The EarthServer Data Space: Analysis-Ready AI-Cubes for the Green Deal

Peter Baumann
Green Deal

- **European Green Deal:** EU’s plan to become climate-neutral by 2050
  - emissions reduction targets in aviation, cars, etc
  - climate-aware rethinking of land use, forestry, agriculture
  - Emissions Trading System

- **Key question:** how is the climate & environment impacted, what does each proposed measure yield quantitatively?

- **Digital Twin of climate & Earth**
  - Actionable integration of all relevant data – an archetypical Big Data challenge!
Green Deal Data Spaces

- Data Space: agglomeration of data relevant for a particular topic
  - Diverse & time-variant
  - Decentralized & autonomous
  - Connected & federated

- Data spaces typically overlapping with other data spaces
We Need to Better Understand Our Data!
Datacubes?

- Sensors & simulations  
  -> gridded ("raster") data

- natural paradigm  
  for spatio-temporal, n-D data

- Avoid undue complexity  
  -> data + service on high semantic level
Homogenized, Analysis-Ready Datacubes

unified view on n-D raster data, such as
- 1-D sensor data
- 2-D imagery
- 3-D x/y/t image timeseries & x/y/z geophysical data
- 4-D x/y/z/t atmospheric data
rasdaman

= „raster data manager“: actionable n-D datacubes
  • pioneered actionable datacubes; 200+ publications, patents

- Big Datacube Management & Analytics engine
  • full-stack implementation, parallel, federated, secured, standards
  • Scaling: nanosat – laptop – cloud – planetary-scale federation

- ISO SQL/MDA standards blueprint, reference implementation
EarthServer

- datacube provider federation
  - 160+ PB location-transparent data space
  - Open standards, zero-coding

- Open, free, transparent, democratic
  - Open & private; free & commercial
  - Have data offerings? Join!

https://earthserver.eu
AI + Datacubes

- **Goal**: seamless integration of ML in datacube engine
  - Tech: extend OGC WCPS via UDF
    
    ```python
    for $c in (Sentinel_2a), $m in (CropModel)
    return encode( nn.predict( $c[...], $m ), “tiff” )
    ```

- **ML + Natural Language Processing in datacube engine**
  - Based on RSVQA by TU Berlin / Begüm Demir
  - Tech: WCPS UDFs
    - Extra string parameter for question text
    - Text output parsed for further processing in query
    
    ```python
    for $c in (Sentinel-2a), $m in (RsvqaModel)
    return rsvqa.predict( $c[...], “Are there artificial areas and water bodies?”, $m )
    ```
Bring Your Favourite Client

- users in comfort zone of well-known clients
  - Map navigation: OpenLayers, Leaflet, ...
  - Virtual globe: NASA WorldWind, Cesium, ...
  - Web GIS: QGIS, ArcGIS, ...
  - Analysis: GDAL, R, python, ...
Actionable AI-Cubes

- Parametrized WCPS UDFs + annotated model libraries
- Admin can build up own libraries, integrate with STAC, etc
- „huggingface with datacubes“
- In prep
Selected Further Projects

- **Societal Benefit:**
  - AgriCube (Germany / Taiwan): climate & vegetation analytics
  - DeepRain (FZ Juelich): ML for Improved local rain prediction for mountaneous areas
  - DynAWI (JKI): Weather Indicators for Extreme Weather Forecasts in Agriculture
  - LANDSUPPORT (U Napoli): Land Management

- **AI:**
  - AI-Cube (TU Berlin): space/time CBIR
  - ML-Cube (NASA): ML on US/EU federated sat datacubes
  - CENTURION (OPT/NET): AI Knowledge Packs on Datacubes