Interactive Semantic Analysis of Earth Observations in The Cloud
- the example of the Sentinel-2 Semantic Data & Information Cube Austria
Conventional non-semantic queries of EO image archives (e.g., USGS Landsat, ESA Sentinel Data Hub). Search by:

- Metadata information.
  - Geographic geographic area (AOI)
  - acquisition time
  - sensor
  - summary quality indexes (e.g., image-wide cloud cover)
- “thumbnail” image preview (RGB image QuickLook)

- no semantic querying
- no analysis capability
- no extraction of higher-level information products through time
The overarching goal is to build an **Austrian data & information cube – first operational semantic EO data cube worldwide**

Sen2Cube.at project is exemplarily showing that it is possible to

- Conduct **semantic content-based image and information retrieval (SCBIR)** through time in big EO databases and

- allow human users to query and analyse EO data on a higher **semantic level** (i.e. based on at least basic land cover units and encoded ontologies).
Yet another data cube?

Source: ODC
a semantic EO data cube was defined as a spatio-temporal data cube containing EO data, where for each observation at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance.

goes beyond state-of-the-art analysis ready data (ARD) and the use of data cubes solely as data storage by incorporating semantic enrichment (i.e. initial, data-driven information extraction)

we aim for a generic, semantic EO data cube concept driven by automatic integration of optical EO data and automatic generic semantic enrichment in contrast to application-driven solutions (e.g. forest application, crop cycles, specific composites etc.).

generic concept enables diverse queries and analysis possibilities directly within data cubes, including semantic queries for replicable extraction of EO-based indicators /EO-based analysis from big EO data.
Key concept of Sen2Cube.at for spatiotemporal analytics of multi-source EO big data

1. Automatic semantic enrichment: Optical satellite image and associated fully automatic data-derived information layers + additional (open) data like e.g. DEMs

2. Data cube technologies: Data cube system storing images and image derived products query-optimised not acquisition-optimised

3. Web-based inference engine: Semantic content-based queries through time and space in user defined AOIs

Semantic enrichment - multi-spectral colour naming

SIAM (Satellite Image Automatic Mapper) multi-spectral colour naming
- fully automated, physical model based
- parameter-free, no samples needed
- near real-time (ca. 5 min. per Sentinel-2 granule)
- scalable and parallelisable
- multi-sensor support (at least TOA calibrated)

*Expert system*

Automatically generated information layers:
- Multi-spectral categories: 18, 33, 48 and 96 colours
- multi-spectral greenness index
- binary vegetation mask
- 5 category haze mask

96 colours (pre-classes) represented by pseudo-colours and associated semantics.

SIAM spectral categorisation

Sentinel-2 scene
(Austrian/German border)
27 August 2016
SIAM spectral categorisation

96 spectral categories
(Austrian/German border)
27 August 2016
- store data **query-optimised, not acquisition-oriented**
- **different** access methods (API, query language) - data cubes as infrastructure
- provide a **logical view** on the data
  - index external files
  - data as multi-dimensional array
- different solutions are considered in the approach (here: Open Data Cube (ODC) selected as implementation platform)
- Inspire conform coordinate reference system (here: EPSG 3035, LAEA Europe)

Server structure

- Client
- Central API server
- Knowledgebase (Expert/User Queries)
- Factbase (data + information layers)
- Future:
  - Auth Server, User Mgmt
  - Notification Server
- An inference engine for enhanced querying as a Web interface in a client-server solution.
- Create, save and share complex semantic queries/decision rules, suitable for spatiotemporal EO image analytics (increasing the knowledgebase)
- Generic Web-Interface: access to different data cubes possible
- "Plug & Play" for data cubes: shared knowledge base
Dirk Tiede | University of Salzburg | dirk.tiede@sbg.ac.at

Demonstrator

Sentinel-2 granule 33UWP, 100km x 100 km (10 000 x 10 000 pixels per images), 3 different orbits, 07/2015-07/2019

⇒ will be scaled up for whole of Austria until 03/2020

~ 8000 data sets a 100km x 100km and 10m resolution
~ > 80 billion observations per year (!) covering 13 spectral bands
~ 10-20 TB data cube accessible "on the fly" (daily growing, semantic enrichment and data cube ingestion is fully automated)
Sentinel-2 Granule 33UWP, 3 different orbits, 07/2015-07/2019, > 250 images
Services based on the generic approach.....

- a broad application range can benefit from Sen2Cube.at's innovation of a generic data & information cube.

- the following service demonstrations are or will be derived and developed from different use cases
Semantic queries for content-based image retrieval & analysis

- develop knowledge-based semantic queries for SCBIR
- search and select Sentinel-2 scenes (several thousand) based on their content – including "dark" data
Spatio-temporal analysis through time

- spatiotemporal analysis through time analysing the semantic categories through time for any user-defined AOI and/or time span.

- spatiotemporal models can be shared in a knowledge base (here: analysis of water occurrence through time)
User-defined cloud-free mosaics and composites

Mosaics and composites

- apply pre-defined semantic queries through time
- user-defined areas-of-interest and timeframes
- better selection of best-suited pixels (from all available data) on the fly using semantics

Source: Sentinel-2
Location-based access

What was the status of Snow during 01.12.2016 and 05.04.2017 at 48.30 / 14.23?

What was the status of Vegetation during 01.12.2016 and 05.04.2017 at 47.90 / 15.26?

What was the status of Vegetation during 01.12.2016 and 05.04.2017 at 47.29 / 12.88?

What was the status of Snow during 01.12.2016 and 05.04.2017 at 47.90 / 15.26?
Per-parcel statistics

- allow user-defined parcel calculations for spectral and semantic profiles through time (upload of user defined geometry)
- particularly relevant for forestry and agricultural domains

- vegetation trends
- event detection
- when snow was last detected
- ...

Source: Sentinel-2
Thank you for your attention!

http://sen2cube.at/

Dirk Tiede, Josef Strobl & contributors: Martin Sudmanns, Hannah Augustin, Luuk van der Meer, Christian Werner, Andrea Baraldi

Department of Geoinformatics - Z GIS, www.zgis.at
University of Salzburg, Austria