FOSS4G and GIScience: trajectories and intersections

Gilberto Camara

https://gilbertocamara.org

image: Alamy

1984

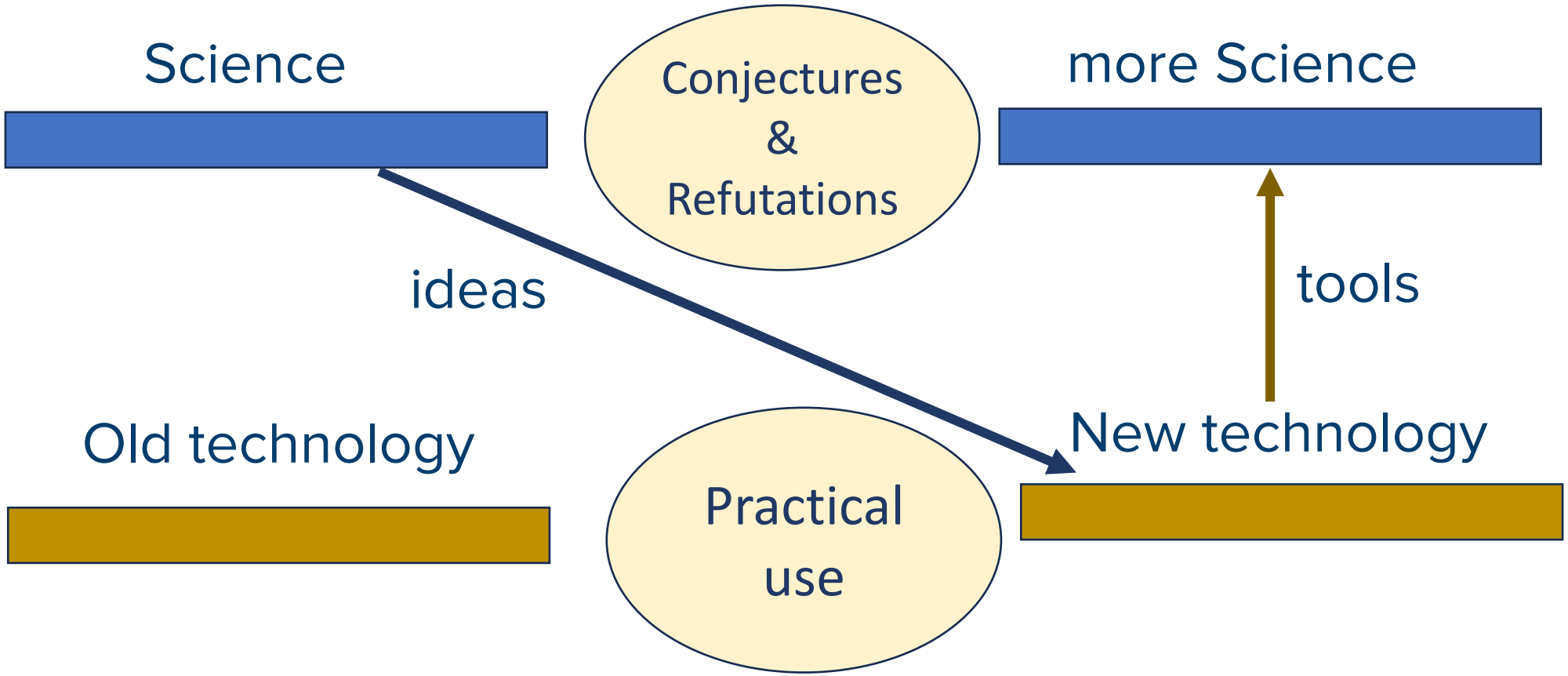
1988

Science and technology

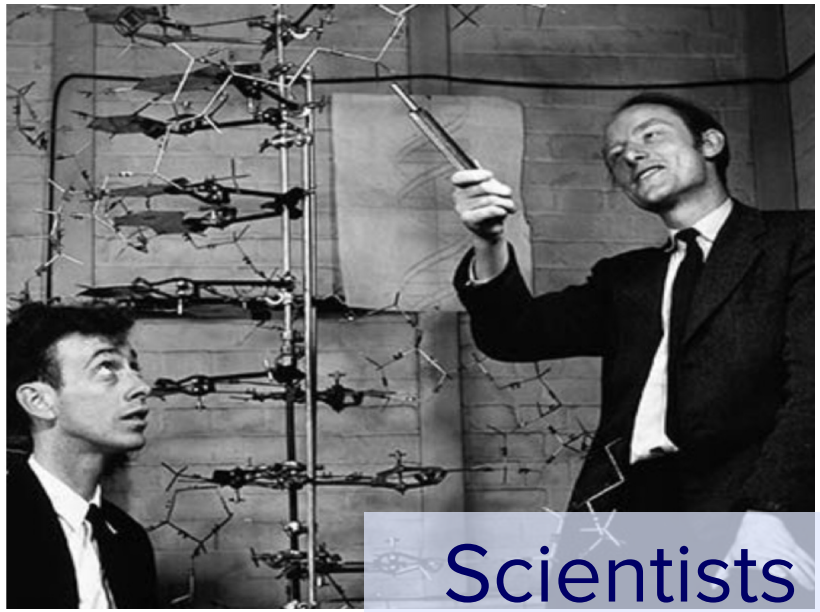
- Science
  - source of ideas for new technology;
  - creation of a knowledge base.

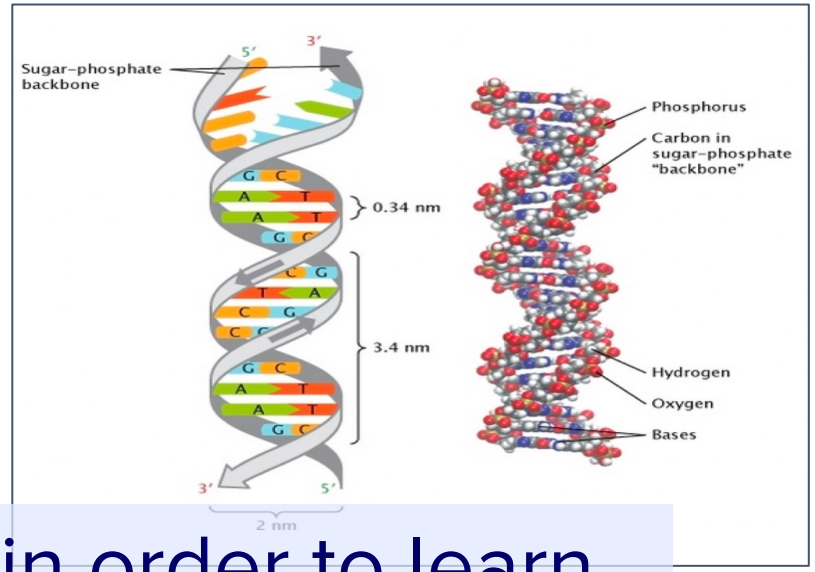
- Technology
  - Use of knowledge to solve problems and use resources;
  - Source of tools and techniques to improve Science

Science and technology



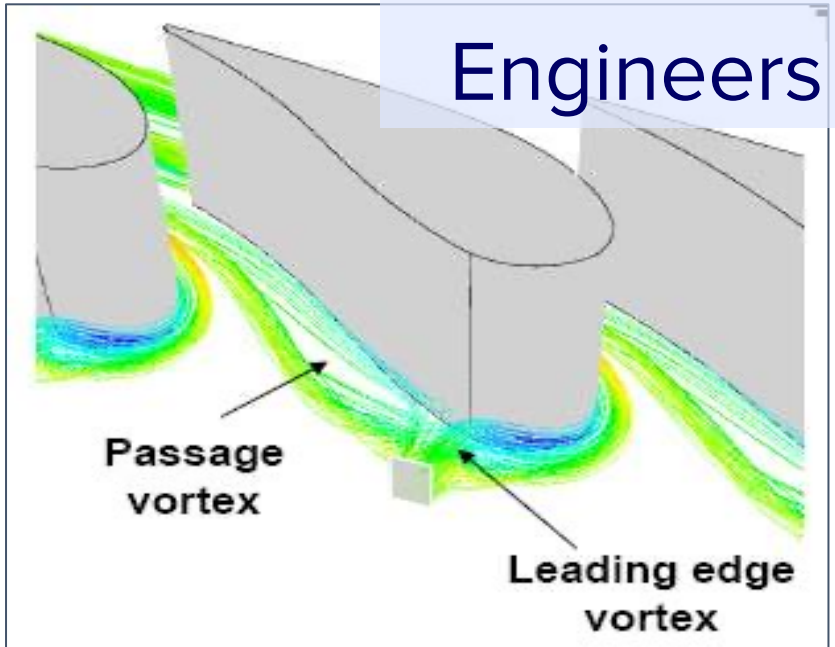
Scientists and engineers





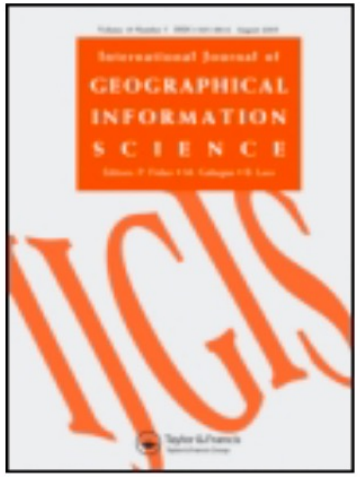
Scientists build in order to learn

Engineers learn in order to build



GIScience and FOSS4G

- Which GIScience ideas have influenced FOSS4G?
- Which FOSS4G tools have enabled better GIScience?
- Which GIScience ideas could be relevant to the future of FOSS4G?



Taylor & Francis
Taylor & Francis Group

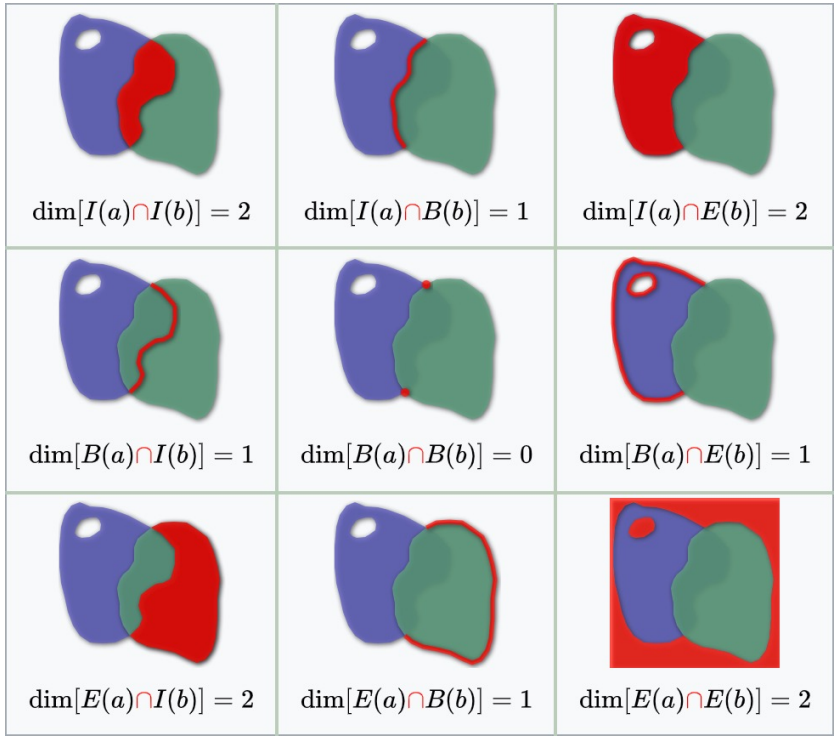
International Journal of Geographical Information System

ISSN: 0269-3798 (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis19

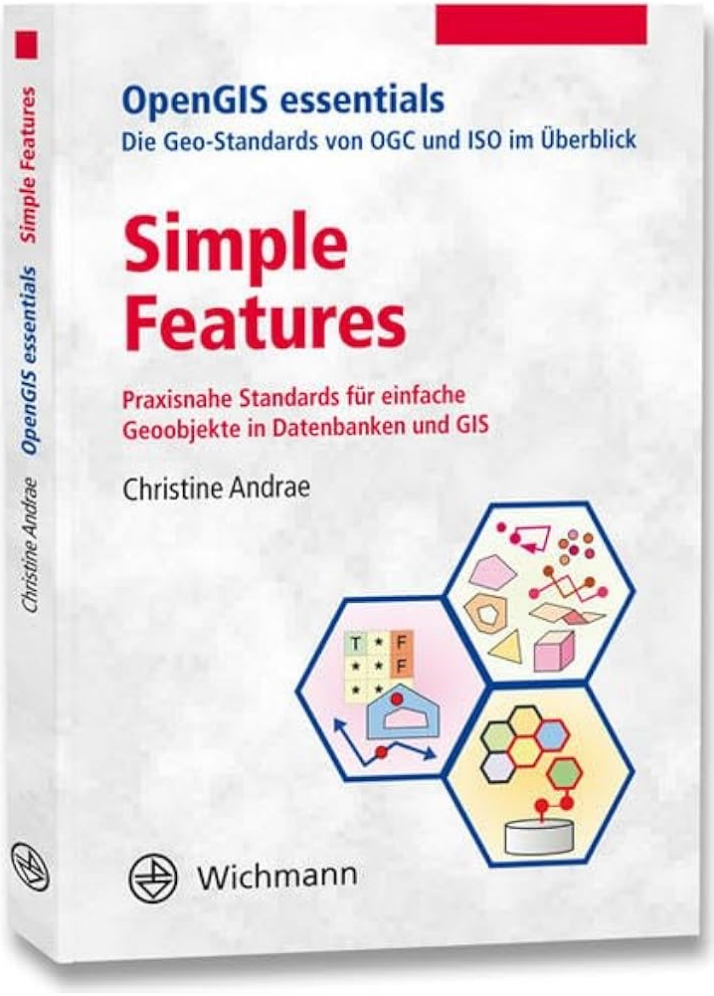
Point-set topological spatial relations

MAX J. EGENHOFER & ROBERT D. FRANZOSA

4-intersection paper

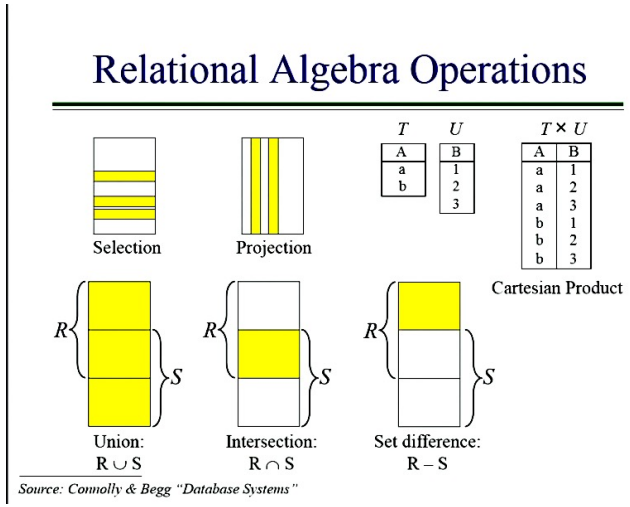


DE-9IM

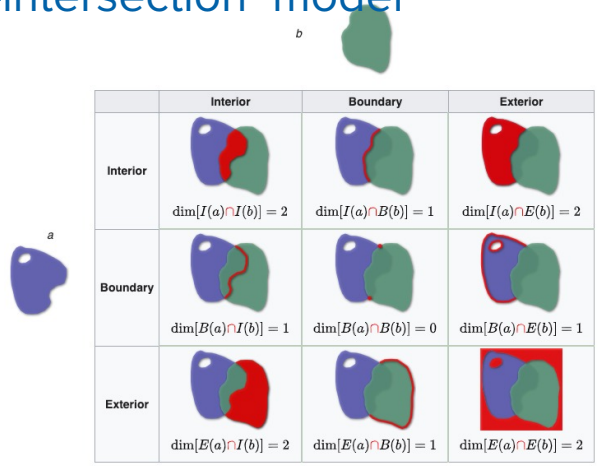


FOSS4G: PostGIS, JTS, R-sf, Geopandas, ...

Standards need theory
(and theory can come from good practices)

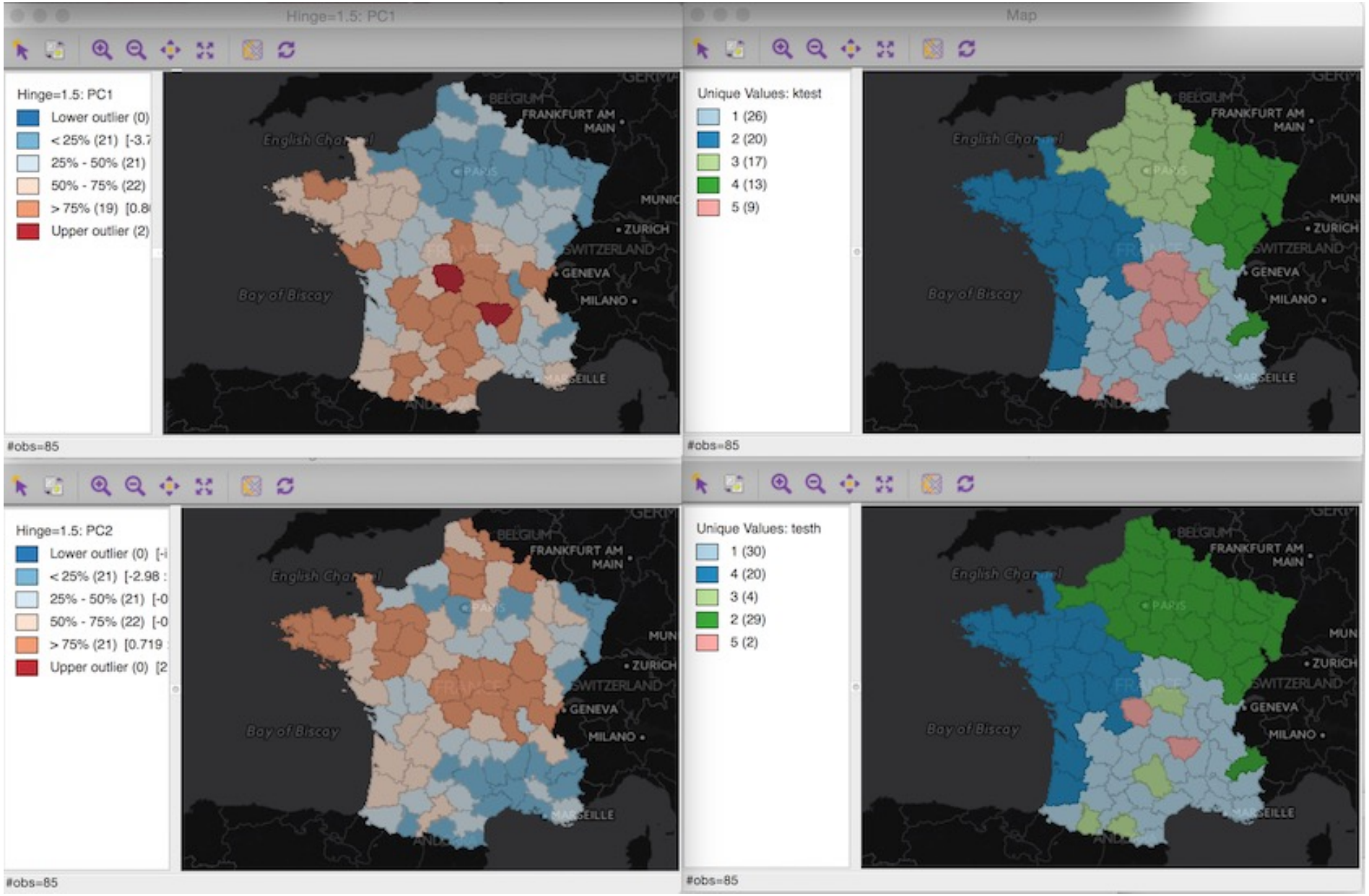


Dimensionally extended
9-intersection model

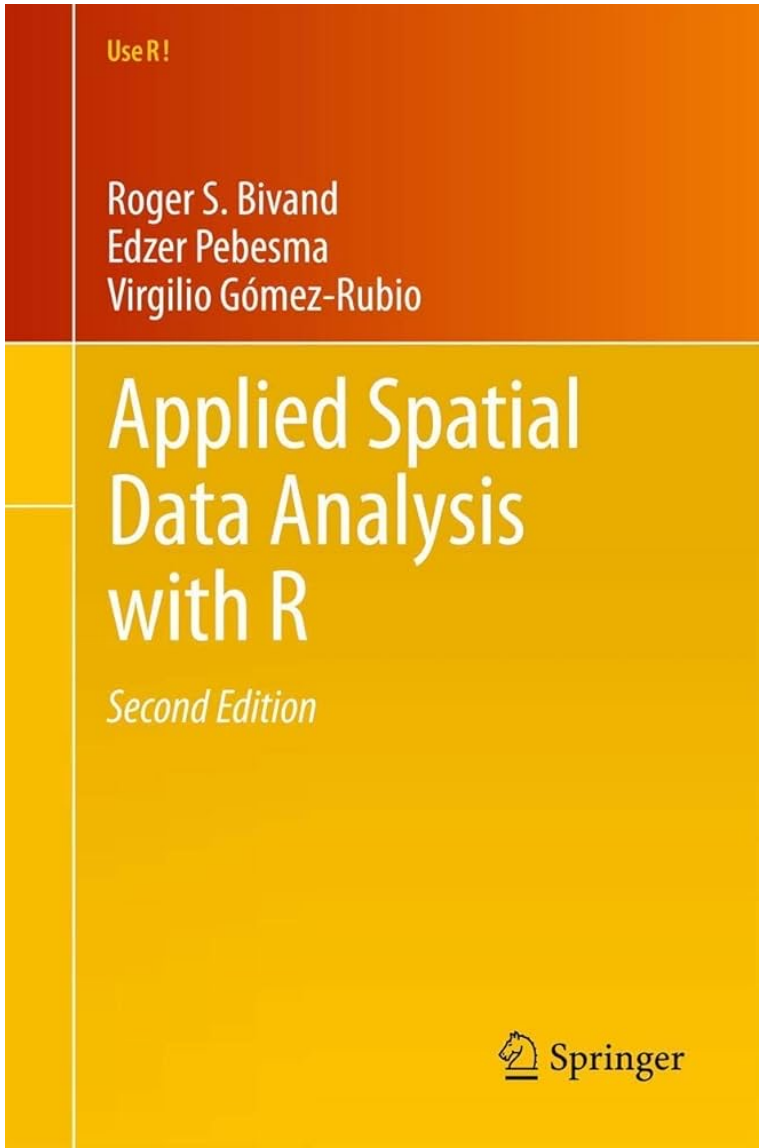


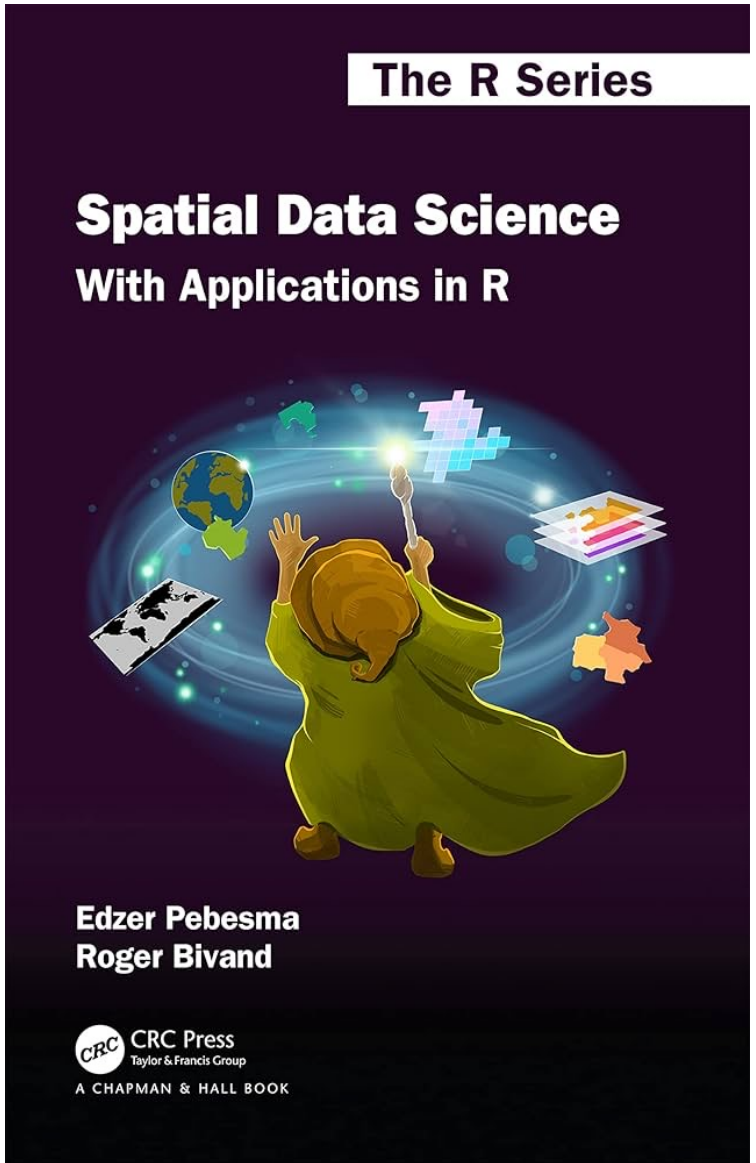
SF-SQL, GeoSPARQL, WFS, …

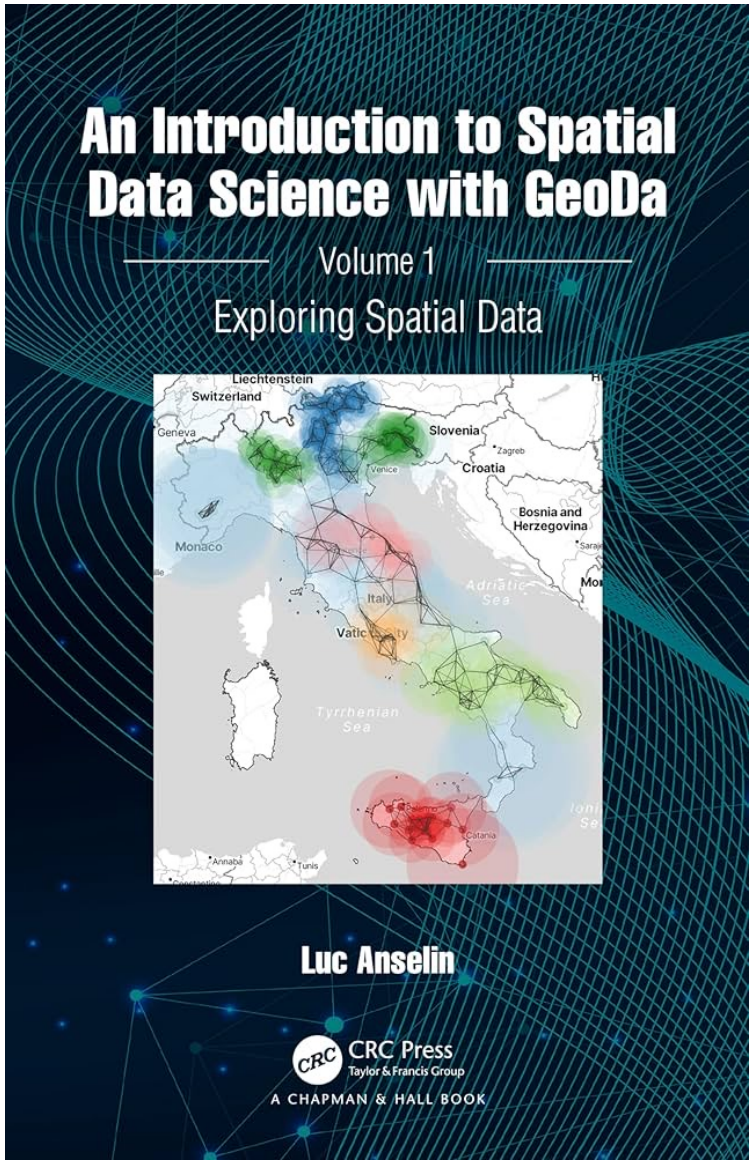
Spatial Analysis



Spatial Analysis: from academia to FOSS4G

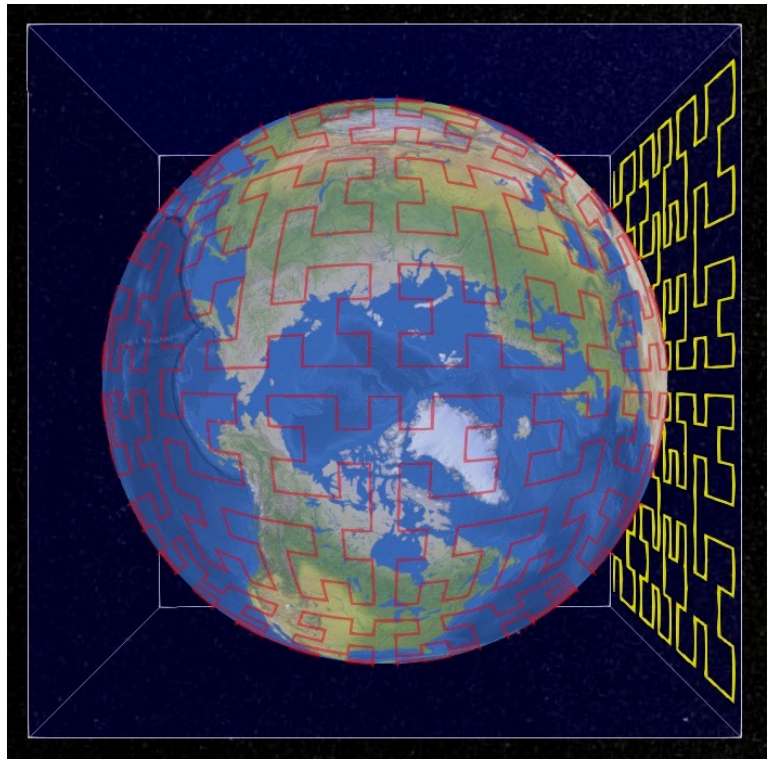






From FOSS4G to academia

Google S2 geometry lib



GIScience and FOSS4G

- What GIScience ideas have influenced FOSS4G?
- What FOSS4G tools have enabled better GIScience?
- What GIScience ideas could be relevant to the future of FOSS4G?

Geospatial Semantics: Why, of What, and How?

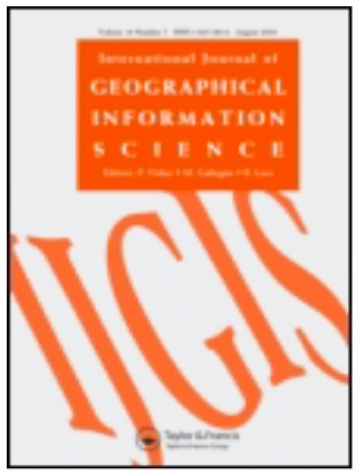
Werner Kuhn
Institute for Geoinformatics, University of Münster,
Robert-Koch-Str. 26-28, D-48151 Münster, Germany
kuhn@uni-muenster.de

Abstract. Why are notions like semantics and ontologies suddenly getting so much attention, within and outside geospatial information communities? The main reason lies in the componentization of Geographic Information Systems (GIS) into services, which are supposed to interoperate within and across these communities. Consequently, I look at geospatial semantics in the context of semantic interoperability. The paper clarifies the relevant notion of semantics and

SPATIAL COGNITION AND COMPUTATION, *4*(1), 1–end

Fields and Objects in Space, Time, and Space-time

Antony Galton
University of Exeter



Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Event-oriented approaches to geographic phenomena

Michael Worboys

One Step up the Abstraction Ladder: Combining Algebras - From Functional Pieces to a Whole

Andrew U. Frank

Department of Geoinformation
Technical University Vienna
Gusshausstr. 27-29, A-1040 Vienna, Austria
frank@geoinfo,tuwien.ac.at

Earth observation data analysis in the age of big data

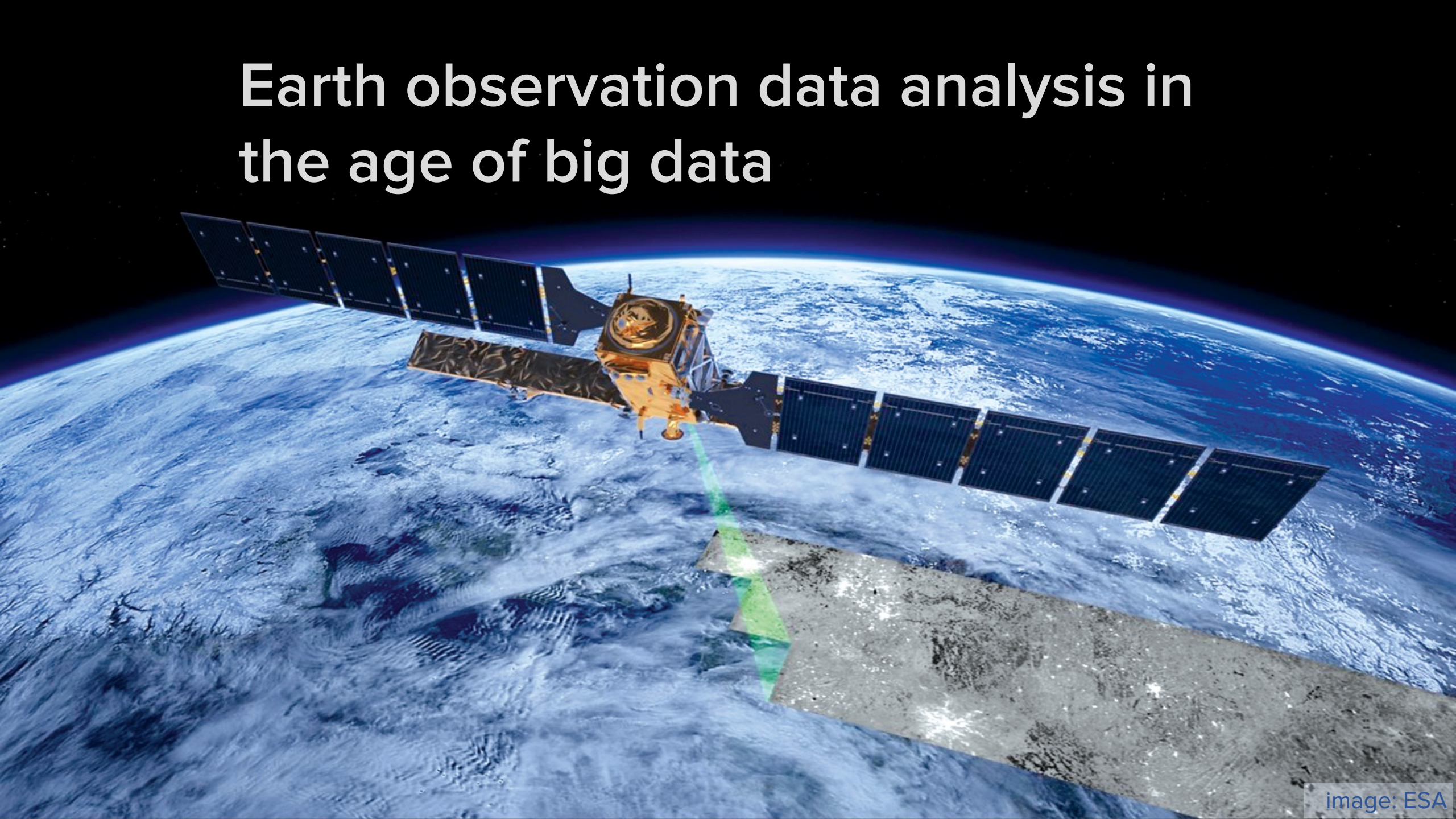
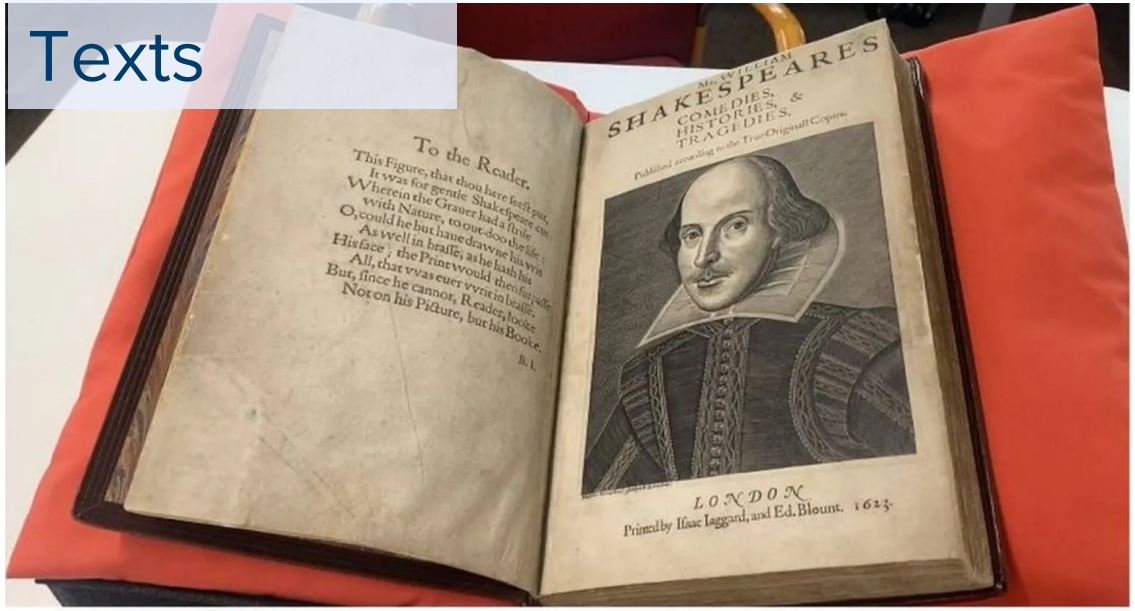
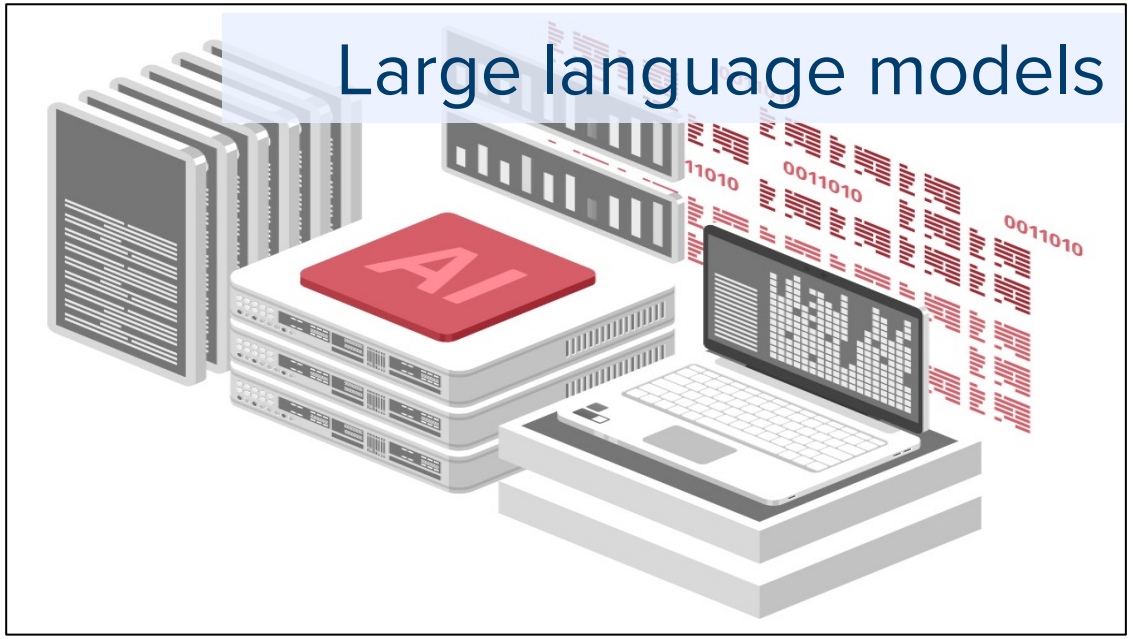


image: ESA

Texts

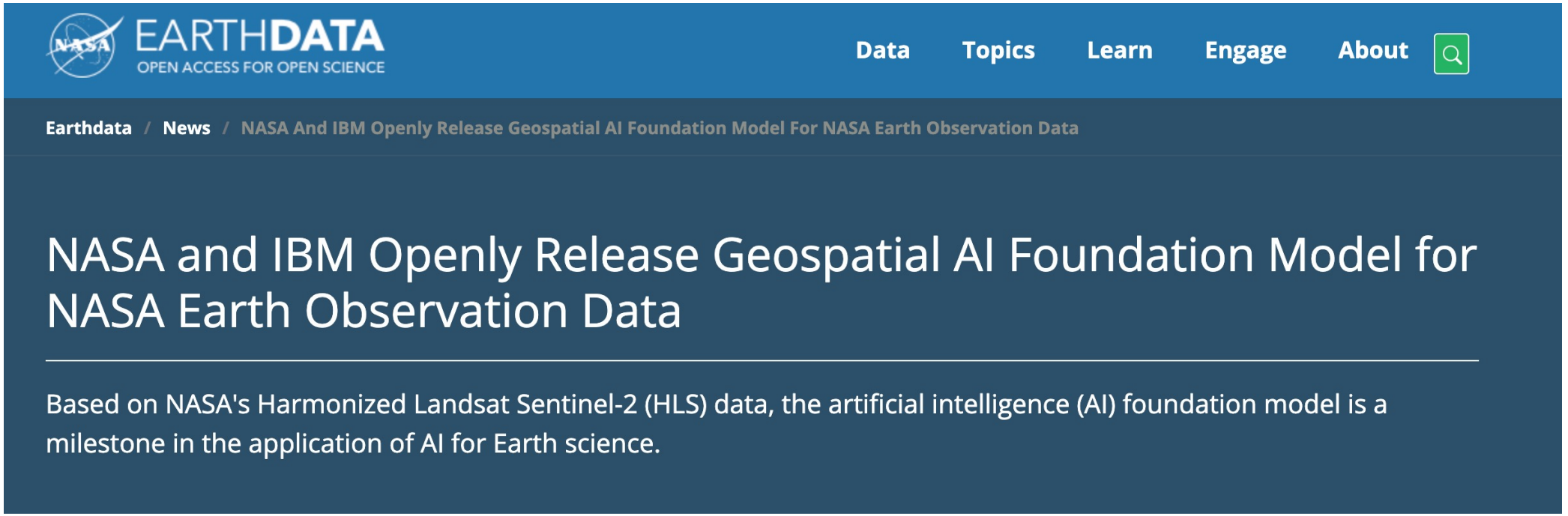


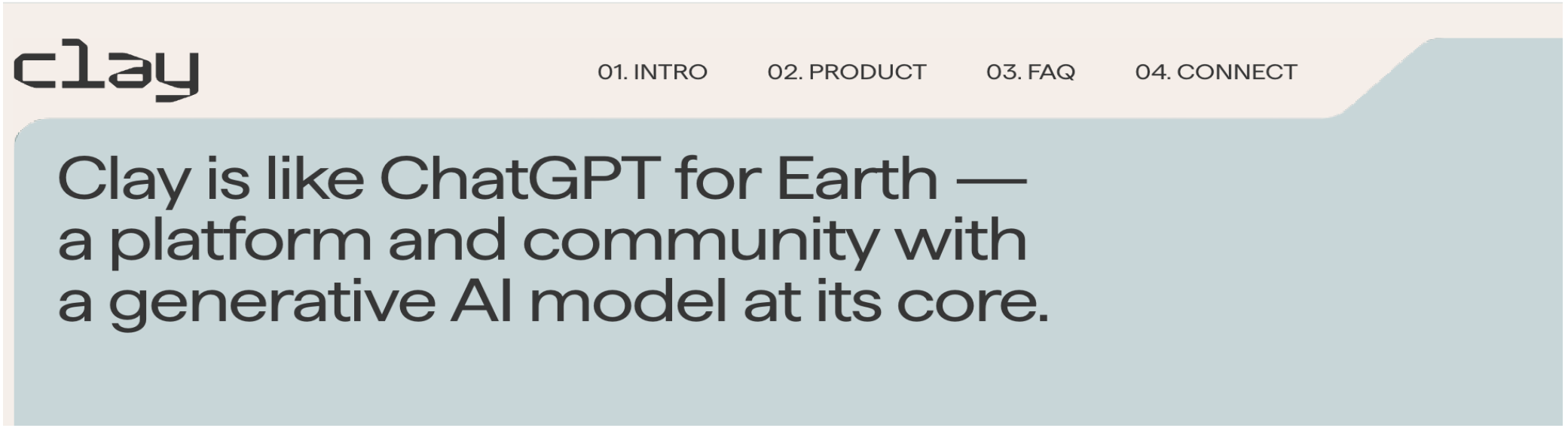
Large language models



Earth observation

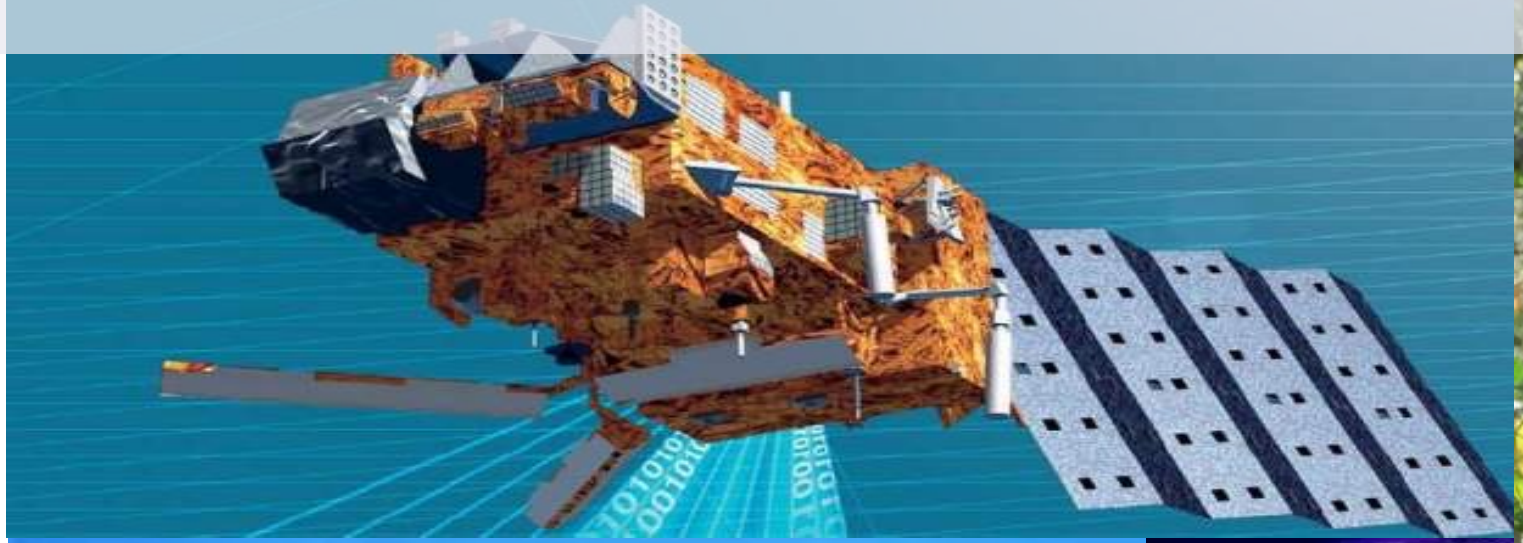
Foundational EO models

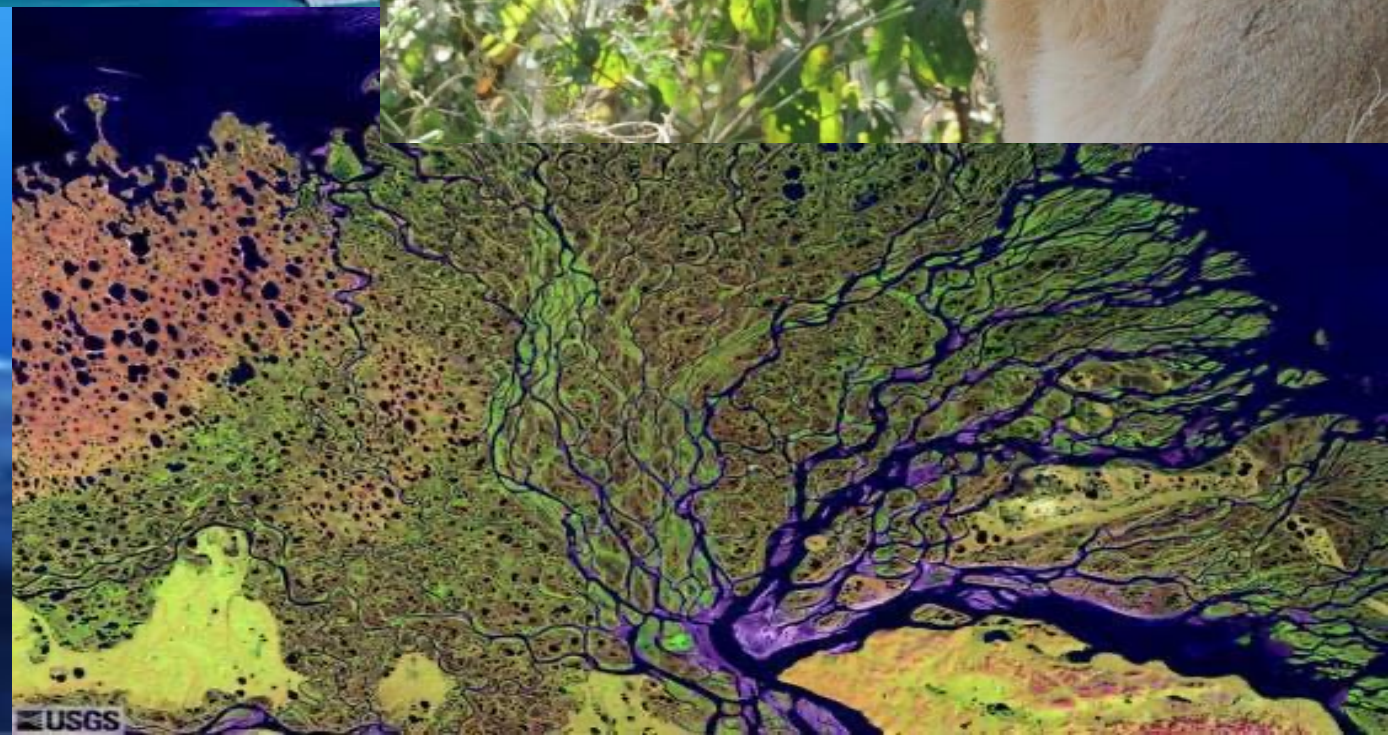

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE
Data
Topics
Learn
Engage
About
Earthdata / News / NASA And IBM Openly Release Geospatial AI Foundation Model For NASA Earth Observation Data
NASA and IBM Openly Release Geospatial AI Foundation Model for NASA Earth Observation Data
Based on NASA's Harmonized Landsat Sentinel-2 (HLS) data, the artificial intelligence (AI) foundation model is a milestone in the application of AI for Earth science.

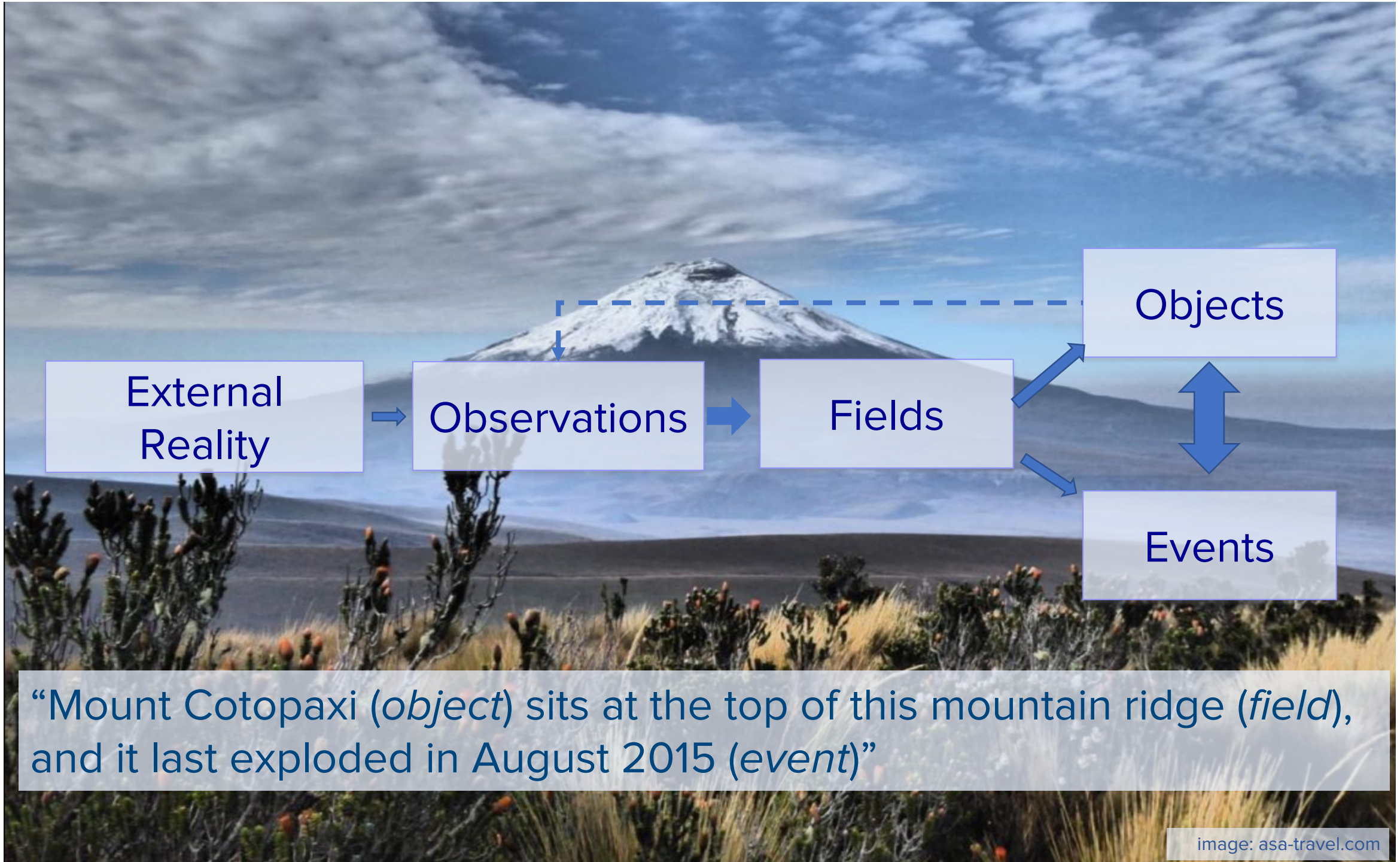

clay
01. INTRO
02. PRODUCT
03. FAQ
04. CONNECT
Clay is like ChatGPT for Earth —
a platform and community with
a generative AI model at its core.

Reality exists independently of human representations

We have access to the world through our observations



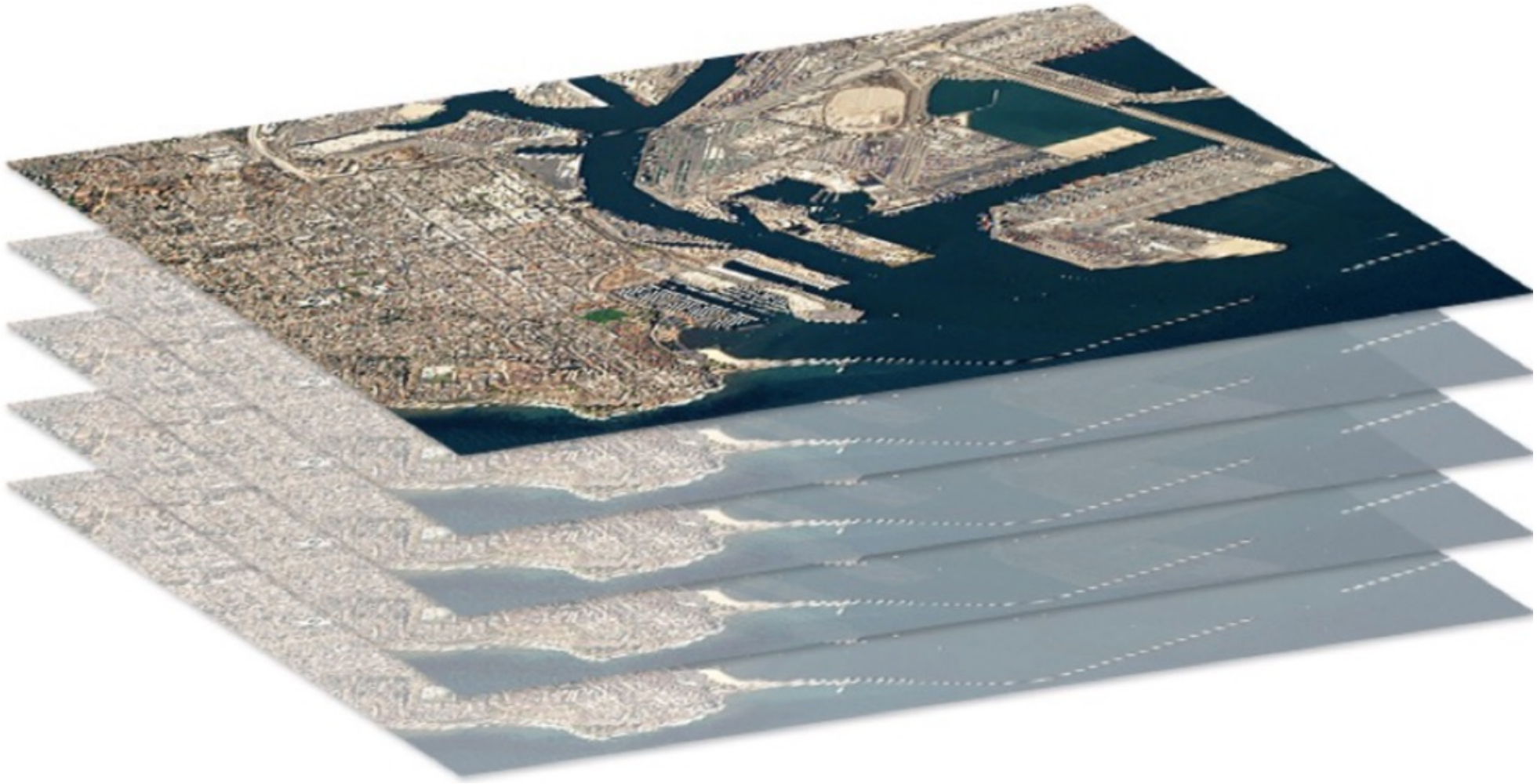




"Mount Cotopaxi (*object*) sits at the top of this mountain ridge (*field*), and it last exploded in August 2015 (*event*)"

image: asa-travel.com

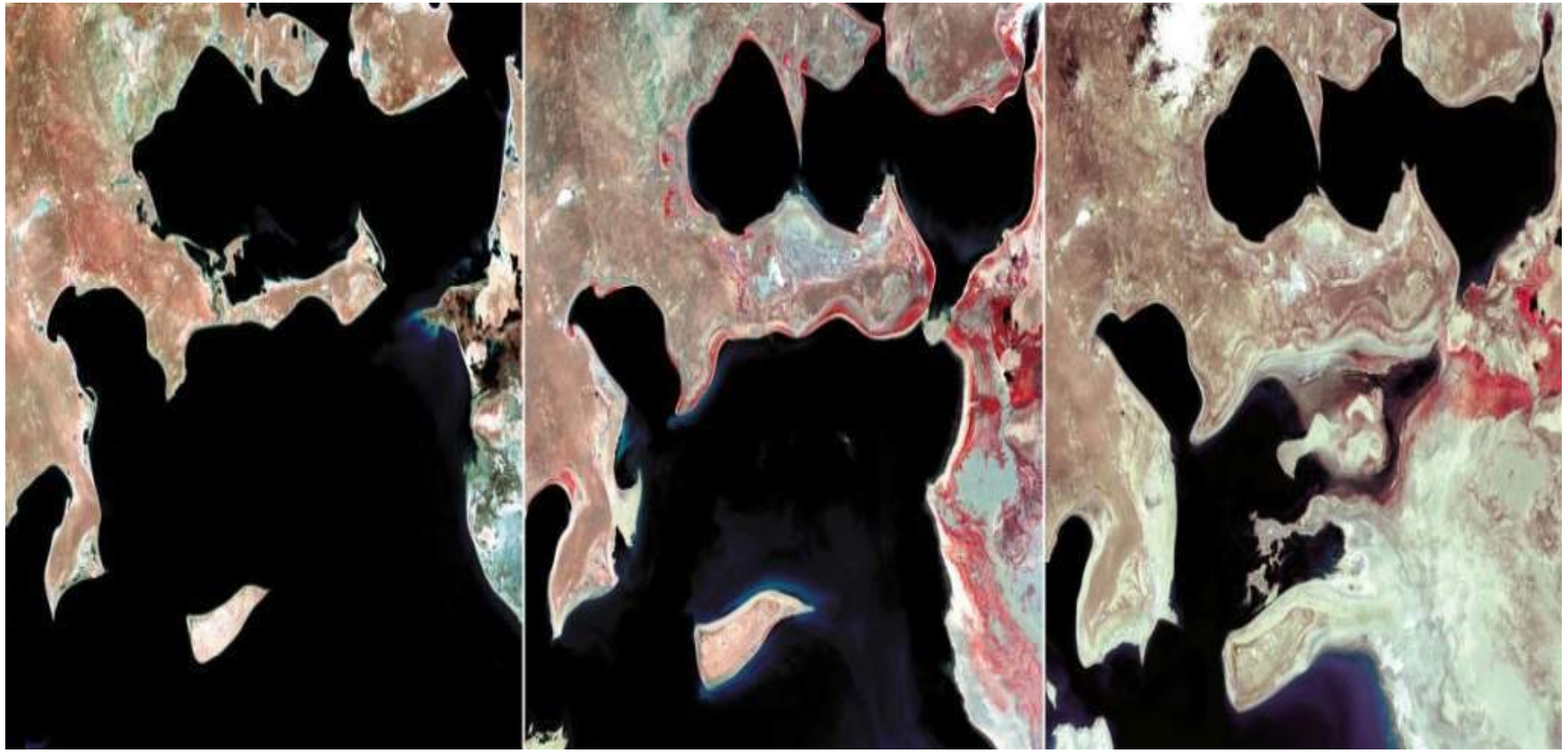
Earth Observation Data Cubes



Same area, multiple times

image: Swiss Data Cube

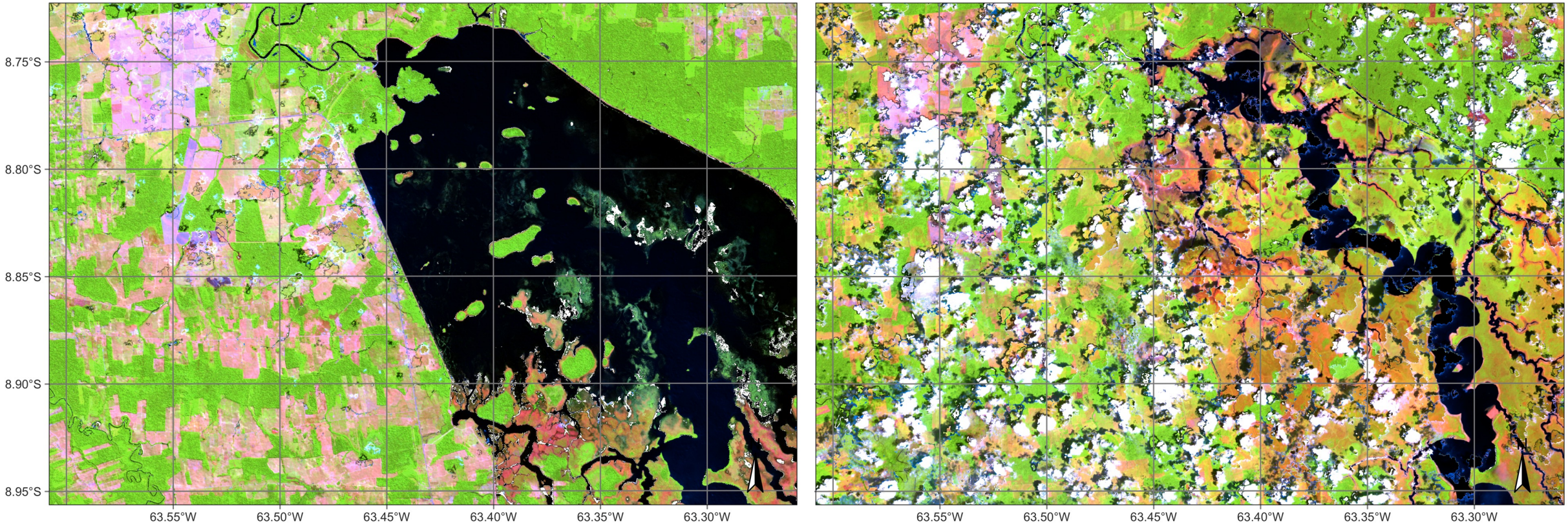
What's in an image?



"A remote sensing image is a measurement that captures snapshots of change trajectories. The focus of the ontological characterization of images should be on searching for changes instead of searching for content."
(Camara et al, COSIT 2001)

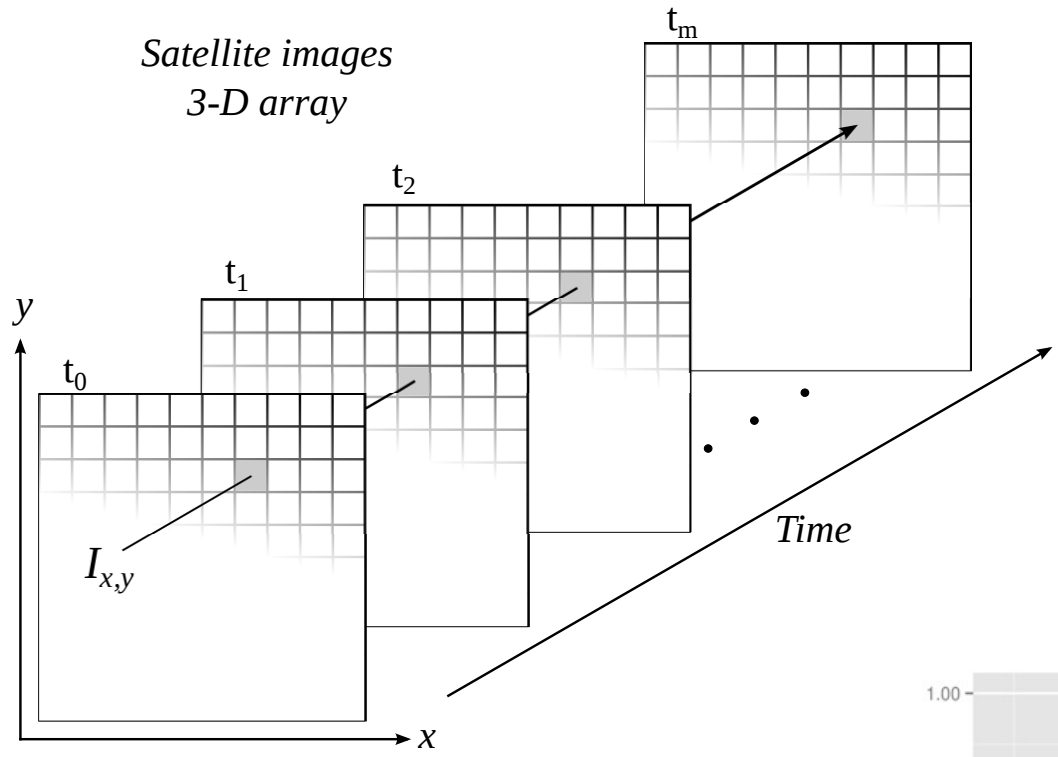
images: USGS

What's in an image time series?



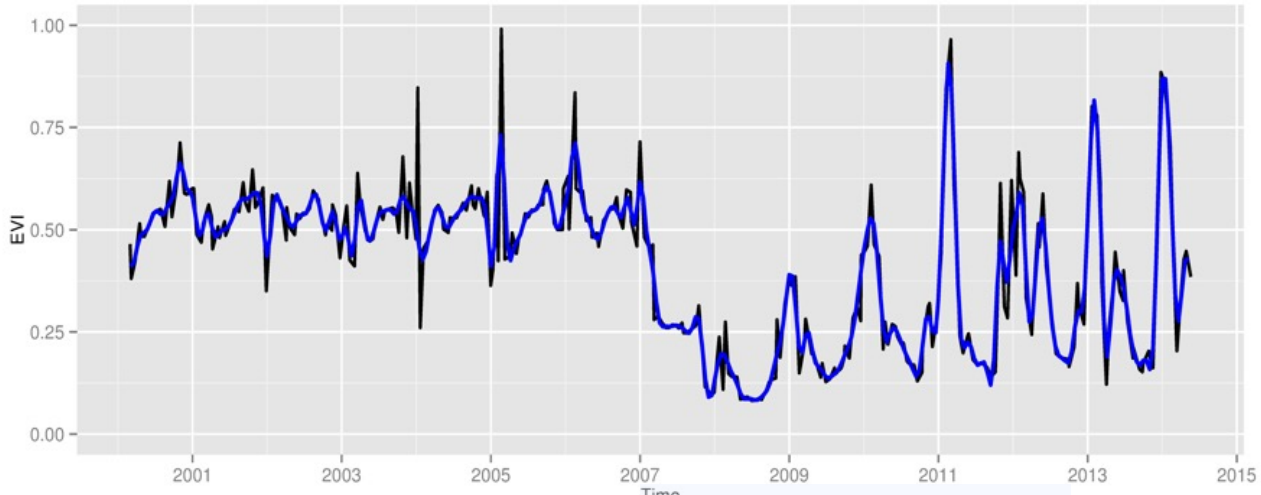
Searching for changes instead of searching for content.
(Camara et al., "What's in an image?", COSIT 2001)

Space first, time later or time first, space later?



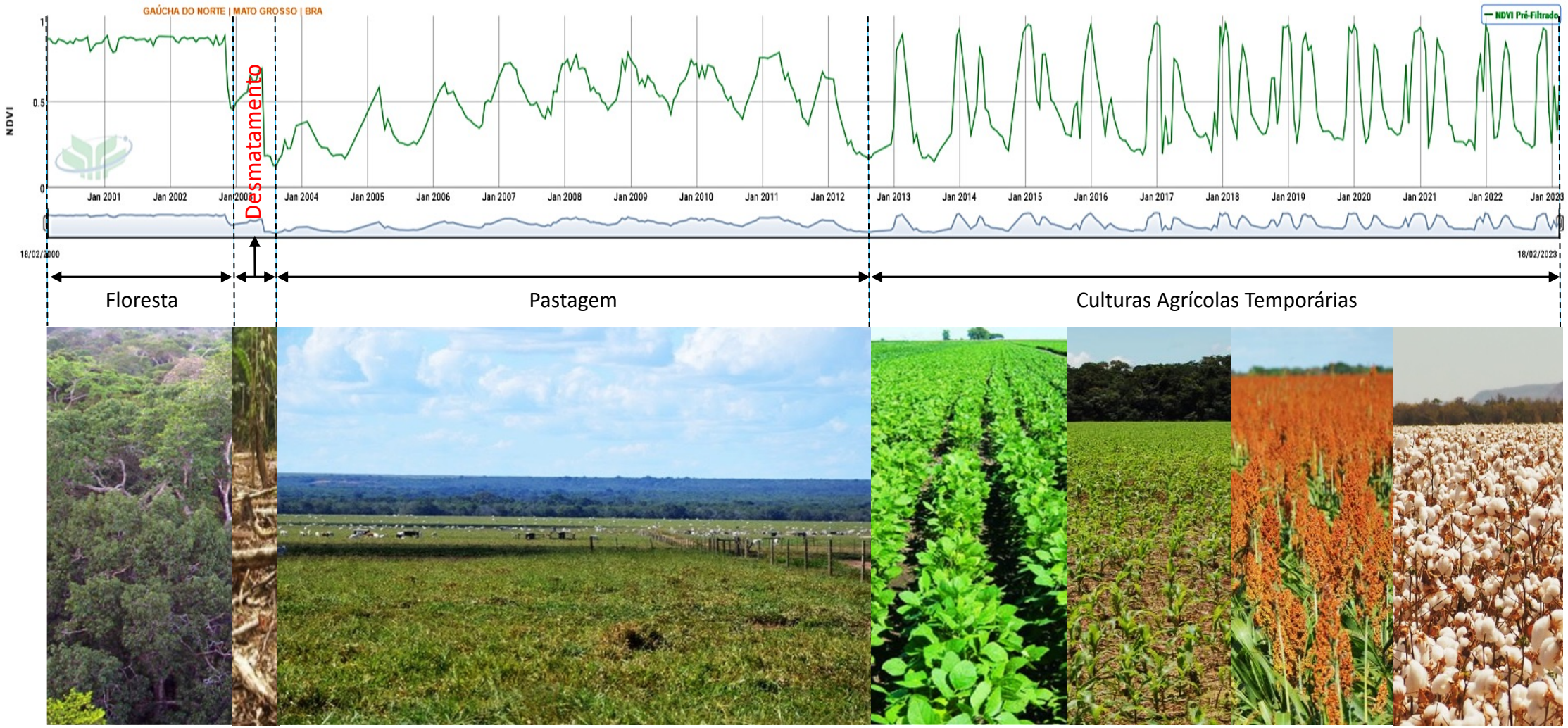
Space first: classify images; compare results in time

Time first: classify time series; join results to get maps



graphics: V Maus, author

Time Series showing changes



Big EO data: access to image time series

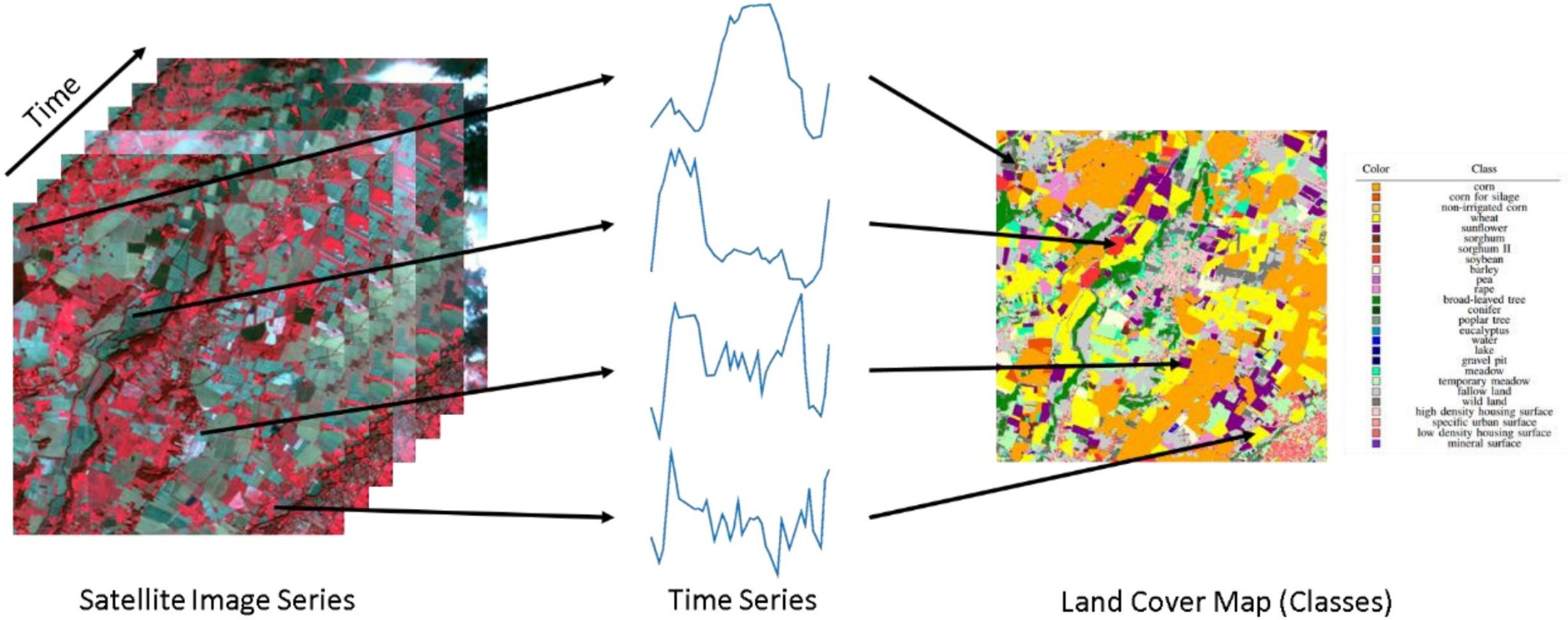
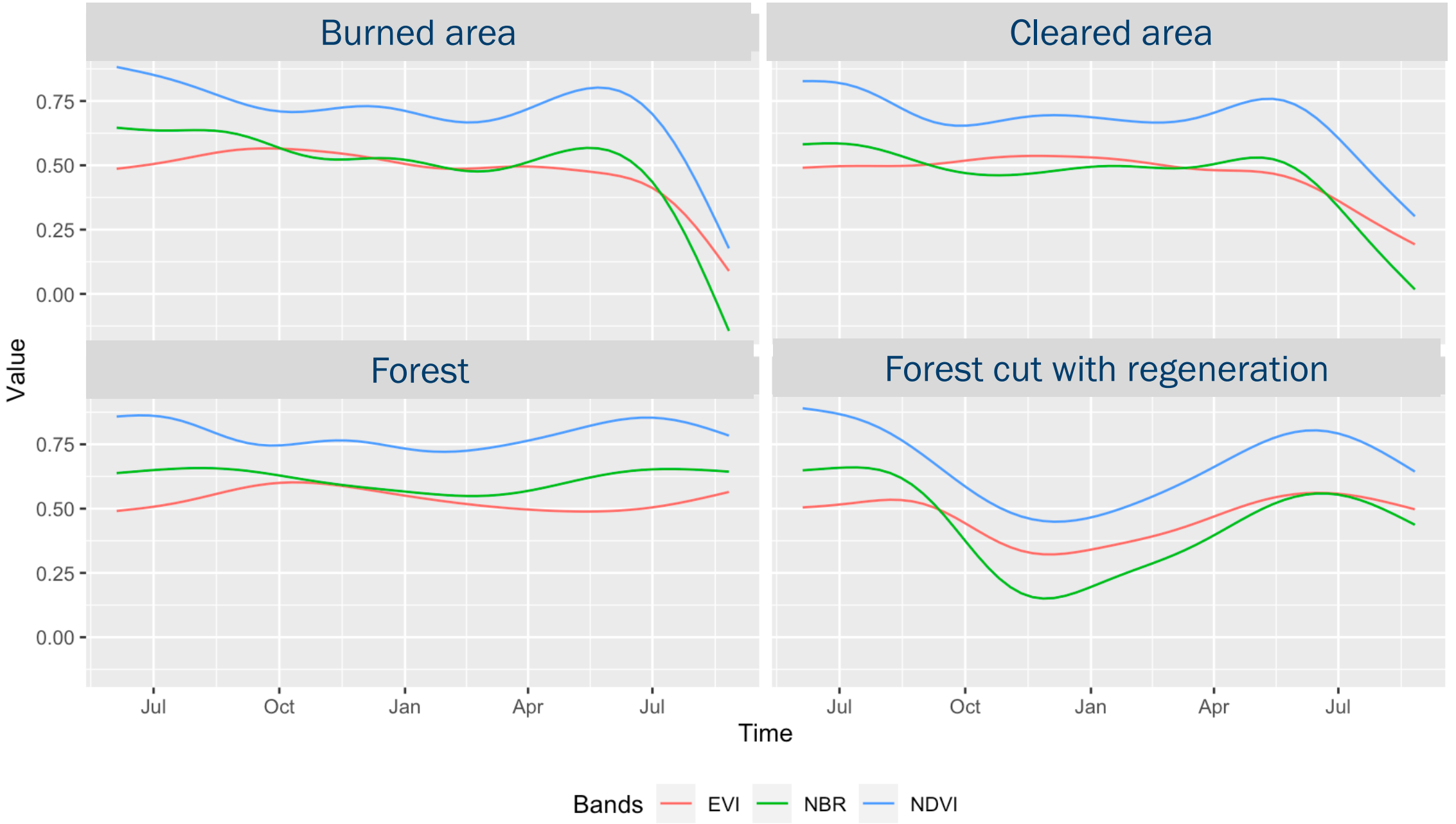
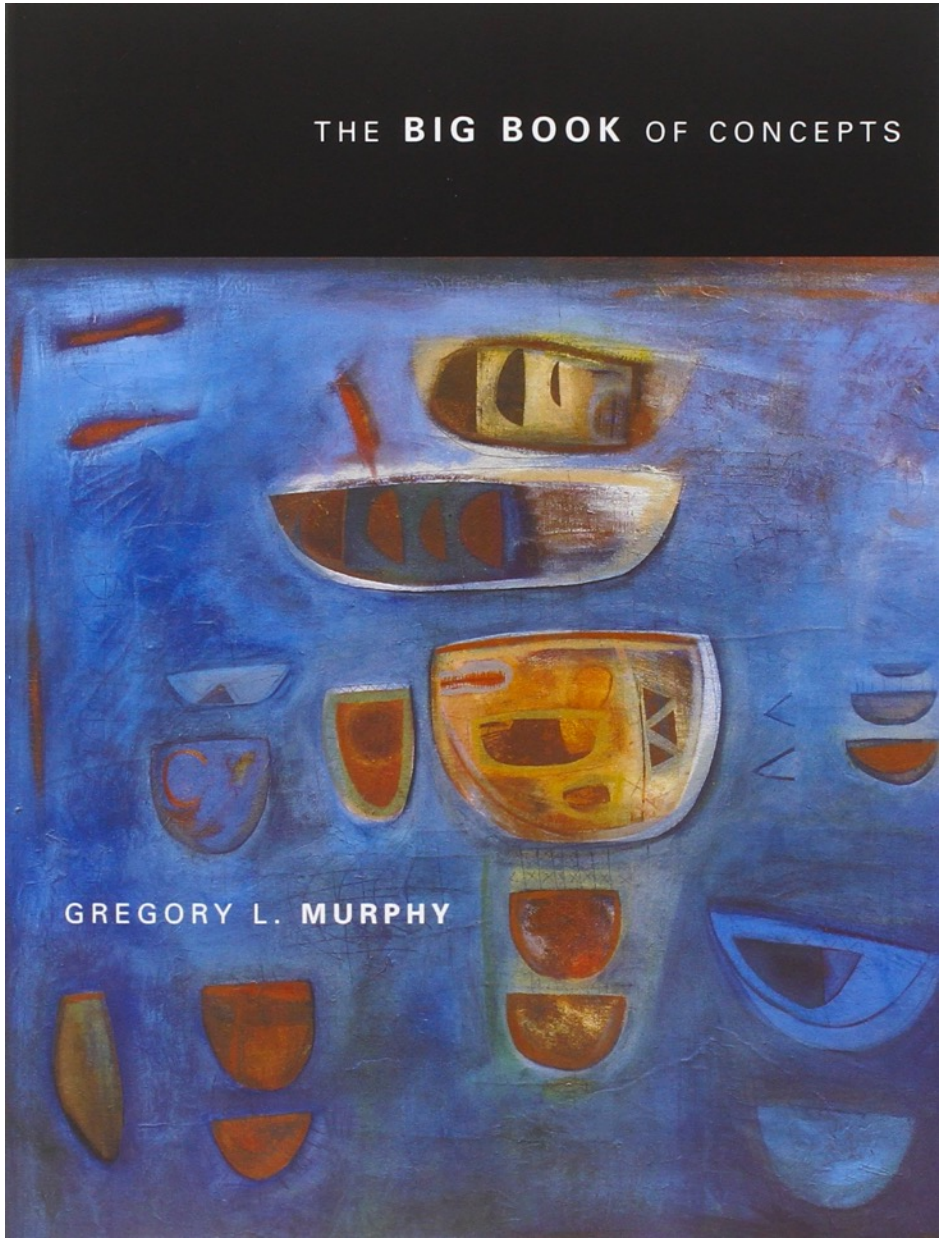


image: F Petitjean, Monash Univ

Event-based samples (model estimates)



Our concepts have limits

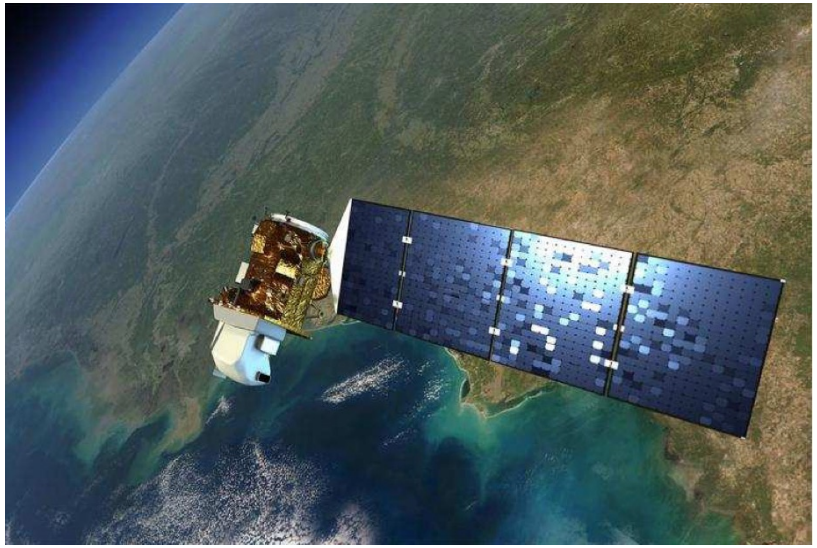


"The gradation of properties in the world means that our smallish number of categories will never map perfectly onto all objects" (Murphy, 2004)

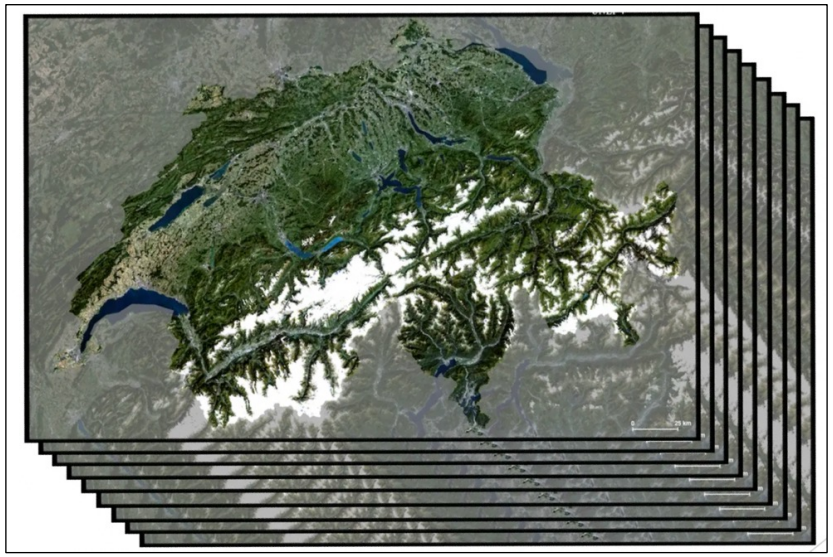
image: Amazon

The elephant in the room

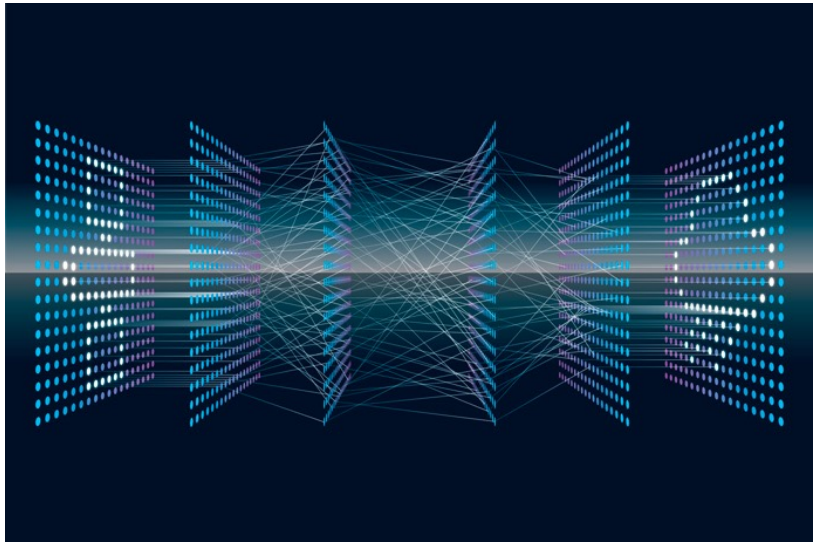
Ecosystems are highly variable
Local knowledge is essential



Big satellite data



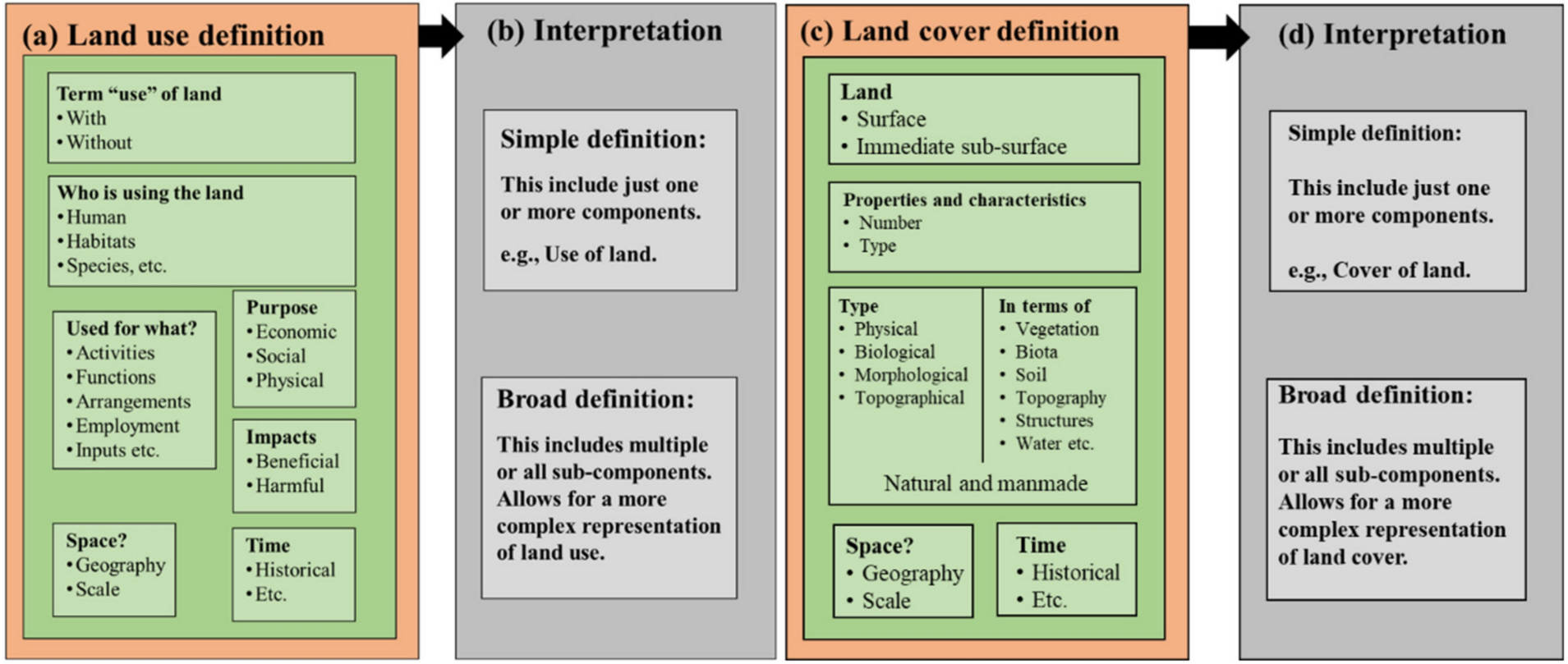
Data cubes



Machine learning

images: Alamy, NASA, Swiss Data Cube, ECMWF

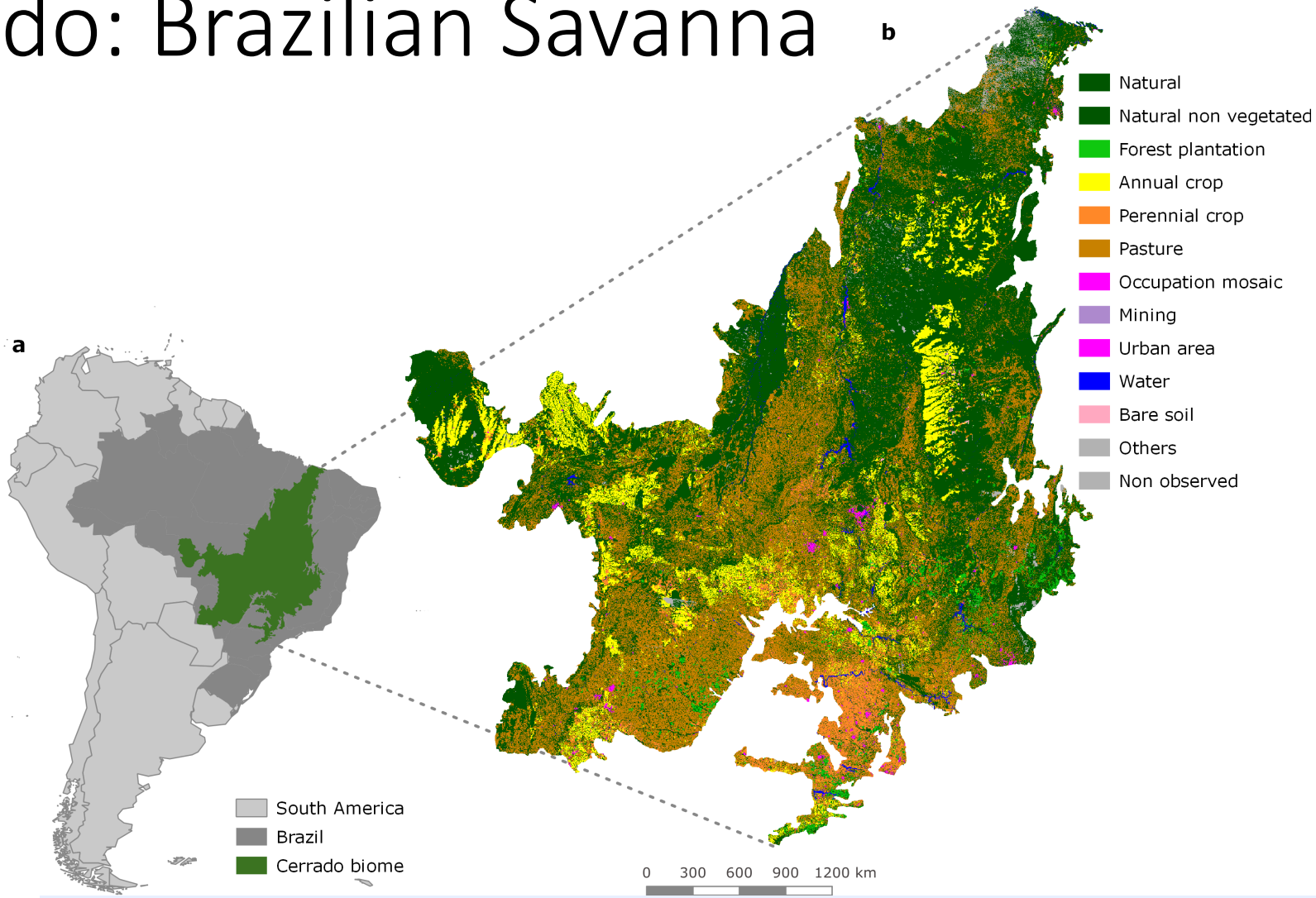
How LUCC ontologies are built?



"Geographical concepts are situated and context-dependent, can be described from different, equally valid, points of view, and ontological commitments are arbitrary" (Janowicz et al., 2012)

image source: Nedd et al., 2021

Cerrado: Brazilian Savanna



covers 2 million km^2, latitude ranges from 5^{0}S to 25^{0}S

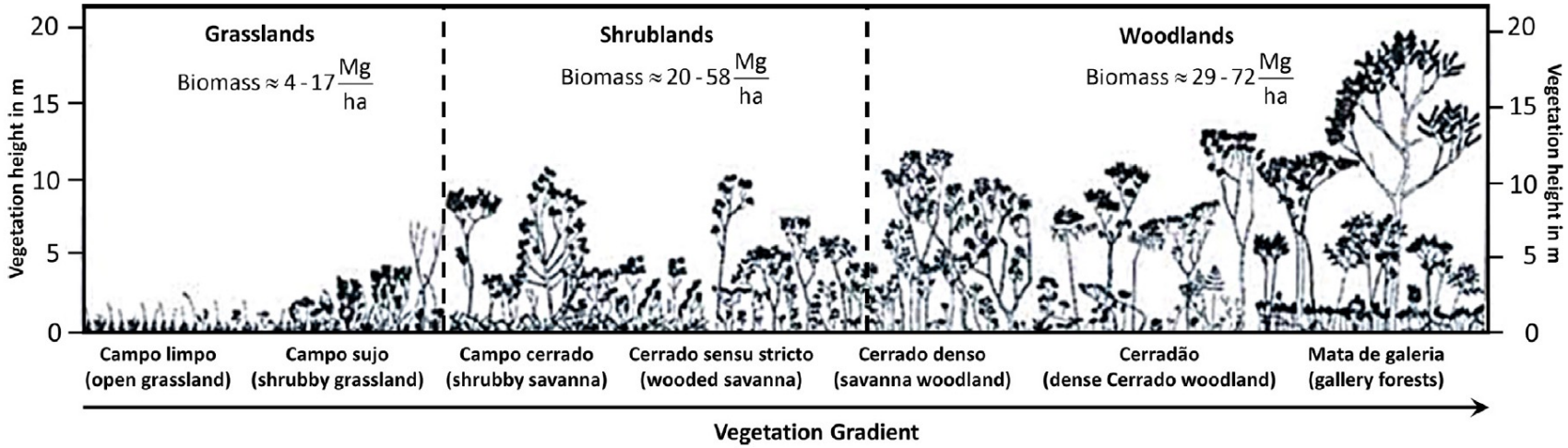
image: TerraClass, INPE/EMBRAPA

Cerrado: Brazilian savanna

Open cerrado

Cerrado stricto sensu

Cerradão (dense woodland)



images: Del-Claro and Silingardi (Annals Brazilian Academy Sci, 2019), Schwieder et al. (IJAEOG, 2016)

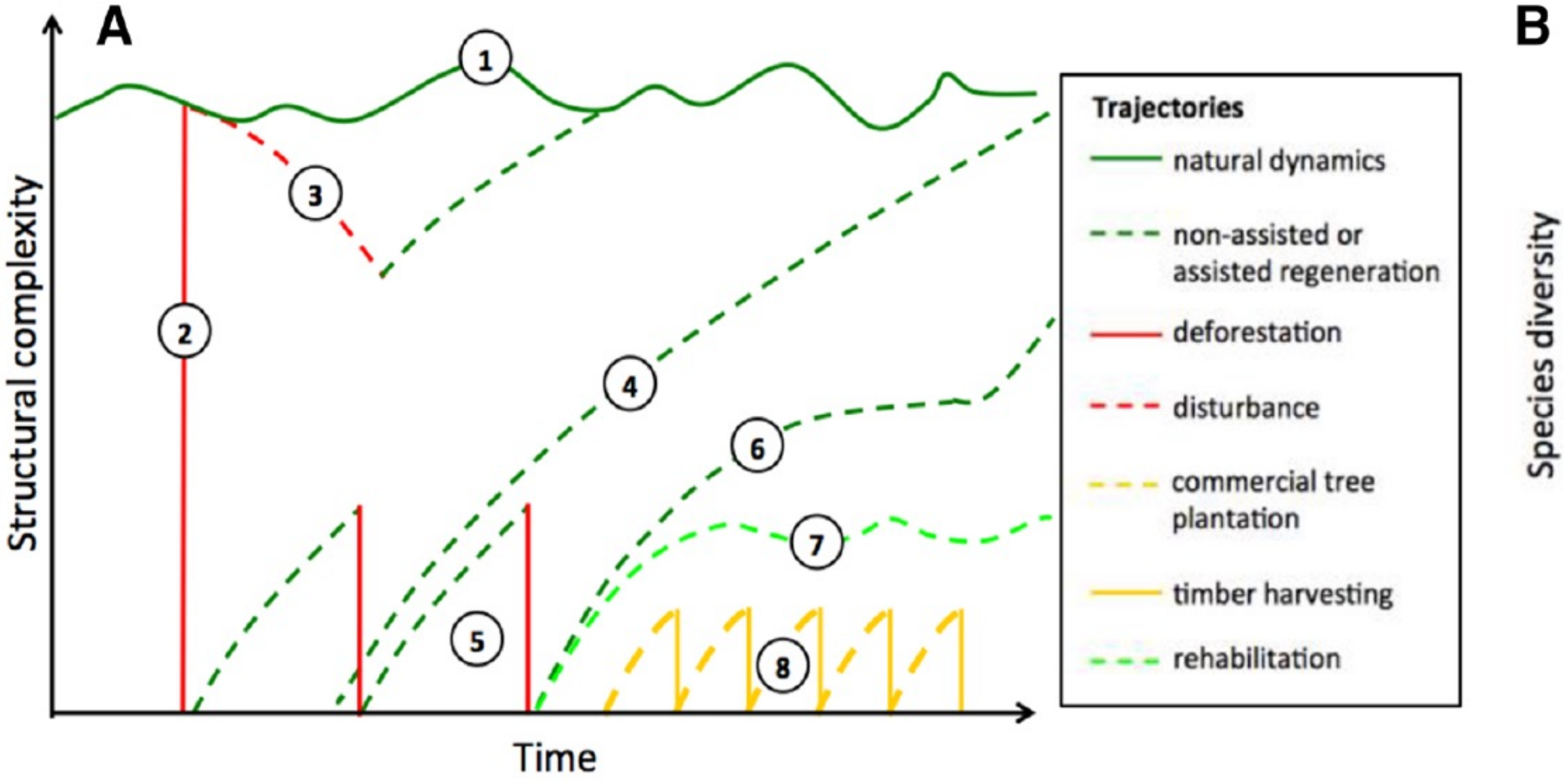
What is a forest?

images: Alamy

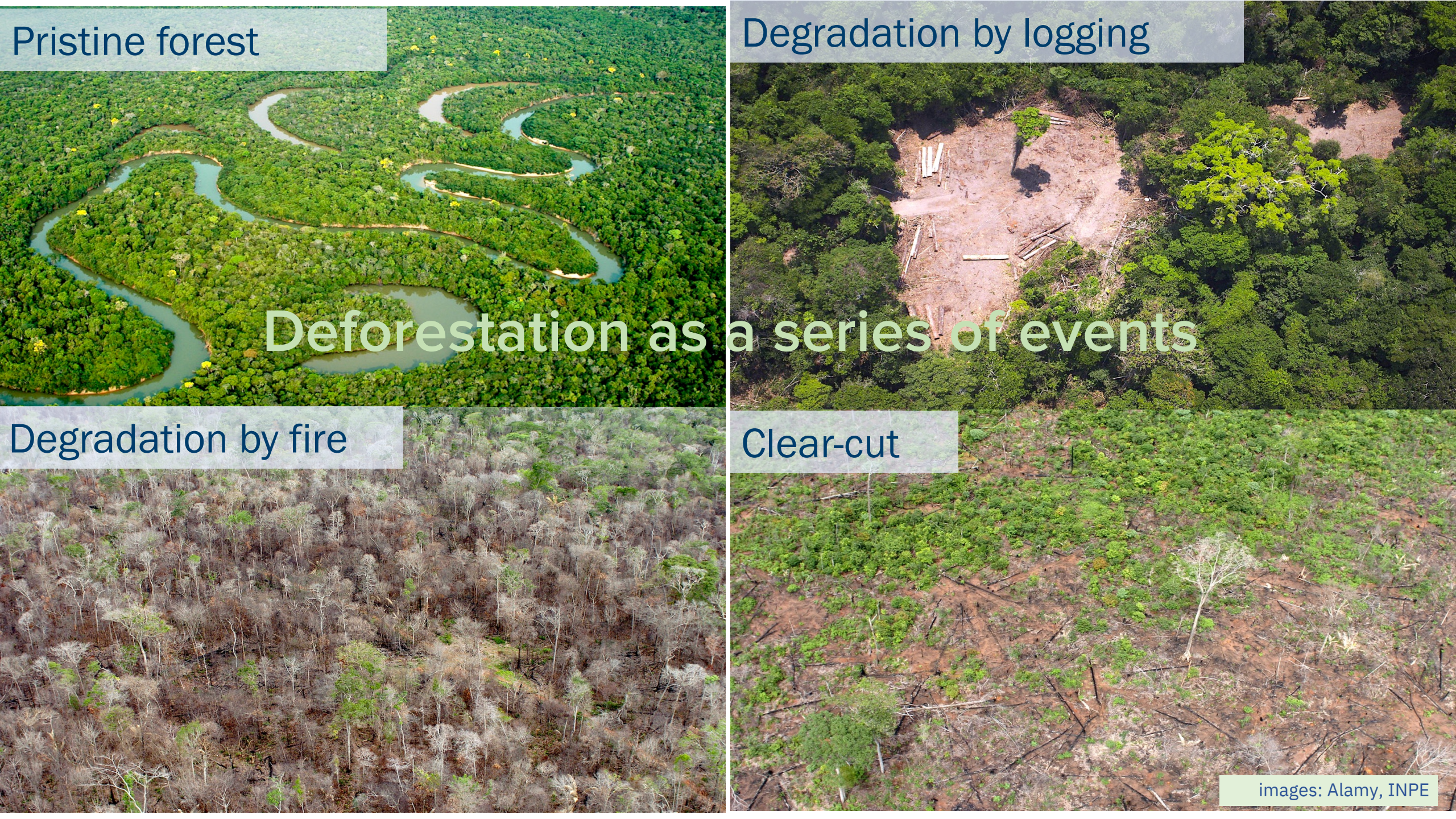
When is a forest not a forest?

images: Alamy

Distinguishing forests by temporal evolution



source: Chazdon et al (2016)


Pristine forest
Degradation by logging
Deforestation as a series of events
Degradation by fire
Clear-cut
images: Alamy, INPE

Natural or man-made landscapes?

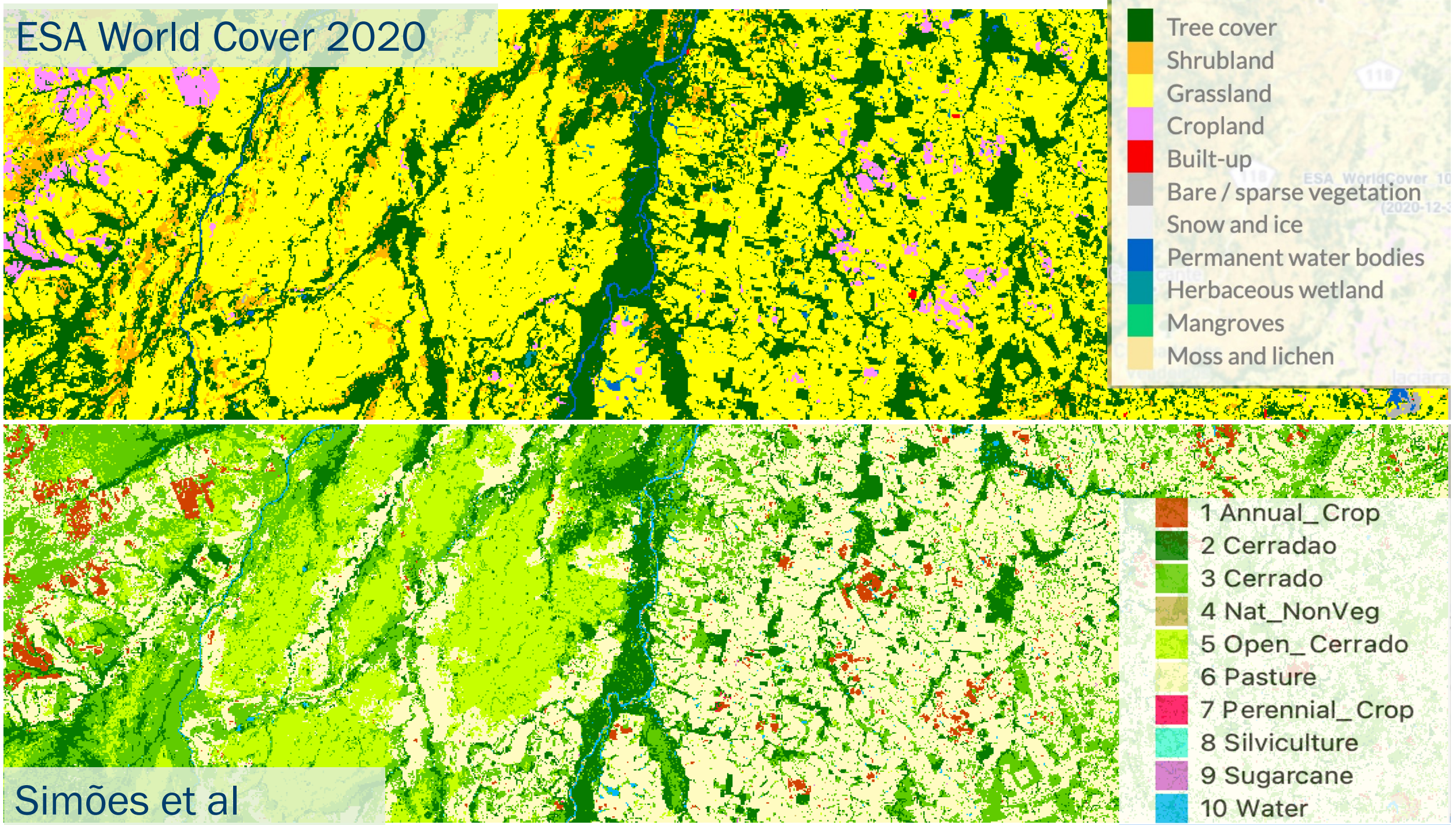
Natural savanna (Brazilian Cerrado)

Cattle in pasture (in the Cerrado)

images: Nature Conserv, Alamy

Locally relevant?

vito remote sensing · Brockmann Consult GmbH · ES GROUP · GAMMA REMOTE SENSING · IIASA · WAGENINGEN UNIVERSITY & RESEARCH


ESA World Cover 2020
Tree cover
Shrubland
Grassland
Cropland
Built-up
Bare / sparse vegetation
Snow and ice
Permanent water bodies
Herbaceous wetland
Mangroves
Moss and lichen
Simões et al
1 Annual_Crop
2 Cerradao
3 Cerrado
4 Nat_NonVeg
5 Open_Cerrado
6 Pasture
7 Perennial_Crop
8 Silviculture
9 Sugarcane
10 Water

Big EO data analysis in practice

BRAZIL DATA CUBE

sits

Obtain and harmonise ARD

Put ARD in a cloud collection

Build a data cube

Select ML method

Collect training data (in-situ)

Quality control on training data

Train data set

Classify data

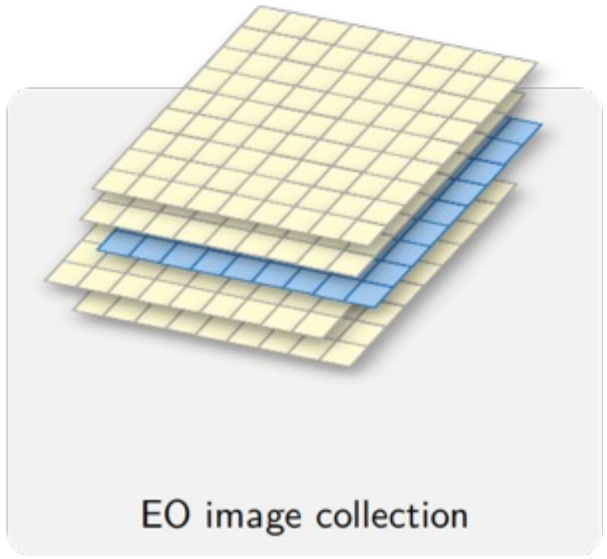
Smoothen data

Label map

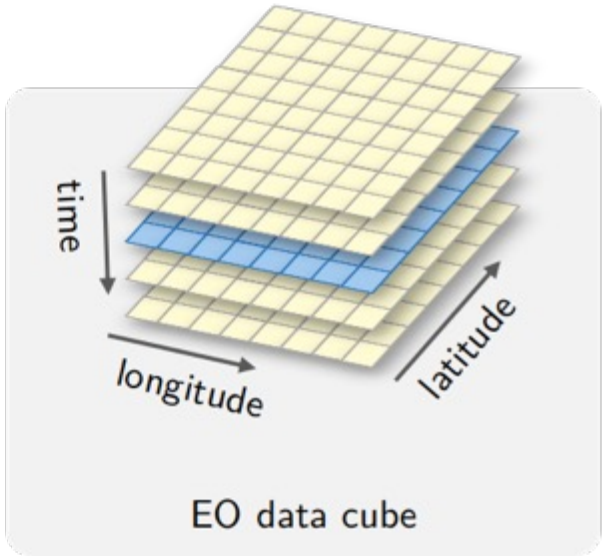
Access accuracy

GEO GROUP ON EARTH OBSERVATIONS

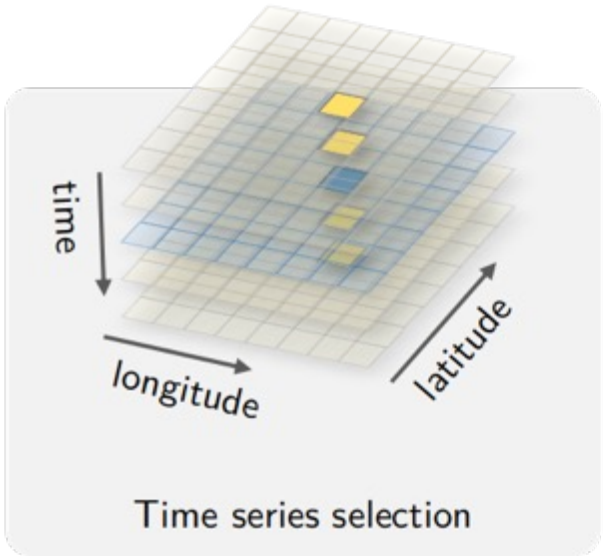
What is an EO data cube?



EO image collection



EO data cube



Time series selection

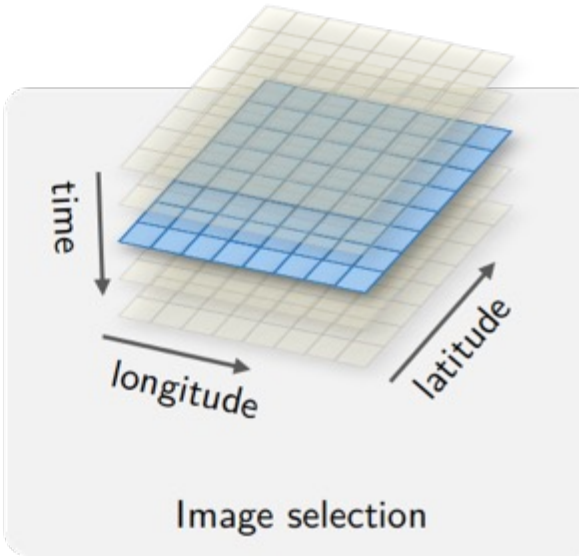


Image selection

Data cube = regular partitions of space and time

Functional programming in big EO data

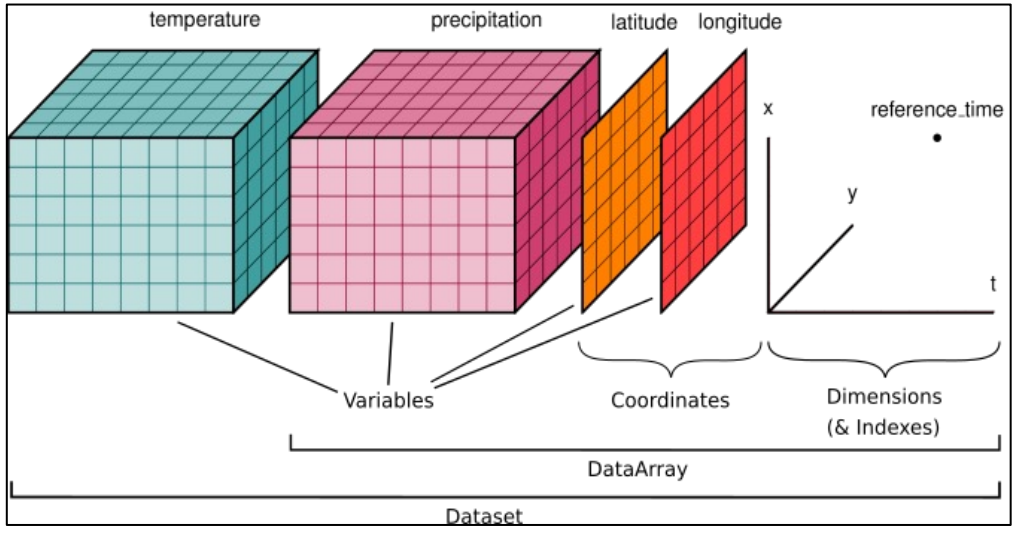
extract (DataCube, pos) → Samples

apply (DataCube, fun) → DataCube

train (Samples, MLalg) → MLmodel

classify (DataCube, MLmodel) → Map

Open-ended scripting or structured APIs?




```
# train model using TempCNN algorithm
tcnn_model <- train(
    data     = samples_cerrado_lc8,
    ml_method = sits_tempcnn())

# classify data cube
probs_cerrado <- classify(
    data     = lc8_cerrado_cube,
    ml_model = cnn_model)

# generate thematic map
map <- label_classification(
    cube = probs_cerrado)
```

Xarrays: maximum flexibility, lack of consensus API, hard for end-users

APIs: user-focused, simple scripts, rapid results, more rigid architecture

Lessons from GEE: build a clean API

```
# define a datacube of 2 million km2
l8_cerrado_cube <- sits_cube(
    source     = "BDC",
    collection = "LC8_30_16D",
    roi        = "./cerrado.shp",
    start_date = "2017-09-01",
    end_date   = "2018-08-31"
)

# obtain the time series
samples_cerrado_lc8 <- sits_get_data(
    cube = l8_cerrado_cube,
    file = "./samples_48K.csv"
)
```

```
# train model using TempCNN algorithm
tcnn_model <- sits_train(
    data      = samples_cerrado_lc8,
    ml_method = sits_tempcnn())

# classify data cube
probs_cerrado <- sits_classify(
    data     = l8_cerrado_cube,
    ml_model = cnn_model)

# generate thematic map
map <- sits_label_classification(
    cube = probs_cerrado)
```

Cerrado biome (200 million ha)

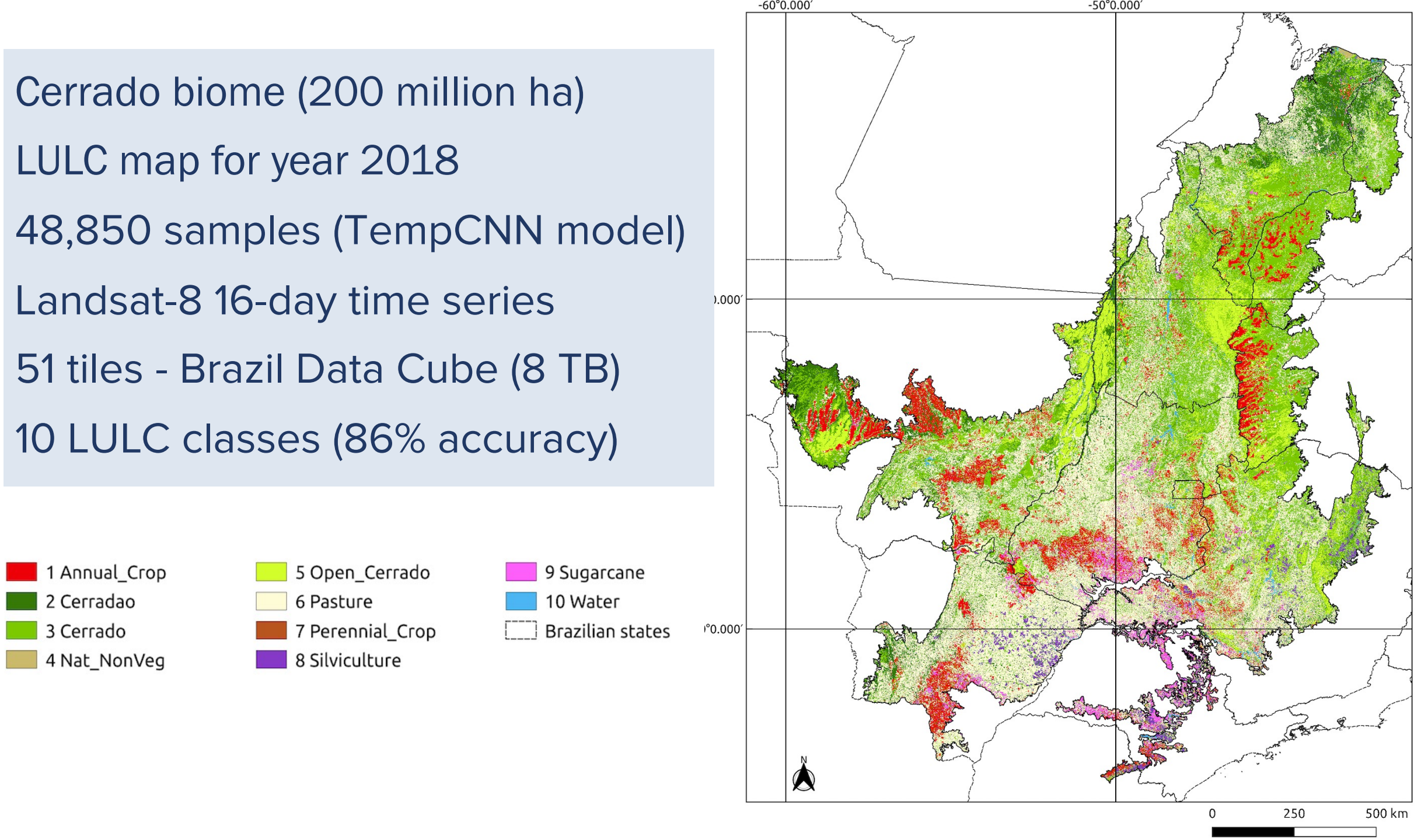
LULC map for year 2018

48,850 samples (TempCNN model)

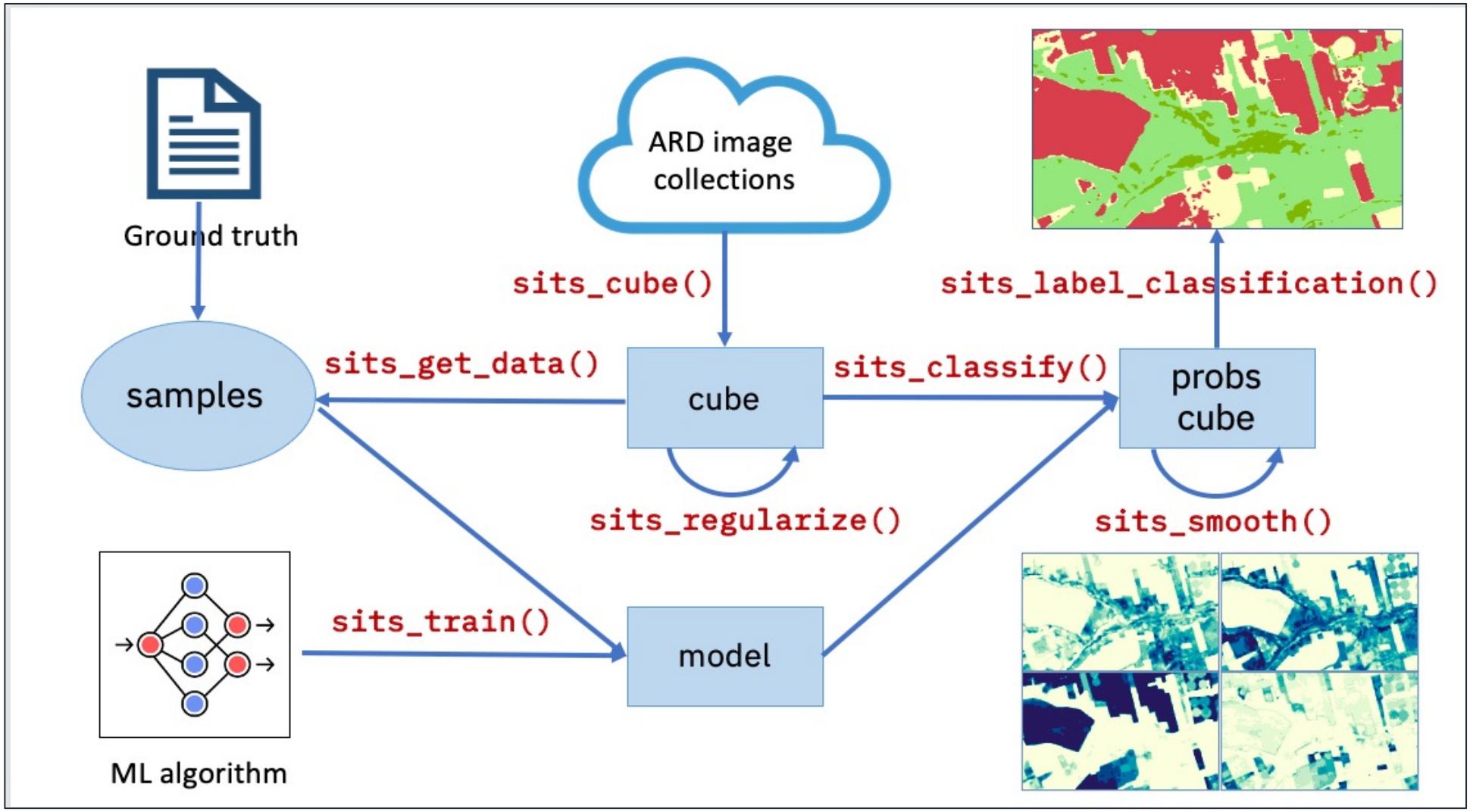
Landsat-8 16-day time series

51 tiles - Brazil Data Cube (8 TB)

10 LULC classes (86% accuracy)



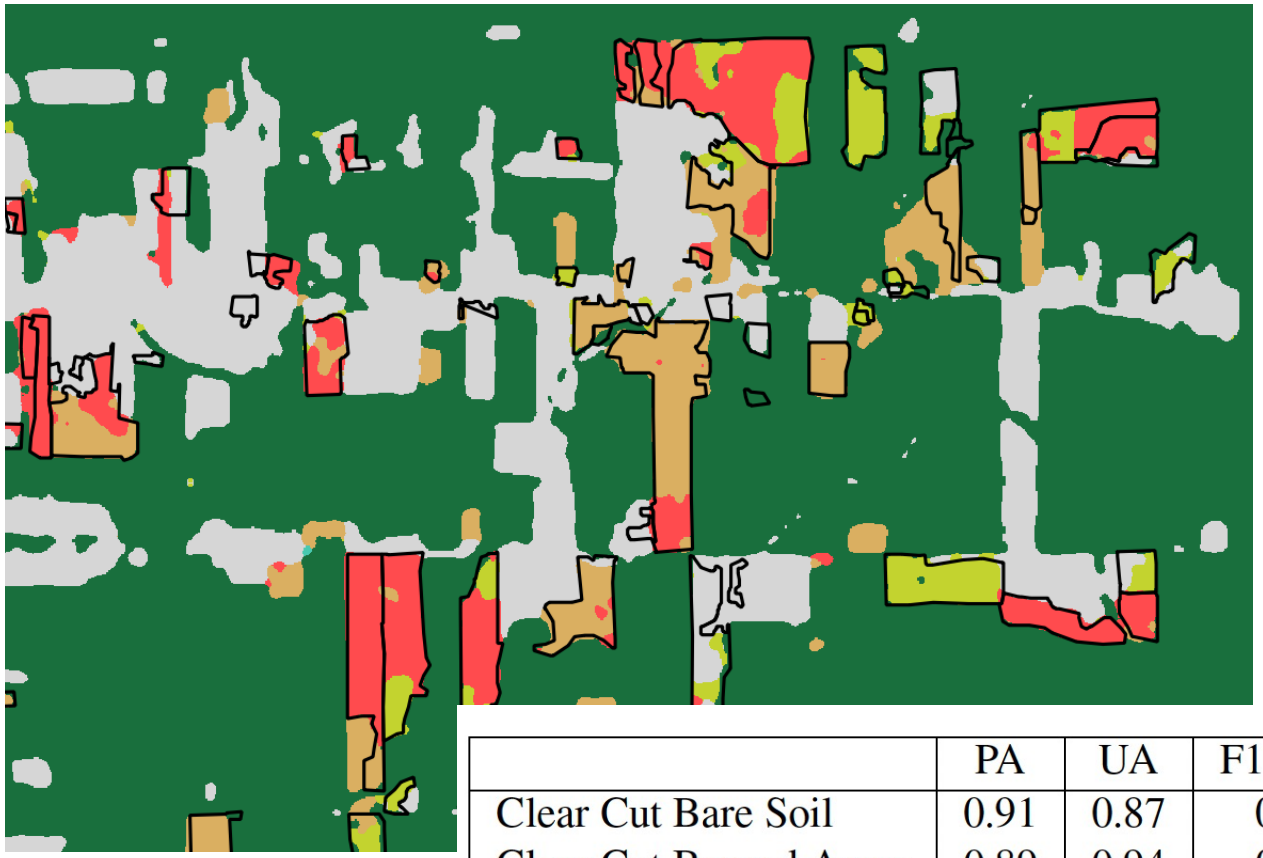
SITS: Easy-to-learn API



Target user community: Earth science experts

Visual interpretation is hard for ML to beat...

Comparison SITS and visual interpretation



- CorteRaso_SoloExposto
- CorteRaso_Queimada
- Floresta
- CorteRaso_Vegetação
- Água
- Áreas úmidas
- Máscara de desmatament
- Não Floresta

	PA	UA	F1 score
Clear Cut Bare Soil	0.91	0.87	0.89
Clear Cut Burned Areas	0.89	0.94	0.91
Clear Cut Vegetation	0.79	0.88	0.83
Forest	0.95	0.86	0.90
Mountainside Forest	0.73	0.97	0.83
Riparian Forest	0.95	0.90	0.93
Seasonally Flooded	0.87	0.87	0.87
Water	0.96	0.95	0.96
Wetlands	0.70	0.93	0.80

	PA	UA	F1 score
Clear Cut	0.94	0.97	0.95
Forest	0.98	0.91	0.94
Wetland	0.84	0.95	0.89
Water	0.96	0.95	0.95

Which GIScience ideas could be relevant to the future of FOSS4G?

- Geospatial ontologies
- Fields, objects and events in space-time
- Geospatial algebras and functional programming