The Austrian semantic Earth observation data cube - Sen2Cube.at
A research idea becomes a national-wide data cube implementation

Martin Sudmanns\textsuperscript{1} & Dirk Tiede\textsuperscript{1}

With contributions from:
Hannah Augustin\textsuperscript{1}, Lucas van der Meer\textsuperscript{1}, Luke McQuade\textsuperscript{1,2}, Andrea Baraldi\textsuperscript{2}

\textsuperscript{1}Department of Geoinformatics - Z_GIS, University of Salzburg, Austria
\textsuperscript{2}Spatial Services GmbH, Salzburg, Austria
Motivation: Big EO Data
Changes in workflows

Big EO Data Analysis Platforms

- **Big EO data analysis platforms**: Often provided by large companies or large research institutions, including collections of PBs of open access data.

- **Ease of use of pre-configured platforms**, with no need to consider difficult data structures, software and hardware requirements.

- **BUT:**
  - The dependency on a specific platform, its data and functionalities, and generally the loss of control may limit the diversity of users, limiting R&D and the sustainability of business models.
  - We argue for the additional possibility of 'local' solutions to use and contribute to big EO data analytics, enable new (different) analyses and product development, otherwise not possible with large-scale systems.

Our approach

Our approach....

▪ Make Earth observation (EO) data easier accessible & allow cloud-based analyses of TBs of data (if possible programming free)

▪ Designing/creating and upscaling our own developed analysis approach for big EO data: semantic EO data cubes

▪ Tackle new research topics that require big data processing e.g., around SDGs

▪ Requirements:
  ▪ Access to computing, storage and big EO data resources and the data management (at least in proof-of-concept size).
  ▪ Flexible/adaptable software
  ▪ Own developments on top

⇒ Possibility to create our own cloud service for big EO data analysis
  
  ... implemented by a small team (5 + students)
Definition of a semantic EO data cube

"A semantic EO data cube or a semantics-enabled EO data cube is a data cube, where for each observation at least one nominal (i.e., categorical) interpretation is available and can be queried in the same instance”

Different data cube software possible; here: Open Data Cube

Temporally stacked EO images (+other geodata), either as view or as physical data structure. Usually coupled with analysis-ready data (ARD). Main goal: Abstracting data storage from users.

We go beyond ARD – starting from semantically enriched data allows programming free access, and an even higher abstraction of data storage from user access.

Key components of a semantic EO data cube

1. Images: All images (here: Sentinel-2), every pixel semantically enriched (fully automated, no training samples) + additional (open) datasets (e.g. DEM)

2. Data cube technology: User-defined areas-of-interests and time intervals

3. Web-based inference engine: High-level semantic querying

Tiede, Dirk; Baraldi, Andrea; Sudmanns, Martin; Belgiu, Mariana; Lang, Stefan (2017): Architecture and prototypical implementation of a semantic querying system for big Earth observation image bases. In European journal of remote sensing 50(1), pp. 452–463. DOI: 10.1080/22797254.2017.1357432
Semantic enrichment

Every (optical) EO image has a semantic skin/layer:

- Reflectance values are associated to **spectral categories**
- **Transferrable** between sensors
- Time series of spectral categories are the **basis for on-demand semantic queries**
Semantic enrichment

SIAM (Satellite Image Automatic Mapper) "multi-spectral colour naming"

- Fully automated, based on a physical model
- No parameter, no training-samples
- Near real-time (< 5 min. for a Sentinel-2 granule)
- Scalable, parallelisable in Docker containers
- Multi-sensor support (at least TOA calibration)

SIAM spectral categorization
2 Data cube technology

- Spectral signature (reflectance)
- Categorical variables
- Continuous variables
- Additional geographic data (e.g. altitude)
High-level semantic querying

past

today

Weak / young vegetation

Strong vegetation

Cloud

Strong vegetation

Bare soil
Input:
Optical EO images, at least top-of-atmosphere calibrated

Method:
Hybrid bottom-up semantic enrichment + top-down semantic querying approach

Output:
Maps, time series, ....

Worldwide applicable semantic enrichment in different granularities
Application-agnostic, generic factbase (EO)
Web-based graphical inference engine and semantic querying language
Shared knowledgebase containing semantic
High-level semantic querying

- An **inference engine for semantic querying** as a Web interface in a client-server solution.
- Different and multiple output types are possible and depending on the query.
- **Generic Web interface**: access to different semantic EO data cubes possible
- Create, save and share semantic queries in a **knowledgebase**
- **Open Source** code: https://github.com/zgis/semantique


Department of Geoinformatics - Z_GIS (University of Salzburg)
High-level semantic querying

- Bridging **image domain** and **semantic domain**
- Conceptually, the two domains are separated from each other
- Implementation is Python-based and **data-cube-software agnostic** (our data cube class: OpenDataCube)
Example: How green is Austria?

National-wide information-layers

Percent of vegetation observations between March and September 2020 (without clouds) for entire Austria

Demo (Recorded)
- **scalable, container-based architecture**
- **Horizontal scaling**
- **Concurrent multi-user access**
- **A layout** describes the semantic EO data cube’s content
- **multiple clients supported through an API**
Clients and interfaces

The semantic EO data cube architecture allows:

- Executing semantic models in the cloud via standard JSON API **on-demand in the cloud**
- Earth observation data can be easily integrated in existing software/clients

---

**Command-line interface**

```bash
$ (sem2cli.exe) sem2cli774e6d80b20f --help
Usage: sem2cli [OPTIONS] COMMAND [ARGS]...

Options:
    --log_file FILE Write log to this file instead of stderr.
    -v, --verbose Verbosity log output. Can be added up to three times for even more verbosity (WARNING, INFO, DEBUG).
    --help Show this message and exit.

Commands:
    inference Display / create / modify inferences
    session Session related commands like `login`
    version Prints program version
```

**Mobile app**

**Full ArcGIS Pro Integration**

**Nature’s Calendar citizen science app**
• EO data cubes can have a significant local impact
• Adding own datasets
• Extending as midstream technology
• Make use of local knowledge and integrate communities
• Main challenge: Interoperability and transferability

Can we have a local impact?
Dr Martin Sudmanns 
Assoc. Prof. Dr Dirk Tiede

Paris Lodron University Salzburg | Department of Geoinformatics – Z_GIS
E-mail: martin.sudmanns@plus.ac.at | dirk.tiede@plus.ac.at
E-mail: info@sen2cube.at
Web: https://sen2cube.at

The projects (Sen2Cube.at, SIMS, SemantiX) associated to this presentation were funded under the Austrian Space Applications Programme (ASAP). An OCRE EO grant is supporting the EO cloud & data infrastructure used.