverse clients accessing rasdaman]
...But That's Not What You Want to See

- Users should stay in their comfort zone
  - navigation: Web client; ...; analytics: python, R, ...

```python
>>> con = Connection(hostname="127.0.0.1", port=7001)
>>> mr = RasCollection(con, "mr")
>>> mr = mr[100,150] # Array Subsetting
>>> mr += 1
>>> mr = mr ** 2 # Square of all elements
>>> mr = mr. filter (oid=2)
>>> mr.query
<RasQueryObject>
>>> str (mr.query)
"Select exp(mr[100,150]+1,2) from mr where oid(mr) = 2"
>>> arr = mr.eval ()
<RasArrayObject>
>>> arr.to_array () # Default conversion : Numpy Array
[[...],[...],[...]]
```
Agri-Cubes :: INSPIRE Conf :: ©2018 rasdaman

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.earthserv.eu,
  www.planetserv.eu
Classical Archive + Metadata = Datacubes?

- **WaterML 2.0: timeseries = time slices**
  - Good performance on x/y slice extraction
  - Disastrous performance on timeseries analysis

- **OGC Coverages:**
  Implementation can choose efficient layout
  - Same performance in all directions, space & time
OGC Datacube Standards: CIS & WCS

- **Coverage Implementation Schema (CIS):** space/time datacube model

- **Web Coverage Service (WCS):** datacube service model
  - **WCS Core:** access to spatio-temporal coverages & subsets
    - \( \text{subset} = \quad \text{trim} \quad | \quad \text{slice} \)
  - **WCS Extensions:** optional facets
    - **Web Coverage Processing Service (WCPS)** geo datacube query language

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

- Web clients (m2m, browser)
- Internet
- rasserver
- geo services
- database
- file system
- external archives

- Optional compression
- distributed query processing
  No single point of failure
- alternative storage

[SSTD 2013]
Spatio-Temporal Datacubes on Virtual Globes
Big Datacube Standards

- **Open Geospatial Consortium (OGC) datacube standards:**
  - Data model: Coverage Implementation Schema (CIS)
  - Service model: Web Coverage Service (WCS)

- **ISO:**
  - TC211: CIS & WCS
  - SC32: SQL/MDA („Multi-Dimensional Arrays“)

- **EU INSPIRE:**
  - coverages & WCS

- **Emerging: IEEE**