

# OPEN DATA SCIENCE EUROPE

## WORKSHOP

# Working with harmonized LUCAS dataset

Instructors: Lukáš Brodský, Martin Landa, Tomáš Bouček

[lukas.brotsky@fsv.cvut.cz](mailto:lukas.brotsky@fsv.cvut.cz)  
[martin.landa@fsv.cvut.cz](mailto:martin.landa@fsv.cvut.cz)  
[tomas.boucek@fsv.cvut.cz](mailto:tomas.boucek@fsv.cvut.cz)



# Draft content

- 1: LUCAS dataset introduction
- 2: Accessing LUCAS dataset using Python API (Jupyter notebooks)
- 3: Harmonized LUCAS dataset
- 4: Land product validation with LUCAS points (use case)
- 5. LUCAS land cover translation to CORINE and classes aggregation
- 6. Accessing LUCAS dataset from QGIS (LUCAS plug-in)
  
- Software requirements:
  - ODSE virtual box with eumap library
  - Jupyter Notebooks with sample data sets

[../ODSE\\_workdir/code/odse-workshop-2021/Python-training/05\\_lucas/\\*.\\*](#)

# LUCAS dataset (Eurostat)

LUCAS stands for the **Land Use and Coverage Area frame Survey**.

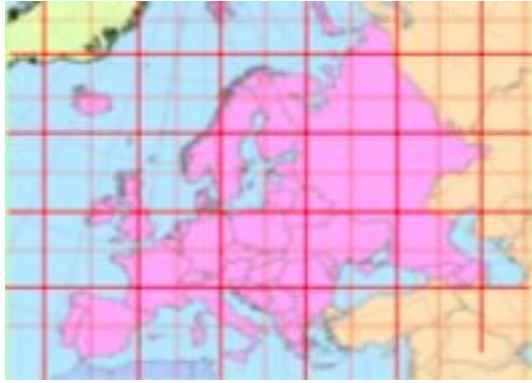
Main goal: **identify changes** in the European Union in land cover and land use.

Such information aim to be used in: nature protection, forest and water management, urban and transport planning, agricultural policy, prevention and mitigation of natural hazards, soil protection and mapping, monitoring climate change, monitoring biodiversity, etc.

---

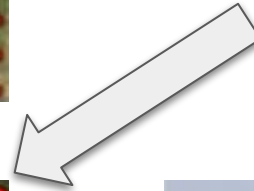
It is a unique **ground-true** dataset for land products (LC) **validation** and new models **calibration**.

# LUCAS survey process



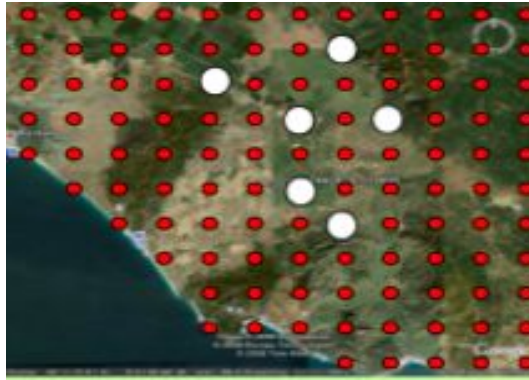
1. phase: sample for stratification from 2 km grid

Land Cover: classes ..

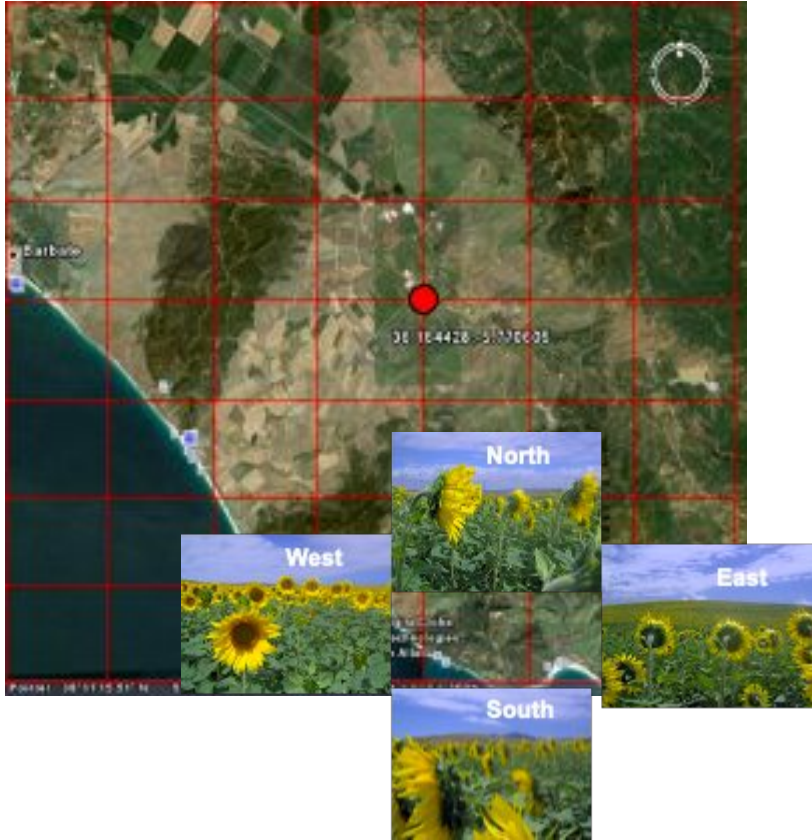


2. phase: in-situ data collection

3. Quality assurance



# LUCAS dataset (Eurostat)



Surveyors examine:

- land cover and land use, management,
- structural elements in the landscape
- and take 500 gram topsoil sample (at one out of 10 points)
- and collect photos.

# LUCAS Land Cover nomenclature

- fully hierarchical (3 levels)
- lc1: 76 classes

Land cover			
A00	ARTIFICIAL LAND	A10	Roofed built-up areas
		A20	Artificial non-built up areas
		A30	Other artificial areas
B00	CROPLAND	B10	Cereals
		B20	Root crops
		B30	Non-permanent industrial crops
		B40	Dry pulses, vegetables and flowers
		B50	Fodder crops
		B70	Permanent crops: fruit trees
		B80	Other permanent crops
C00	WOODLAND	C10	Broadleaved woodland
		C20	Coniferous woodland
		C30	Mixed woodland
D00	SHRUBLAND	D10	Shrubland with sparse tree cover
		D20	Shrubland without tree cover
E00	GRASSLAND	E10	Grassland with sparse tree/shrub cover
		E20	Grassland without tree/shrub cover
		E30	Spontaneously re-vegetated surfaces
F00	BARE LAND AND LICHENS/MOSS	F10	Rocks and stones
		F20	Sand
		F30	Lichens and moss
		F40	Other bare soil
G00	WATER AREAS	G10	Inland water bodies
		G20	Inland running water
		G30	Transitional water bodies
		G40	Sea and ocean
		G50	Glaciers, permanent snow
H00	WETLANDS	H10	Inland wetlands
		H20	Coastal wetlands



# LUCAS Land Use nomenclature

- lu1: 41 classes

Land use			
<b>U100</b>	PRIMARY SECTOR	<b>U110</b>	Agriculture
		<b>U120</b>	Forestry
		<b>U130</b>	Aquaculture and fishing
		<b>U140</b>	Mining and quarrying
		<b>U150</b>	Other primary production
<b>U 200</b>	SECONDARY SECTOR	<b>U210</b>	Energy production
		<b>U220</b>	Industry and manufacturing
<b>U300</b>	TERTIARY SECTOR, TRANSPORT, UTILITIES & RESIDENTIAL	<b>U310</b>	Transport, communication networks, storage, protection works
		<b>U320</b>	Water and waste treatment
		<b>U330</b>	Construction
		<b>U340</b>	Commerce, financial, professional and information services
		<b>U350</b>	Community services
		<b>U360</b>	Recreation, leisure, sport
		<b>U361</b>	Residential
<b>U400</b>	UNUSED AND ABANDONED AREAS	<b>U410</b>	Abandoned areas
		<b>U420</b>	Semi-natural and natural areas not in use

LUCAS access: <https://ec.europa.eu/eurostat/web/lucas>

The screenshot shows the Eurostat website interface for the LUCAS database. At the top, there is a header with the Eurostat logo, navigation links (Legal notice, Cookies, Links, My alerts, Contact), a language selector (English), and a Translate button. Below the header is a search bar with the placeholder text "Type a keyword, a publication title, a dataset title...". The main navigation bar includes links for News, Data, Publications, About Eurostat, and Help. The breadcrumb trail indicates the current location: European Commission > Eurostat > Land cover/use statistics > Data > Database. The left sidebar contains a menu with links to Overview, Data, DATABASE (highlighted in red), LUCAS Grid, Primary data, Lucas photo viewer, Order form, Publications, Methodology, Use cases, and Links. The main content area displays the "LAND COVER/USE STATISTICS (LUCAS)" database, listing several datasets with icons and brief descriptions: "Land cover and land use, landscape (LUCAS) (lan)", "Land cover overview by NUTS 2 regions (lan\_lcv\_ow)", "Land covered by artificial surfaces by NUTS 2 regions (lan\_lcv\_art)", "Land covered by artificial surfaces - index (lan\_lcv\_arti)", "Land use overview by NUTS 2 regions (lan\_use\_ow)", "Land cover for FAO Forest categories by NUTS 2 regions (lan\_lcv\_fao)", and "Settlement area (lan\_settl)". At the bottom, there is a section titled "OTHER RELATED EUROSTAT DATA" with links to Agricultural statistics, Agri-environmental indicators, Environmental Statistics, and Forestry statistics.

Sign In | Register

Legal notice | Cookies | Links | My alerts | Contact | English | Translate

Type a keyword, a publication title, a dataset title...

News | Data | Publications | About Eurostat | Help

European Commission > Eurostat > Land cover/use statistics > Data > Database

LAND COVER/USE STATISTICS (LUCAS) | DATABASE

Overview

▲ Data

**DATABASE**

LUCAS Grid

▼ Primary data

▲ Lucas photo viewer

Order form

Publications

Methodology

Use cases

Links

Land cover and land use, landscape (LUCAS) (lan)

Land cover overview by NUTS 2 regions (lan\_lcv\_ow)

Land covered by artificial surfaces by NUTS 2 regions (lan\_lcv\_art)

Land covered by artificial surfaces - index (lan\_lcv\_arti)

Land use overview by NUTS 2 regions (lan\_use\_ow)

Land cover for FAO Forest categories by NUTS 2 regions (lan\_lcv\_fao)

Settlement area (lan\_settl)

OTHER RELATED EUROSTAT DATA

► Agricultural statistics


► Agri-environmental indicators


► Environmental Statistics


► Forestry statistics



# <https://ec.europa.eu/eurostat/web/lucas/data/primary-data/2018>

**eurostat**  
Your key to European statistics

Legal notice | Cookies | Links |  My alerts | Contact | English | [Translate](#)

Type a keyword, a publication title, a dataset title... 

[News](#) | [Data](#) | [Publications](#) | [About Eurostat](#) | [Help](#)

[European Commission](#) > [Eurostat](#) > [Land cover/use statistics](#) > [Data](#) > [Primary data](#) > 2018

LAND COVER/USE STATISTICS (LUCAS)

Overview

▲ Data

- Database
- LUCAS Grid

▲ Primary data

- 2018**
- 2015
- 2012
- 2009
- 2006

▲ Lucas photo viewer

- Order form

Publications

Methodology

Use cases





























Links

PRIMARY DATA 2018


### LUCAS micro data 2018


The LUCAS micro data 2018 can be downloaded below. Each file contains the data for one country. The explanations and instructions can be downloaded from the column on the right.


The LUCAS 2018 photos can be ordered here:  
<https://ec.europa.eu/eurostat/web/lucas/data/primary-data/order-form>.


 Belgium	 Bulgaria	 Czechia	 Denmark
 Germany	 Estonia	 Ireland	 Greece
 Spain	 France	 Croatia	 Italy
 Cyprus	 Latvia	 Lithuania	 Luxembourg
 Hungary	 Malta	 Netherlands	 Austria
 Poland	 Portugal	 Romania	 Slovenia
 Slovakia	 Finland	 Sweden	 United Kingdom


SEE ALSO


 [B4: LUCAS 2018 Soil and Grassland sample \(EN\)](#)

 [C1: LUCAS 2018 Instructions \(EN\)](#)

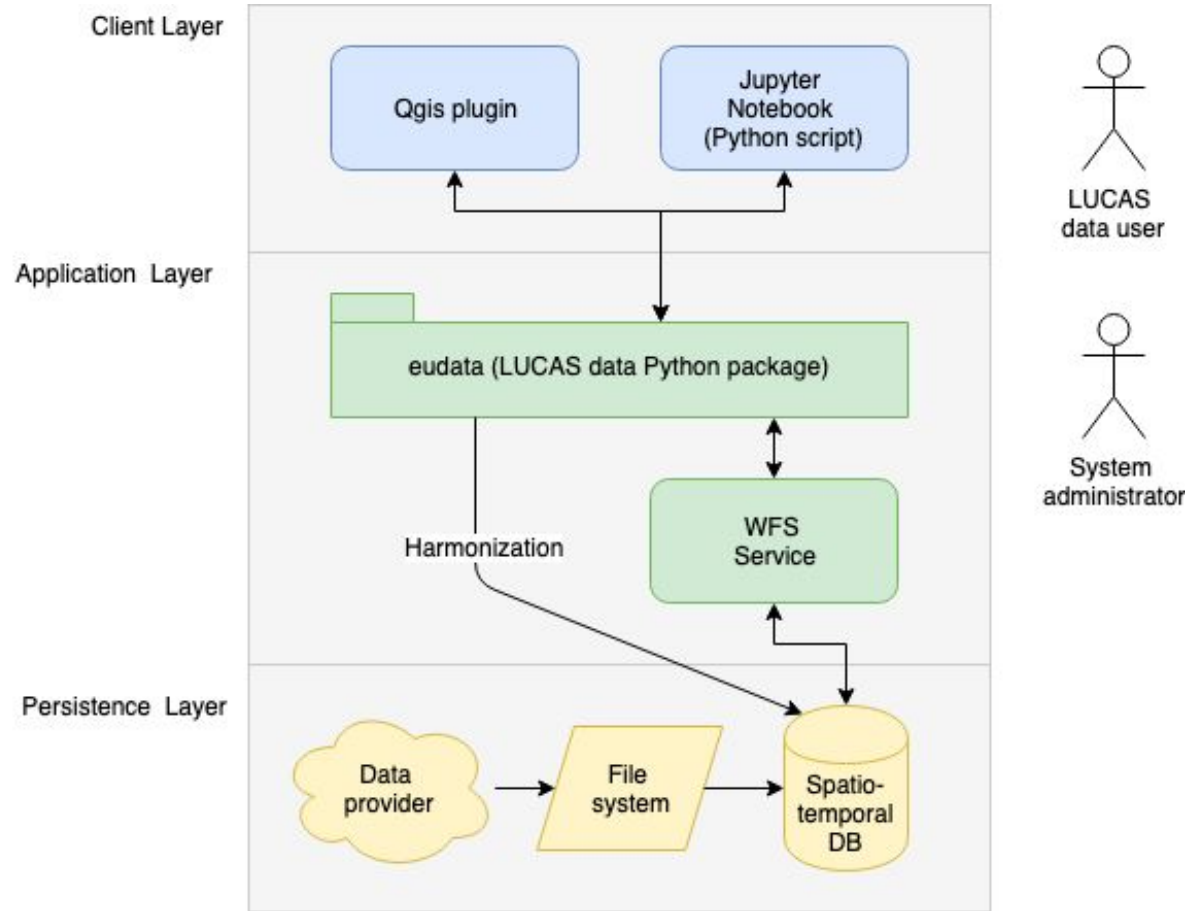
 [C2: LUCAS 2018 Field Form + Ground Document \(template\) \(EN\)](#)

 [C3: LUCAS 2018 Classification \(EN\)](#)

 [C4: LUCAS 2018 Quality Control \(EN\)](#)

 [C5: LUCAS 2018 Plant Identification Guide \(EN\)](#)

# Hands-on #1: accessing LUCAS dataset using Python API



# Hands-on #1: accessing LUCAS dataset using Python API

Hands-on session using Jupyter notebook

[../01\\_lucas\\_access.ipynb](#)

## LUCAS Samples

First of all, let's import `eumap` library.

```
In [1]: # To work with local eumap code
import sys
sys.path.append('../..')

from eumap.datasets.lucas import LucasRequest, LucasIO, LucasClassTranslate
```

## Usage

### 1. Define request

Request is defined by `LucasRequest` object. Bbox filter can be enabled by `bbox` property. Currently only [EPSG:3035](#) is supported.

For testing purpose a request can be created by `build()` method.

# LUCAS dataset grouped thematically

**lc\_lu:** land cover and land use

**lc\_lu\_so:** land cover and land use & soil

**fo:** forestry

**co:** copernicus

**in:** inspire

LUCAS facts		Year	EU countries	Points
<ul style="list-style-type: none"> <li>Covers EU countries</li> <li>Grid 2x2 km</li> <li>Total 1.3 MIO points!</li> <li>2006 -&gt; 2018</li> </ul>		2006	11	168 402
		2009	23	234 623
		2012	27	270 272
		2015	28	339 696
<b>LUCAS project evolves</b> since 2006; => need of harmonization!		2018	28	337 854

# LUCAS dataset

What to harmonize?

- Attributes (new, removed)
- Attribute names (e.g. lc1\_pct vs. lc1\_perc)
- New definitions (continuous variables vs. categorical)
- Data types (string vs. integer)
- No valid data: 8

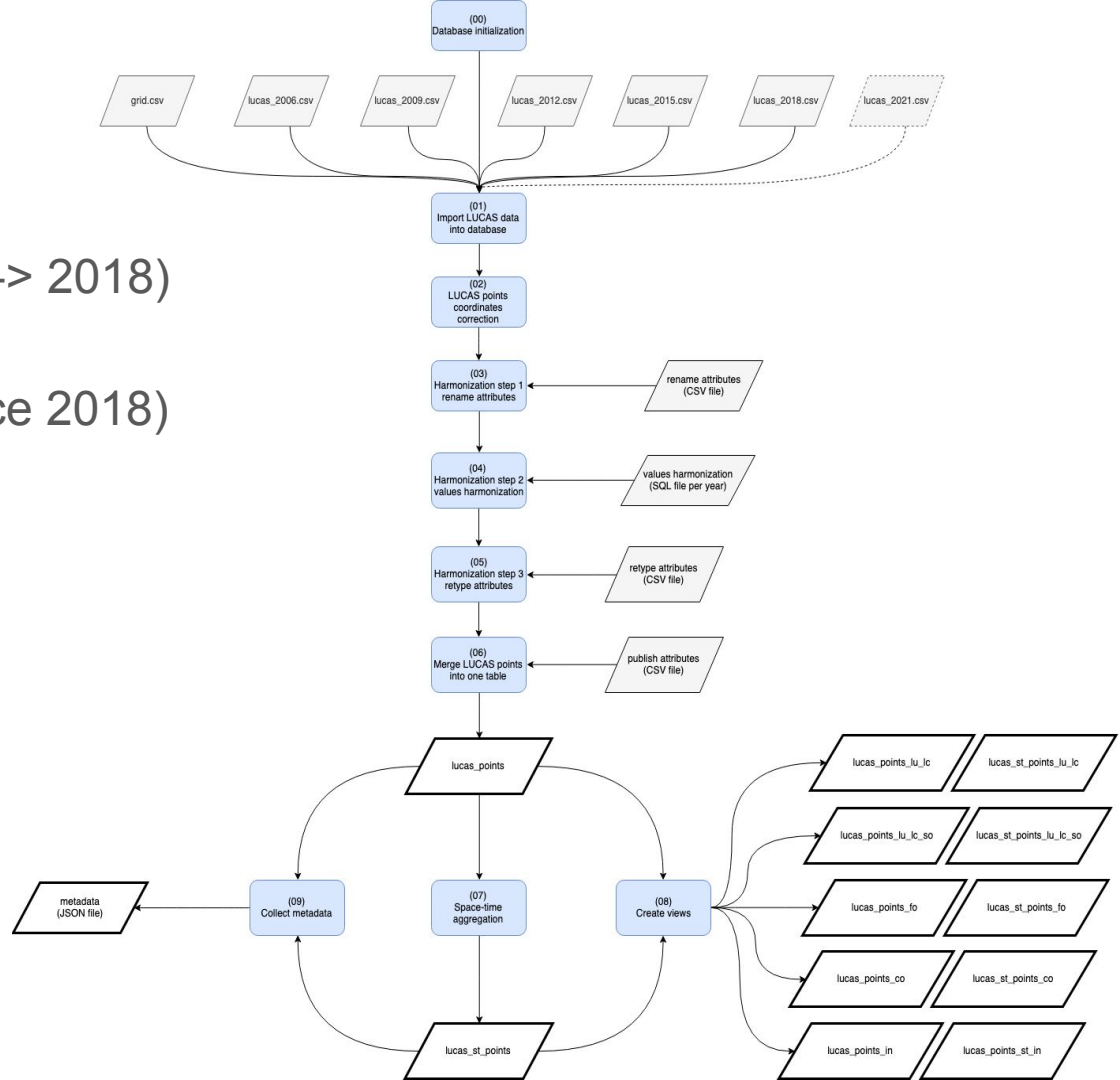
Year	Attributes
2006	20
2009	44
2012	46
2015	59
2018	97



## LUCAS harmonization:

1. import primary data (2006 -> 2018)
2. coordinates correction
3. rename attributes (reference 2018)
4. values harmonization
5. re-type values
6. merge into one table
7. space-time aggregation

=> distribute



# Hands-on #2: Validation with LUCAS datasets

General steps:

1. prepare land cover product for validation: `./sample_land_cover/cz_land*.tif`
2. prepare LUCAS points for the AOI: `./sample_land_cover/cz_lucas*.gpkg`
3. run validation in Jupyter notebook: `./02_land_cover_validation.ipynb`
4. discuss the results;

## Land Cover Validation with LUCAS dataset

This is an example of the land cover product validation using LUCAS points. The process is using class `Validator` to perform the main validation steps.

```
In [1]: import os
import sys
import yaml

from osgeo import gdal
from osgeo import gdalconst
```

# Hands-on #3: Nomenclature translation

Run nomenclature translation in Jupyter notebook: [./03\\_lucas\\_analyze.ipynb](#)

LUCAS to CORINE land cover legend

## Nomenclature translation

```
In [5]: from eumap.datasets.lucas import LucasClassTranslate
```

```
lucastrans = LucasClassTranslate(lucasio.data)  
lucastrans.set_translations("clc3")  
lucastrans.apply()
```

```
In [10]: df = lucasio.to_geopandas()  
# df[(df["clc3"] == None)]
```

```
Out[10]:
```

point_id	nuts0	nuts1	nuts2	nuts3	survey_date	car_latitude	car_longitude	car_ew	gps_proj	...	photo_north	photo_east	photo_south	photo_west	crop_i
----------	-------	-------	-------	-------	-------------	--------------	---------------	--------	----------	-----	-------------	------------	-------------	------------	--------

0 rows x 109 columns

# Hands-on #4: Nomenclature aggregation

Run class aggregation in Jupyter notebook: [./03\\_lucas\\_analyze.ipynb](#)

## Apply aggregation

LUCAS points are obtained with level 3 lucas land cover information. With `LucasClassAggr` method you can get information about level 2 or level 1. New column will be added to the end of attribute table.

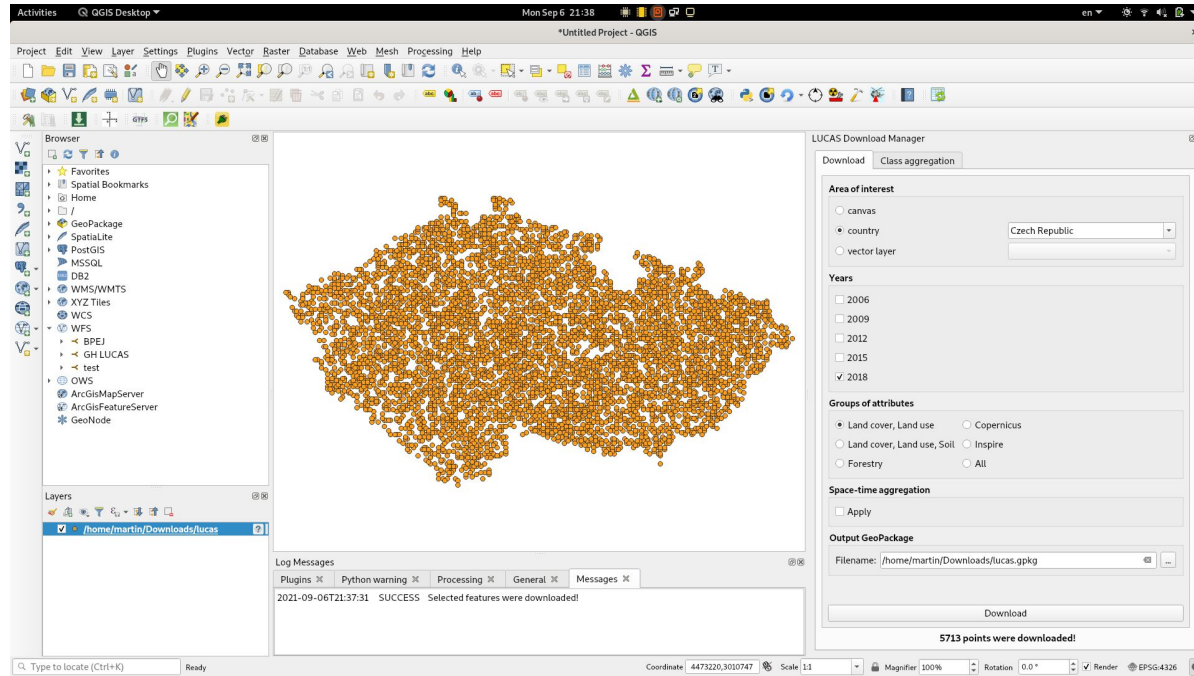
```
In [4]: lcl_to_level1 = {
    "A00": ["A11", "A12", "A13", "A21", "A22", "A30"],
    "B00": ["B11", "B12", "B13", "B14", "B15", "B16", "B17", "B18", "B19", "B21", "B22", "B23", "B31", "B32", "B33", "E
    "Bx1": ["Bx1"],
    "Bx2": ["Bx2"],
    "C00": ["C10", "C21", "C22", "C23", "C31", "C32", "C33"],
    "D00": ["D10", "D20"],
    "E00": ["E10", "E20", "E30"],
    "F00": ["F10", "F20", "F30", "F40"],
    "G00": ["G11", "G12", "G21", "G22", "G30", "G40", "G50"],
    "H00": ["H11", "H12", "H21", "H22", "H23"]
}

lucasaggr = LucasClassAggregate(lucasio.data, mappings=lcl_to_level1)
# lucasaggr = LucasClassAggregate(lucasio.data, mappings_file='aggregation_lcl_h_to_level1.json')
lucasaggr.apply()
```

# Hands-on #4: LUCAS QGIS plug-in

[./lucas\\_download\\_manager.zip](#)

Report issues on Gitlab: [https://gitlab.com/geoharmonizer\\_inea/qgis-lucas-plugin](https://gitlab.com/geoharmonizer_inea/qgis-lucas-plugin)



# Wrap-up LUCAS session

Learned how to ..

- Access the harmonized LUCAS dataset with Python API and QGIS;
- Filter the subsets;
- Use LUCAS for land cover validation;
- .. and it can be used the same way for land cover model calibration!;
- Translate LUCAS nomenclature to e.g. CORINE;
- Aggregate the land cover classes before validation / calibration;

in order to perform land cover research with open data and open software!

**Thanks for your attention!**



Geospatial system for LUCAS data harmonization  
and distribution

to be published as Open Source!