

Semantic Earth Observation Data Cube as New Access to AVHRR Data for Land Ice and Snow Cover Monitoring

Hannah Augustin[1], Martin Sudmanns[1], Dirk Tiede[1], Stefan Wunderle[2], Christoph Neuhaus[2], Andrea Baraldi[3]
[1]Department of Geoinformatics – Z_GIS, University of Salzburg, AUT
[2]Institute of Geography, University of Bern, SUI
[3]Italian Space Agency (ASI), ITA

Contact: hannah.augustin@sbg.ac.at

Context

- Implement a technology increasing in other EO domains to AVHRR
- Improve access to existing archives of data and derived information
  - Compared to purely file-based approaches, an EO data cube is a new way of organising EO data
- Project proposal – currently no dedicated funding
  - Currently only implemented for one month of data for testing (May 2015)
  - Please share any challenges you face dealing with archives, access and storage!

Yet another data cube?



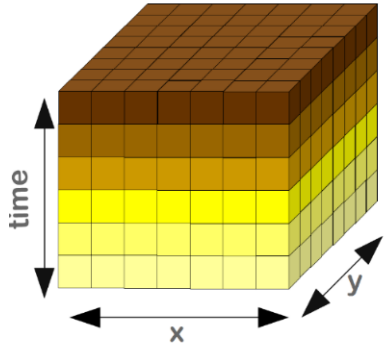
Source: OpenDataCube Initiative

What is a data cube?

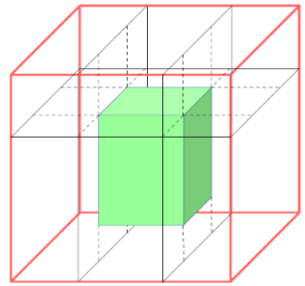
- Currently no standardised definition…
  - a technology/tool for **regular gridded data**
  - a **multi-dimensional** structure with at least one non-spatial dimension
  - a **software** that translates the geographic and temporal positions to the internal file structure, collects the images, and presents them in a way that is convenient, showing the pixels at their geographic position
  - …
- allow researchers to interact with data using given **dimensions** rather than files (e.g. spatial, temporal)

Geospatial Data Cubes

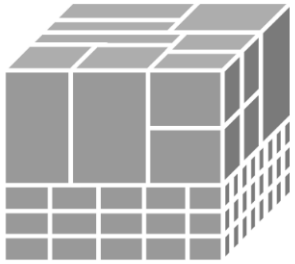
- store data **query-optimised, not acquisition-optimised**
- different access **methods (API, query language)** with data cubes as infrastructure
- provide a **logical view** on the data
  - index external files
  - data as multi-dimensional array
- different solutions exist (e.g. rasdaman, OpenDataCube)



1.)
Voxels arranged as dense temporal stack



2.)
Data access using declarative query language



3.)
Regular or irregular partitioning (tiling)

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.

What is a semantic EO data cube?

- Sensory data are not information and have no inherent meaning without some form of interpretation.

- Concept:
  - „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”

- allow researchers to interact with data using given dimensions **and/or interpreted image content** rather than files (e.g. spatial, temporal)

Augustin, H., Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2019). Semantic Earth observation data cubes. *Data*, *4*(3), 102.

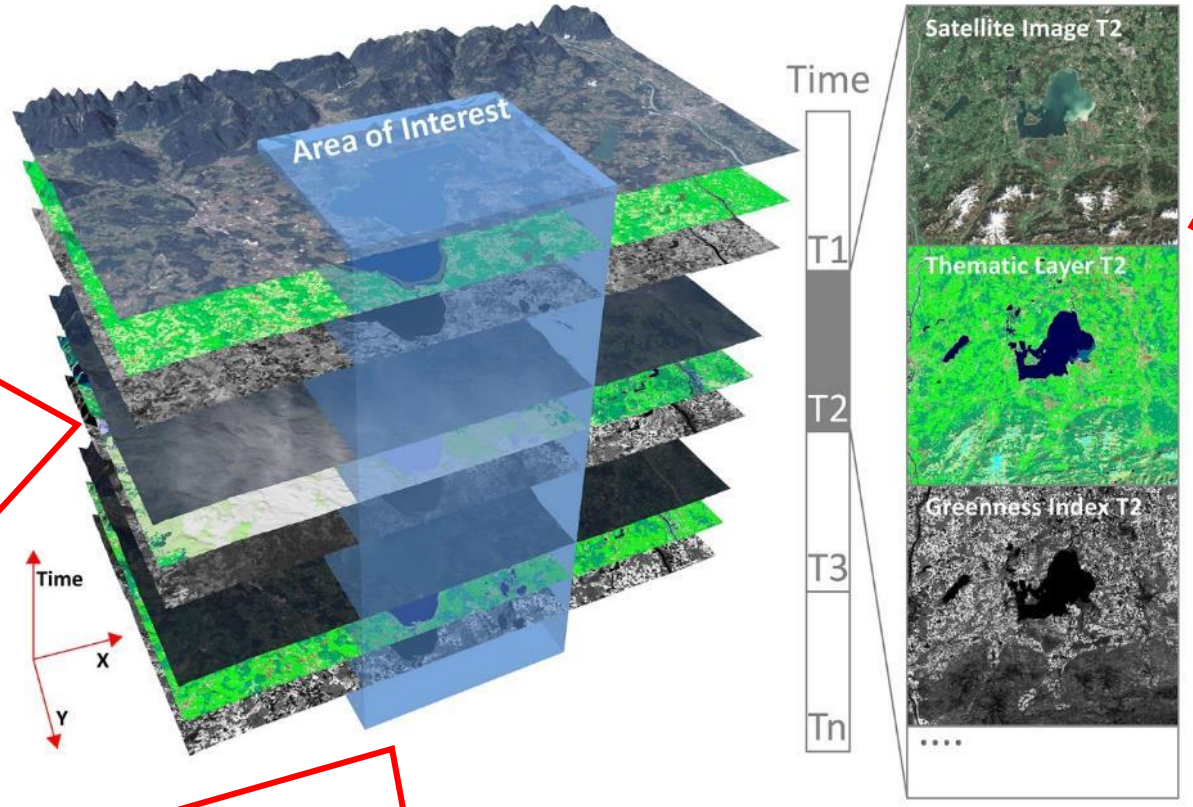
What is a semantic EO data cube?

- Semantic EO data cubes…
  - *…enable semantic analyses that can be incorporated into simple rule-sets in domain language*
  - *…enable users to develop increasingly expressive, comprehensive rule-sets and queries*
  - *…are better positioned to handle current and upcoming big EO data challenges (e.g. sensor agnostic semantic categories)*

Key concept of a semantic EO data cube for spatiotemporal analytics of multi-source EO big data

2

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



1

Automatic semantic enrichment: *fully automatic data*-derived information layers + additional (open) data like e.g. DEMs, probability cloud masks ...

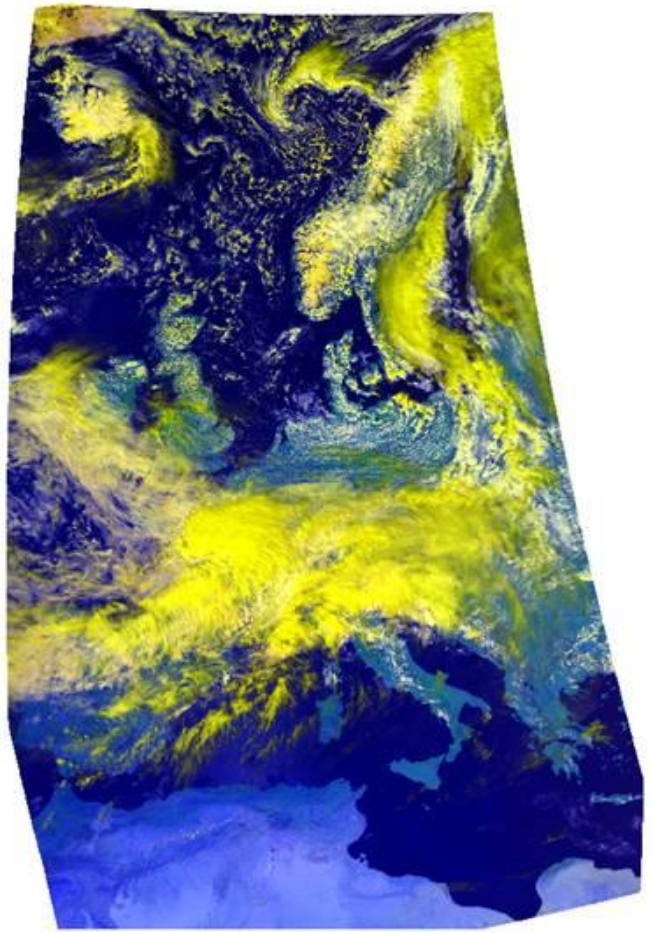
3

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

from Tiede, D., Baraldi, A., Sudmanns, M., Belgiu, M., Lang, S., 2017. Architecture and Prototypical Implementation of a Semantic Querying System for Big Earth Observation Image Bases. Eur. J. Remote Sens. 50, 452–463. doi:10.1080/22797254.2017.1357432

Example queries

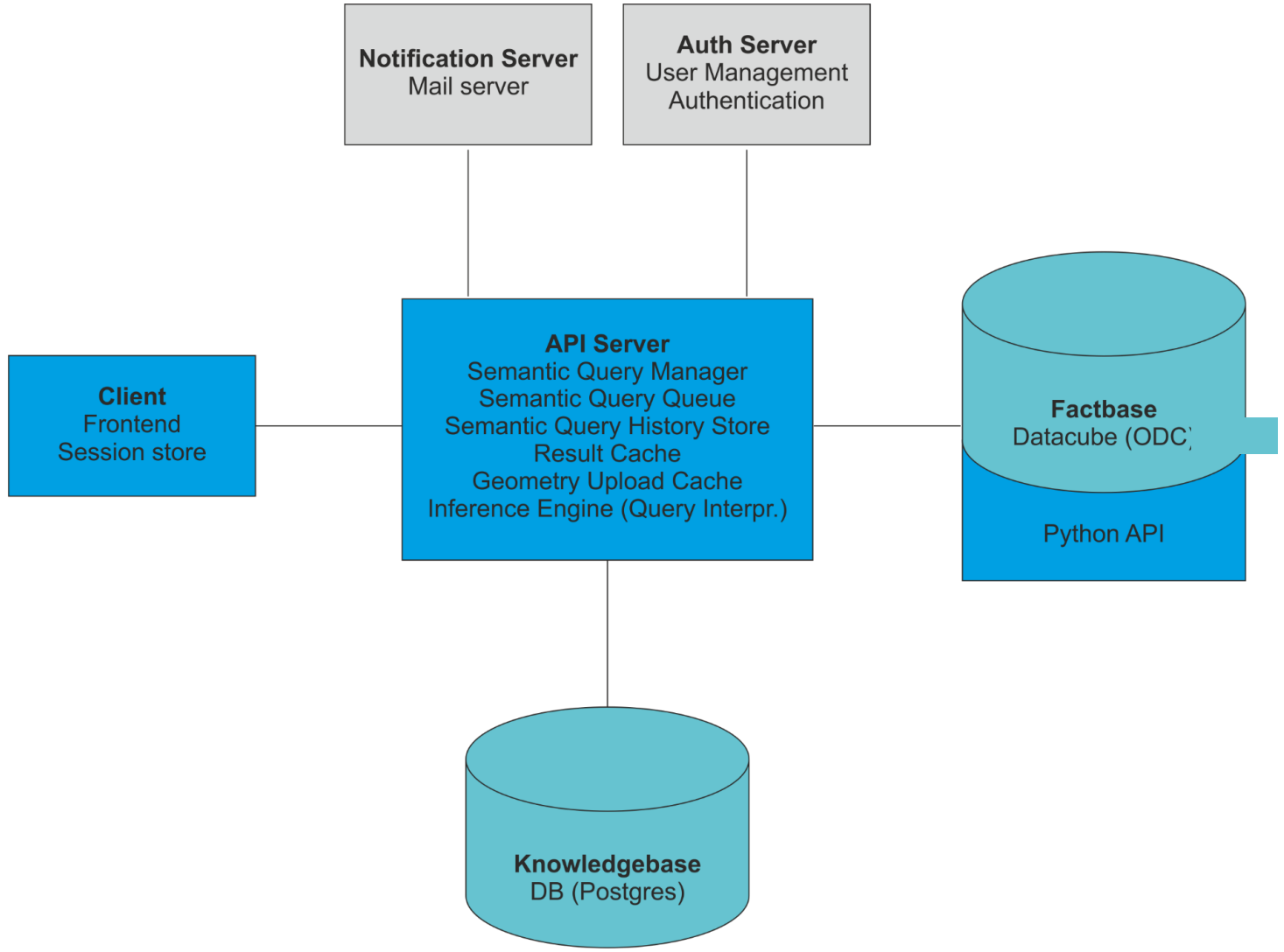
- For multiple study areas…
  - Access all available data with <5% cloud cover for those areas
  - Access all available data with more than 90% snow cover, excluding clouds
    - Returns timestamps as well
  - Access the first observation each year where over 10% snow cover was observed
  - …



Web-based inference engine

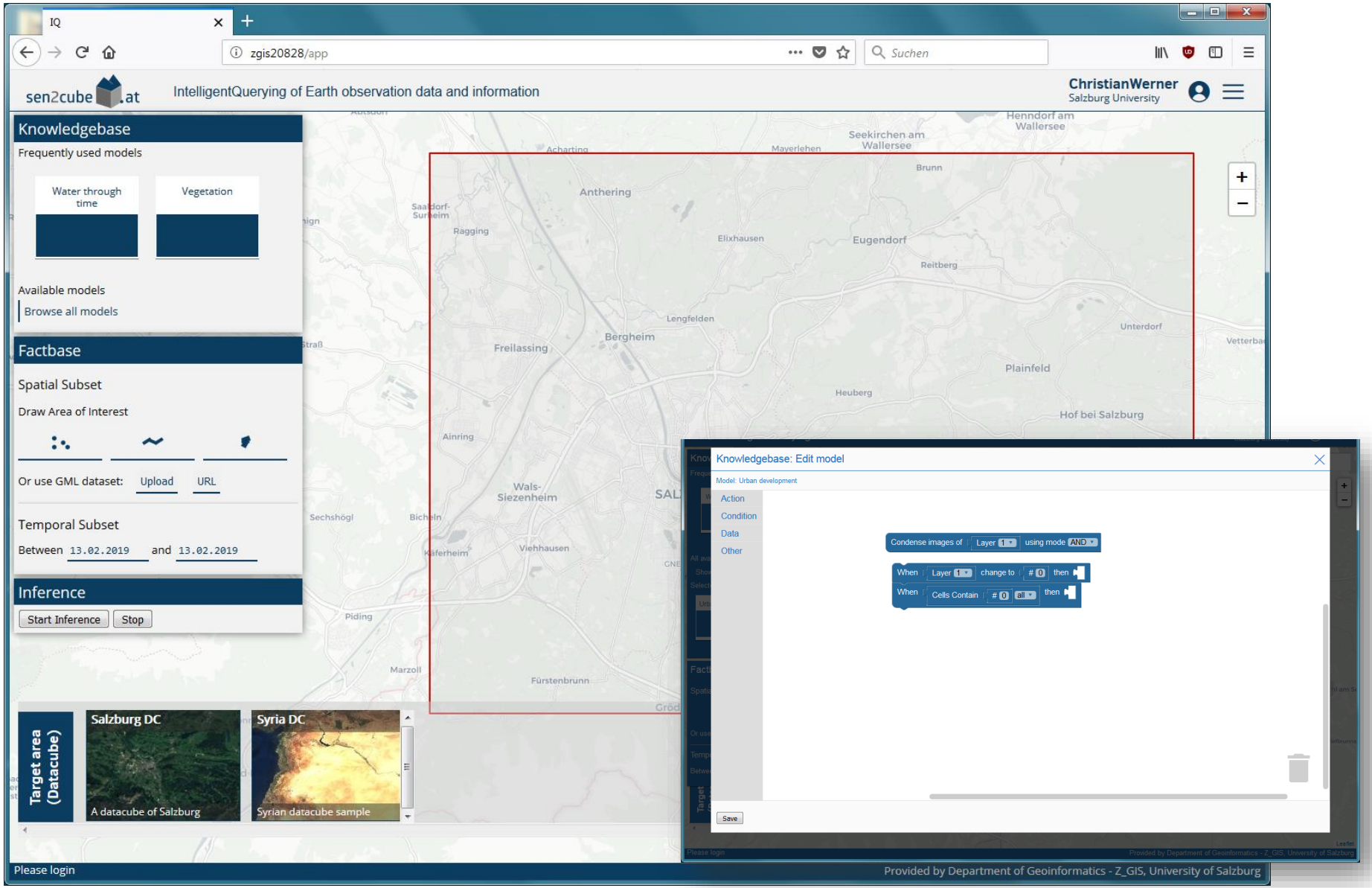
Server structure

- Client
- Central API server
- Knowledgebase (Expert/User Queries)
- Factbase (data + information layers)



Web-based inference engine

Generic Web-Interface (access to different data cubes possible)



Apply to AVHRR data and derived information

- Entire AVHRR data archive calibrated to TOA
- Derived probability cloud masks from University of Bern
- Automatically generated colour categories (e.g. SIAM)
- Combine with existing open datasets (e.g. DEM)
- … ability to document and reproduce analyses given an existing cube and code

Access benefits

- Jupyter notebook access for coding in various languages
  - Julia, python, R …
- Ability to host:
  - locally
  - on a portable SSD
  - on a private server
  - in or linked to various cloud services (e.g. DIAS, AWS buckets)
- Can be implemented well in an educational setting
  - Spin up docker containers such that each participant gets their own mini-data cube, or has access to the same one but runs analysis locally

Connect

In [1]:
```
%matplotlib inline
import datacube
```

In [4]:
```
dc = datacube.Datacube()
dc.list_products()
```

Out[4]:

id	name	description	label	format	time	product_type	creation_time	platform	lat	instrume
2	avhrr3_noaa19_3035	AVHRR3 data from the NOAA19 satellite mission ...	None	GeoTiff	None	AVHRR3	None	NOAA19	None	AVHRI
1	avhrr3_noaa19_masks	Collection of information layers (IL) and mask...	None	GeoTiff	None	INFORMATIONMASKS	None	AVHRR3_IL	None	AVHRR3_

Show measurement details

```
In [4]: dc.list_measurements()
```

Out[4]:

product	measurement	name	dtype	units	nodata	aliases
avhrr3_noaa19_3035	B01	B01	uint16	RTOA [x 1000]	65535	[band_01, B01, Band1]
	B02	B02	uint16	RTOA [x 1000]	65535	[band_02, B02, Band2]
	B03	B03	uint16	BT [K x 10]	65535	[band_3B, B03, Band3B]
	B04	B04	uint16	BT [K x 10]	65535	[band_04, B04, Band4]
	B05	B05	uint16	BT [K x 10]	65535	[band_05, B05, Band5]
	B06	B06	uint16	[x 1000]	65535	[band_3BREF, B06, Band6]
avhrr3_noaa19_masks	pcmmask	pcmmask	uint16	cloud_category	0	[pcm_mask, cloudmask]
	pcmprob	pcmprob	uint16	percent [0-100, 100-200, 200-300]	0	[pcm_probability, cloudprob]

Python xarray object

```
In [5]: study_area = dc.load(
            product = 'avhrr3_noaa19_3035',
            x = (3000000, 4000000),
            y = (4000000, 3000000),
            crs = 'EPSG:3035')

        print(study_area)
```

```
<xarray.Dataset>
Dimensions:  (time: 45, x: 1000, y: 1000)
Coordinates:
  * time     (time) datetime64[ns] 2015-05-01T13:07:00 ... 2015-05-31T14:14:00
  * y        (y) float64 4e+06 3.998e+06 3.998e+06 ... 3.002e+06 3.000e+06
  * x        (x) float64 3.000e+06 3.002e+06 3.002e+06 ... 3.998e+06 4e+06
Data variables:
    B01      (time, y, x) uint16 82 102 120 127 115 121 ... 773 772 772 774 776
    B02      (time, y, x) uint16 56 75 92 99 86 91 ... 648 644 640 637 638 639
    B03      (time, y, x) uint16 2865 2869 2876 2881 ... 2935 2941 2943 2944
    B04      (time, y, x) uint16 2793 2793 2788 2777 ... 2592 2599 2602 2606
    B05      (time, y, x) uint16 2783 2784 2780 2770 ... 2584 2592 2596 2601
    B06      (time, y, x) uint16 51 54 61 71 67 69 ... 159 164 170 174 175 175
Attributes:
    crs:      EPSG:3035
```

List available images

```
In [9]: datasets = dc.find_datasets(product='avhrr3_noaa19_3035')
        for i,image in enumerate(datasets):
            print ("Image #{}".format(i))
            print (image)
```

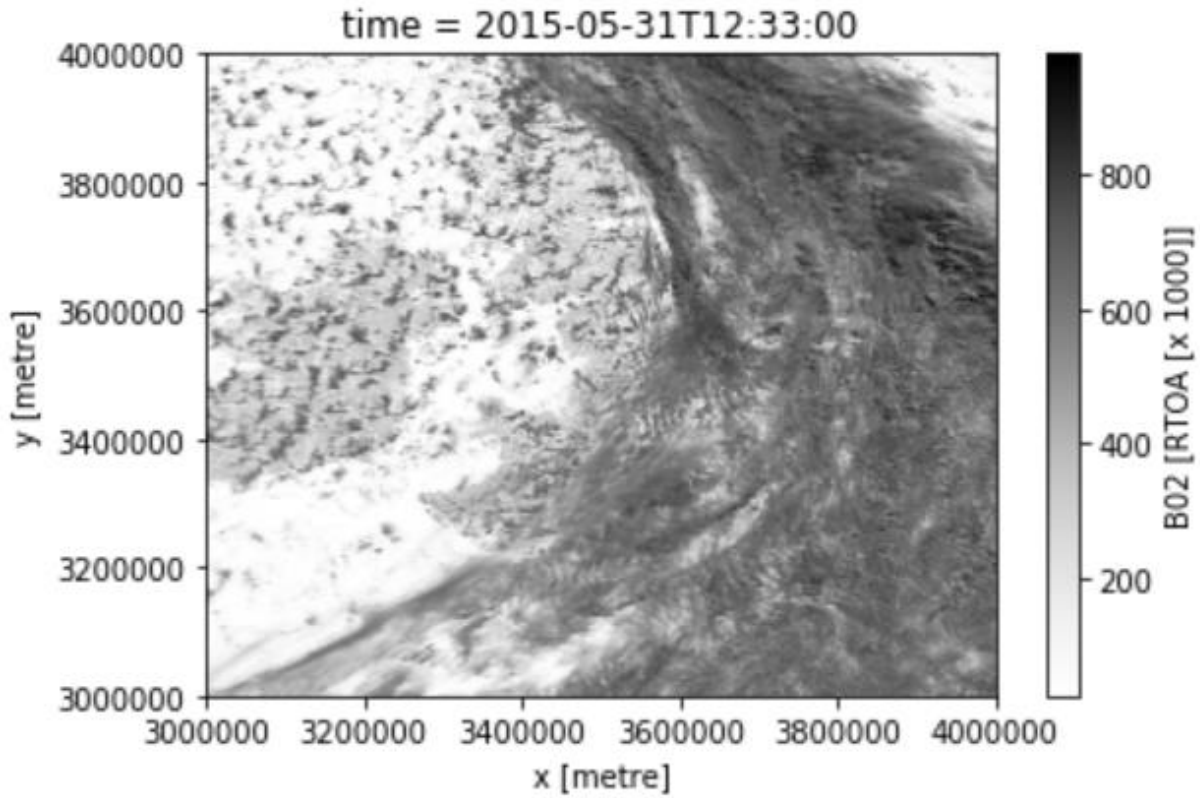
```
roducts/images/avhrr3/noaa19/20150502_1256_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #2
Dataset <id=29bc773b-28b2-4ab0-a523-2bb48282f3d7 type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150503_1245_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #3
Dataset <id=c2b785fa-603c-49ad-8c86-d294098fc3e5 type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150504_1234_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #4
Dataset <id=5ce7555a-9591-4cec-bac8-a605877366ff type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150505_1222_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #5
Dataset <id=b0d4b075-06fb-4916-9951-fc95133307d4 type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150505_1403_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #6
Dataset <id=c2d64846-edd8-40ec-aa2b-d558e331938f type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150506_1211_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
Image #7
Dataset <id=4f1ea994-2f3c-493f-9b50-af5511020f2f type=avhrr3_noaa19_3035 location=file:///p
roducts/images/avhrr3/noaa19/20150506_1352_avhrr3_noaa19_v03/PROC_DATA_3035/metadata.yaml>
```

```
In [7]: test = study_area.B01.loc['2015-5-04':'2015-5-10']
        print(test.shape)

        (11, 1000, 1000)
```

```
In [41]: study_area.B02.isel(time = 43).plot(cmap = "Greys")

Out[41]: <matplotlib.collections.QuadMesh at 0x7f04a04c3198>
```



AVHRR Data & Information Cube

- goes **beyond state-of-the-art 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.

SIAM: multi-spectral colour naming

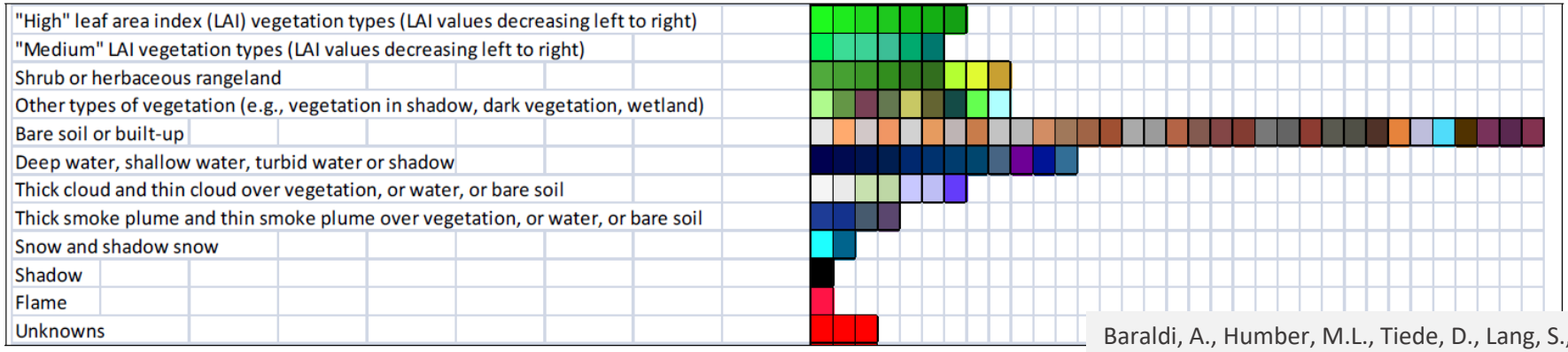
SIAM (Satellite Image Automatic Mapper) multi-spectral colour naming

- fully automated, physical model based
- parameter-free, no samples needed
- near real-time (ca. 24 seconds. per AVHRR image of Europe)
- scalable and parallelisable
- multi-sensor support (at least TOA calibrated)

➔ **Expert system - GOFAI**

Automatically generated information layers:

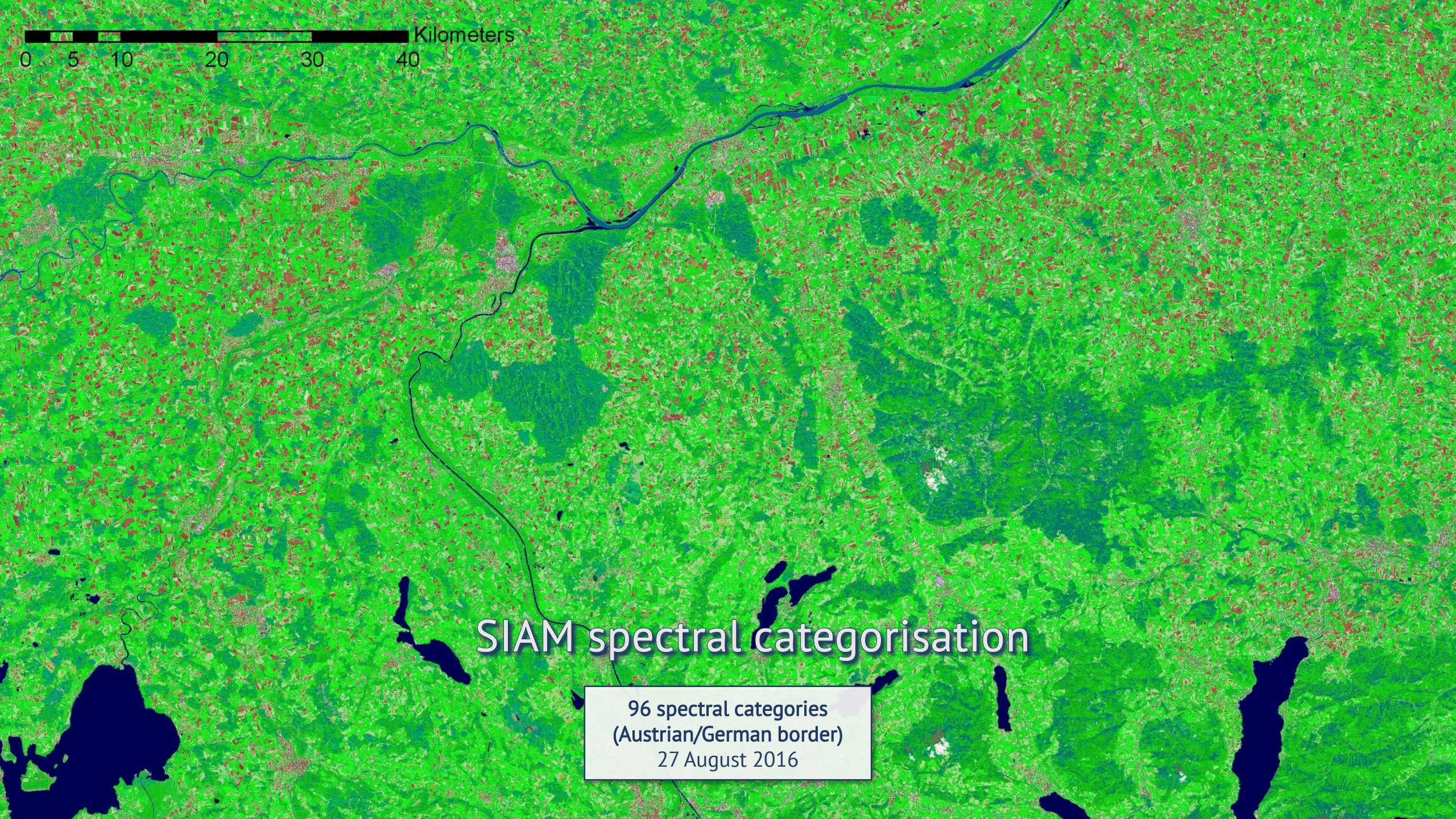
- **Multi-spectral categories**: 17, 33, 43 and 83 colours
- multi-spectral **greenness index**
- binary **vegetation mask**
- …

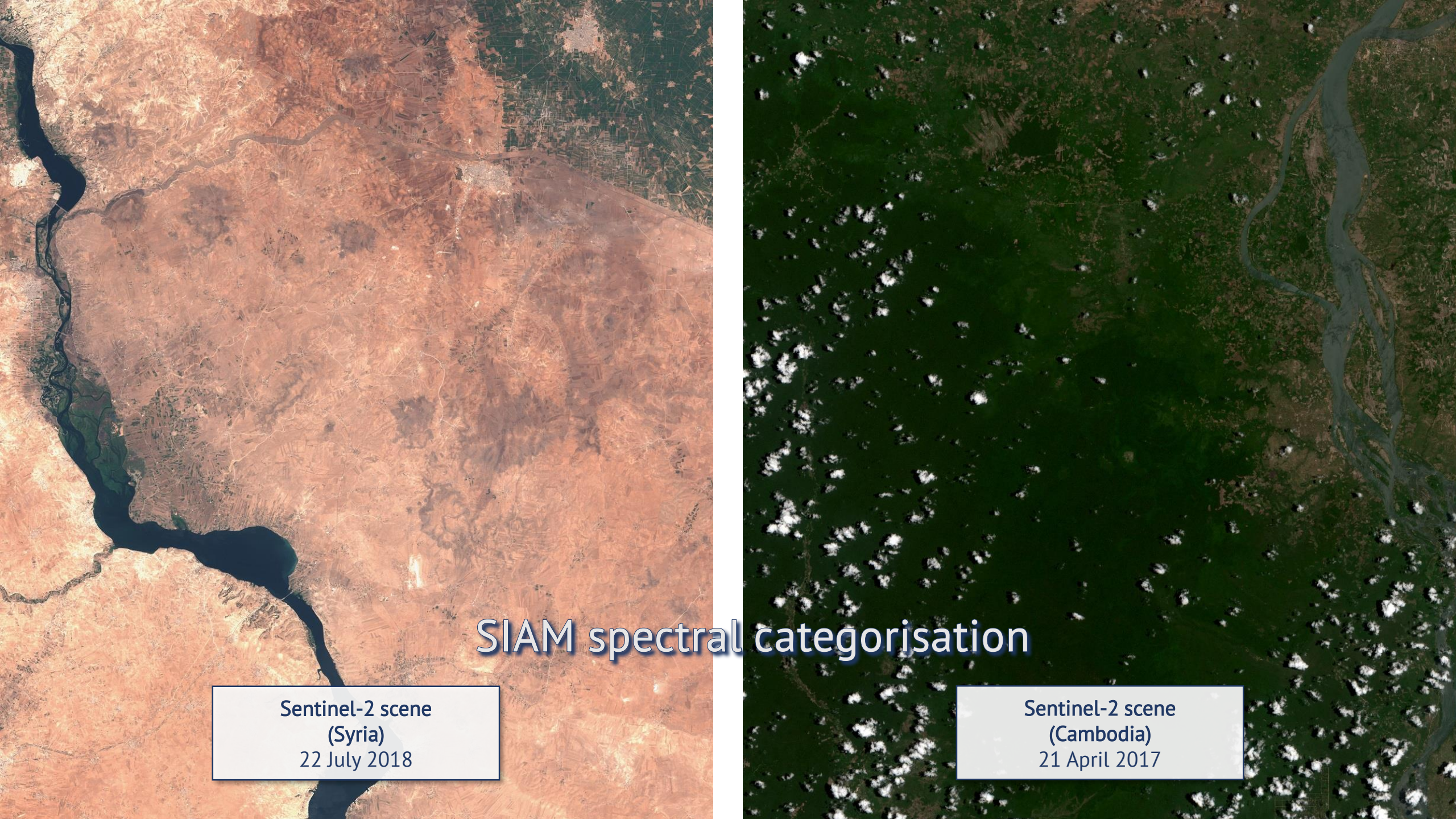


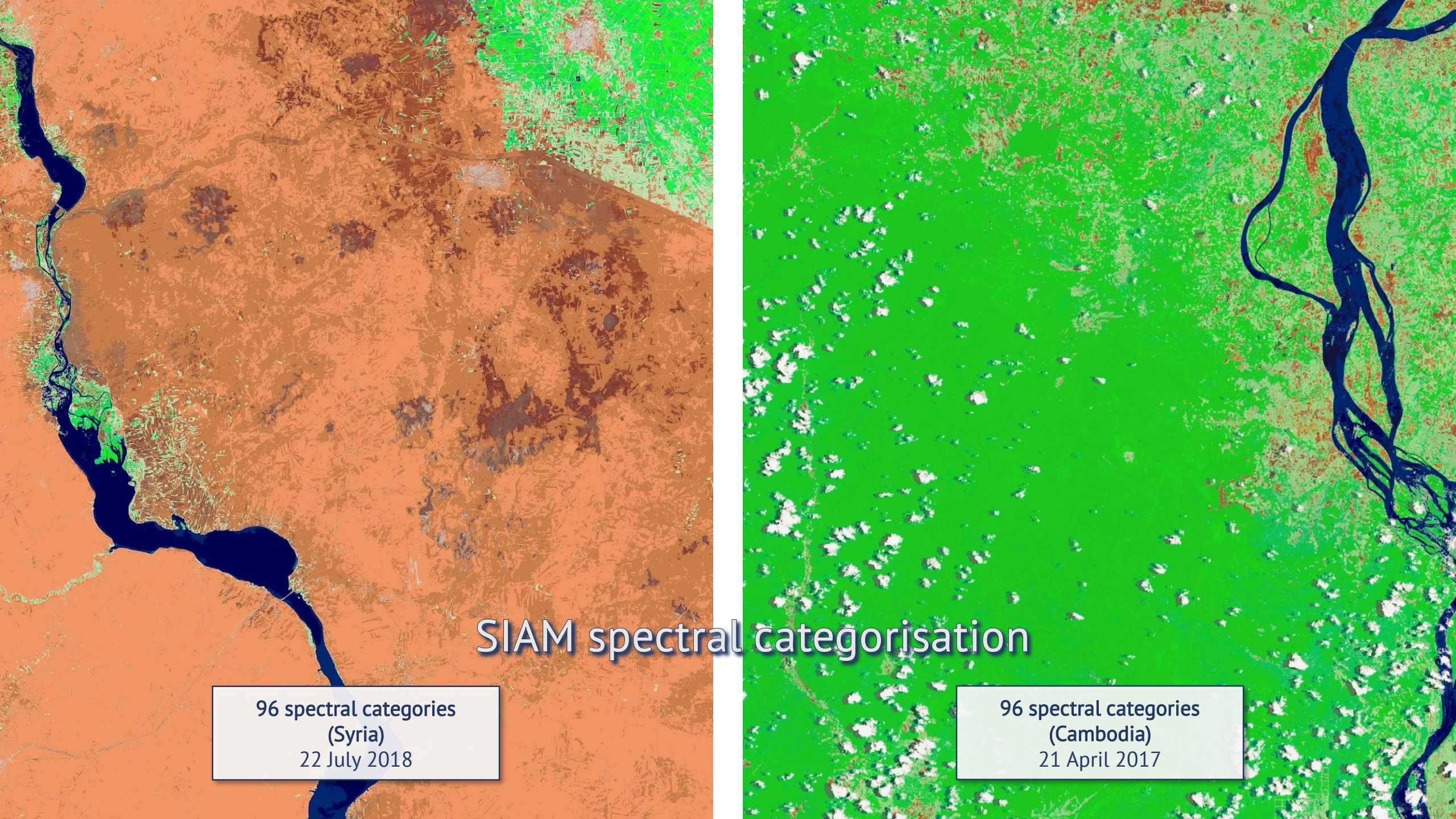
96 colours (pre-classes) represented by pseudo-colours and associated semantics for Landsat-like images.

Baraldi, A., Humber, M.L., Tiede, D., Lang, S., 2018. GEO-CEOS stage 4 validation of the Satellite Image Automatic Mapper lightweight computer program for ESA Earth observation level 2 product generation – Part 2: Validation. Cogent Geosci. 4, 1–52. https://doi.org/10.1080/23312041.2018.1467254


Kilometers
0 5 10 20 30 40
SIAM spectral categorisation
Sentinel-2 scene
(Austrian/German border)
27 August 2016

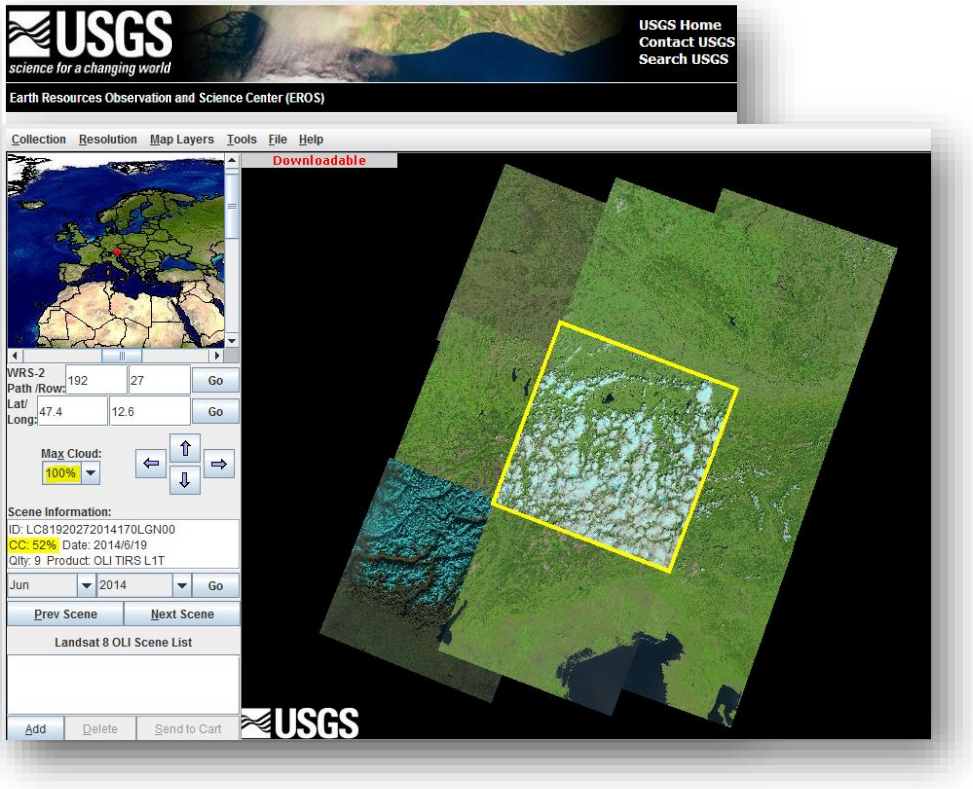

Kilometers
0 5 10 20 30 40
SIAM spectral categorisation
96 spectral categories
(Austrian/German border)
27 August 2016


SIAM spectral categorisation
Sentinel-2 scene
(Syria)
22 July 2018
Sentinel-2 scene
(Cambodia)
21 April 2017

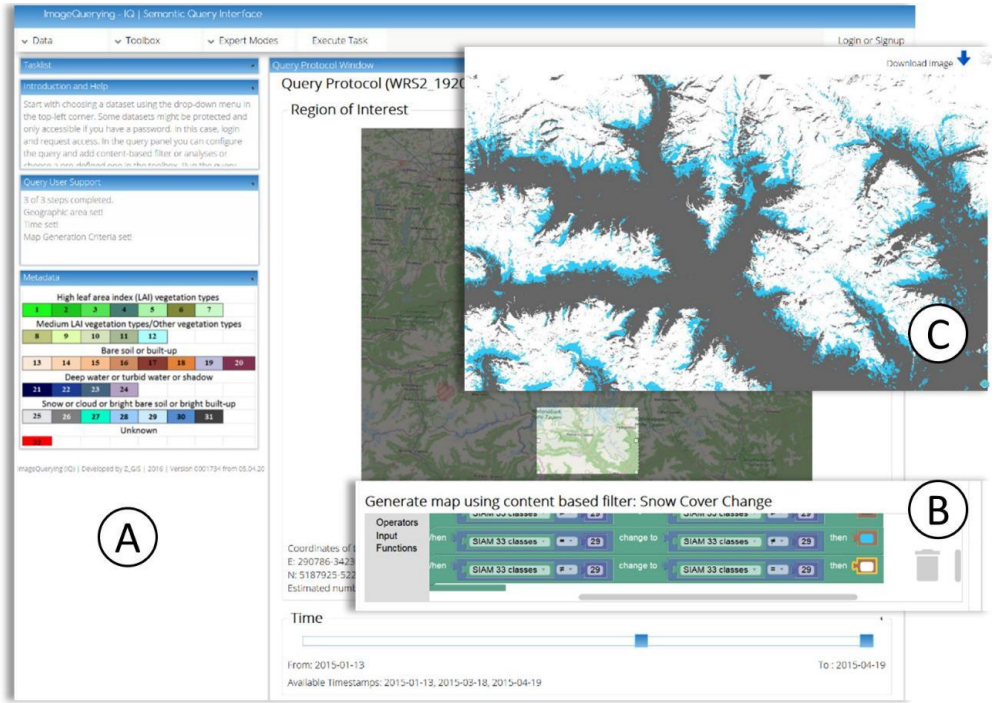

SIAM spectral categorisation
96 spectral categories
(Syria)
22 July 2018
96 spectral categories
(Cambodia)
21 April 2017

Semantic queries for content-based image retrieval

- develop knowledge-based semantic queries for SCBIR
- search and select image data (several thousand) based on their content – including "dark" data



An inference engine for enhanced querying will be programmed as a Web interface in a client-server solution.



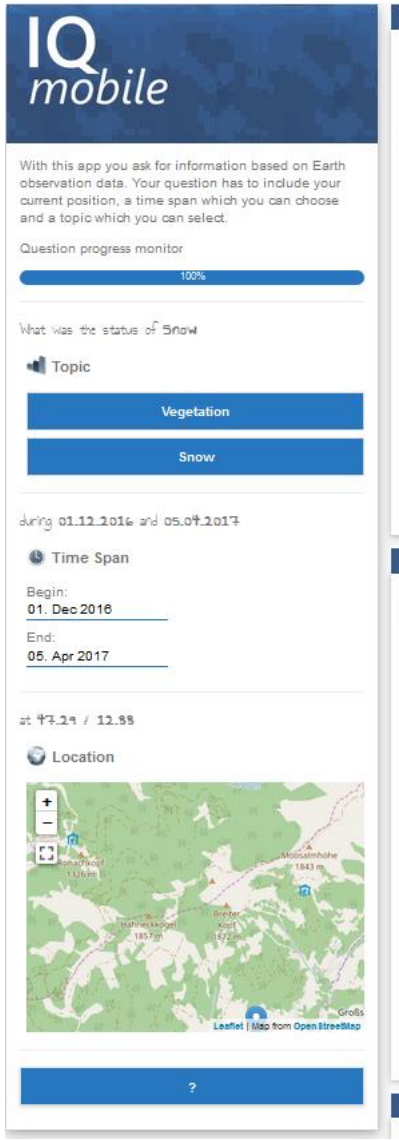
User-defined composites

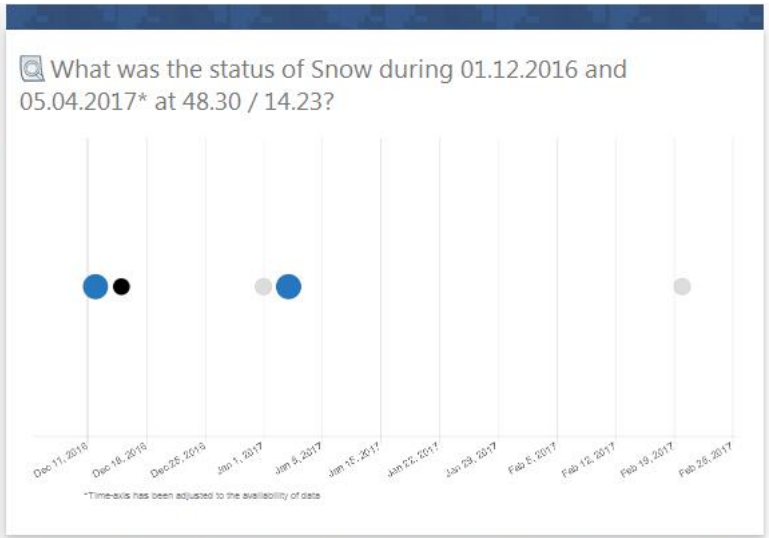
- apply pre-defined semantic queries through time
- user-defined areas-of-interest and timeframes
- better selection of best-suited pixels (from all available data) on the fly using semantics

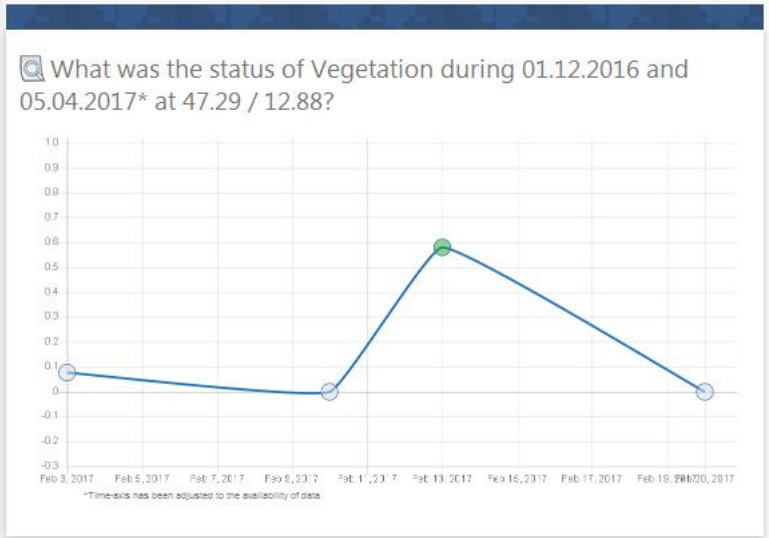


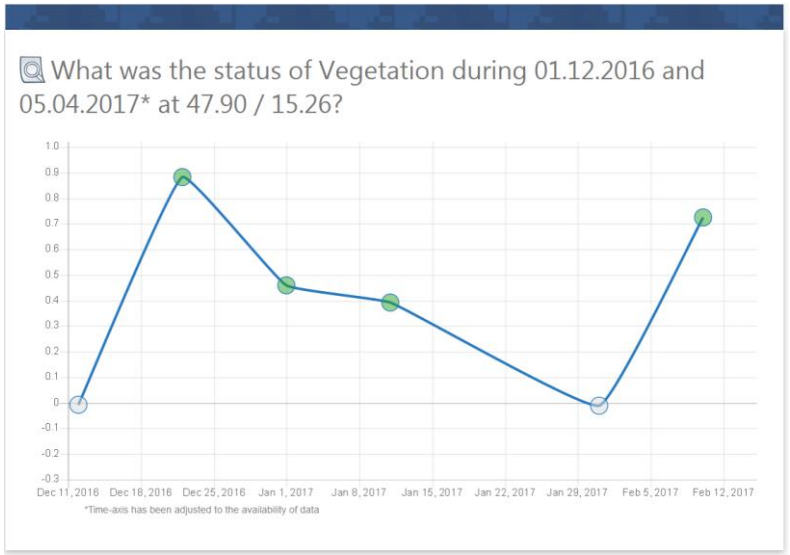
Source: Sentinel-2

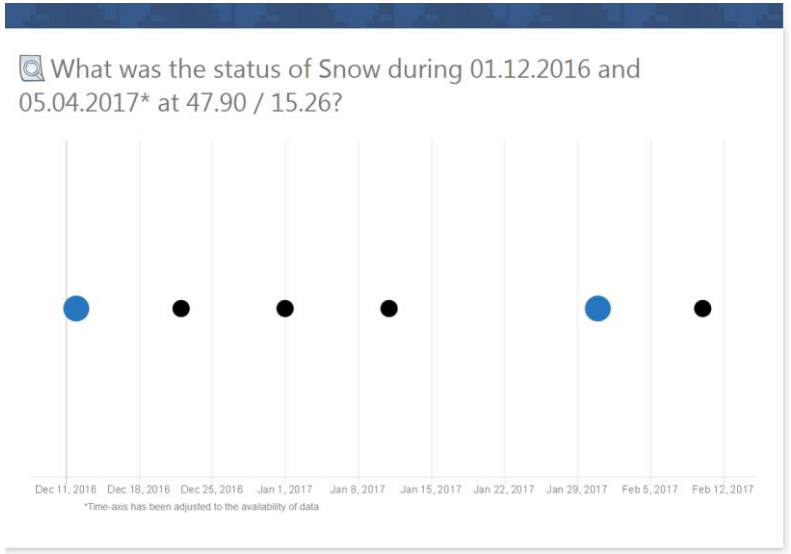
location-based access











Thank you for your attention!

http://sen2cube.at/

Hannah Augustin
hannah.augustin@sbg.ac.at
PhD Researcher
Department of Geoinformatics - Z GIS, www.zgis.at
University of Salzburg, Austria

Sen2Cube.at is a project funded under the Austrian Space Applications Programme (ASAP 14)