SWEDISH SPACE DATA LAB

# Experience from using the Swedish Space Data Lab

**György Kovács**
Luleå University of Technology

# Overview

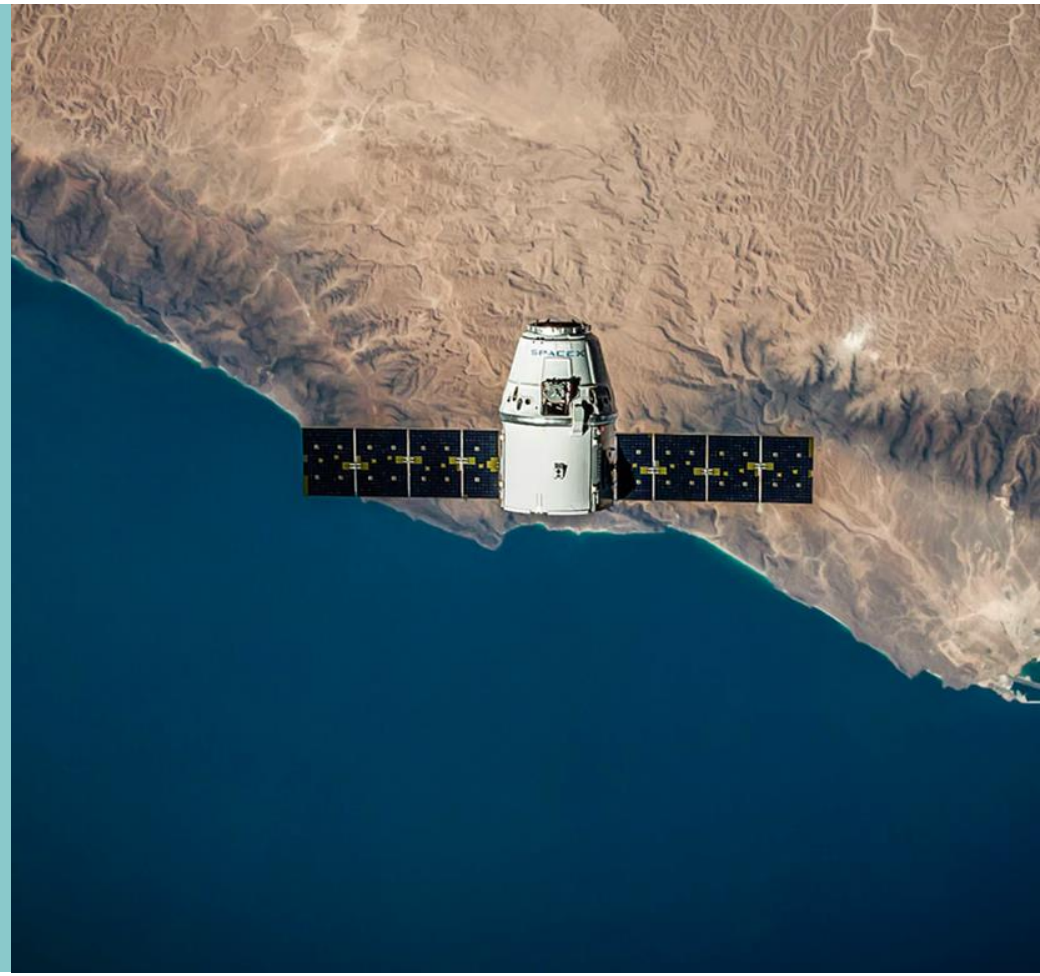
Introduction

- Who we are
- What we do

Experience working with…

- Data
- Algorithms
- Beneficiaries

Questions



# Who we are

LULEÅ TEKNISKA UNIVERSITET

VINNOVA
Sweden's Innovation Agency

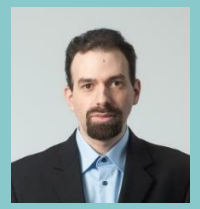
SMHI

RI. SE

AI SWEDEN

Rymdstyrelsen
Swedish National Space Agency





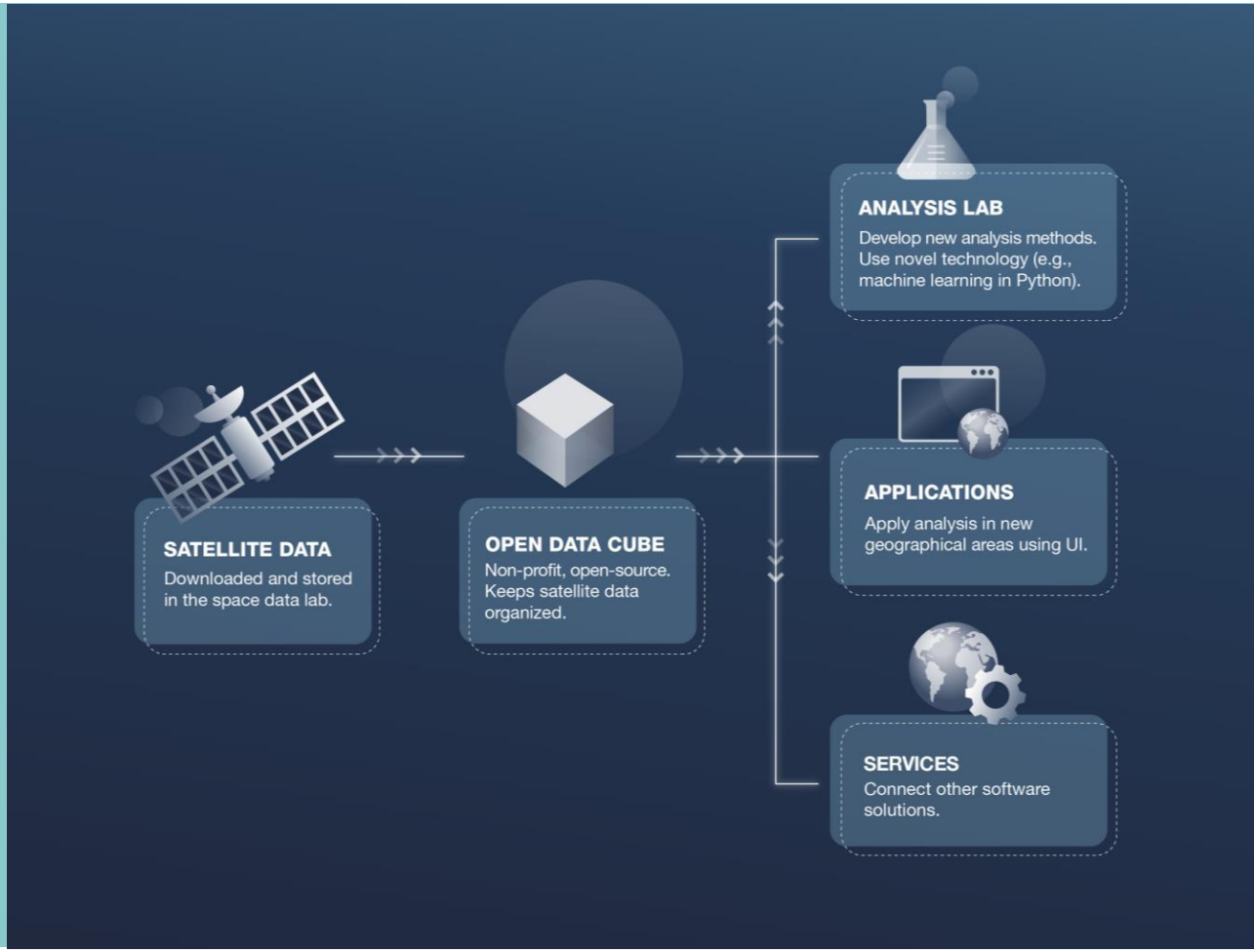



# What we (plan to) do at the SDL

# SDL pipeline

Have a pipeline for various needs

- Everything starts with Satellite data (Sentinel)
- Data accessible through ODC
  - For developing new analysis methods
  - For applying the methods in new geographical areas using a User Interface
  - For providing services, and connecting them to already existing solutions



# Pilots in SDL

1. Drought in Mälardalen
2. Water levels in Vänern
3. Marine Habitats
4. Coastal Zones
5. AI for good
6. API
7. Where and when is the water
8. AI Areal
9. AI Shallow water

Länsstyrelsen Västra Götaland
Länsstyrelsen Västmanlands län
Länsstyrelsen Kalmar län
Länsstyrelsen Västerbotten
Swedish Agency for Marine and Water Management
GÖTEBORGS UNIVERSITET
TRAFIKVERKET
MSB Swedish Civil Contingencies Agency
SKOGSSTYRELSEN
NATURVÅRDSVERKET
SCB Statistiska centralbyrån Statistics Sweden
Jordbruksverket

# Facilitate AI development - Hackathons

*”The hackathon format is a fantastic way to create value in a short amount of time. It is also a way to gather and engage the AI-community, and give the opportunity for like-minded peers to meet and learn from each other.”*

Participated in the organisation of two hackathons

- Space Data Hackathon 2020
- Copernicus Hackathon Sweden

Support provided in both cases

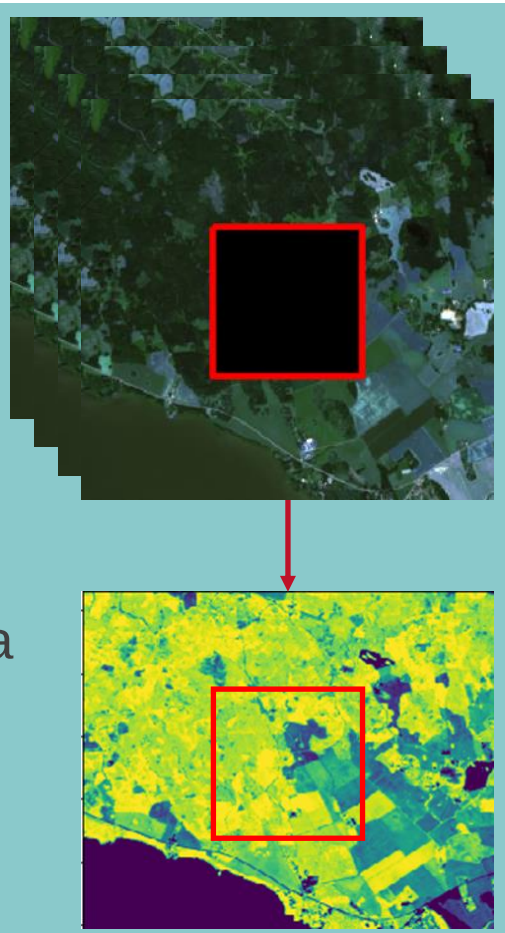
- Task description
- Mentoring (if required)
- Satellite data and computational platform
- Jury

Difference: provided labeled validation and test data in one case

# Space Data Hackathon 2020

## Two-fold task

- Simple time-series prediction
  - Input: a series of dates with average NDVI values
  - Output: average NDVI for a date in the series with missing value
- Spatio-temporal prediction
  - Input: a series of satellite images with an area masked in one
  - Output: prediction of NDVI values for all pixels in the masked area

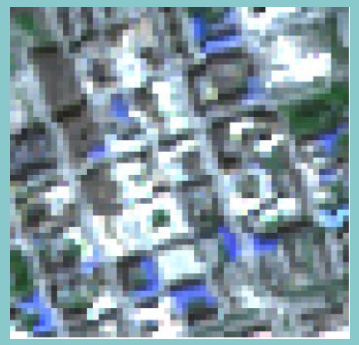


## Take-away:

- The unreasonable effectiveness of permanence model

# Copernicus Hackathon Sweden

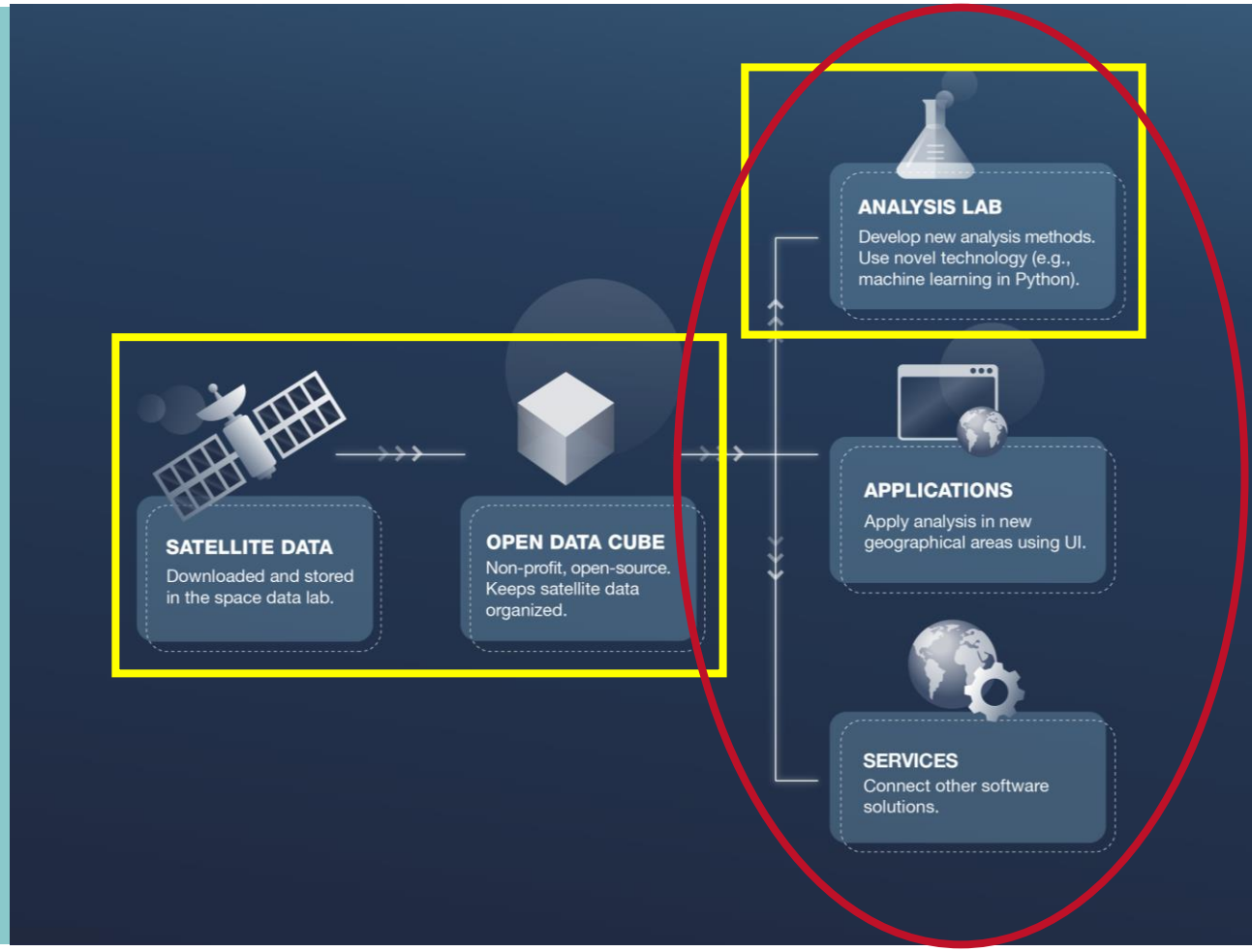
- Snow/Cloud detection
  - Climate change monitoring
  - Differentiation between cloud cover and snow
  - Based on Sentinel-2 data
- Classification of areas
  - Size: 500m X 500m
  - From COVID-19 perspective
  - Create a useful categorization
- Take-away
  - Less defined tasks, more freedom, more deep learning (ResNet and Unet)



# Experience from using the SDL

Different lessons learned from each part of the pipeline

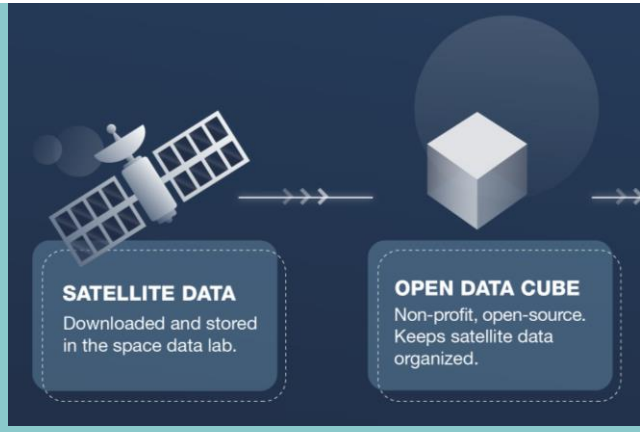
- Working with the data
- Working with algorithms
- Working with users



# Working with the data – infrastructure

The underlying infrastructure is crucial

- Large demand on IO-performance
- Many potential bottlenecks in the system
  - Hard drives
  - Networking
  - Load balancing
  - CPU
- Has to be designed ground-up with these considerations in mind

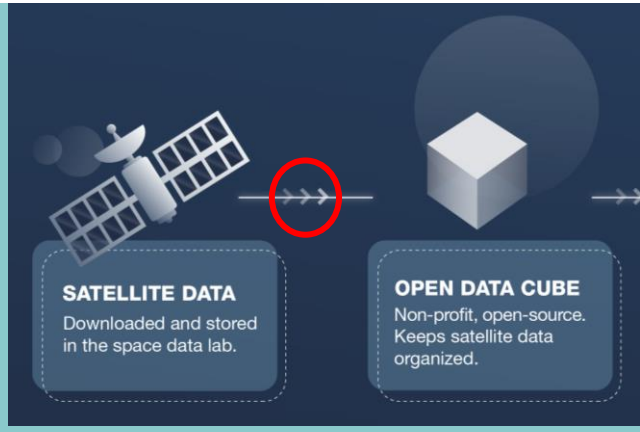



kubernetes

# Working with the data – leap to ODC

## Satellite to ODC: more than a simple step

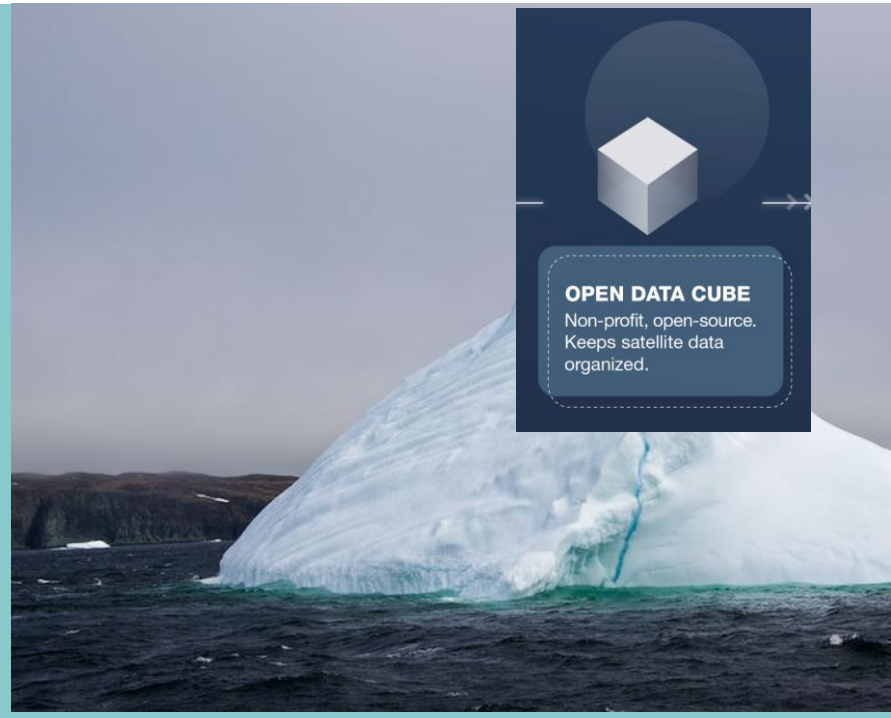
- For data science, this is the first step, but for data engineering, the last step of a long process
- Much work goes into handling the data up until it is indexed and ready for use (long processing chain)
  - For one product maybe 5 minutes, but we want 100Ks of products
  - The process may have to be repeated
  - Parallelization is important
  - Automatization is a must (manual handling of data does not scale well)



# Working with the data – beyond ODC

## The ODC is just the tip of the iceberg

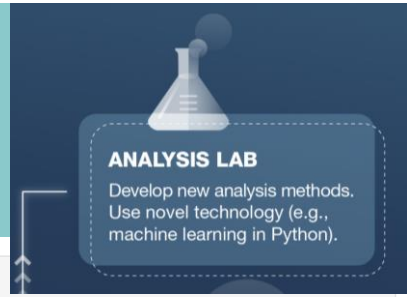
- Metadata
  - Each satellite source needs its own
  - Manual production unfeasible, scripts not fully available
- Coordinate-systems
  - Want to store our data in SWEREF99 TM when possible
- File format
  - Specific file format needed to increase the performance
- The ODC does not handle this, we have to produce the workflows to take care of it



# Working with algorithms – data access

## Can be done through JupyterLab

- Makes analysis interactive
- Possible to visualize and manually check images
- Facilitates the use of Python developed libraries, such as gpytorch, tensorflow, keras




```
[3]: lat = (65.590962, 65.638668)
     lon = (22.043322, 22.175764)

     query = {
         "product": "sentinel_s2_l2a_zips",
         "output_crs": "epsg:3006",
         "lat": lat,
         "lon": lon,
         "resolution": (10, -10),
         "time": ["2019-07-01", "2019-08-01"],
         "measurements": [
             "B02_10m",
             "B03_10m",
             "B04_10m",
             "B08_10m"
         ]
     }
```

### Load data

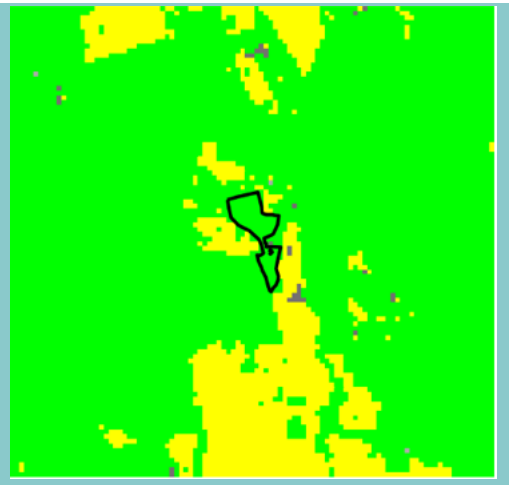
Once the datacube knows *what* data, *where* and *when*, we can use the `dc.load()` function to load the data of interest. The cell below shows an example of how to load using a progress bar. Another option would be tu run the line `ds = dc.load(progress_cbk = progress, **query)` only, and wait until the cell status changes from `[*]` to `[number]`. It may take some time.

```
[ ]: def progress(processed, total):
         print(f'\rProcessing {processed} of {total}...', end='', flush=True)
     ds = dc.load(progress_cbk = progress, **query)
```

```
Processing 51 of 76...
```

# Working with algorithms – ground truth

- Creating ground truth data for algorithms
  - Can create a widget for it
  - Visualize the data
  - Manual labelling based on visual information and automatic suggestion

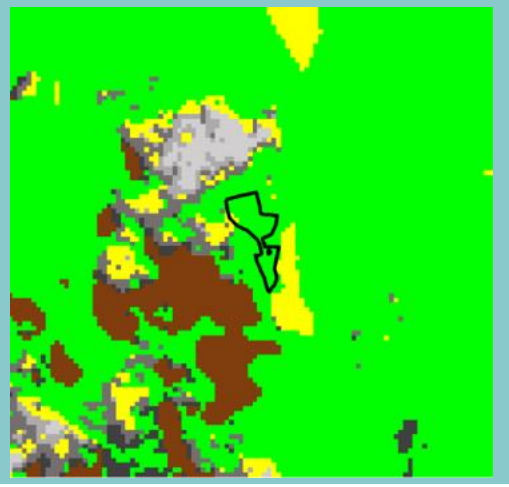


2020-08-09 10:22:22

Select score:

- Reject
- Accept
- N/A

Suggest: Accept



2020-07-03 10:32:20

Select score:

- Reject
- Accept
- N/A

Suggest: Accept

# Working with algorithms – future work

- The data available for algorithms is still limiting what can be done (Sentinel-2 coverage for Sweden from 2018)
- Data handling by the OpenDataCube library
  - Using specific data structures
  - Create example notebooks/documentation to help new users
  - Collaborate with other datacubes on the algorithms
- Use pilot outcomes as basis for development

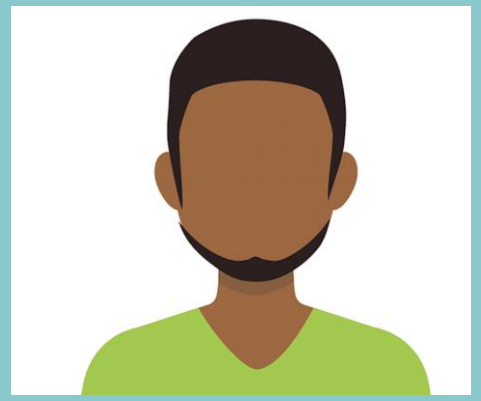
OPEN DATA CUBE

# Experience working with beneficiaries

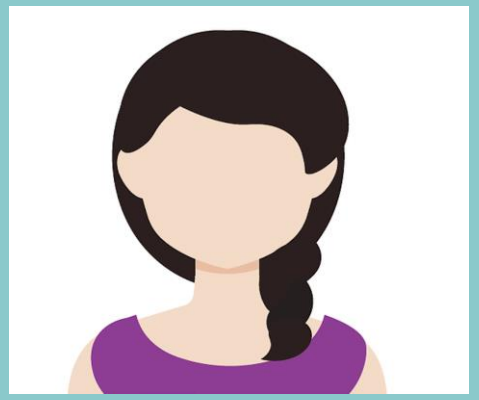
Difficulties when technical staff directly has to work with beneficiaries

Help of specialized background knowldge

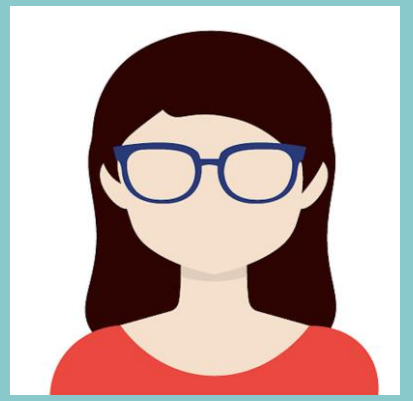
Identified 3 prominent roles or ”personas” of interest



Albert
Preparedness Coordinator
County Board



Camilla
Consultant
Enterprise

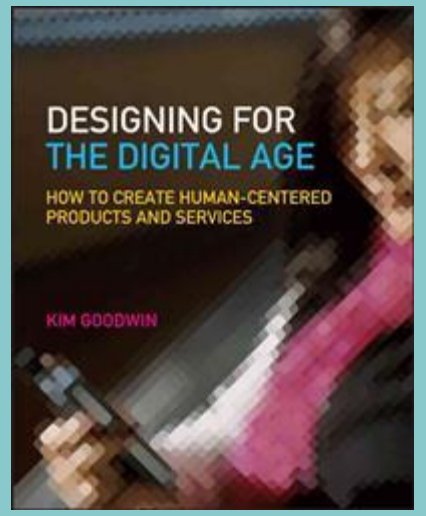


Ebba
Geographic Information Officer
Governmental Agency

# Experience working with beneficiaries - methodology

## Distinction between people interacting with SDL

- Beneficiaries: take decisions to adopt new solutions
- Users: work hands-on with satellite data




## Working with beneficiaries

- Following Kim Goodwin's guidelines
- 10 in-depth interviews conducted
- 3 personas as result, representing attributes and behaviours

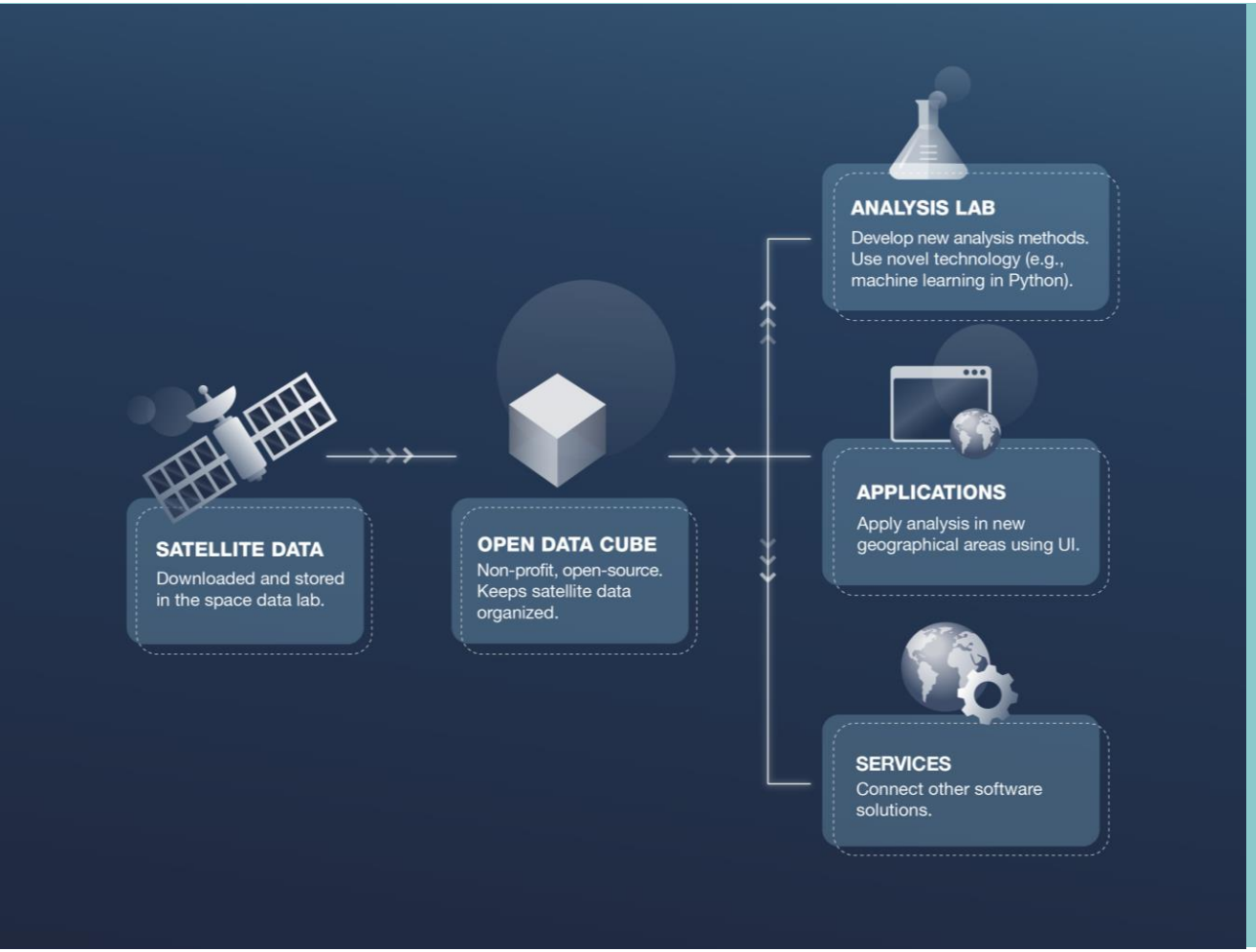
# Experience working with beneficiaries - personas

- Coordinator on county board
- Focus on specific phenomena (climate adaptation, flooding, etc.)
- Geographical interest in own county
- Interested in change detection, long-term temporal analysis
- Wants to keep some data private
- Some technical skills

- Consultant/business associate
- Focus on specific phenomena meeting end-user needs
- Geographical interest extensive
- Interested in the technical infrastructure of SDL
- The more data the better
- Identifies as ”space actor”, having the most technical skills of the three

- Geographic Information Officer
- Focus on natural resources and their use
- Geographical interest is Sweden
- Interested in change detection, long-term temporal analysis
- The more data the better
- Advanced technical skills

# Experience working with beneficiaries – access point for all beneficiaries




## Analysis lab

- For Python programmers
- Take advantage of ML methods

## GUI

- Readily computed data models

## Integration services

- Connect EO data with other software

# Thank you - questions