DOMAIN KNOWLEDGE FOR REMOTE SENSING

DR. MARIANA BELGIU

UNIVERSITY
OF TWENTE.

THE FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION (ITC)

- Established in 1950 and became a faculty of the University of Twente since 2010.
- Capacity development is ITC's **mission**, where we apply share and facilitate the effective use of knowledge and tools for tackling global wicked problems. We offer education and conduct research.
- Main **domains of expertise** include Resource Security, Disaster Resilience, Geo-health and Geo-Artificial Intelligence.

UNIVERSITY OF TWENTE.

MY PROFILE

Ph.D. in Remote Sensing: “**Formal ontologies** for extracting information from remotely sensed data”

Assistant Professor at ITC/University of Twente

Machine Learning | Deep learning| Multi-Temporal Image analysis

Agriculture mapping & monitoring | Hidden Hunger

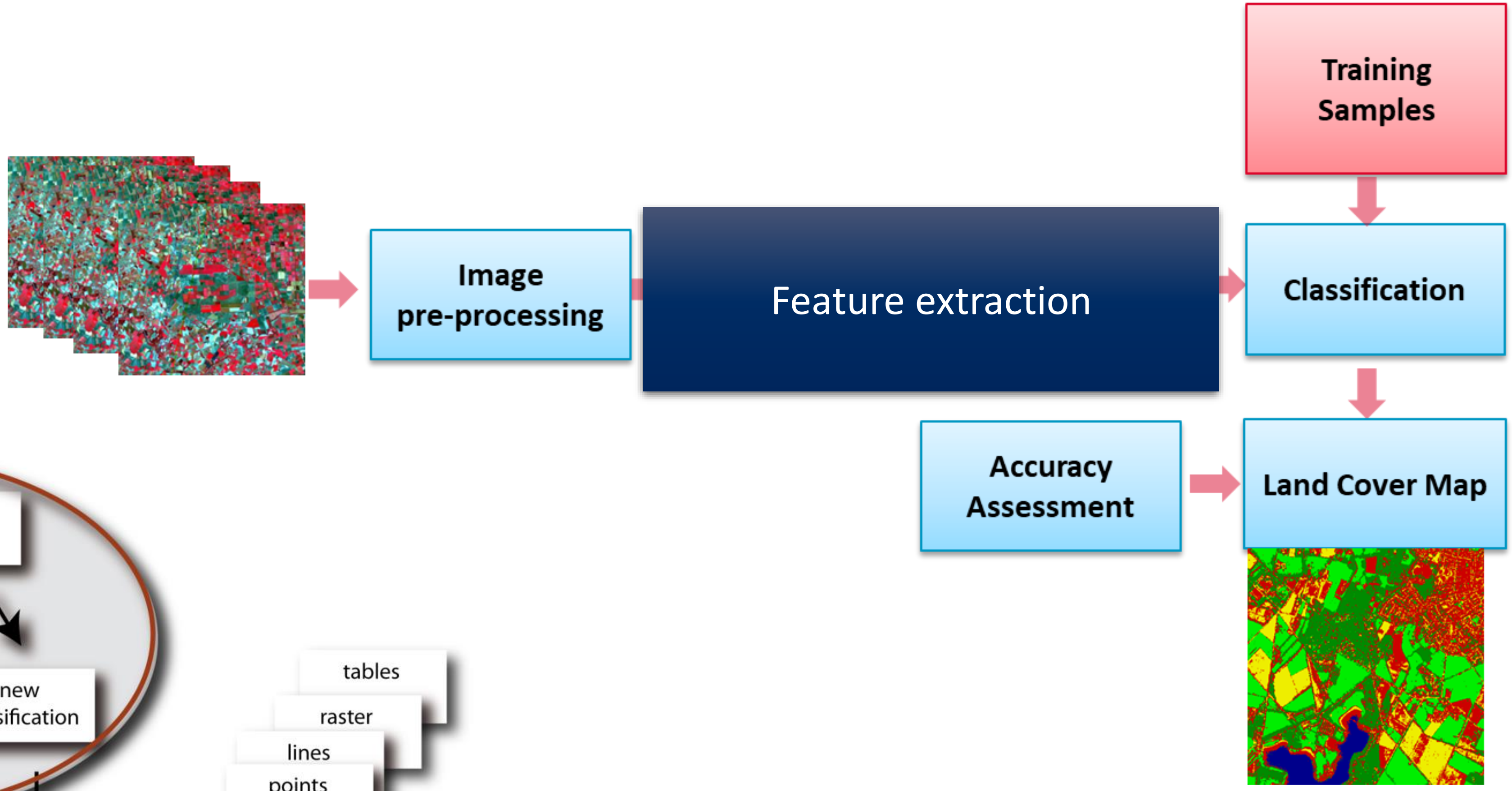
Associate Editor of the ISPRS Journal of Photogrammetry and Remote sensing

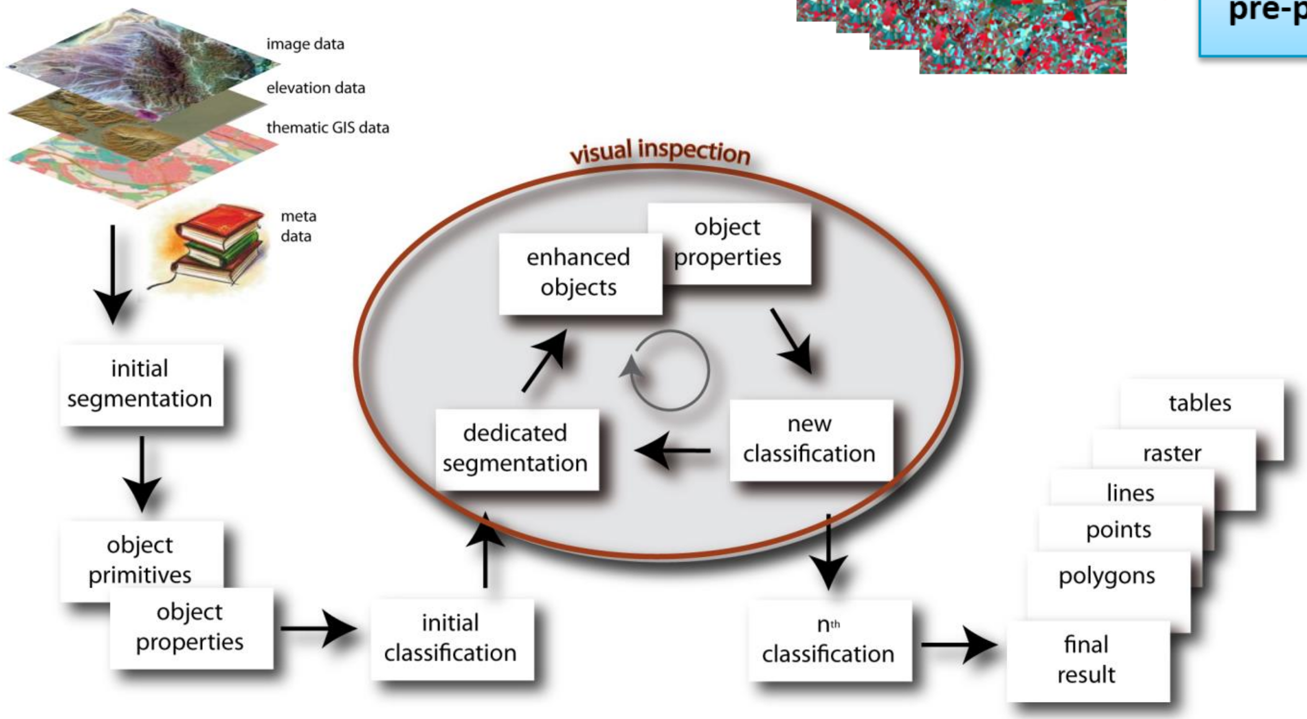
UNIVERSITY
OF TWENTE.

CONTENT

- (Some) Challenges related to remote sensing image analysis
- How can ontologies help?
- How do the current machine learning/deep learning-based approaches account for the identified challenges?

RULE-BASED | MACHINE LEARNING | DEEP LEARNING





Source: P. Hofmann (image used with his permission)

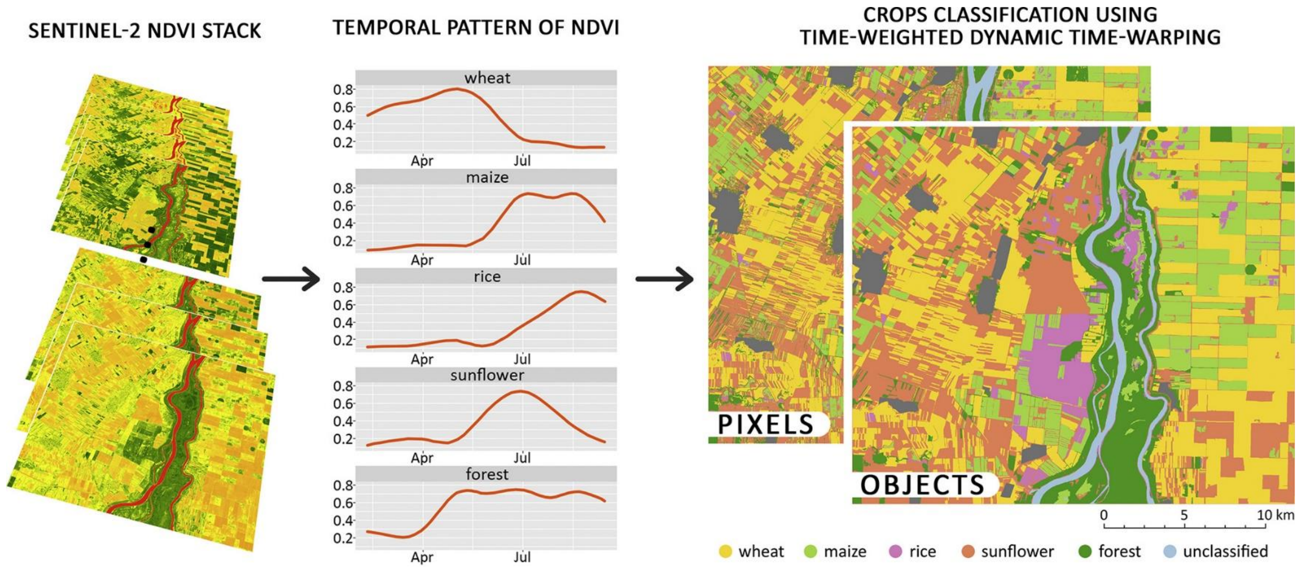
UNIVERSITY OF TWENTE.

DOMAIN KNOWLEDGE FOR REMOTE SENSING ANALYSIS

- Crop semantics
- Agroecological conditions, management practices, and crop calendar
- Types of remote sensing data
- Representative features
- Valid rulesets for the target crops

OR

- Representativeness of the samples
- Spatial and class distribution of labeled samples



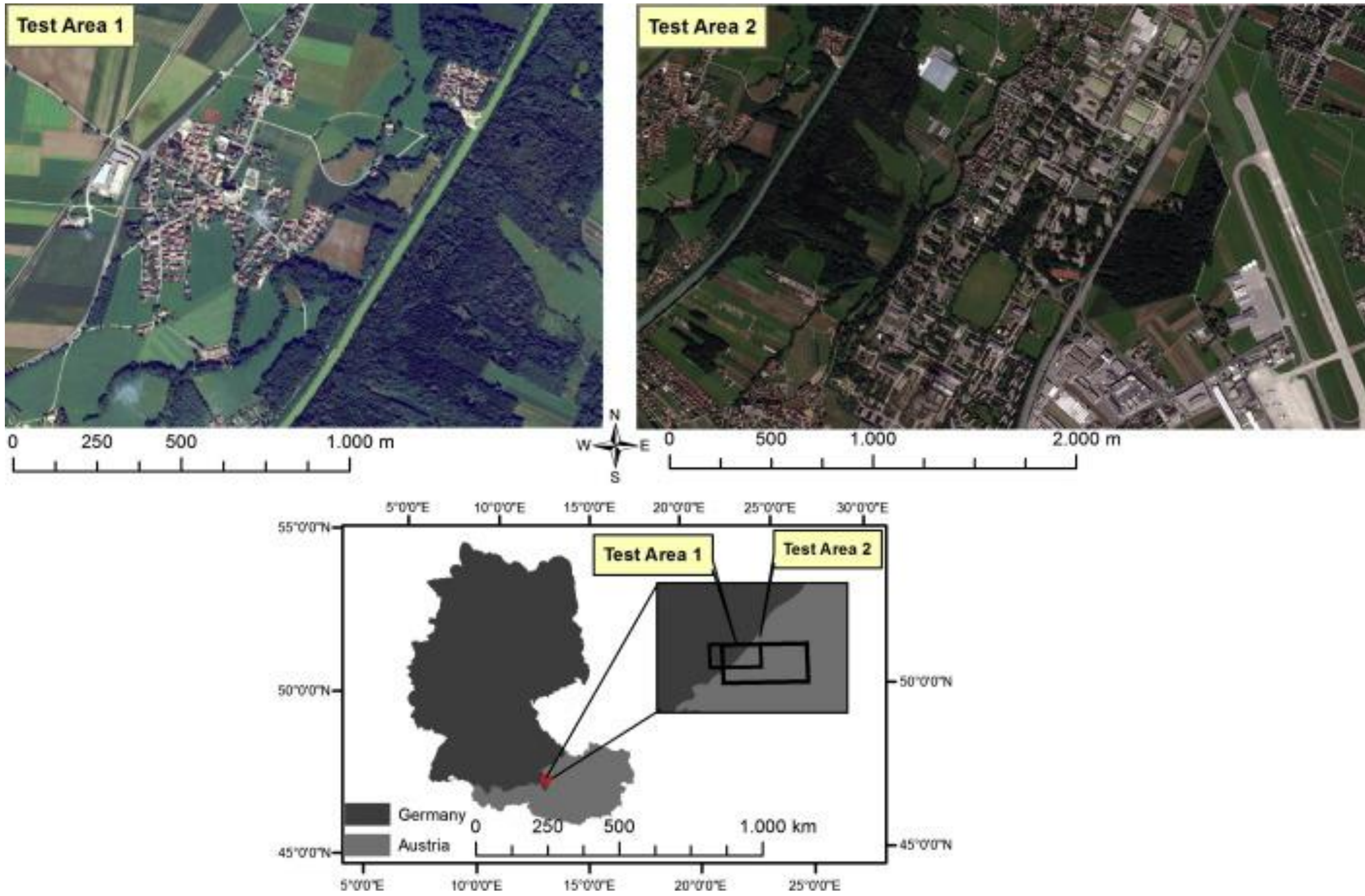
Source: Belgiu et al., 2018

IMAGE ANALYSIS = HIGHLY SUBJECTIVE TASK

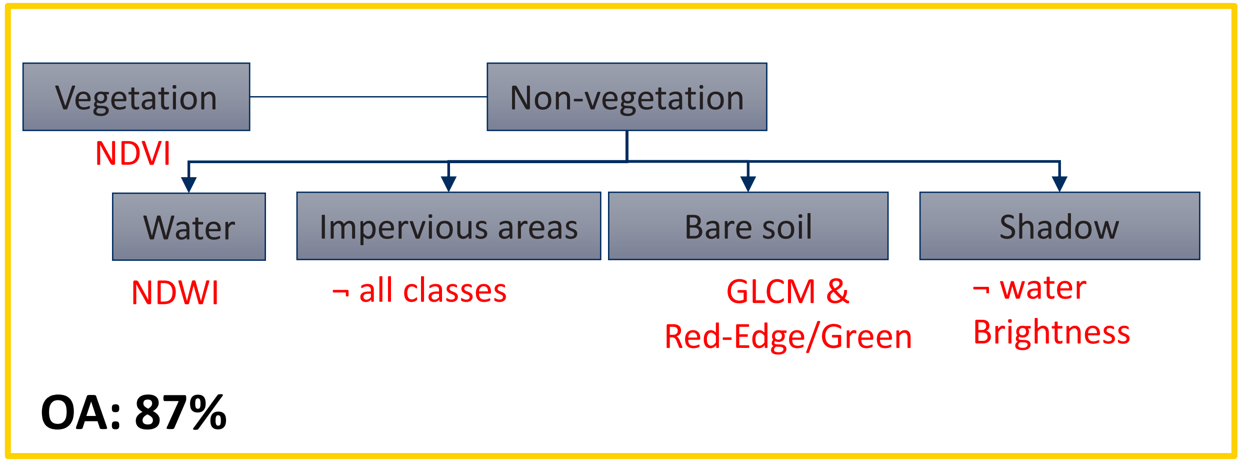
Classification task:
- Vegetation
- Water
- Impervious areas
- Bare soil

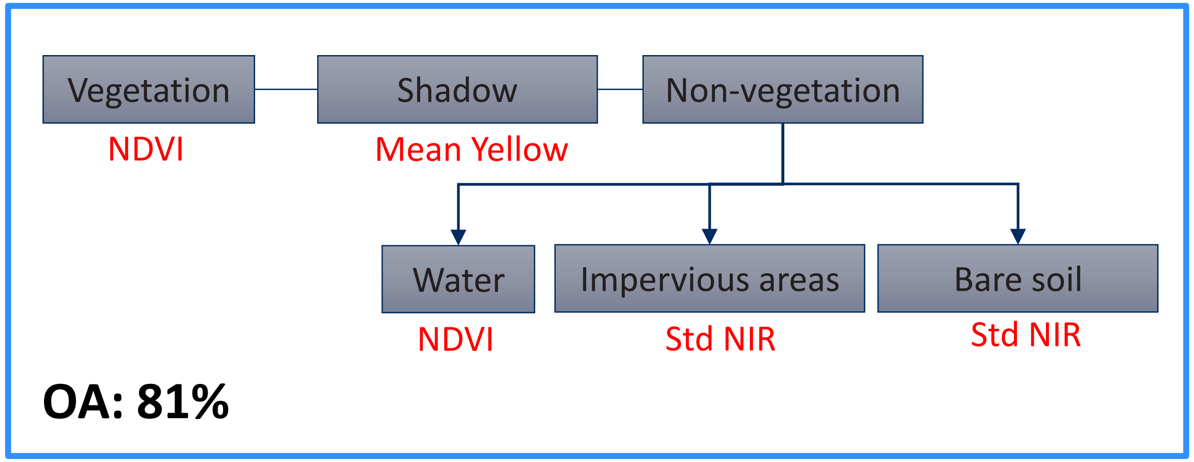
Data: WorldView-2 image (2 m resolution)

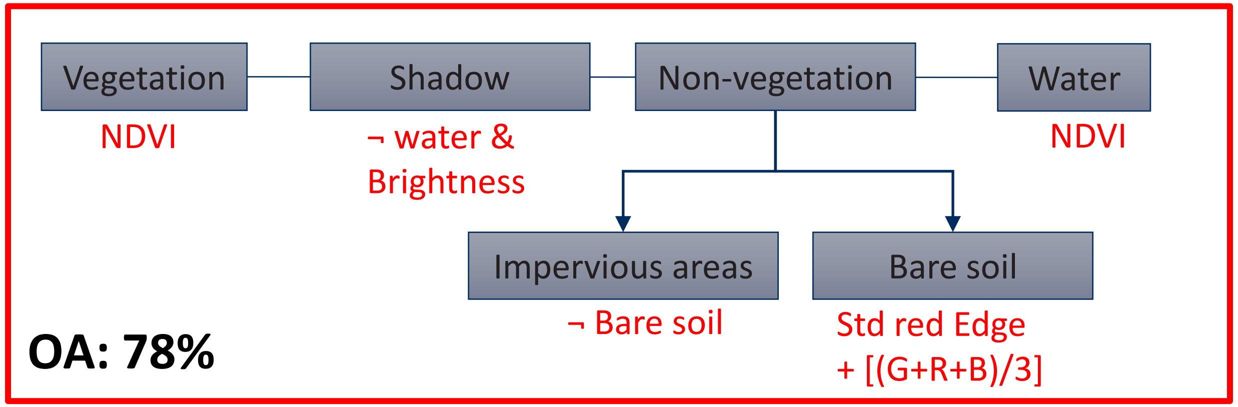
Four experts carried out the classification

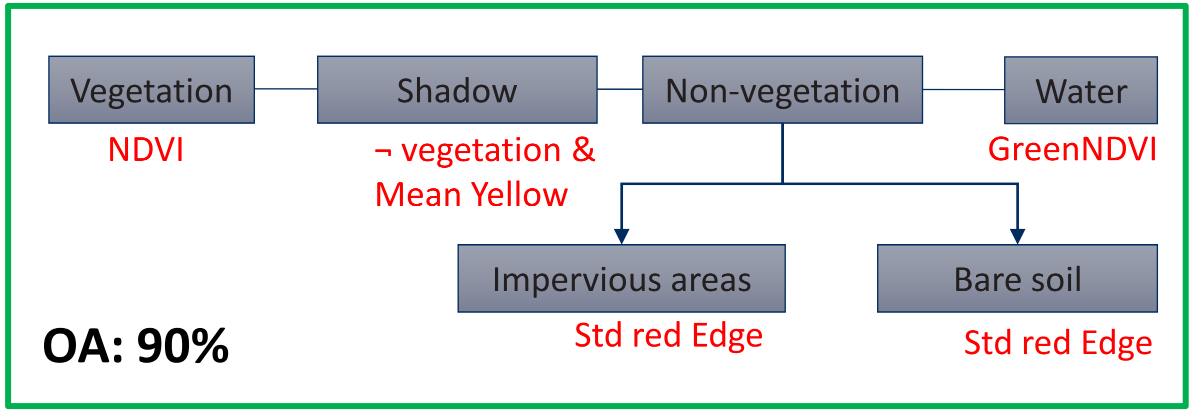


UNIVERSITY OF TWENTE.


Vegetation
NDVI
Non-vegetation
Water
NDWI
Impervious areas
¬ all classes
Bare soil
GLCM &
Red-Edge/Green
Shadow
¬ water
Brightness
OA: 87%


Vegetation
NDVI
Shadow
Mean Yellow
Non-vegetation
Water
NDVI
Impervious areas
Std NIR
Bare soil
Std NIR
OA: 81%


Vegetation
NDVI
Shadow
¬ water &
Brightness
Non-vegetation
Water
NDVI
Impervious areas
¬ Bare soil
Bare soil
Std red Edge
+ [(G+R+B)/3]
OA: 78%


Vegetation
NDVI
Shadow
¬ vegetation &
Mean Yellow
Non-vegetation
Water
GreenNDVI
Impervious areas
Std red Edge
Bare soil
Std red Edge
OA: 90%

RESULTS

Different features used to define the class rulesets

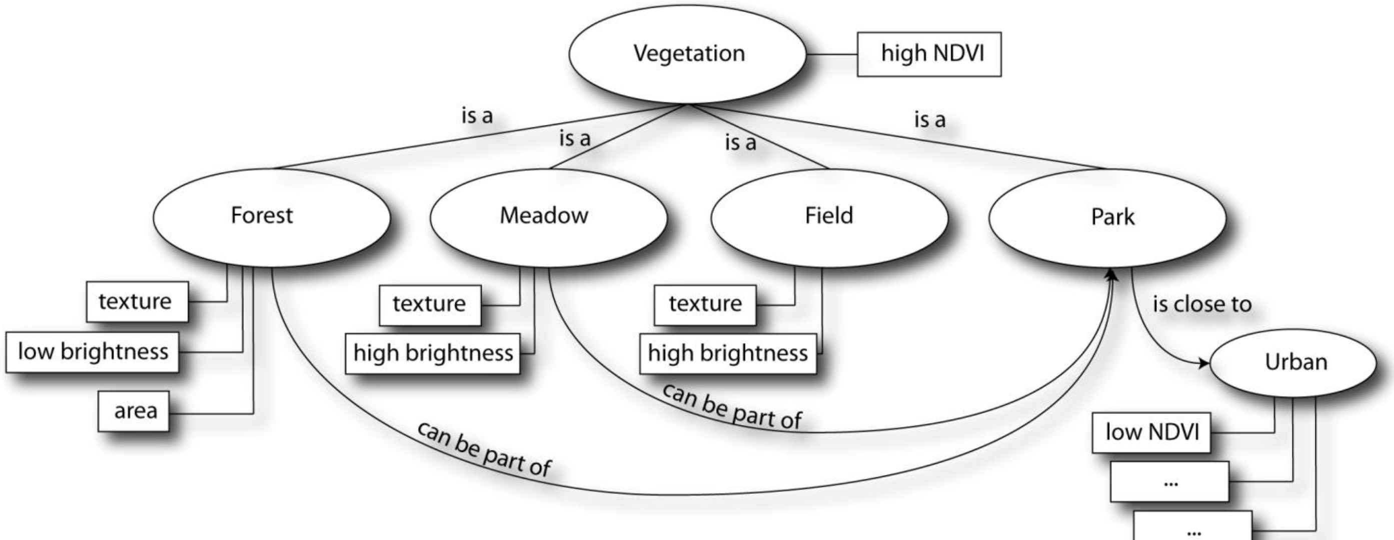
Different thresholds for the common features

Different allocation of classes to the designed hierarchical classification levels

VAGUENESS OF GEOGRAPHIC CONCEPTS

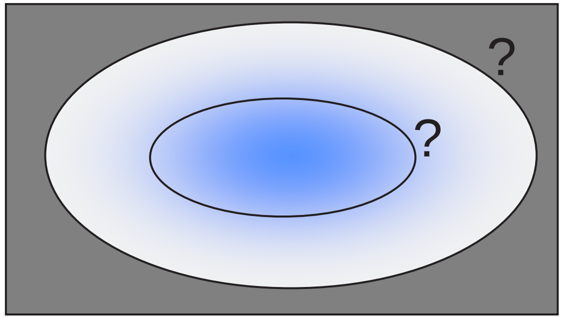
- **Partiality:** one identified feature might not be enough to identify a target class

All types of forests?



Source: P. Hofmann (image used with his permission)

- **Threshold vagueness:**
  - User-defined/context-specific threshold: extensional uncertainty



UNIVERSITY OF TWENTE.

AMBIGUITIES OF GEOGRAPHIC CONCEPTS/SEMANTIC GAP

Forest in Brazil = an area greater than 1 hectare, with more than **30% canopy cover** and a minimum tree height of **5 meters**

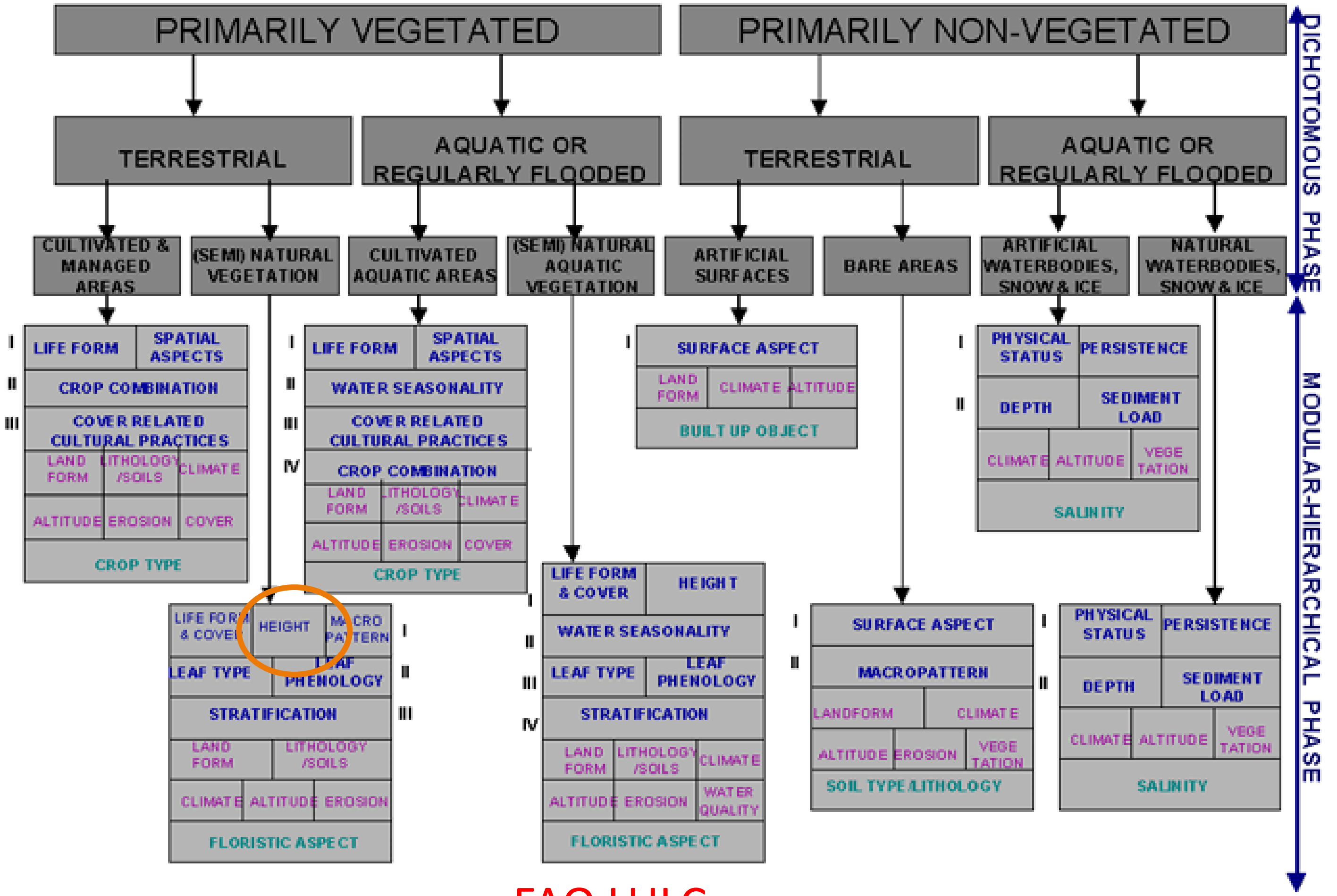
Forest in China = land with a minimum area of 0.67 ha, minimum **20% crown cover** and a minimum tree height of **2 meters**

ESRI 2020 Land cover – 10 m/2020
Trees: Any significant clustering of tall **(~15-m or higher)** dense vegetation, typically with a closed or dense canopy

FROM-GLC10
Forest: Trees observable in the landscape from the images. Tree cover percentage classification to **>15%** and tree height classification to **>3m**

SENSORY GAP

Gap between the entity in the world and the information derived from an image (Smeulders et al. 2000)



FAO LULC

SEMANTIC HETEROGENEITY LULC CLASSIFICATIONS

Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC): 10 m/2017
Cropland = land that has clear traits of **intensive human activity.** Pasture could be transitional from croplands to natural grasslands. Lands for rice cultivation, arable and tillage lands, **greenhouse farming** (build up in WorldCover).

ESRI 2020 Land cover – 10 m/2020
Cropland = human planted/plotted cereals, grasses, and crops not at tree height; examples: corn, wheat, soy, fallow plots of structured land, **oil palm plantation**

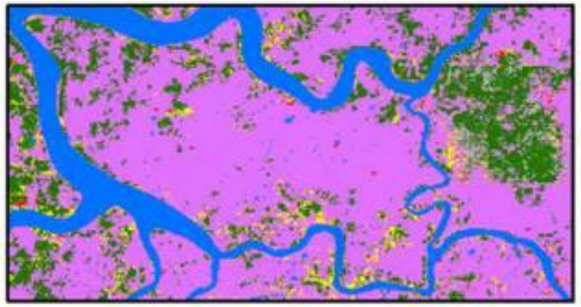
WorldCover 2020
Cropland: land covered with annual cropland that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. Greenhouses are considered as built-up. It includes also rice paddies and irrigated /inundated agriculture (flooded vegetation in Esri 2020 Land cover). **Oil palm plantations are not included.**

DIFFERENT DEFINITIONS → DIFFERENT CLASSIFICATIONS

Rajang river delta, Malaysi

ESRI 2020 Land cover

FROM-GLC-2017



WorldCover 2020

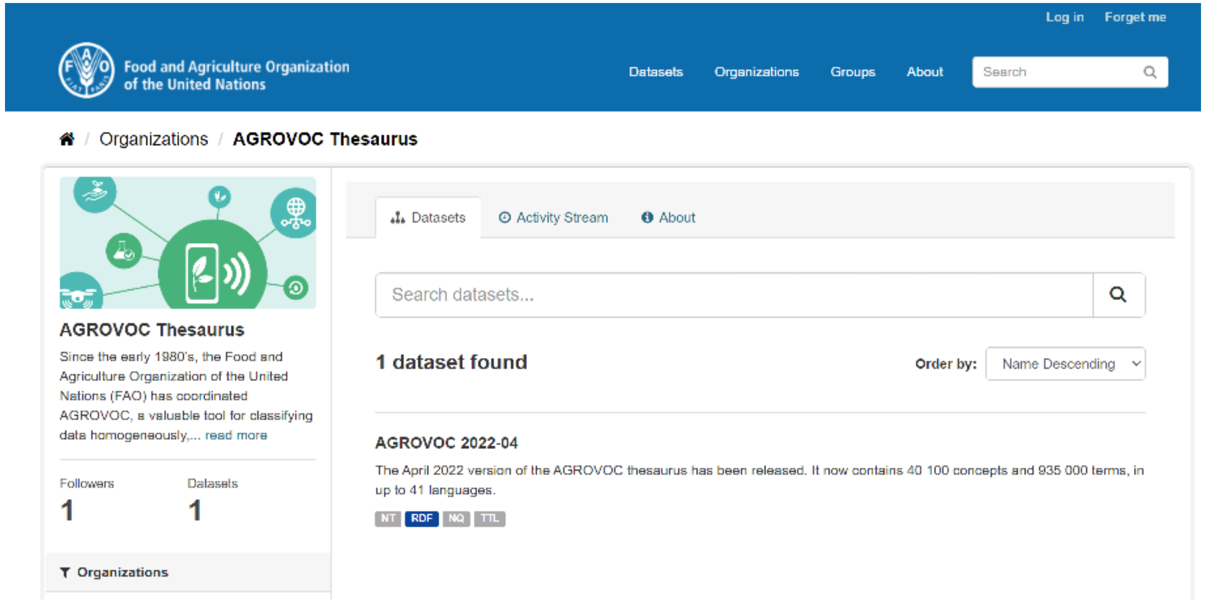
Tree cover	
Shrubland	
Grassland	
Cropland	
Built-up	
Bare / sparse vegetation	
Snow and Ice	
Permanent water bodies	
Herbaceous wetland	
Mangroves	
Moss and lichen	

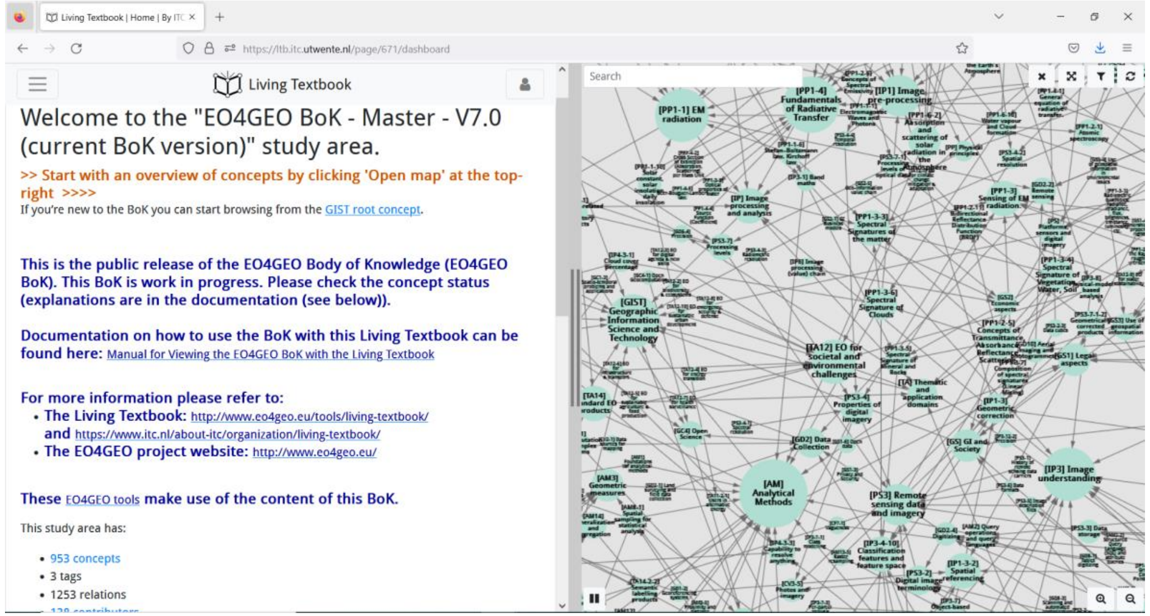
UNIVERSITY OF TWENTE.

ONTOLOGY FOR REMOTE SENSING APPLICATIONS

Definition: Formal explicit specification of a shared conceptualization (Gruber 1995)

- Simple Knowledge Organization System (SKOS)
- Resource Description Framework (RDF)
- Web Ontology Language

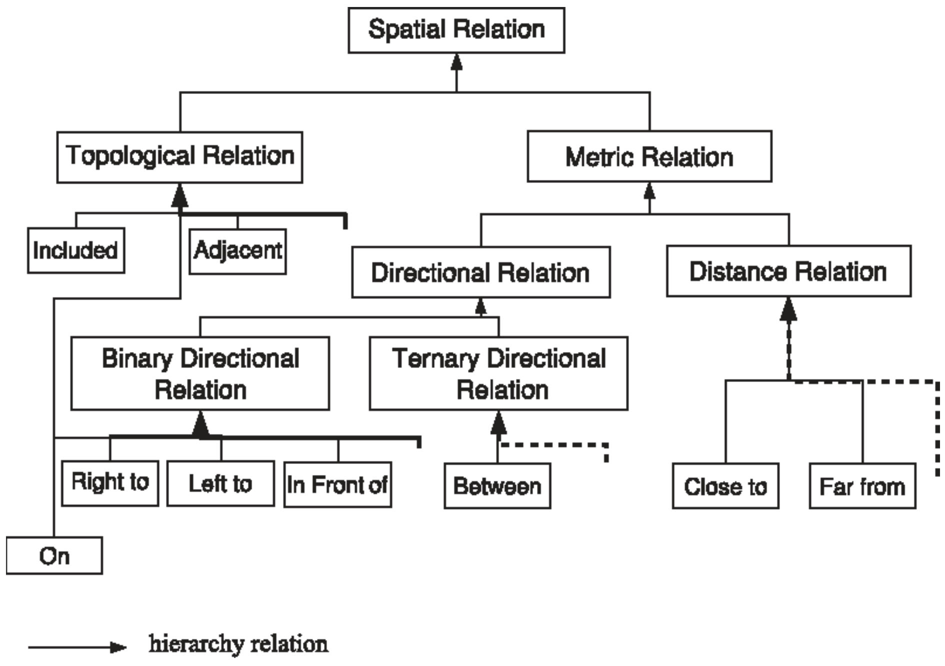




UNIVERSITY OF TWENTE.

ONTOLOGY FOR REMOTE SENSING APPLICATIONS

- Subjectivity
- Vagueness issue:
  - Threshold vagueness: fuzzy ontology
- Ambiguity: explicit specifications of definitions
- Sensory gap: making it explicit
- Semantic heterogeneity: interoperability



Source: Hudelot et al, 2008

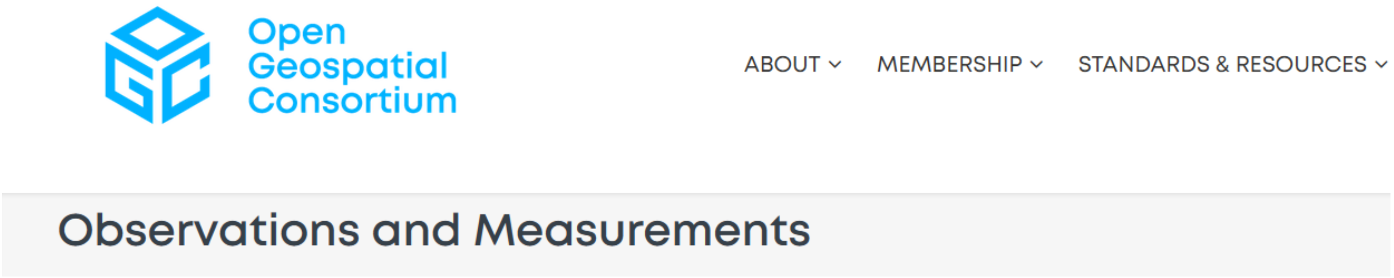
ONTOLOGY FOR REMOTE SENSING APPLICATIONS

Content-based image retrieval

Interpretation of satellite images

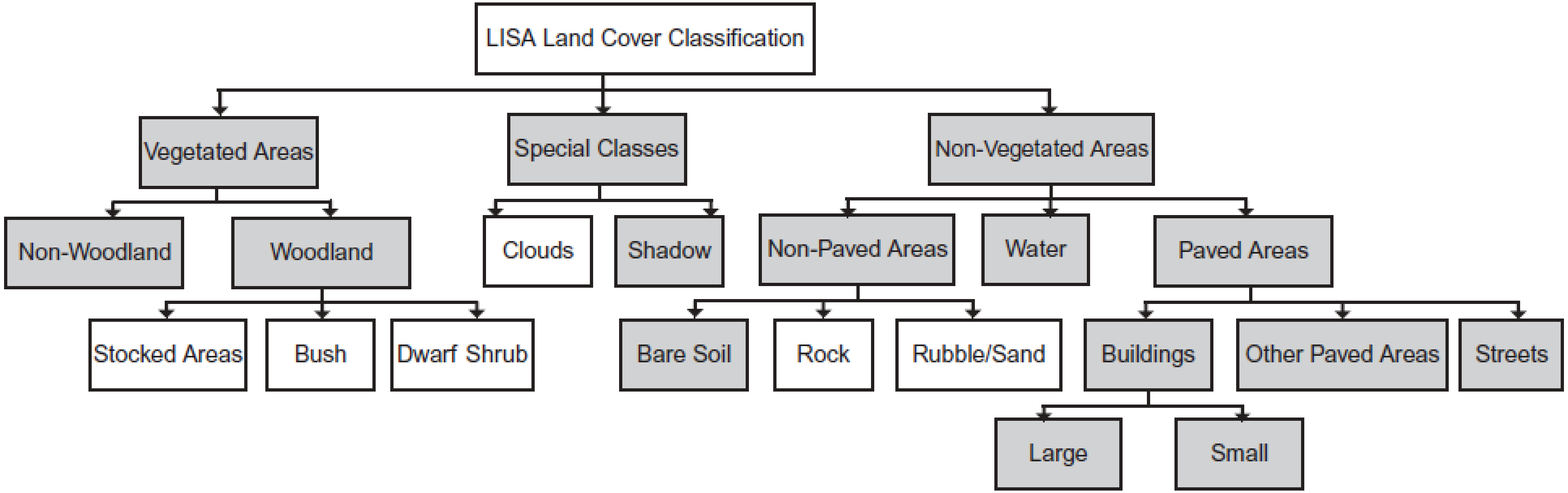
How to develop an ontology?

- Top-down
- Bottom-up (Janowicz, 2012)



USING ONTOLOGIES IN THE OBIA FRAMEWORK (I)

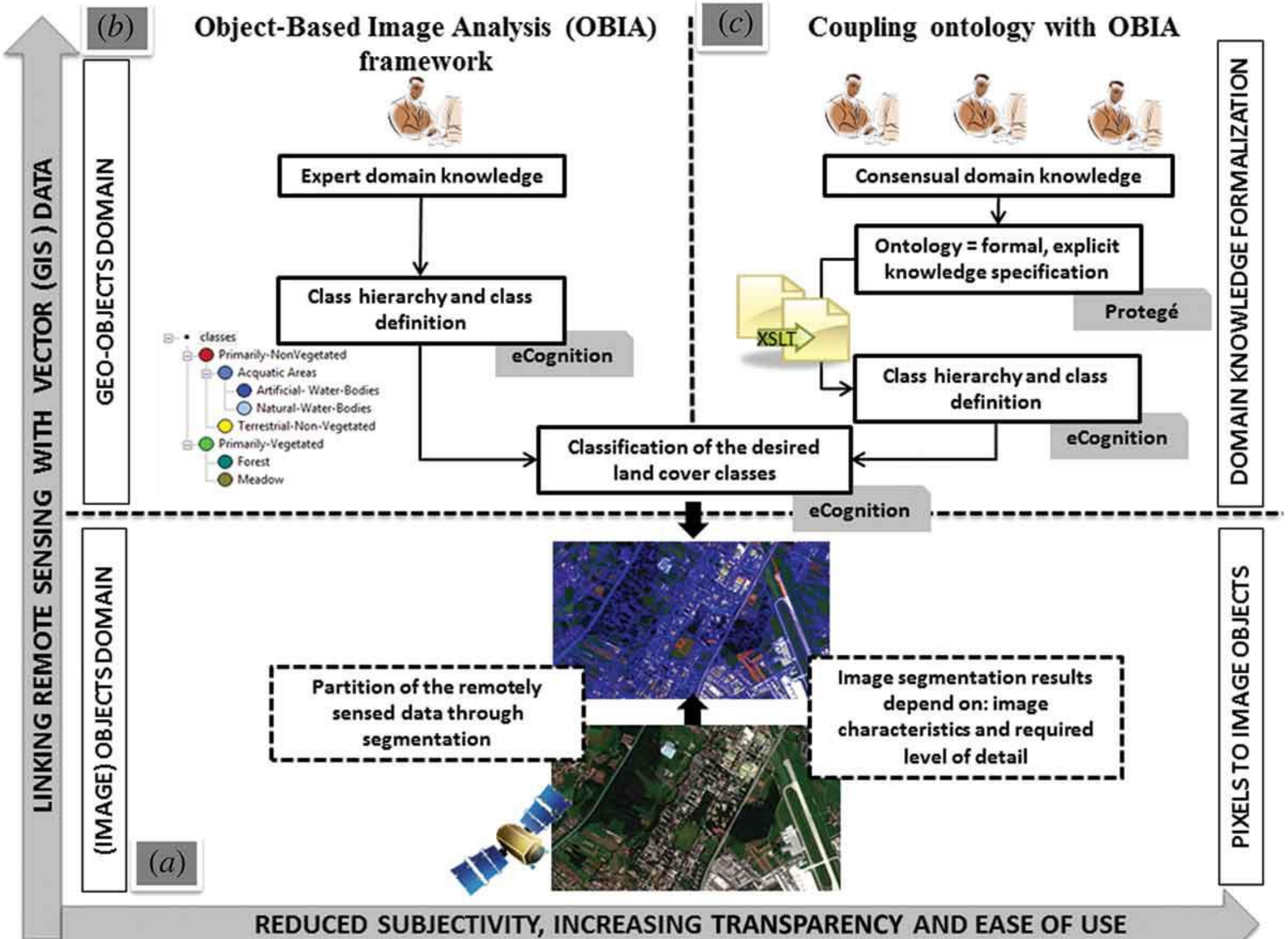
- Land Information System Austria (LISA)



Source: Belgiu et al., 2014

- Symbolic knowledge extended with numeric knowledge elicited from literature

USING ONTOLOGIES IN THE OBIA FRAMEWORK (II)

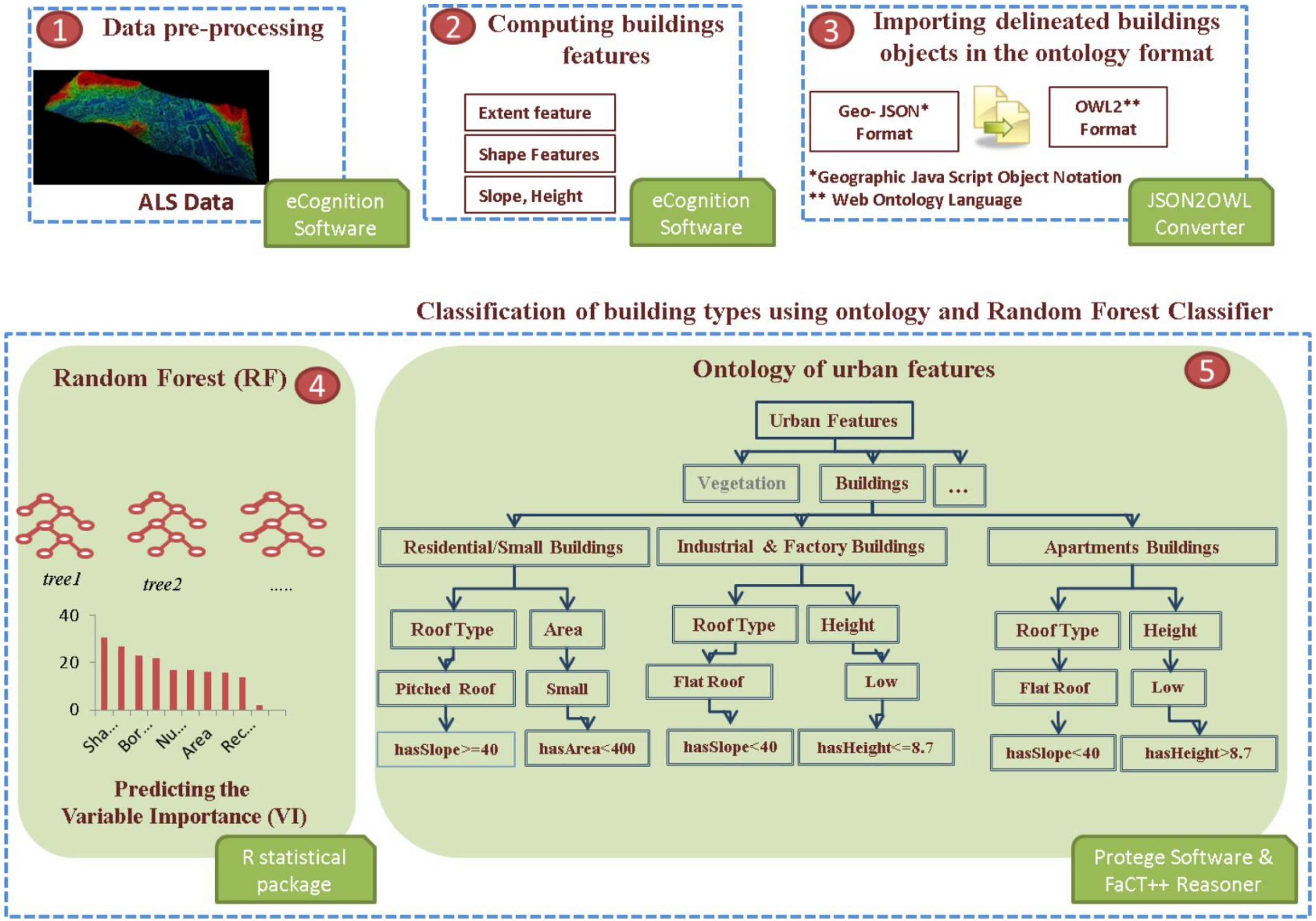


UNIVERSITY OF TWENTE.

ADVANTAGES OF INTEGRATING ONTOLOGIES IN OBIA

- Reducing subjectivity
- Vagueness issue:
- Ambiguity/Semantic gap: explicit specifications of LULC definitions and their representation in the image
- Sensory gap
- Semantic heterogeneity/reusability across communities

BUILDING TYPE CLASSIFICATION USING AIRBORNE LASER SCANNING DATA



UNIVERSITY OF TWENTE.

BUILDING TYPE CLASSIFICATION USING AIRBORNE LASER SCANNING DATA

Conceptualization

Formalization

ResidentialSmall−Buildings≡Buildings∩hasRoofType.PitchedRoof∩hasArea.Small AreaApartment Buildings≡Buildings∩hasRoof.FlatRoof∩hasHeight.Low
Industrial and Factory Buildings≡Buildings∩hasRoof.FlatRoof∩hasArea.LargeArea ∩hasHeight.Low

UNIVERSITY OF TWENTE.

ADVANTAGES OF BOTTOM-UP APPROACHES

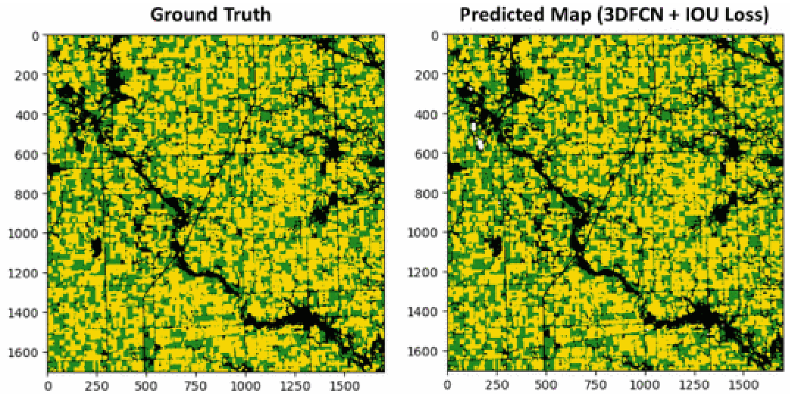
- Reducing subjectivity
- Vagueness issue
- Ambiguity/Sematic gap: explicit specifications of building type definitions and their representation in the image
- Sensory gap: making it explicit
- Semantic heterogeneity/reusability across communities

DOMAIN KNOWLEDGE AND MACHINE LEARNING

Vagueness: partiality

- **Intra-class heterogeneity =** do the annotated labels capture it?

Fairness = are the annotated labels representative of all classes?



ELSEVIER

Remote Sensing of Environment
Volume 233, November 2019, 111375

Needle in a haystack: Mapping rare and infrequent crops using satellite imagery and data balancing methods

François Waldner [a], Yang Chen [b], Roger Lawes [c], Zvi Hochman [a]

Vagueness: Threshold

- Fuzzy Random Forest

Ambiguities/Semantic gap= Do the annotated labels used to train a machine learning/deep learning model matches the semantics of the target class?

UNIVERSITY OF TWENTE.

DOMAIN KNOWLEDGE AND MACHINE LEARNING

Transferability = where to transfer to?

ELSEVIER

International Journal of Applied Earth Observation and Geoinformation

Volume 114, November 2022, 103054

Assessing the generalization capability of deep learning networks for aerial image classification using landscape metrics

Caroline M. Gevaert, Mariana Belgiu

DO WE (STILL) NEED ONTOLOGIES?

Data-centric AI (Andrew Ng)

1. Using the appropriate labels
2. Getting rid of the noisy data instances
3. Data augmentation
4. Analysis of error
5. Employing domain experts to identify the accuracy or inaccuracy in data points
6. engineering the data needed to build a successful AI system