

living planet | BONN symposium | 23-27 May 2022

TAKING THE PULSE
OF OUR PLANET FROM SPACE



EUMETSAT



ECMWF



Active learning and visual data mining for the generation of Earth observation image benchmarks

Wei Yao, Mihai Datcu

2022.05.27

Main goal

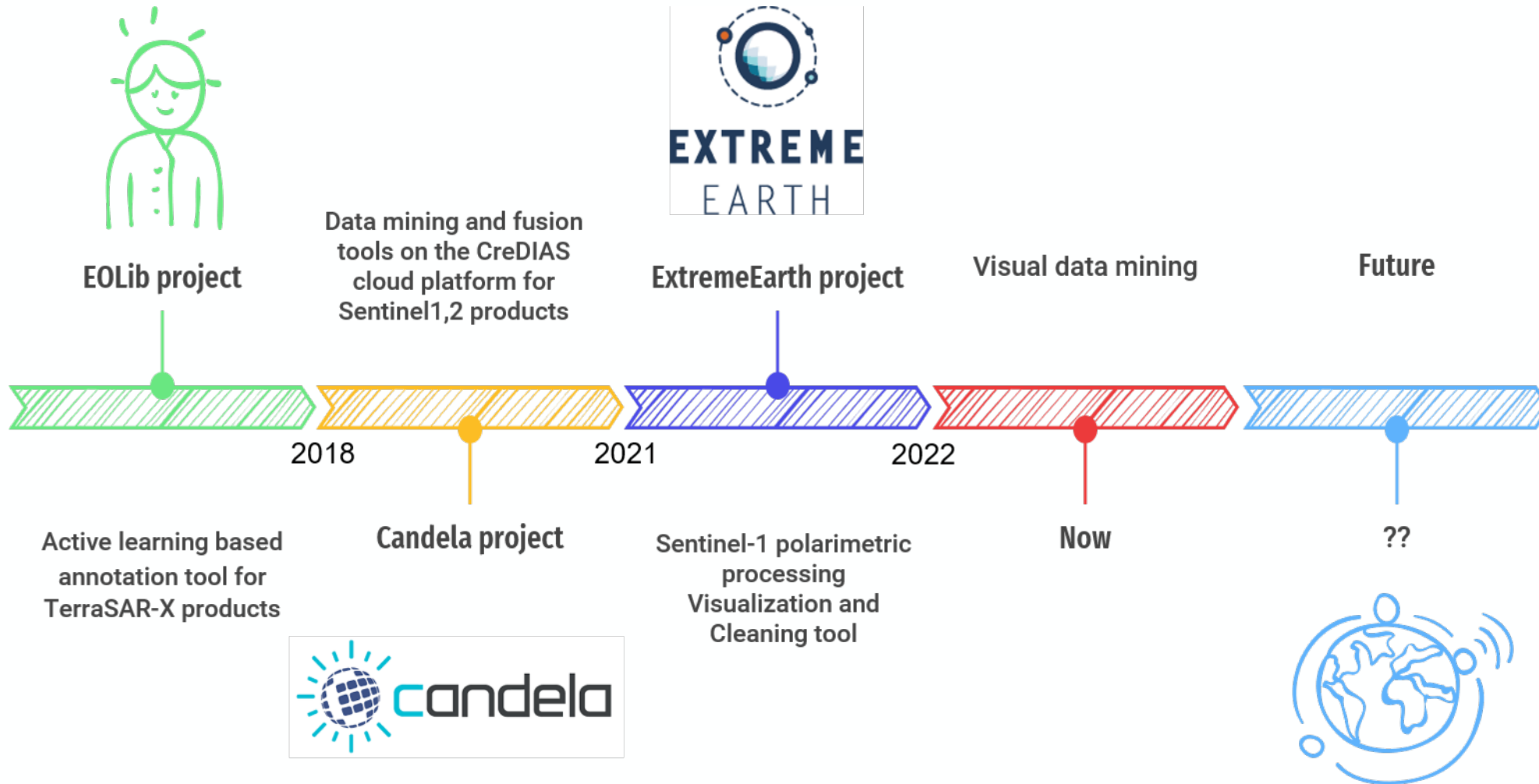
- Generate the best datasets at minimal cost.

Challenges

- How to **foster** the generation of EO image datasets?
- How to **validate** the generation of EO image datasets?
- How to **visualise** the datasets?



Evolution of the development



Active learning: machine is interactively inquiring a user for new knowledge.

Setup: Given prior knowledge, choose where to collect more data.

- Access to unlabelled data
- Make a query to obtain labels
- Add “informative” labels or correct wrong labels

Output: A labelled dataset learned from a few samples.

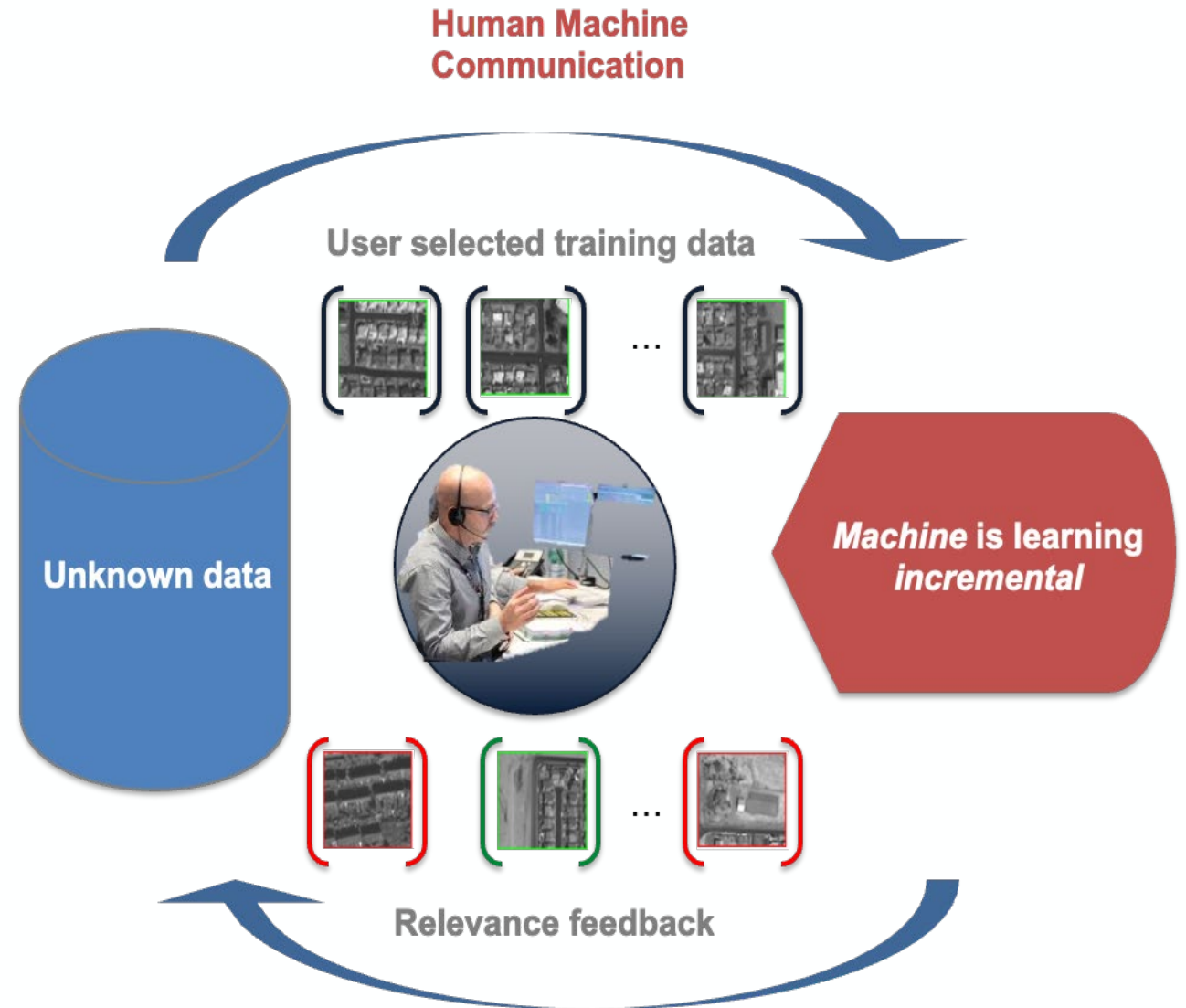


Image annotation tool: S2

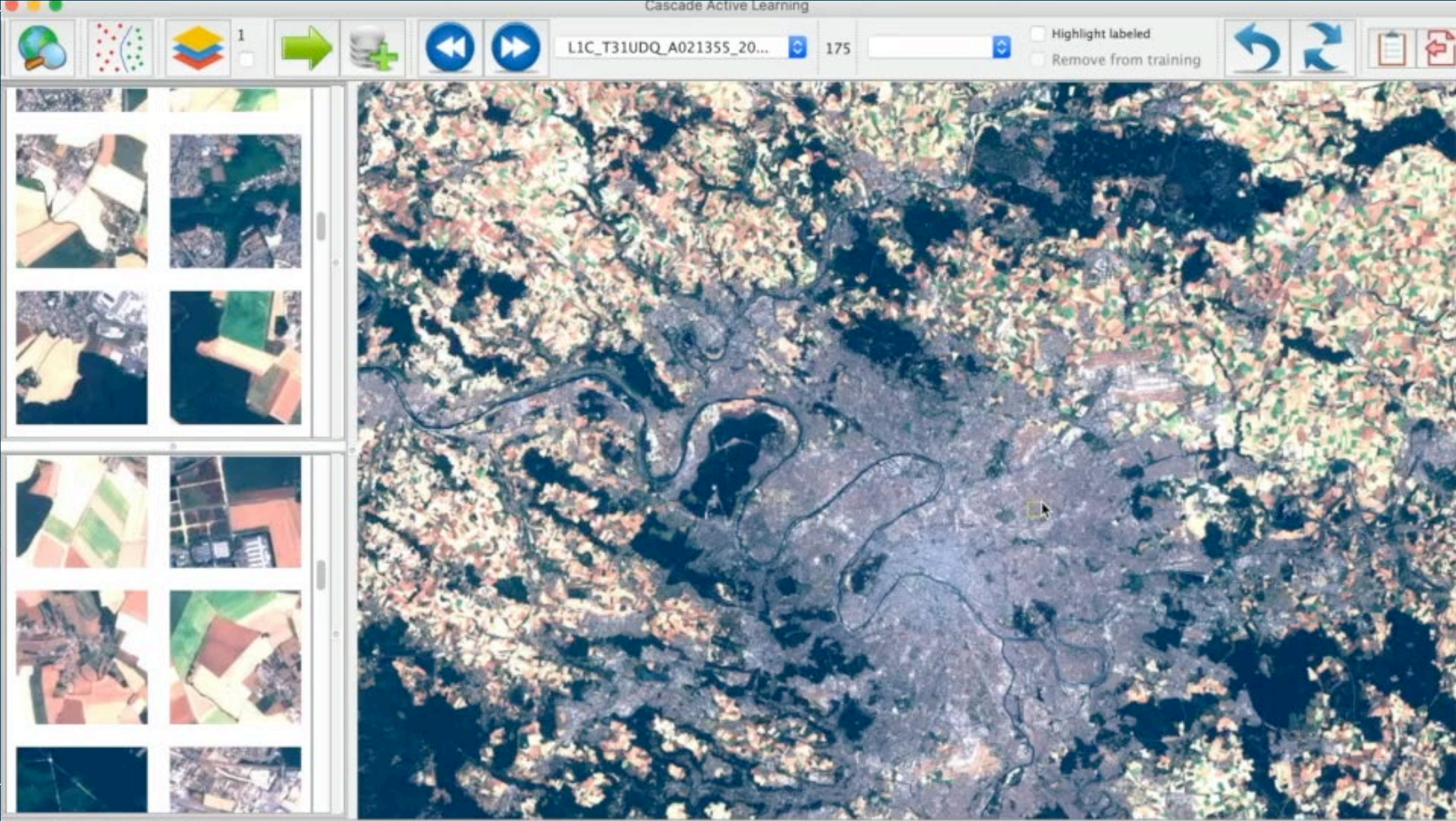


Image annotation tool on a cloud platform

Data mining / fusion architecture

Cloud platform layer

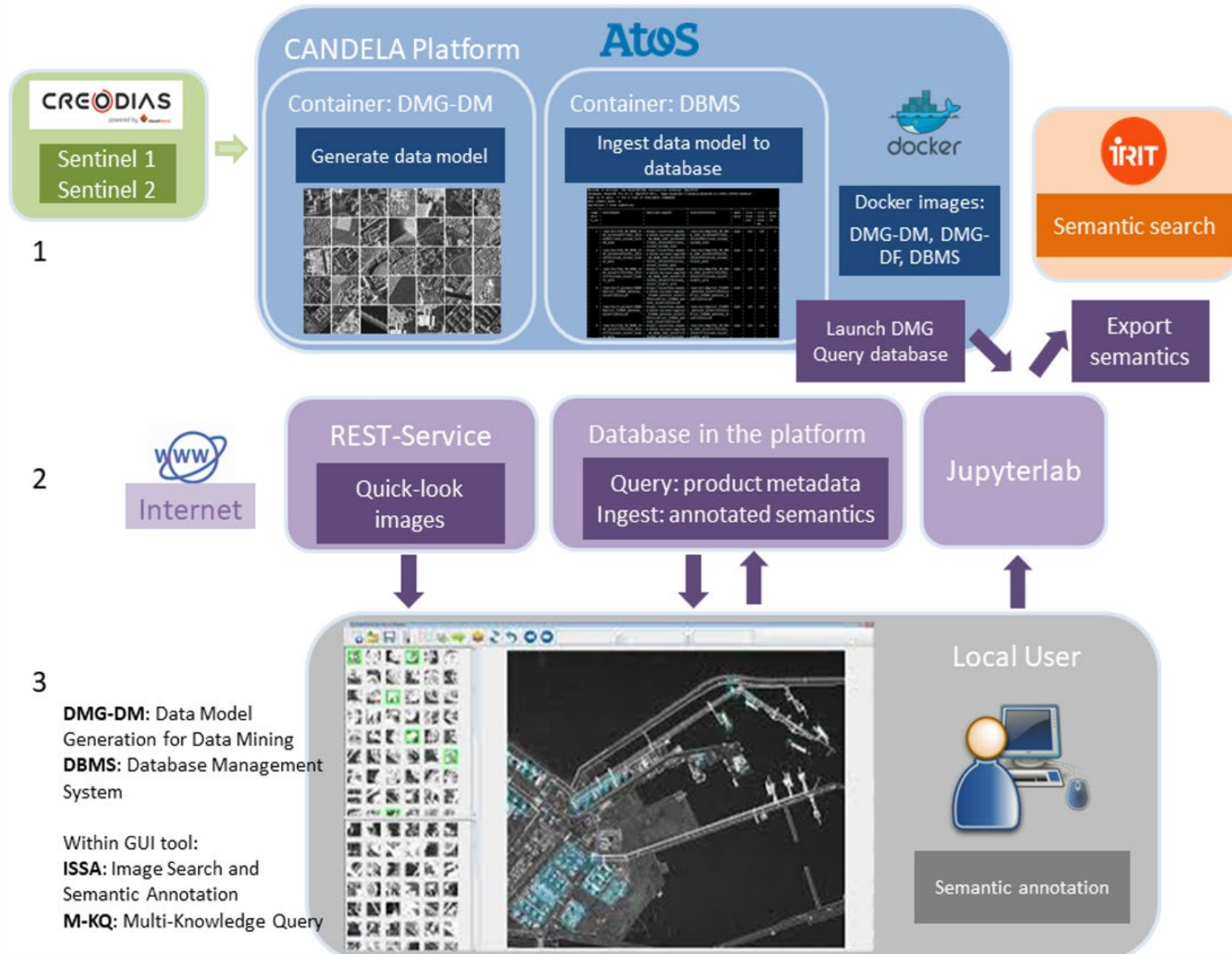
- CreoDIAS connector to EO products
- Atos cloud platform
- IRIT semantic search tool

Internet transfer layer

- Jupyterlab as UI
- transfer quicklooks via REST API
- query and ingest in remote database

User frontend layer

- perform semantic annotation
- multi-knowledge query

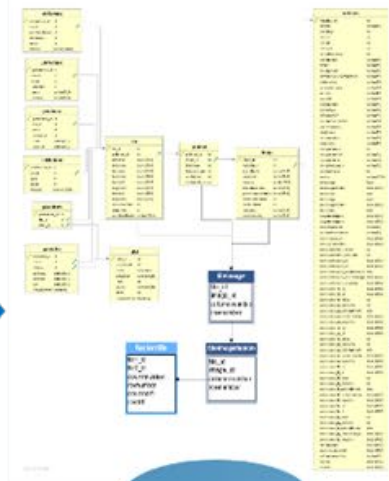
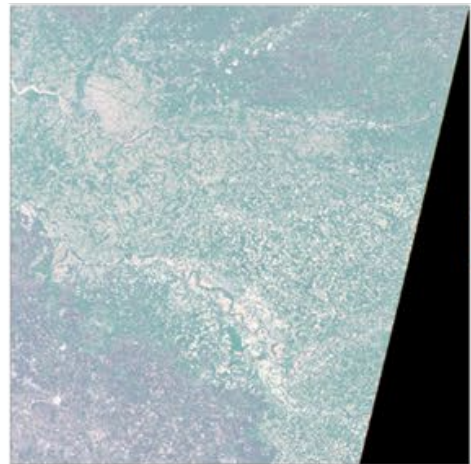


Data fusion: S1 + S2

Data Mining Data Base
SQL model: EO semantics

Interface with IRIT tool

GeoTiff: labels
and coordinates



```
Draw annotation

[1]: from DrawAnnotation import DrawAnnotation

Run the DrawAnnotation class

[2]: print('Please make sure that you have used DLR frontend to perform annotation...')
print('Please correct the tags to your data in the function load_params...')
print('Ingest_id is the id of your processed product in the MonetDB database...')

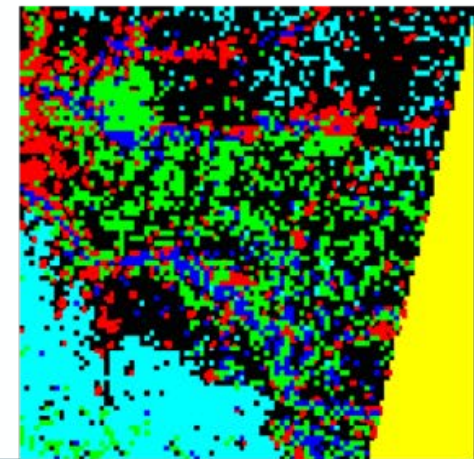
paramStruct = {}
paramStruct['Dataset_Category'] = "LC_DM"
paramStruct['Product_id'] = "/eodata/Sentinel-2/MSI/L1C/2017/04/09/S2A_MSI_L1C_20170409T105651_H0204_R094_T30TYQ_20170409T110529_SA"
paramStruct['Product_id'] = "/eodata/Sentinel-1/SAR/GRD/2019/07/18/S1A_IW_GRDH_1SDV_20190718T051950_20190718T052015_028167_032E7E"
paramStruct['Ingest_id'] = "45"
paramStruct['Ingest_id'] = "16"

DrawAnnotation.run(paramStruct)

print('Done!')
print('The output is: \noutfile_mono: the gray level annotation')

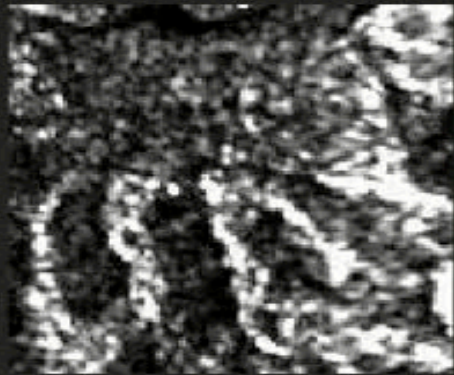
Please make sure that you have used DLR frontend to perform annotation...
Please correct the tags to your data in the function load_params...
Ingest_id is the id of your processed product in the MonetDB database...
Done!
The output is:
outfile_mono: the gray level annotation

[3]:
```



What about “noisy” labels?

Data cleaning: manual



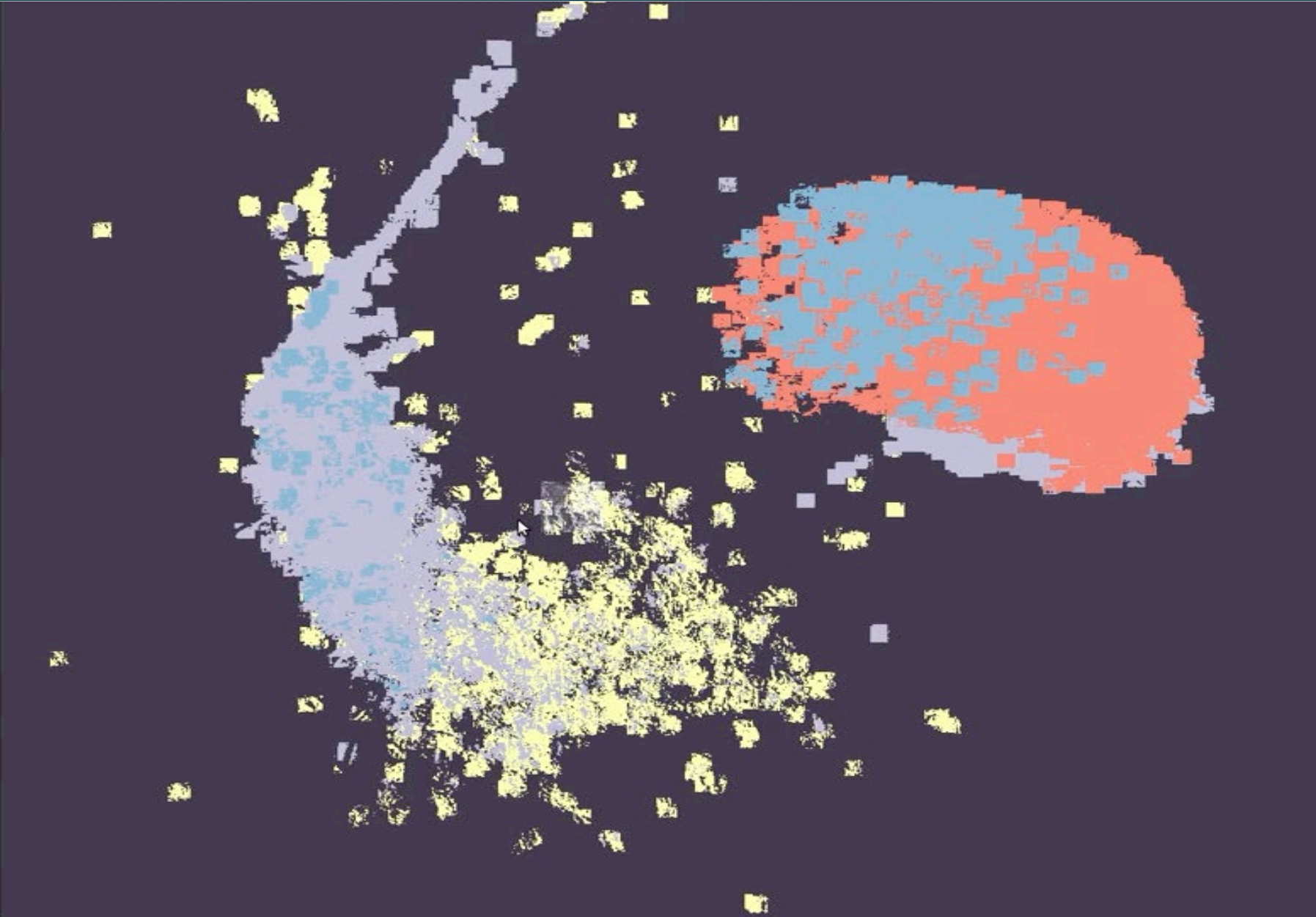
Label: mixed urban
Index: 12627

Nearest neighbors:



Legend

- mixed forest
- mixed urban
- cropland
- sea



Do we need to visualise the whole dataset?

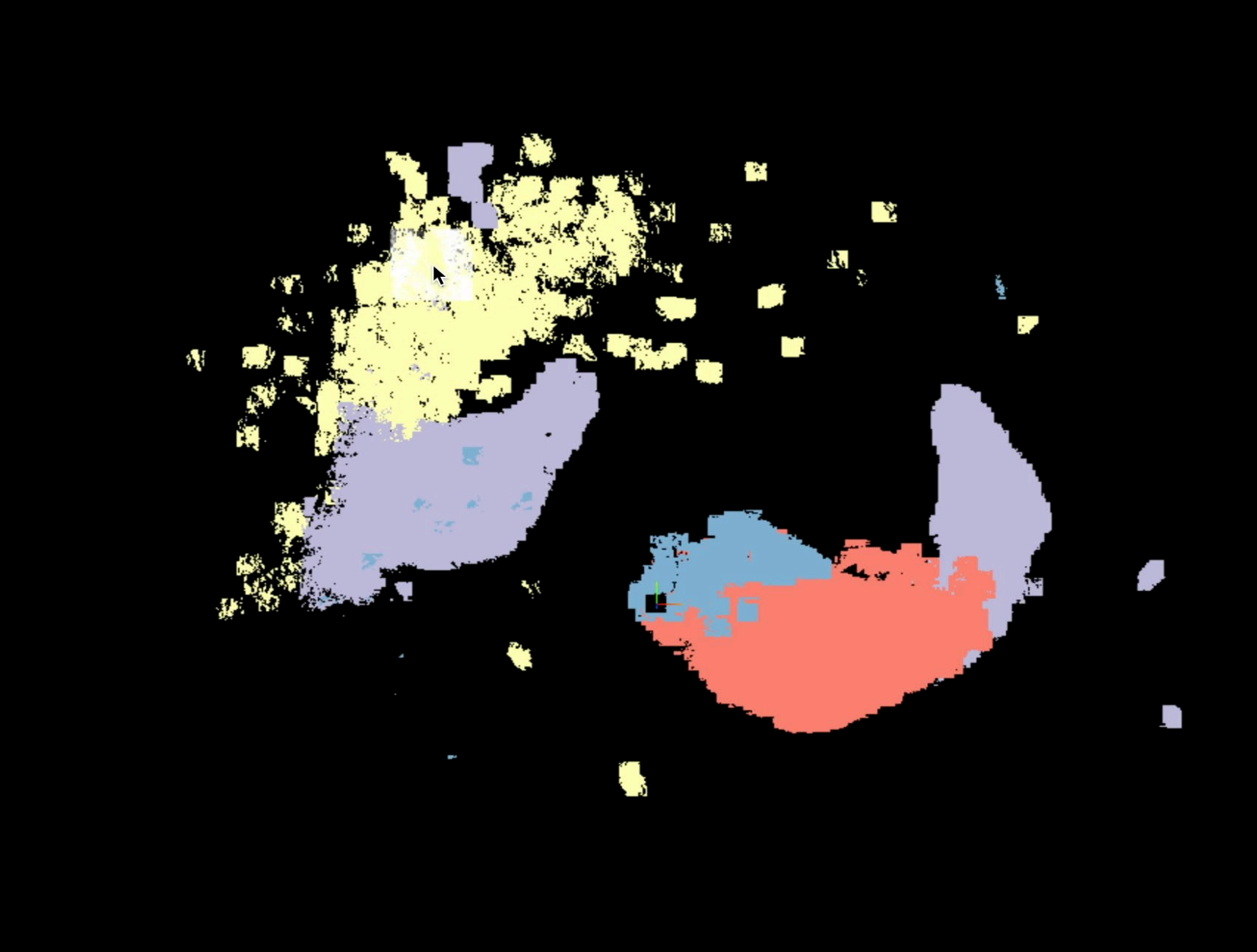
Data cleaning: automatic



Label: mixed forest

Index: 2709

Nearest neighbors:



Little random walk...

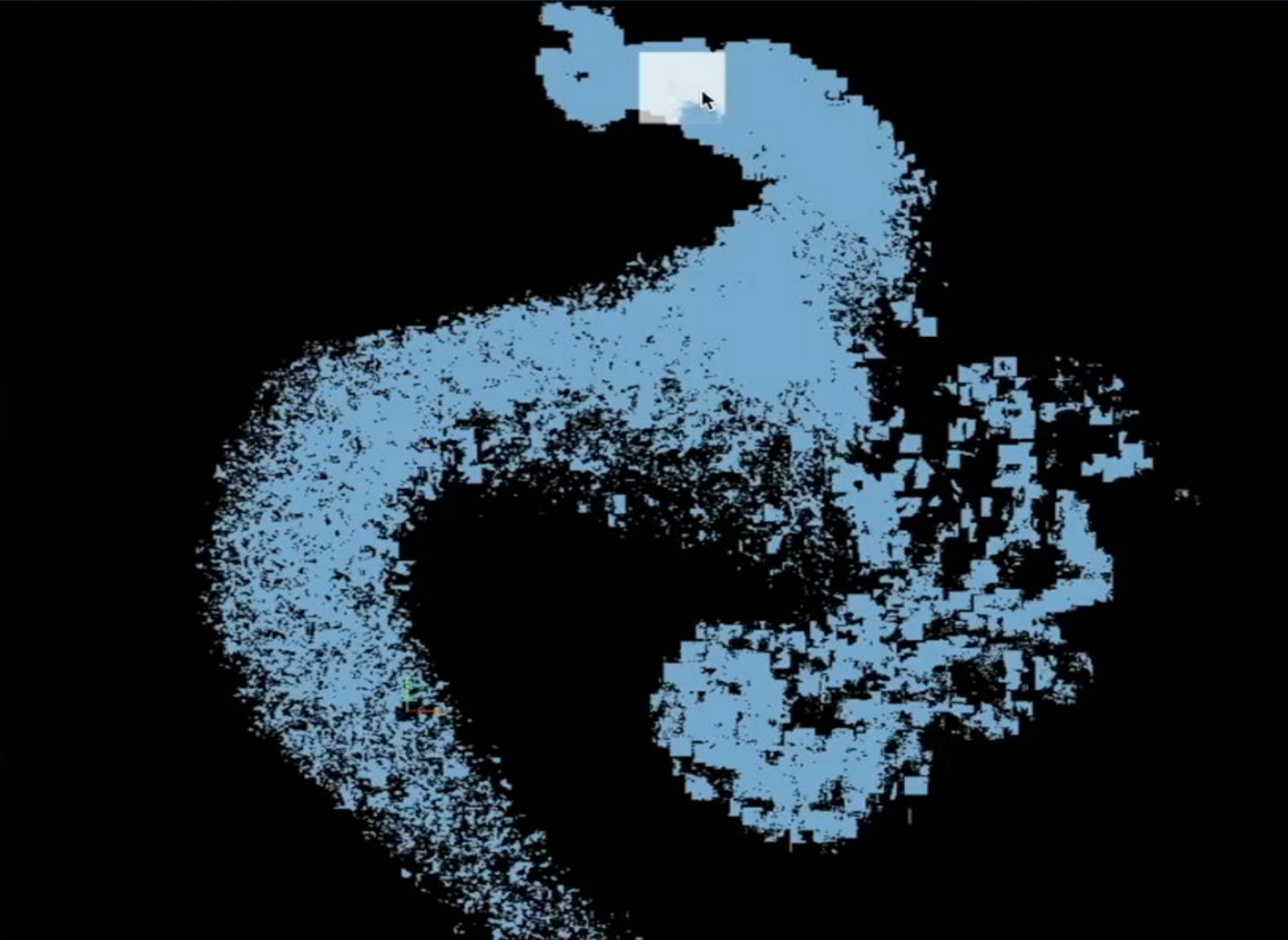




Label: Broad-leaved forest
Sea and ocean

Index: 9054

Nearest neighbors:



EOLib: An Image Information Mining Project, Datcu, Mihai und Espinoza-Molina, Daniela und Dumitru, Corneliu Octavian und Schwarz, Gottfried und Reck, Christoph und Manilici, Vlad (2016) EOLib: An Image Information Mining Project. EO Open Science 2016, 12-14 Sep 2016, Frascati, Italy.

An active learning tool for the generation of Earth observation image benchmarks, Wei Yao, Corneliu Octavian Dumitru, Mihai Datcu, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Brussels, Belgium, July 11-16, 2021.

A pattern analysis image validation tool for the generation of reliable Earth observation image benchmarks, Wei Yao, Gottfried Schwarz, Mihai Datcu, IEEE International Geoscience and Remote Sensing Symposium (IGARSS), to be published, 2022.

Thank you!

