Big data science for Earth observation: large scale visual analytics and knowledge discovery

Mihai Datcu

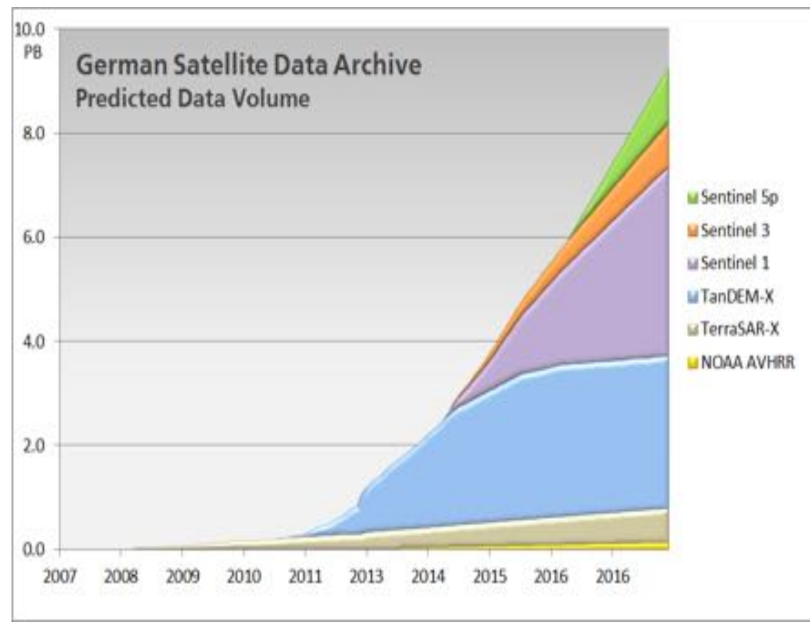
Motivation

Big Earth Observation Data

ENVISAT

TerraSAR-X and **TanDEM-X**

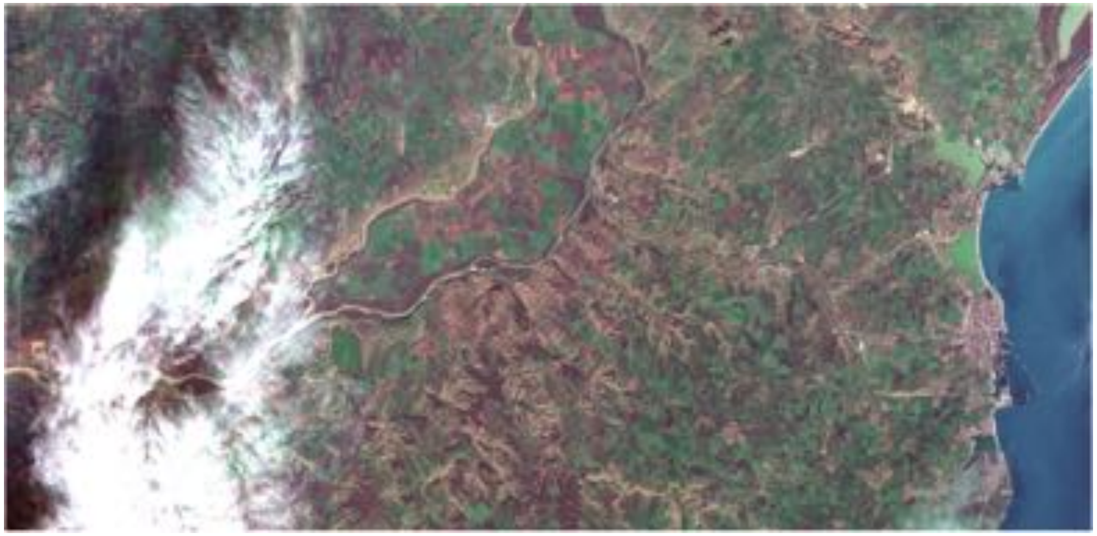
Copernicus Sentinel 1 and 2



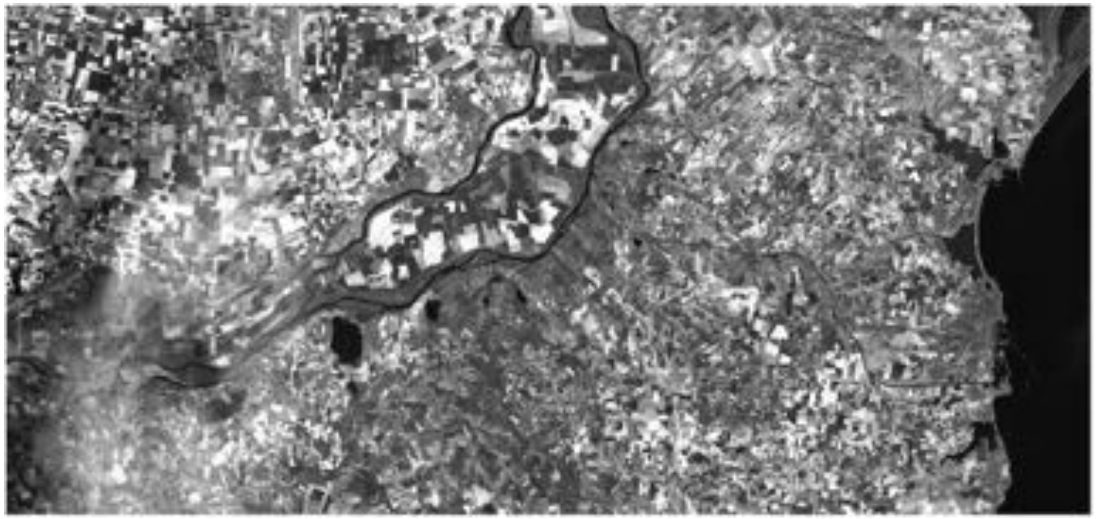
The DLR - DFD EO Digital Library

DLR

Multispectral sensors: Sentinel 2



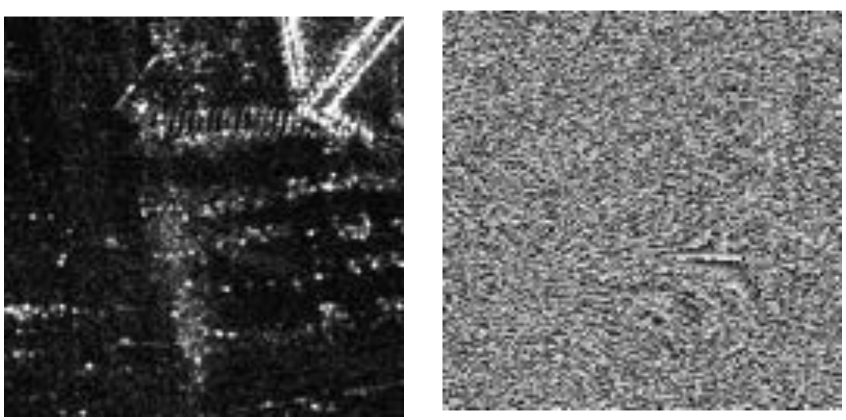
RGB Image



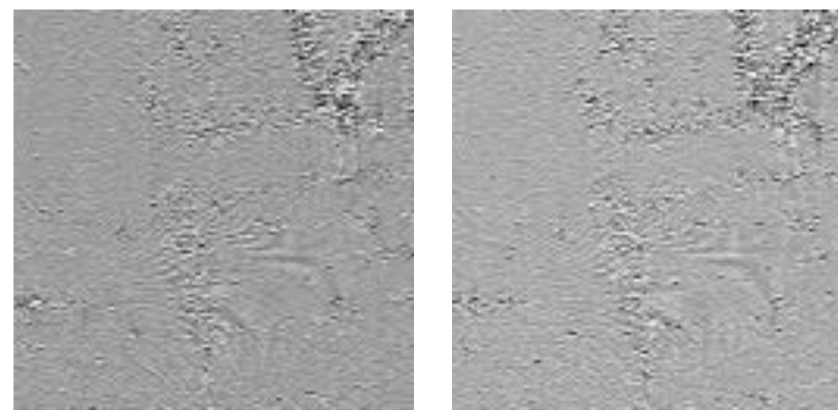
IR Image

SAR vs. optical images

TerraSAR-X Single Look Complex Image



Amplitude and Phase



Real and Imaginary components

The EO context

Earth Observation data is always **used jointly** with information extracted from **other sources** such as GIS, in-situ observations, or maps.

The goal is the exploration of these data and the **timely delivery** of focused **information and knowledge** in a simple understandable format.

The data volumes, their heterogeneity, unstructured nature and their complexity have become a major Big EO Data challenge for all applications.

(a) SAR (b) Multispectral (c) Map - OpenStreetMap (d) LUCAS

EO data particularities

- EO images: multisensory, eg. MS, SAR, altimeter, etc.
- These are multidimensional signals, acquired by sensors or instruments
- Sensor data carry physical meaning, radiation level, wavelength, etc.
- They are measuring land, ocean, or atmospheric parameters
- The VHR EO images observe detailed spatial structures and objects
- Satellite Image Time Series observe evolution processes over long period of time.

- An important particularity of EO images should be considered, is their "**instrument**" nature, i.e. they are **sensing physical parameters, and they are often sensing outside of the visual spectrum.**

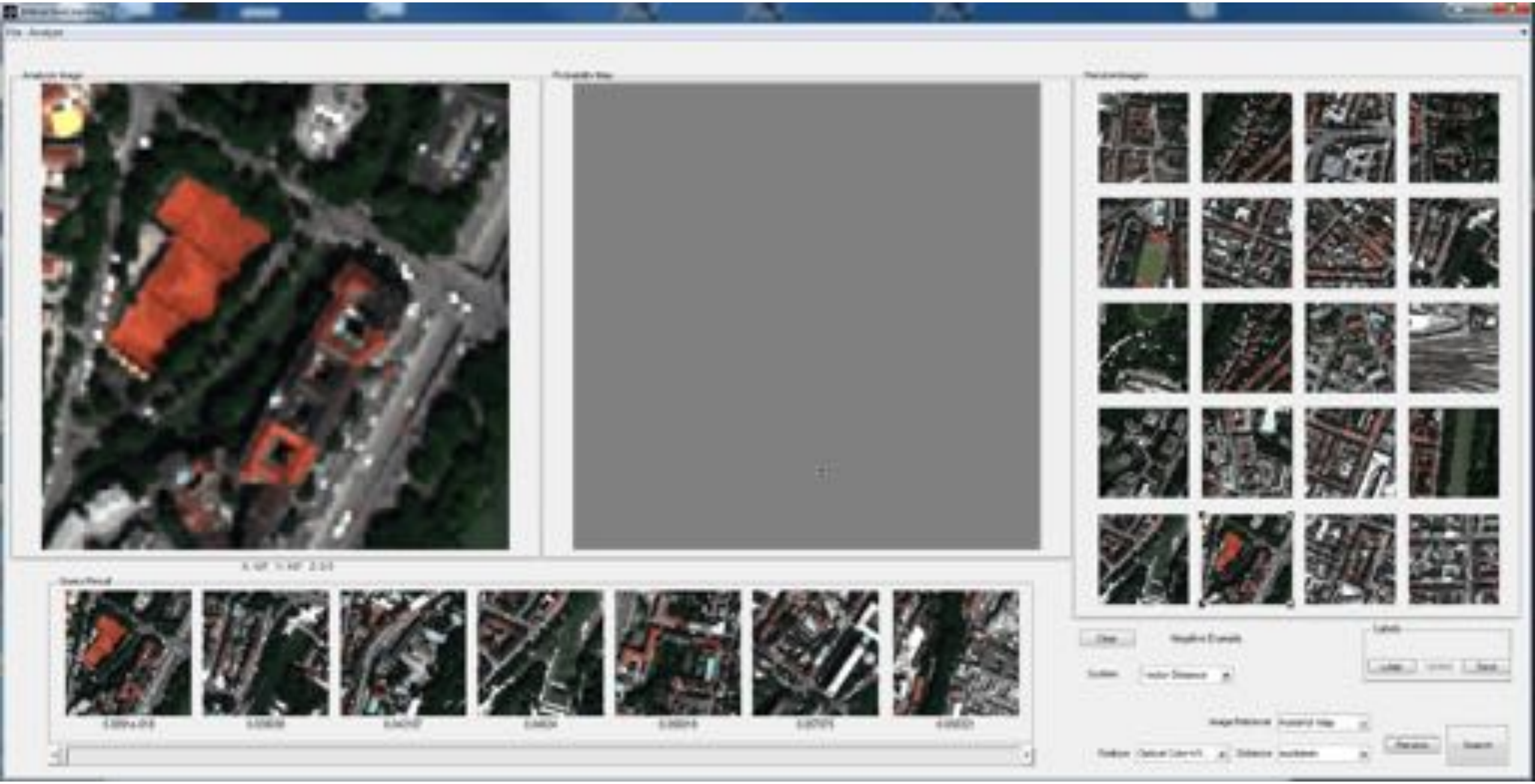
Big EO Data Analytics

- The today techniques, methods, and tools, for automated data analysis are insufficient for the analysis and information extraction from EO data sources.
- A new goal has become the gathering of the user's interest, together with the transformation of the data into reduced information and knowledge items, and adaptation to direct and easy understanding.
- The capability of retrieving information interactively and the use of data-driven paradigms are now more than ever necessary due to the huge data volumes being involved.

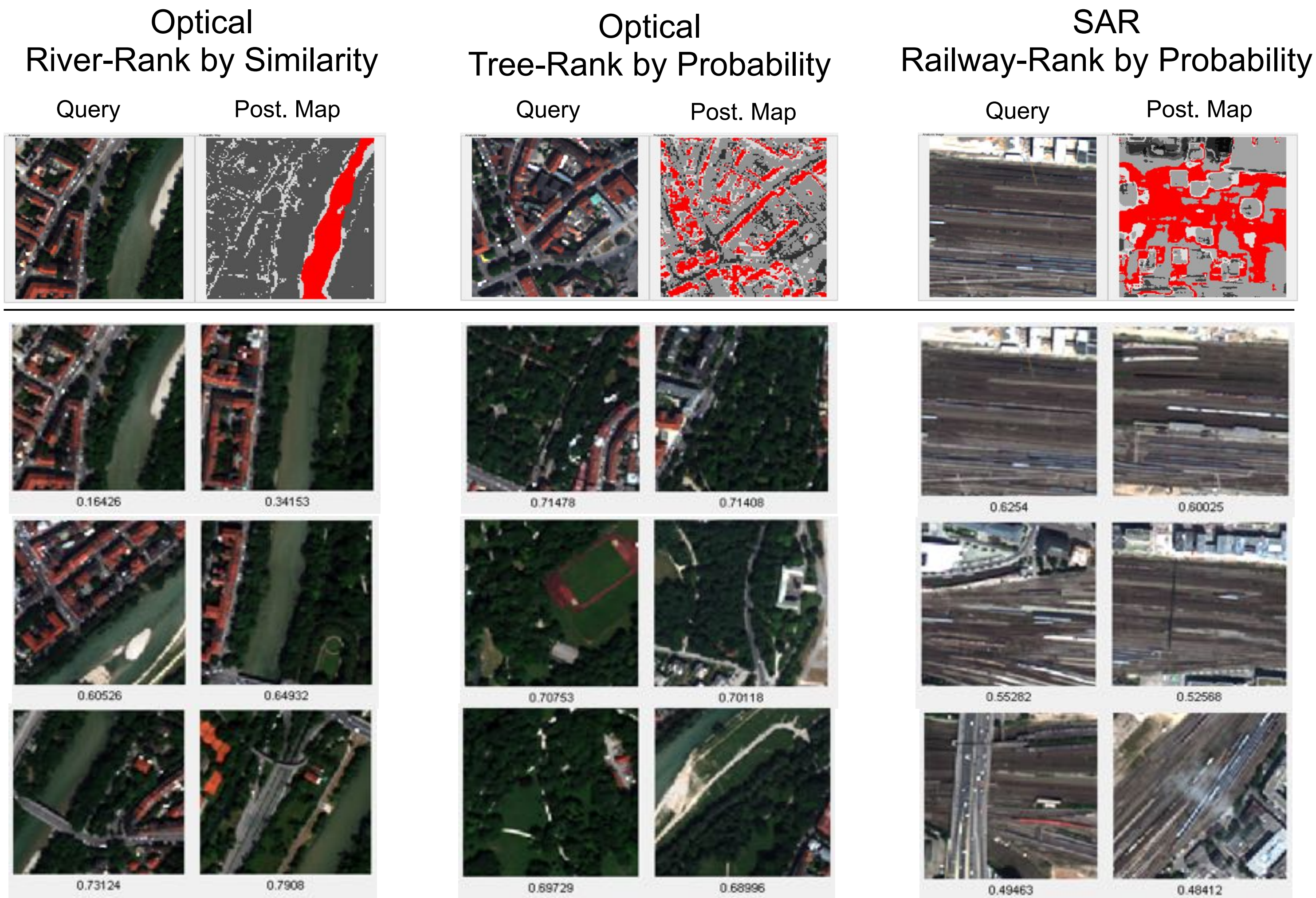
DLR

Examples

Multisesnor search engine

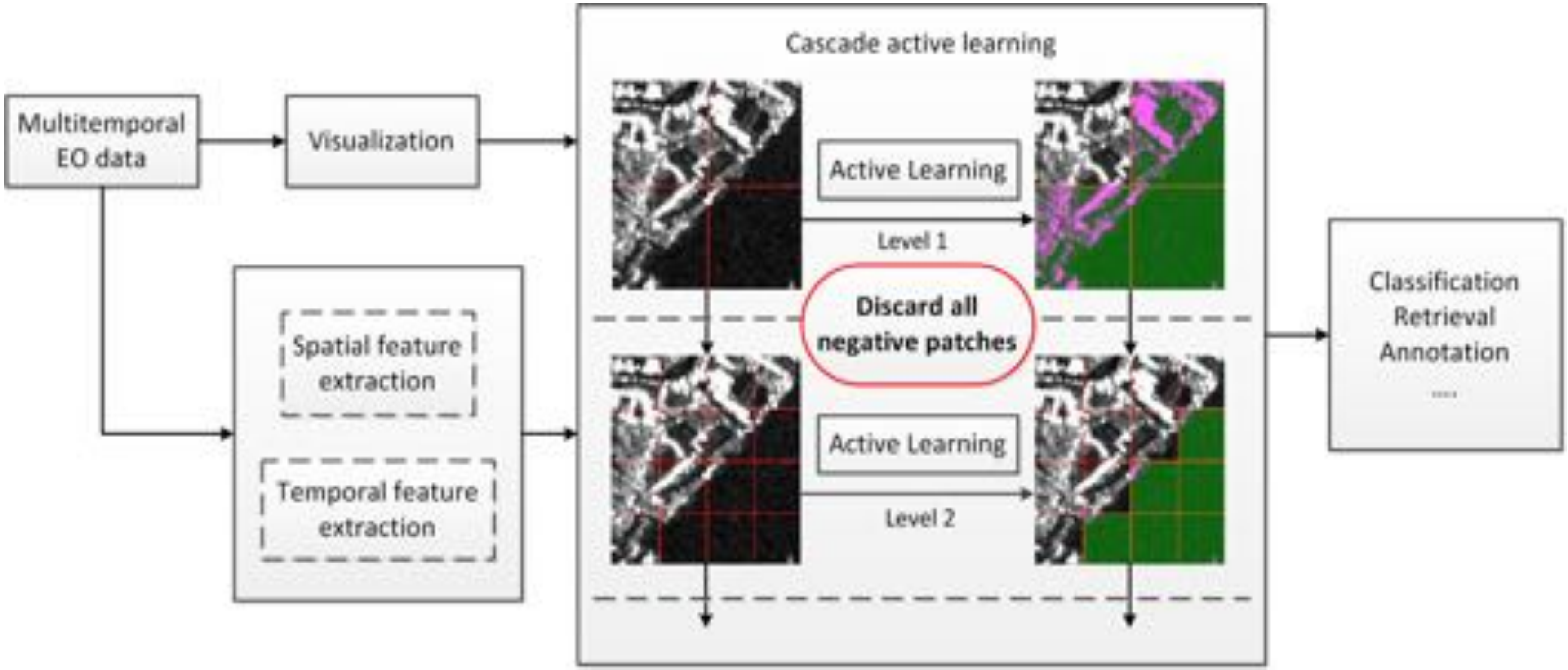


Query Results

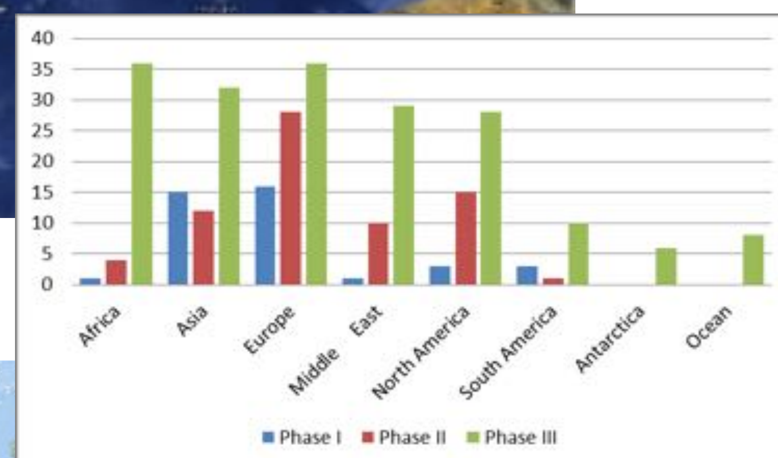


Cascaded active learning

- Two main components: Feature extraction and Learning

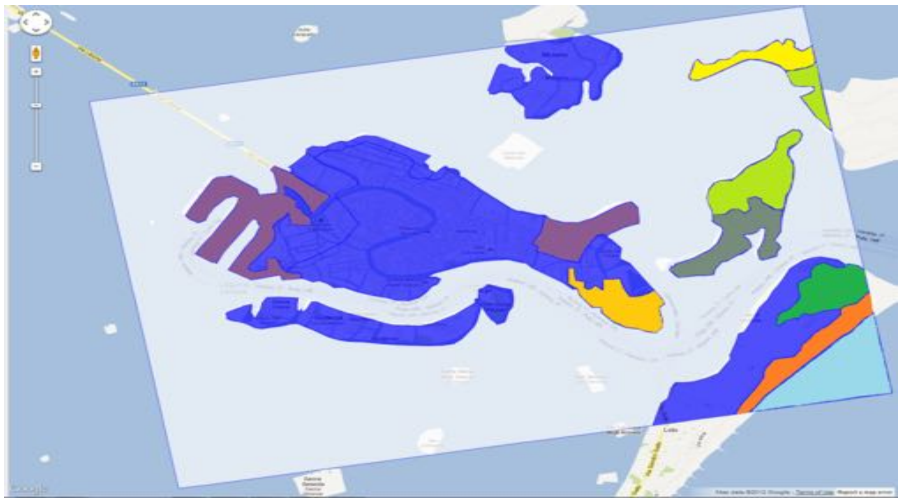


Semantic annotation



The location of the 300 TerraSAR-X scenes and the distribution of the scenes

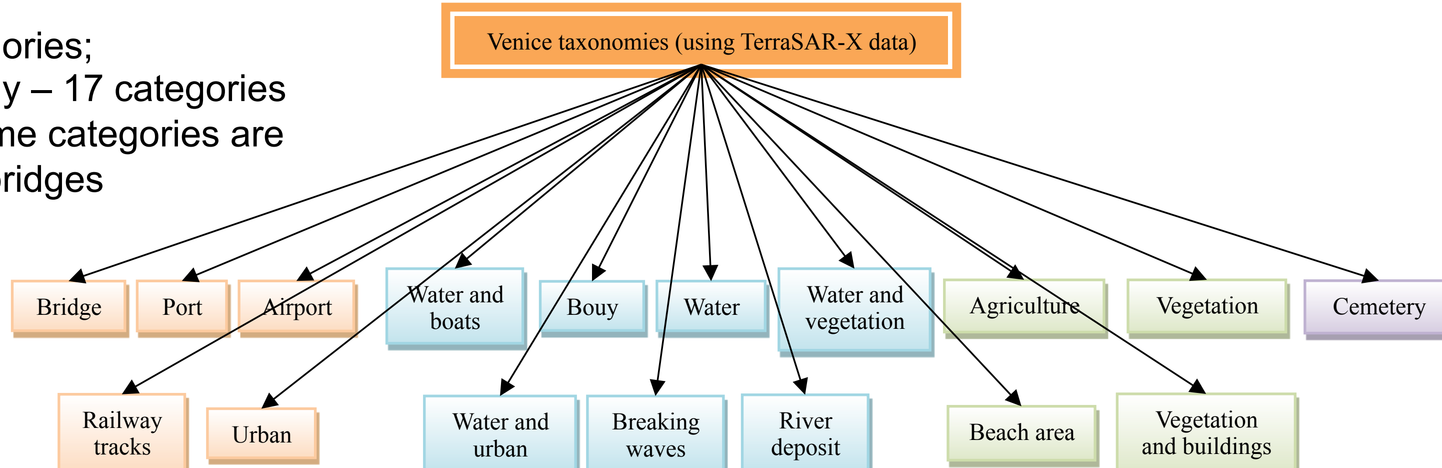
From CLC to our semantic taxonomy



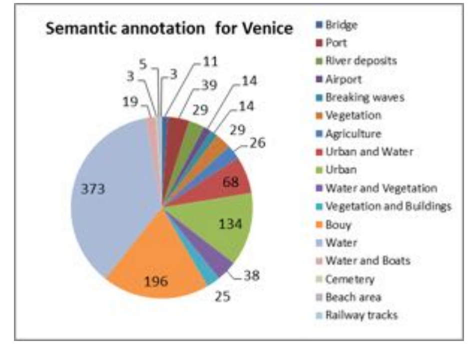
Legend - categories defined for Venice using CORINE Land Cover nomenclature:

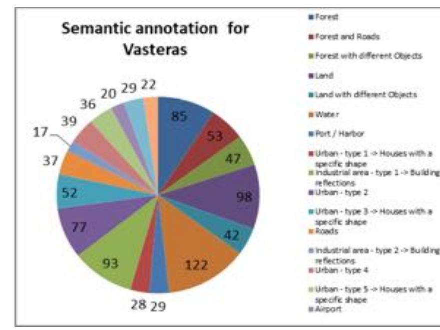
- Marine waters – coastal lagoons
- Marine waters – sea and ocean
- Urban fabric
- Pastures
- Forest
- Heterogeneous agricultural areas
- Open spaces with little or no vegetation
- Industrial, commercial and transport units
- Open spaces with little or no vegetation
- Artificial, non-agricultural vegetated areas

- Using: CLC – 10 categories;
  our methodology – 17 categories
- In the case of CLC some categories are mixed together (e.g., the bridges are included in marine waters – coastal lagoons)

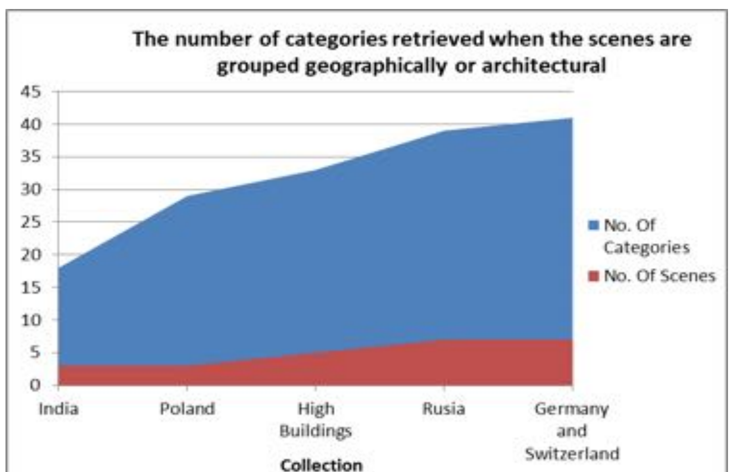


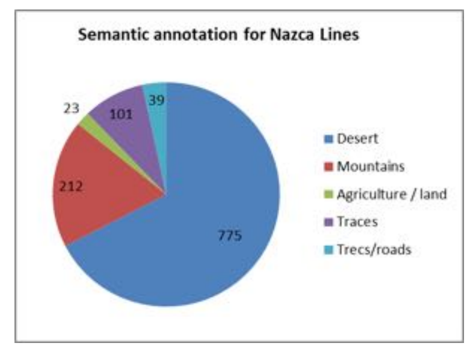
Semantic catalogues





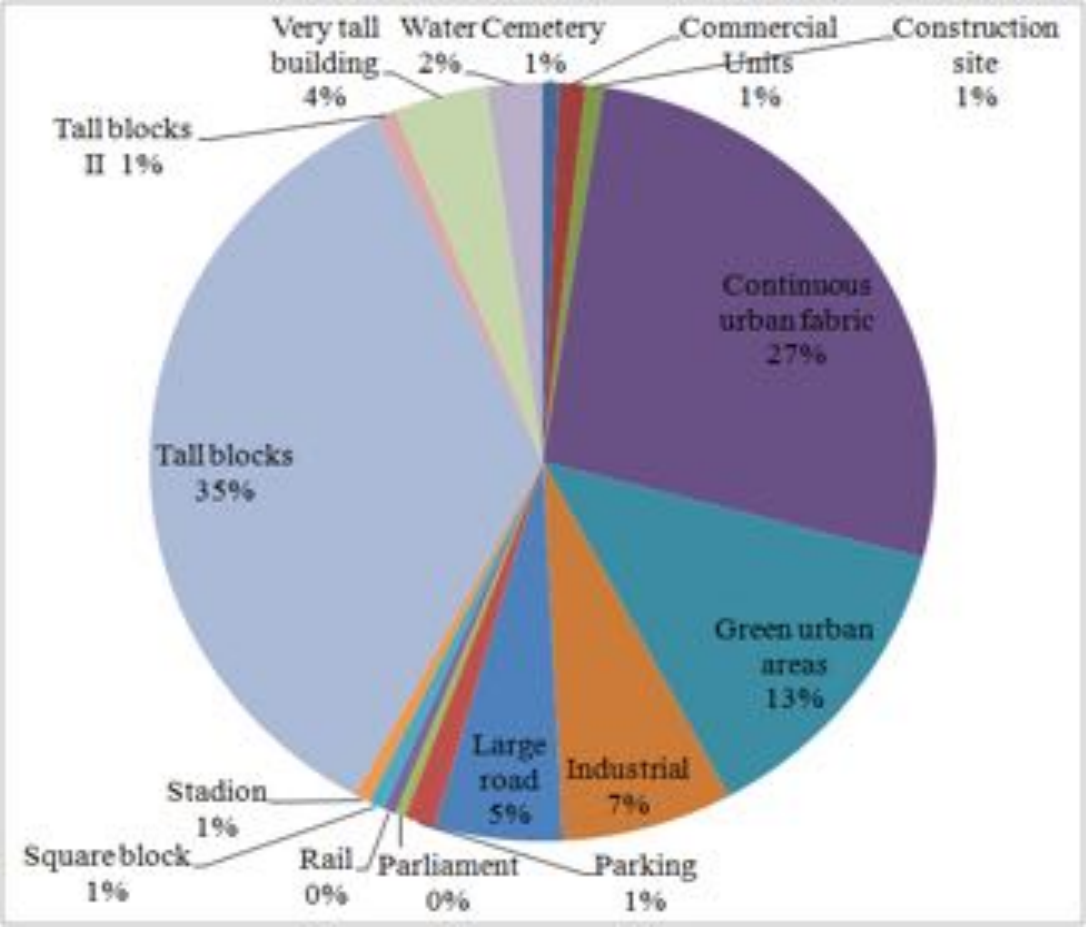
- Bangkok (Thailand);
- Shenyang (China);
- Nazca Lines (Peru);
- Havana (Cuba);
- Venice (Italy);
- Vasteras (Sweden);
- Oran (Algeria);
- Bogota (Columbia).





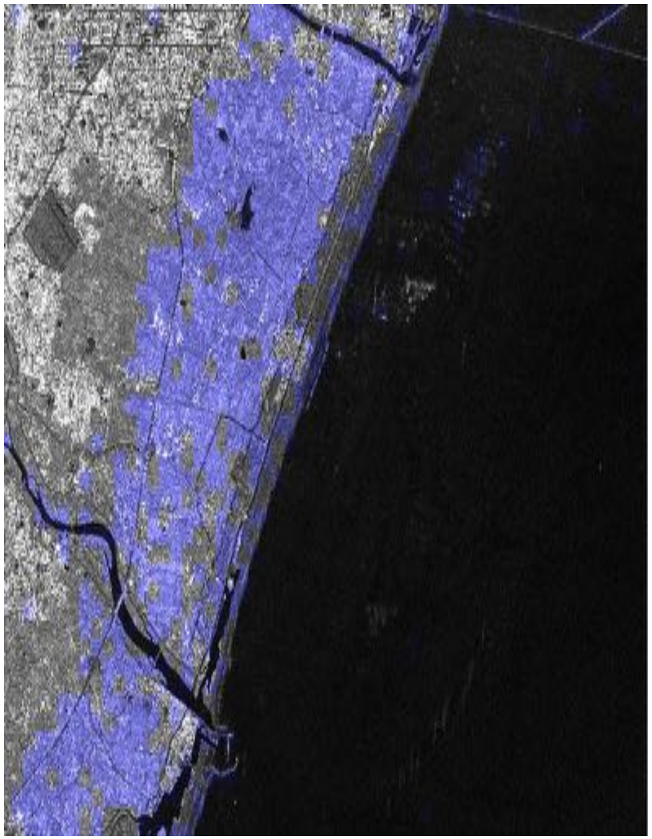
SCENE CATEGORIES & INFORMATION CONTENT

1 HS TerraSAR-X Scene =
up to10 000 image patches
(100 x 100 m)

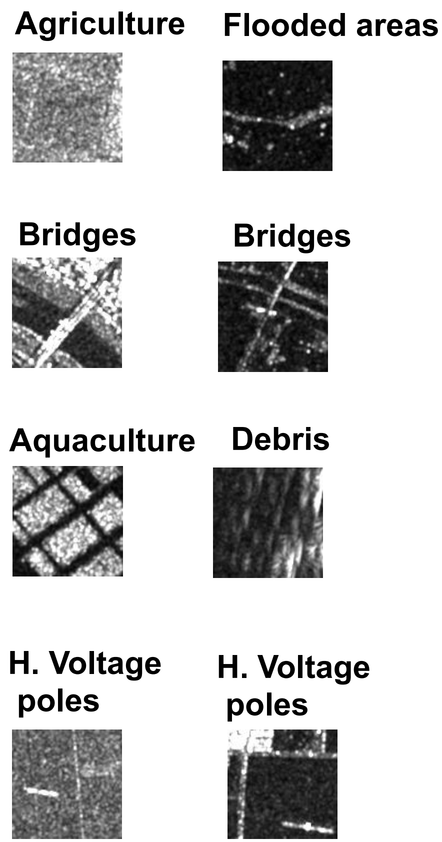


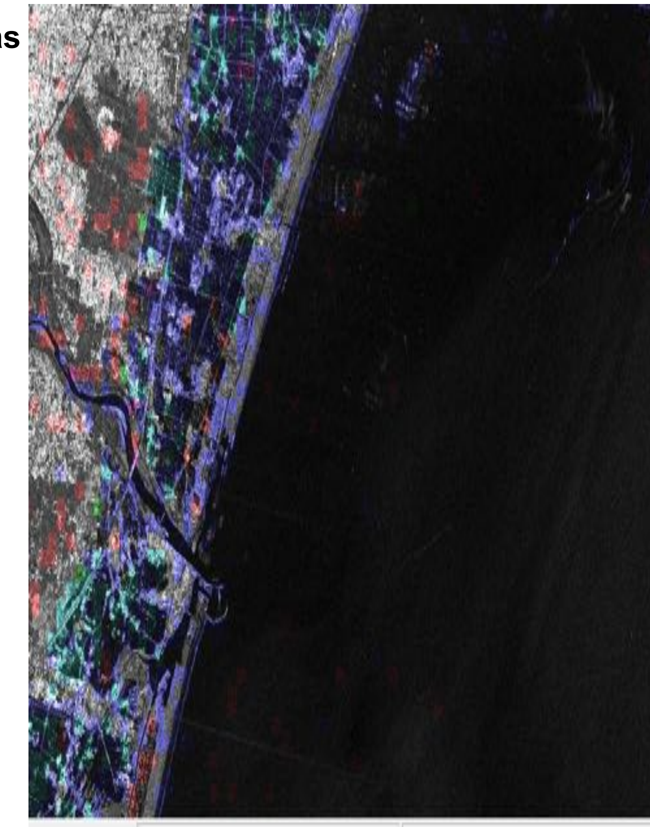
Knowledge Discovery application example

The damages in the agriculture can be clearly seen by comparing the classification in pre disaster image (left figure) with the post disaster image (right figure).



TerraSAR-X scene **before** Tsunami 20.10.2010

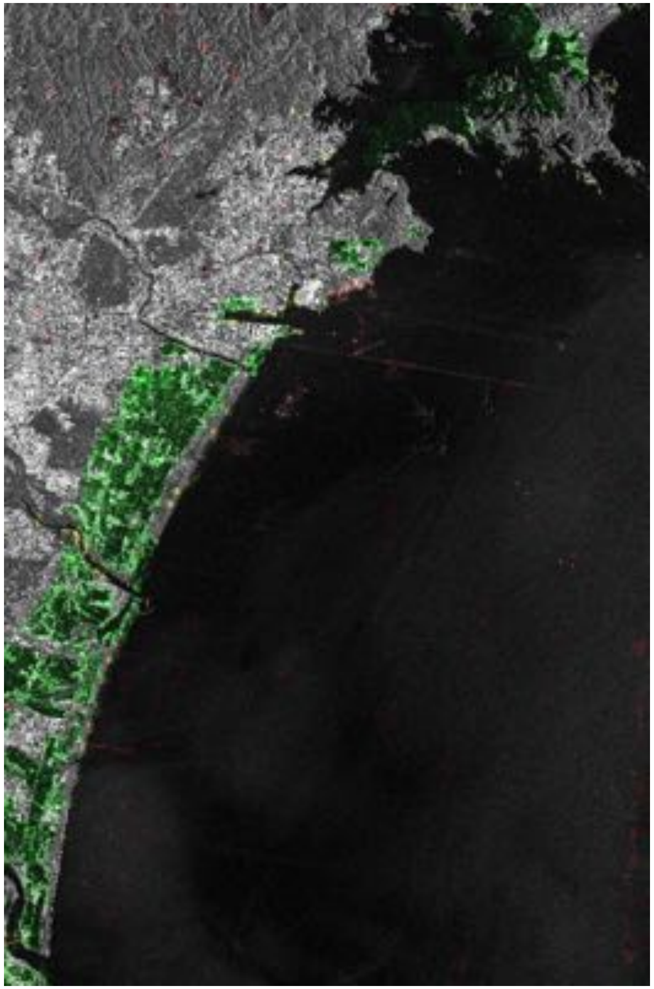


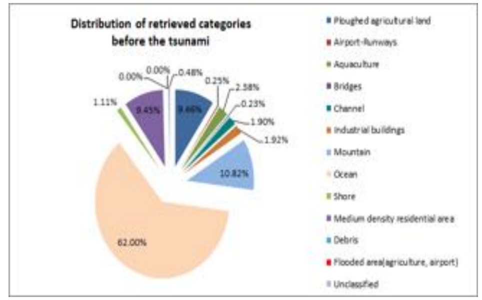


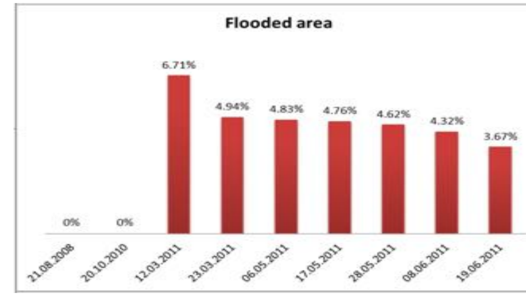
TerraSAR-X scene **after** Tsunami 12.03.2011

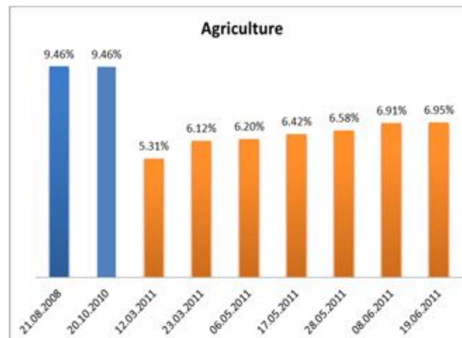
DLR

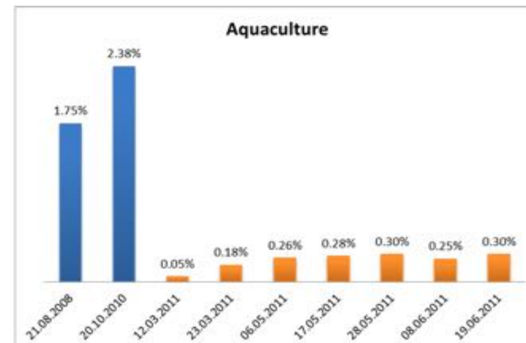
Data Analytics: Tsunami effects assessment

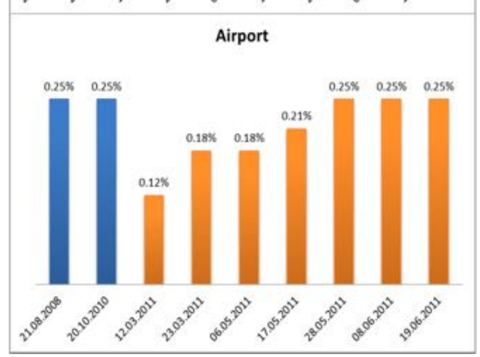


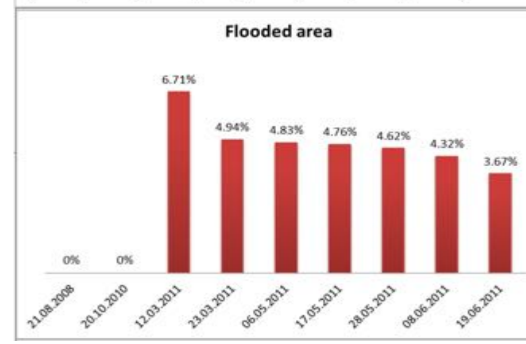






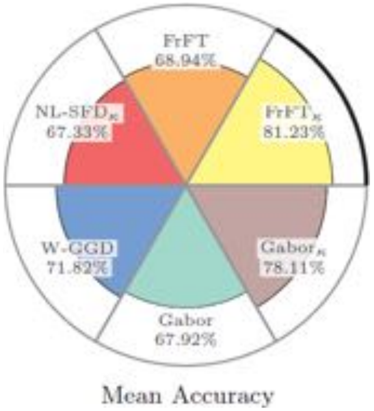


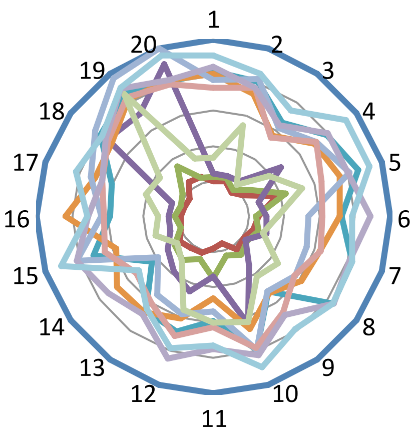


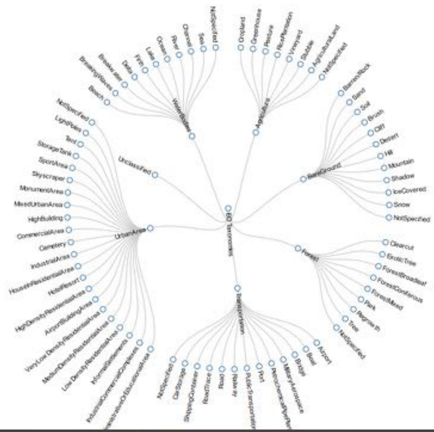


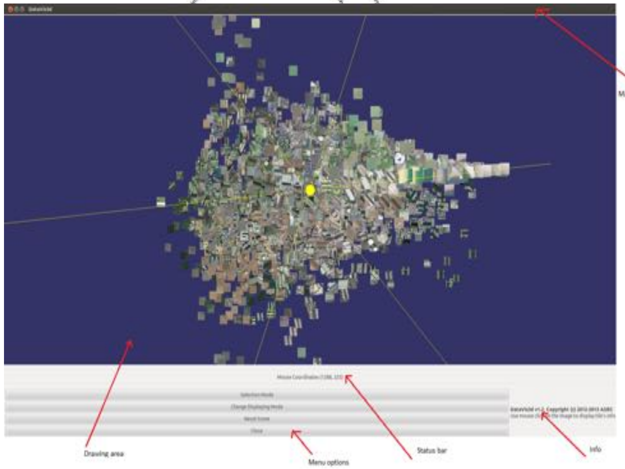
DLR

Leading Edge: Big Data Analytics



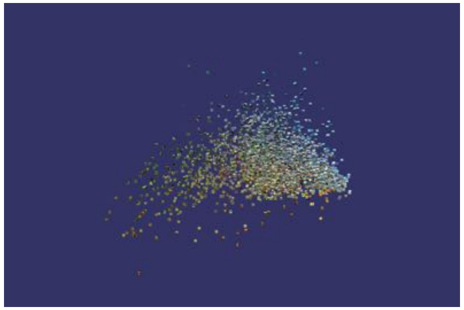


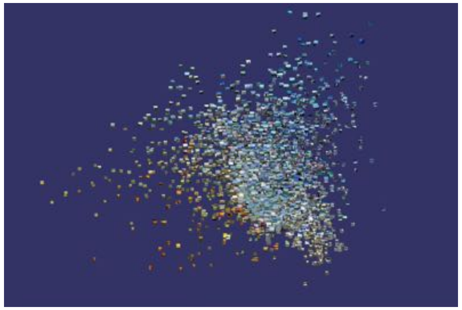


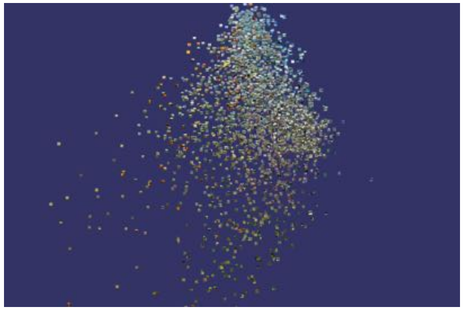


Immersive Visual Information Mining for EO image archives

Navigation inside the EO image collections using the CAVE automatic virtual environment

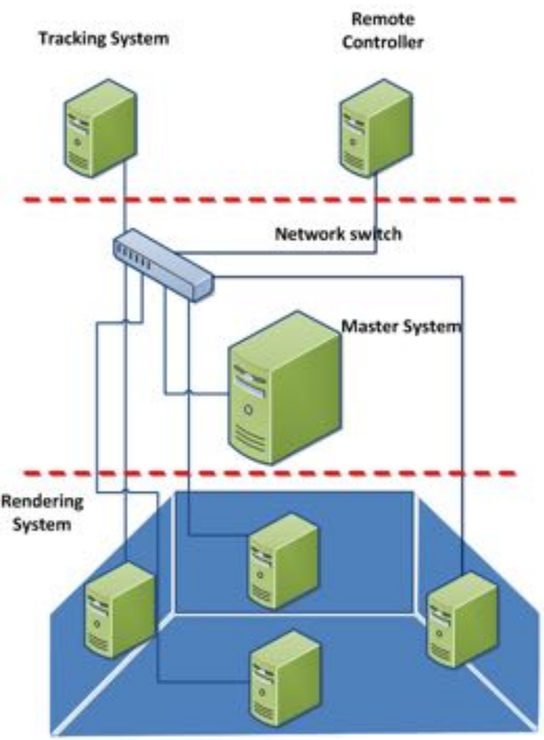


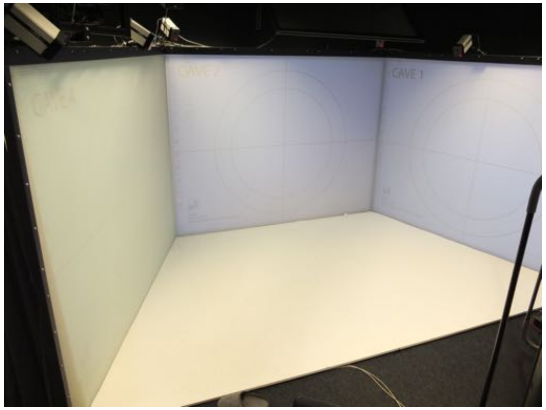




DLR

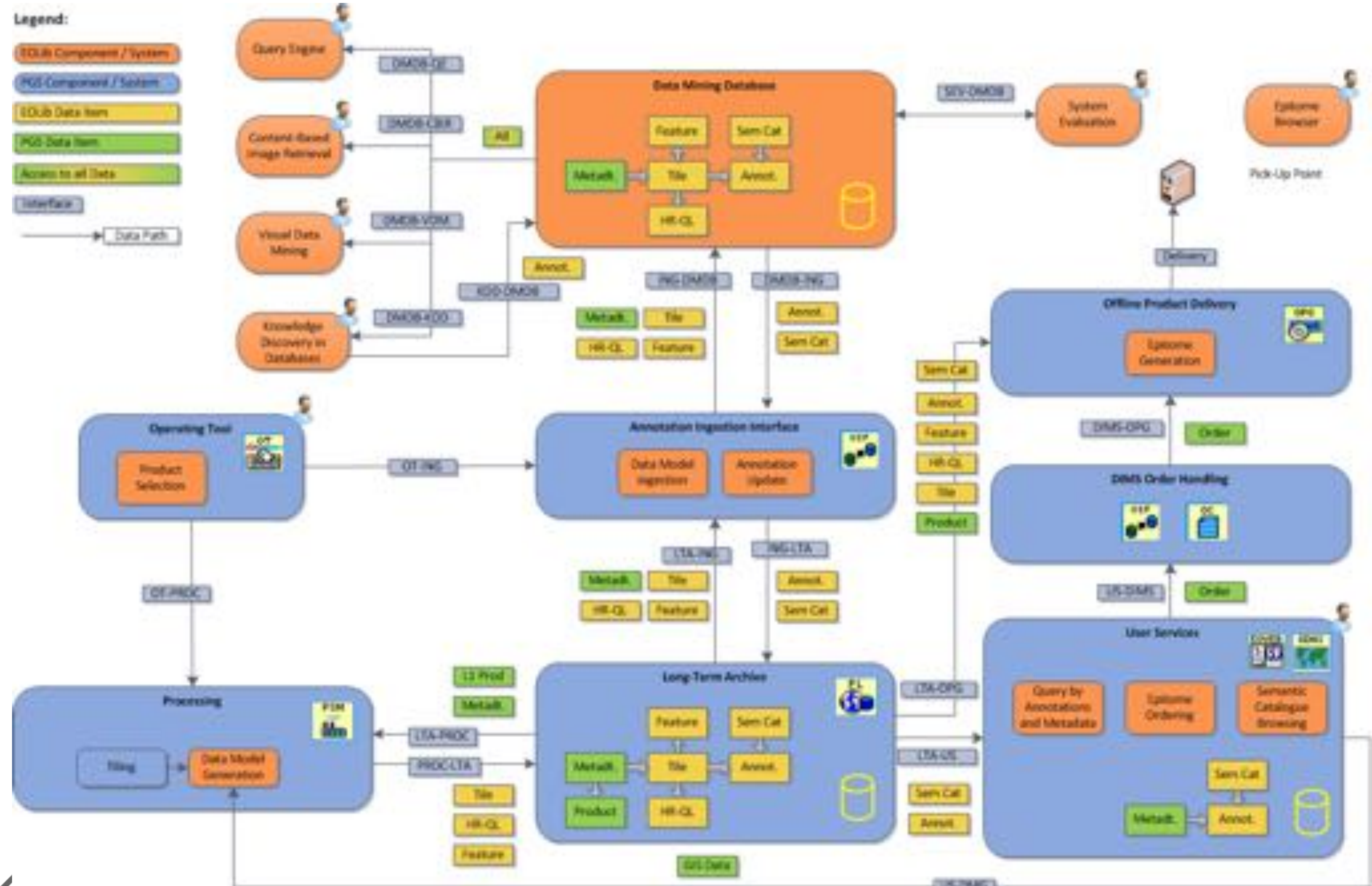
Immersive Visual Information Mining for EO image archives





EOLIB – Earth Observation Image Librarian

- Big Data Mining in the TerraSAR-X Ground Segment System



Selected references

Babaee, M.; Tsoukalas, S.; Rigoll, G.; Datcu, M., "Immersive visualization of visual data using nonnegative matrix factorization." in Neurocomputing, vol.173, pp. 245-255, 2016.

Alonso, K.; Datcu, M., "Accelerated Probabilistic Learning Concept for Mining Heterogeneous Earth Observation Images," in Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol.8, no.7, pp.3356-3371, 2015.

Espinoza-Molina, D.; Nikolaou, C.; Dumitru, C.O.; Bereta, K.; Koubarakis, M.; Schwarz, G.; Datcu, M., "Very-High-Resolution SAR Images and Linked Open Data Analytics Based on Ontologies," in Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol.8, no.4, pp.1696-1708, 2015.

Blanchart, P.; Ferecatu, M.; Shiyong Cui; Datcu, M., "Pattern Retrieval in Large Image Databases Using Multiscale Coarse-to-Fine Cascaded Active Learning," in Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol. 7, no.4, pp.1127-1141, 2014.

Datcu, M.; D'Elia, S.; King, R.L.; Bruzzone, L., "Introduction to the Special Section on Image Information Mining for Earth Observation Data," in Geoscience and Remote Sensing, IEEE Transactions on , vol.45, no.4, pp.795-798, 2007.

Datcu, M.; Seidel, Klaus, "Human-centered concepts for exploration and understanding of Earth observation images," in Geoscience and Remote Sensing, IEEE Transactions on , vol.43, no.3, pp.601-609, 2005.

DLR