

WORKING WITH SPATIAL DATA

WORKSHOP @ THE PLANT
FASOS
APRIL 2023

Dr. Monika Barget (Maastricht/Mainz)



OVERVIEW



- 1) What are "geodata" and why do they matter in research?
- 2) Where do you find geodata and what processing is required?
- 3) What tools are available for editing spatial data?
- 4) Interactive session: geocoding with Python and the Geonames API

1) What are "geodata"?



DEFINITION "GEODATEN"

"Geodata is information assigned to a spatial location (georeference). 90% of all data can be attributed to a location or have a spatial reference and are therefore geodata."

<https://www.geoportal.rlp.de/>

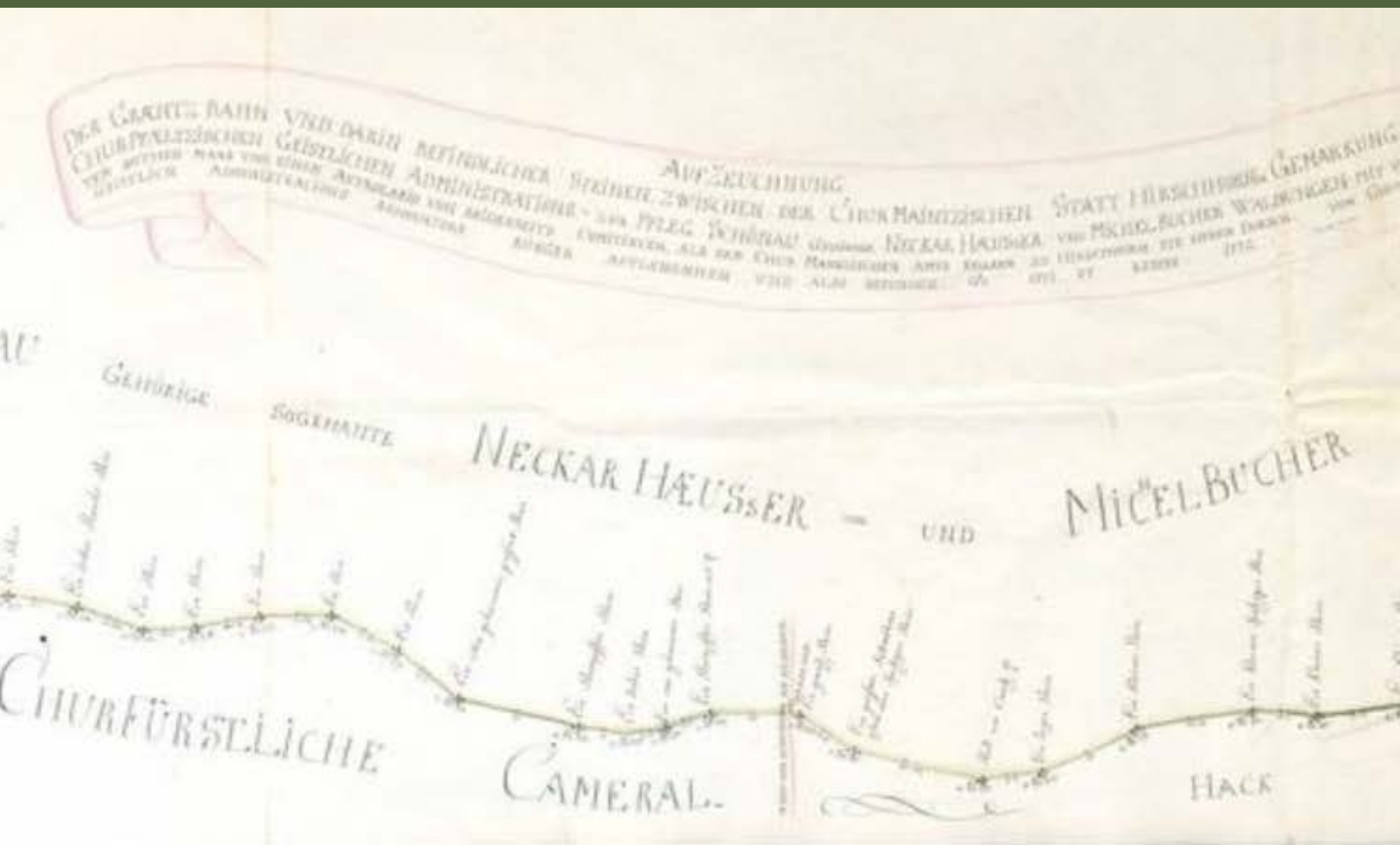
"Historical geodata offer the same research possibilities as current geo-related data. As data protection concerns relating to persons do not have to be taken into account when working with historical data, personal data can also be visualised in geoinformation systems."

<https://www.leo-bw.de/themenmodul/sudwestdeutsche-archivalienkunde/archivaliengattungen/raumbezogene-abbildungen/geodaten>

GEODATA IN RESEARCH

PRIMARY GEODATA

- explicit place names in gazetteers
- place names in encyclopaedias (e.g. Zedler's Lexicon)
- travel descriptions
- place data in (hand-drawn or printed) maps
- place names on letterheads, on telegrams, etc.



SECONDARY GEODATA

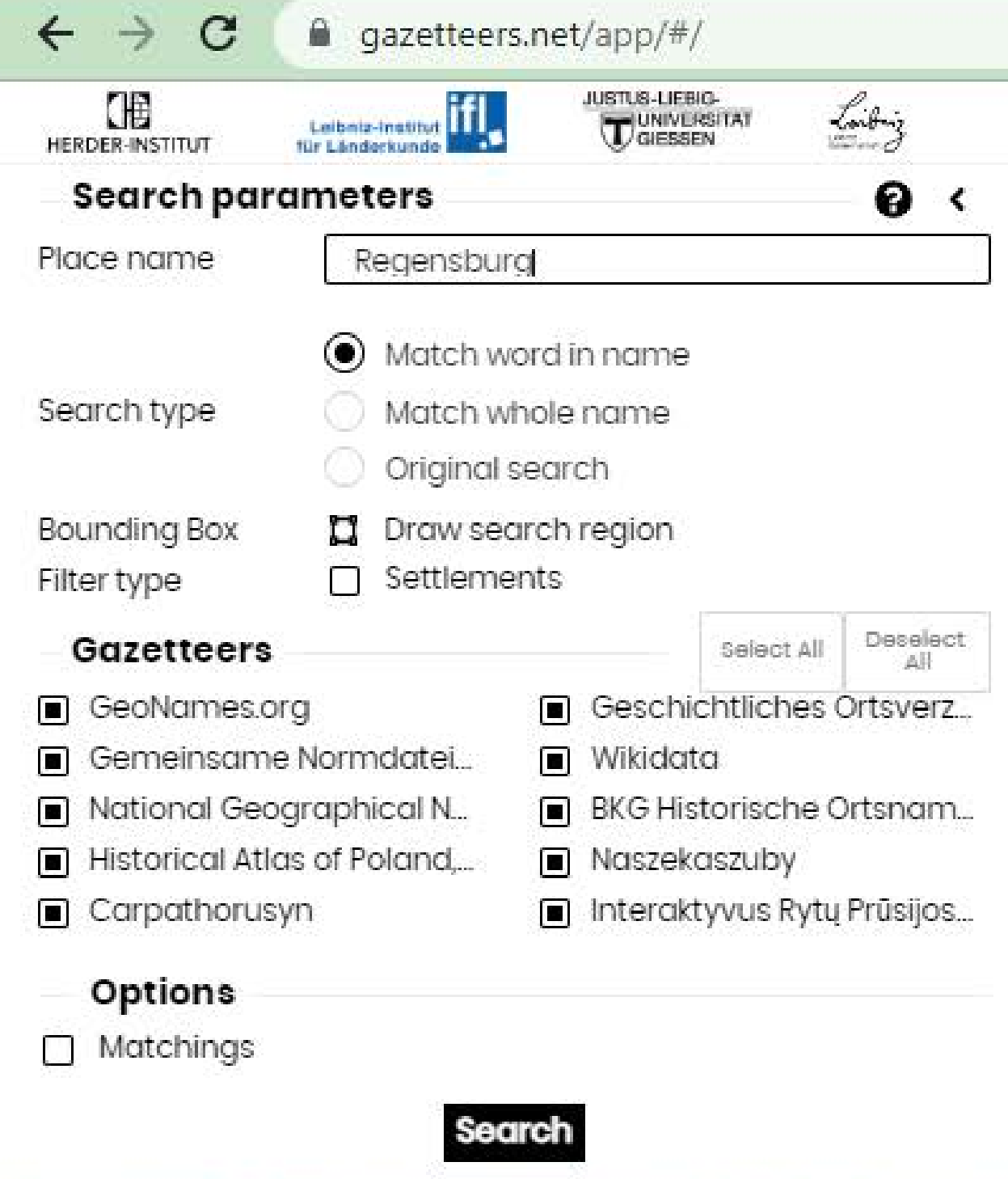
- reconstructed location data from available metadata (often based on results of provenance research)
- location data derived from biographies
- location data derived from institutional information
- location data derived from archaeological finds



EXAMPLE: "DIGIKAR"

- 3-year cooperation project of the Leibniz Institute of European History Mainz (IEG), the Leibniz Institute for Regional Geography Leipzig (IfL), the Leibniz Institute for East and Southeast European Research Regensburg (IOS), the Johannes Gutenberg University Mainz (JGU) and the École des Hautes Études en Sciences Sociales Paris, France (EHESS).
- Alternative and experimental collection, modelling and visualisation of historical location data from Electoral Mainz and Electoral Saxony (17th & 18th c.)

<https://digikar.eu/>

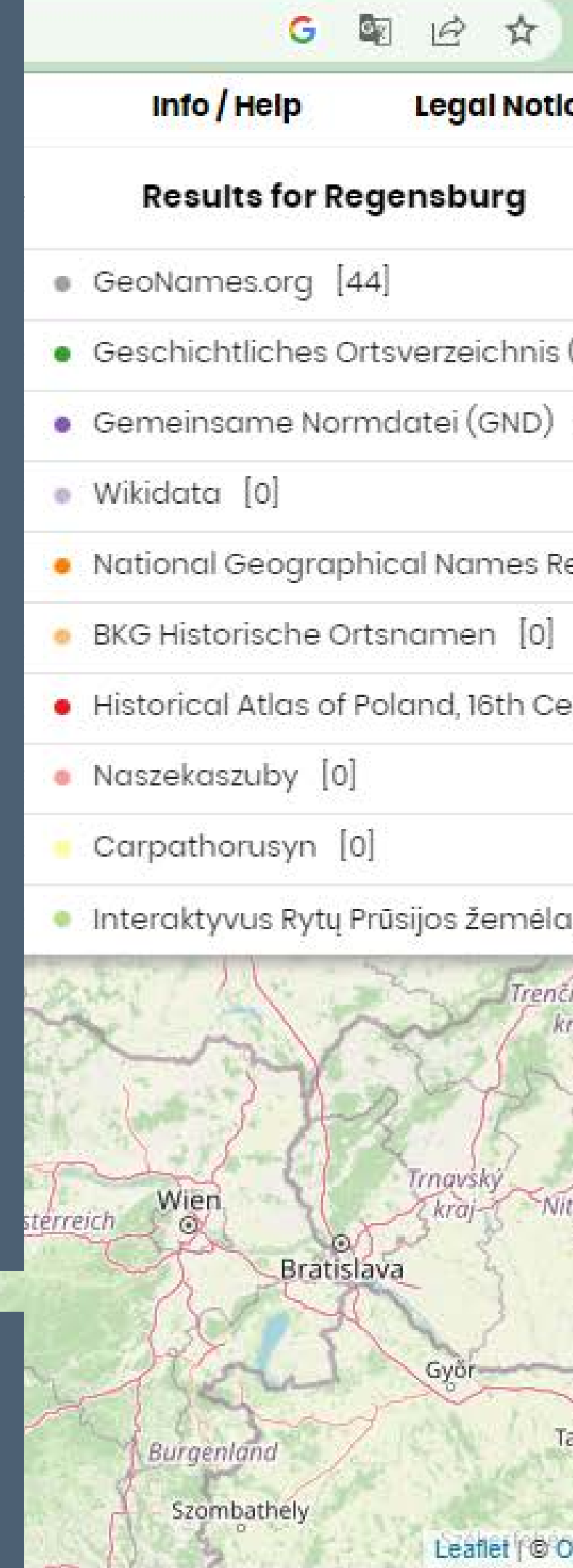


EXAMPLE: "GAZETTEERS NET"

- "The gazetteers.net web application is developed as part of the Gazetteer research project by the Herder Institute (HI), the Institute for Regional Geography (IfL) and the Justus Liebig University Giessen (JLU). The application is intended to support users in working with different digital gazetteers, and to help them explore their content and metadata structure."
- Search for place names and links to existing entries from national and international (standard) databases (e.g. GND and Wikidata)



<https://gazetteers.net/app/#/>





EXAMPLE: "WHG"

Amsterdam

Search Index 🔗

Search Database 🗄️

pre-filters

PLACE CATEGORIES [CHECK ALL](#) ?

- administrative entities (A)
- cities, towns, hamlets... (P)
- sites, buildings, complexes... (S)
- roads, routes, rail... (R)
- regions, landscape areas (L)
- terrestrial landforms (T)
- water bodies (H)

TEMPORAL (DEFAULT)

1700

SPATIAL

Region, country

[reset search](#)

- "World Historical Gazetteer (WHG) is providing a collection of content and services that permit world historians, their students, and the general public to do spatial and temporal reasoning and visualization in a data rich environment at global and trans-regional scales."
- Data contributions in two formats: Linked Places and Linked Traces
- in-built analysis and visualisation for contributors
- linking individual data sets (from current research projects) with existing data for future research and teaching
- focus on time-dimensions / place developments



UNION INDEX SEARCH RESULTS (1)

List may include records with no geometry

Title ⚡	Linked ⚡	Countries ⚡	Type(s)
Amsterdam	4	NL	inhabited p

<https://whgazetteer.org/>



Time
Machine

ABOUT US

BUILDING A TIME MACHINE

TIME

EXAMPLE: "TIME MACHINE"

- "Time Machine is aiming to join Europe's rich past with up-to-date digital technologies and infrastructures, creating a collective digital information system mapping the European economic, social, cultural and geographical evolution across times."
- Leaders = Time Machine Organisation (TMO) Wien
- Members: ca. 800 cultural institutions, universities, tech companies etc. worldwide
- Funding: Austrian government & Horizon2020

<https://www.timemachine.eu/>

ADDING A NEW DIMENSION TO THE PAST

2) Where to find geodata...



CHALLENGES

- Geodata exist in various types of sources from handwritten archival records to printed maps.
- Geodata have not always been systematically recorded in archival history (e.g. in metadata).
- Existing structured geodata are often time-dependent (see references to political structures of the 1930s in the metadata of the Mainz Birth and Teaching Letters).
- Extracting and modelling geospatial data from historical sources is time-consuming and often requires difficult decisions in the area of normalisation and data uncertainty.
- **IMPORTANT:** Place names per se are not sufficient!



FROM PLACE NAME TO GEO INFORMATION

NAME



identification as a "place name": e.g. Paris as a city vs. Paris as a mythological figure



Time-dependent "place types"
(if necessary, orientation
towards existing ontologies):

e.g. a village that develops into
a city and thus also receives
new rights.

Time-dependent geocoding
(mostly manual correction of
automatic geocoding):

e.g. a monastery that moved its
location after the Thirty Years'
War.

FORMATS FOR INTEROPERABLE GEODATA

Geodata cannot simply be collected as lists in WORD files. Even the collection in tables is usually only the first step towards further processing. Widely used formats for geodata:

Shapefiles

(developed by ESRI for ArcView):

- .shp (geometries)
- .dbf (attributes)
- .shx (index to create links)

GeoJSON

(standardised since 2015 by the Geographic JSON working group of the Internet Engineering Task Force):

- .geojson (or .json, joint recording of geometries and attributes in a file)

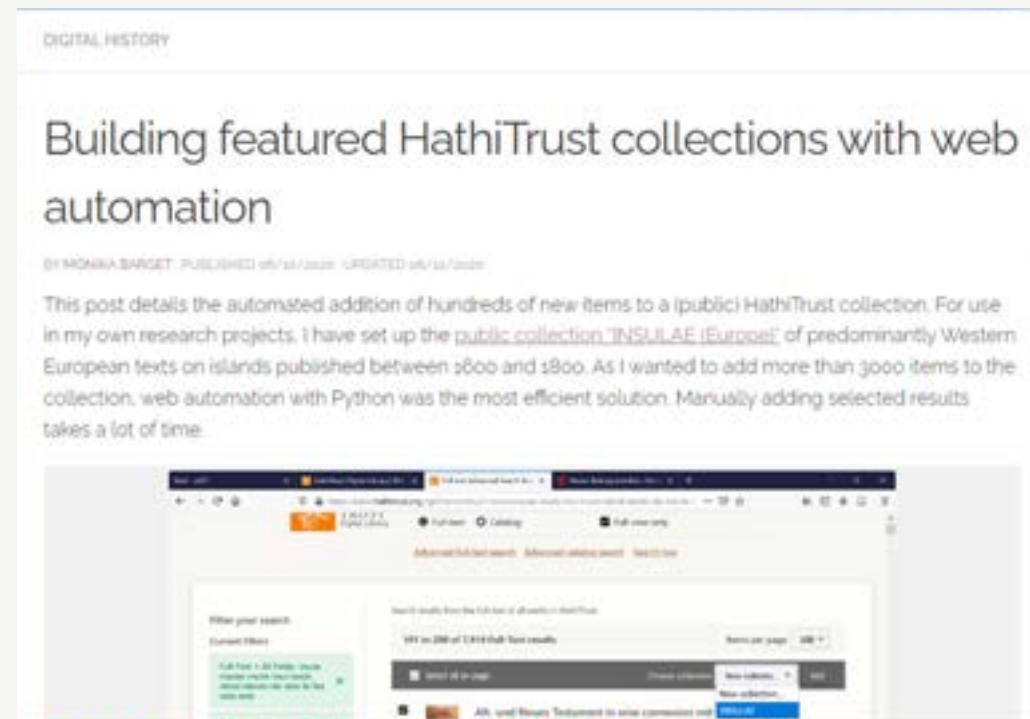
EXTRACTING SPATIAL DATA

Calling data via API

```
<change>
  <date type="creation" n="1432309523" value="2015-05-22">
  <date type="lastmod" n="1587714958" value="2020-04-24">
</change>
</revisionDes>
</teiHeader>
<text>
</body>
</div>
<listPerson>
  <person xml:id="cd515562-be2e-40a0-8427-ac3179667e0d">
    <persName>
      <roleName type="honorific" full="abb">Prof. Dr. med.</roleName>
      <forename>Friedrich</forename>
      <addName>Wilhelm</addName>
      <surname>Brussatis</surname>
      <reg>Brussatis, Friedrich</reg>
    </persName>
    <sex value="1">male</sex>
    <birth when="1919-04-01">
      <date>01.04.1919</date>
      <placeName>
      </birth>
    <death when="1989-01-26">
      <date>26.01.1989</date>
      <placeName>
      </death>
    <faith>0</faith>
  </person>
</listPerson>
</text>
</div>
```

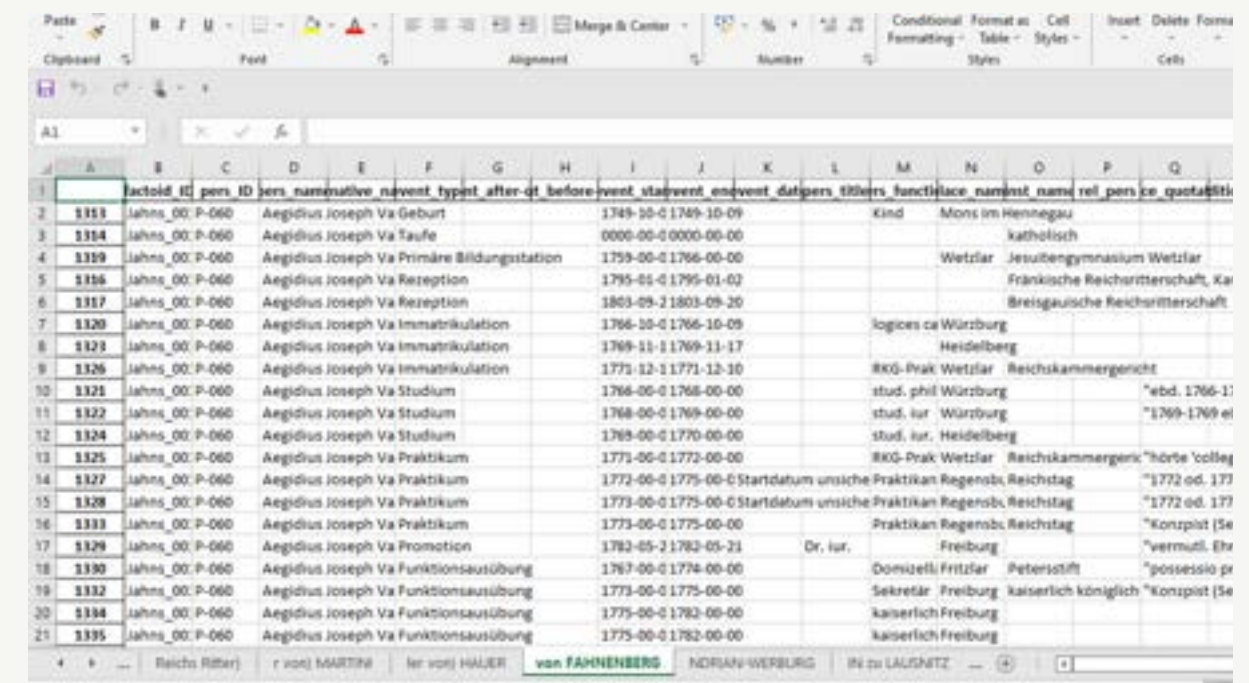
<http://gutenberg-biographics.ub.uni-mainz.de/api/items/persons/>

Webscraping
(e.g. Metadata Harvesting)



<https://insulae.hypotheses.org/169>

Data collection in table format (e.g. EXCEL and CSV)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q		
1		factoid ID	pers ID	pers name	navevent	typent	after-ot	before-vent	start-vent	end-vent	datpers	titlers	functiplace	naminst	name	rel	pers	ce	quotatiti
2	1113	Jahns_00-P-060	Aegidius Joseph Va	Geburt				1749-10-0	1749-10-09				Kind	Mons im Hennegau					
3	1114	Jahns_00-P-060	Aegidius Joseph Va	Taufe				0000-00-0	0000-00-00					katholisch					
4	1119	Jahns_00-P-060	Aegidius Joseph Va	Primäre Bildung	station			1759-00-0	1766-00-00				Wetzlar	Jesuitengymnasium Wetzlar					
5	1116	Jahns_00-P-060	Aegidius Joseph Va	Rezeption				1795-01-0	1795-01-02					Fränkische Reichsritterschaft, Ka					
6	1117	Jahns_00-P-060	Aegidius Joseph Va	Rezeption				1803-09-2	1803-09-20					Breisgauische Reichsritterschaft					
7	1120	Jahns_00-P-060	Aegidius Joseph Va	Immatrikulation				1766-10-0	1766-10-09				logices ca Würzburg						
8	1121	Jahns_00-P-060	Aegidius Joseph Va	Immatrikulation				1769-11-1	1769-11-17				Heidelberg						
9	1126	Jahns_00-P-060	Aegidius Joseph Va	Immatrikulation				1771-12-1	1771-12-10				IKG-Prak Wetzlar	Reichskammergericht					
10	1121	Jahns_00-P-060	Aegidius Joseph Va	Studium				1766-00-0	1766-00-00				stud. phil Würzburg						"ebd. 1766-1
11	1122	Jahns_00-P-060	Aegidius Joseph Va	Studium				1768-00-0	1769-00-00				stud. iur Würzburg						"1769-1769 ei
12	1124	Jahns_00-P-060	Aegidius Joseph Va	Studium				1769-00-0	1770-00-00				stud. iur. Heidelberg						
13	1125	Jahns_00-P-060	Aegidius Joseph Va	Praktikum				1771-00-0	1772-00-00				IKG-Prak Wetzlar	Reichskammergericht	"hörte 'colleg				
14	1127	Jahns_00-P-060	Aegidius Joseph Va	Praktikum				1772-00-0	1775-00-0	C	Startdatum unsiche	Praktikan	Regensb.	Reichstag					"1772 od. 177
15	1128	Jahns_00-P-060	Aegidius Joseph Va	Praktikum				1773-00-0	1775-00-0	C	Startdatum unsiche	Praktikan	Regensb.	Reichstag					"1772 od. 177
16	1133	Jahns_00-P-060	Aegidius Joseph Va	Praktikum				1773-00-0	1775-00-00				Praktikan	Regensb.	Reichstag				"Konzipit [Se
17	1129	Jahns_00-P-060	Aegidius Joseph Va	Promotion				1783-05-2	1783-05-21				Dr. iur.	Freiburg					"vermutl. Ehz
18	1130	Jahns_00-P-060	Aegidius Joseph Va	Funktionsausübung				1767-00-0	1774-00-00				Donszell; Fritzlar	Petersstift					"possessio pr
19	1132	Jahns_00-P-060	Aegidius Joseph Va	Funktionsausübung				1773-00-0	1775-00-00				Sekretär	Freiburg	kaiserialich königlich	"Konzipit [Se			
20	1134	Jahns_00-P-060	Aegidius Joseph Va	Funktionsausübung				1773-00-0	1782-00-00				kaiserialich	Freiburg					
21	1135	Jahns_00-P-060	Aegidius Joseph Va	Funktionsausübung				1775-00-0	1782-00-00				kaiserialich	Freiburg					

Example of "factoid" table structure used in the DigiKAR project

EXAMPLE: EXTRACTION OF LOCATION DATA FROM PRINTED MAPS WITH THE OCR WORK TIME TRANSKRIBUS LITE

The screenshot displays the Transkribus Lite web interface. At the top, the navigation bar includes the Transkribus Lite logo, links for Home, Jobs, and Info, and a user profile for m.barget@maastrichtuniversity.nl with options for Credits and EN. Below this, a breadcrumb trail shows 'Collections > MAPS > Topographia1646_maps > Page 6'. The main content area is split into two panels. The left panel shows a historical map with various place names, including 'Westerburg WALD.', 'Dalem', 'Hollach', and 'Hademer'. A small rectangular box highlights a specific location on the map. The right panel features a control bar with an 'Annotation' toggle, font size controls (A+, A-), and a dropdown menu set to 'In Progress'. Below the control bar, a list of extracted location names is displayed, with 'Dalem' highlighted in a grey bar. The list includes: Ludesdorff / Wied / Isenberg, Delem, Waldmanshausen, Hammerstein, Hollach, Stockhausen / LHON, Andernach / Enger, Hademer, Leune, and Montabour. A vertical 'Feedback' button is visible on the far right edge.

Transkribus
lite

Home Jobs Info

m.barget@maastrichtuniversity.nl Credits EN

< 6 6 Go > Collections > MAPS > Topographia1646_maps > Page 6

IN_PROGRESS | m.barget@maastrichtuniversity.nl

Annotation A+ A- In Progress

Ludesdorff / Wied / Isenberg

Dalem

Waldmanshausen

Hammerstein

Hollach

Stockhausen / LHON

Andernach / Enger

Hademer

Leune

Montabour

Feedback

3) Tools for editing
spatial data...




```

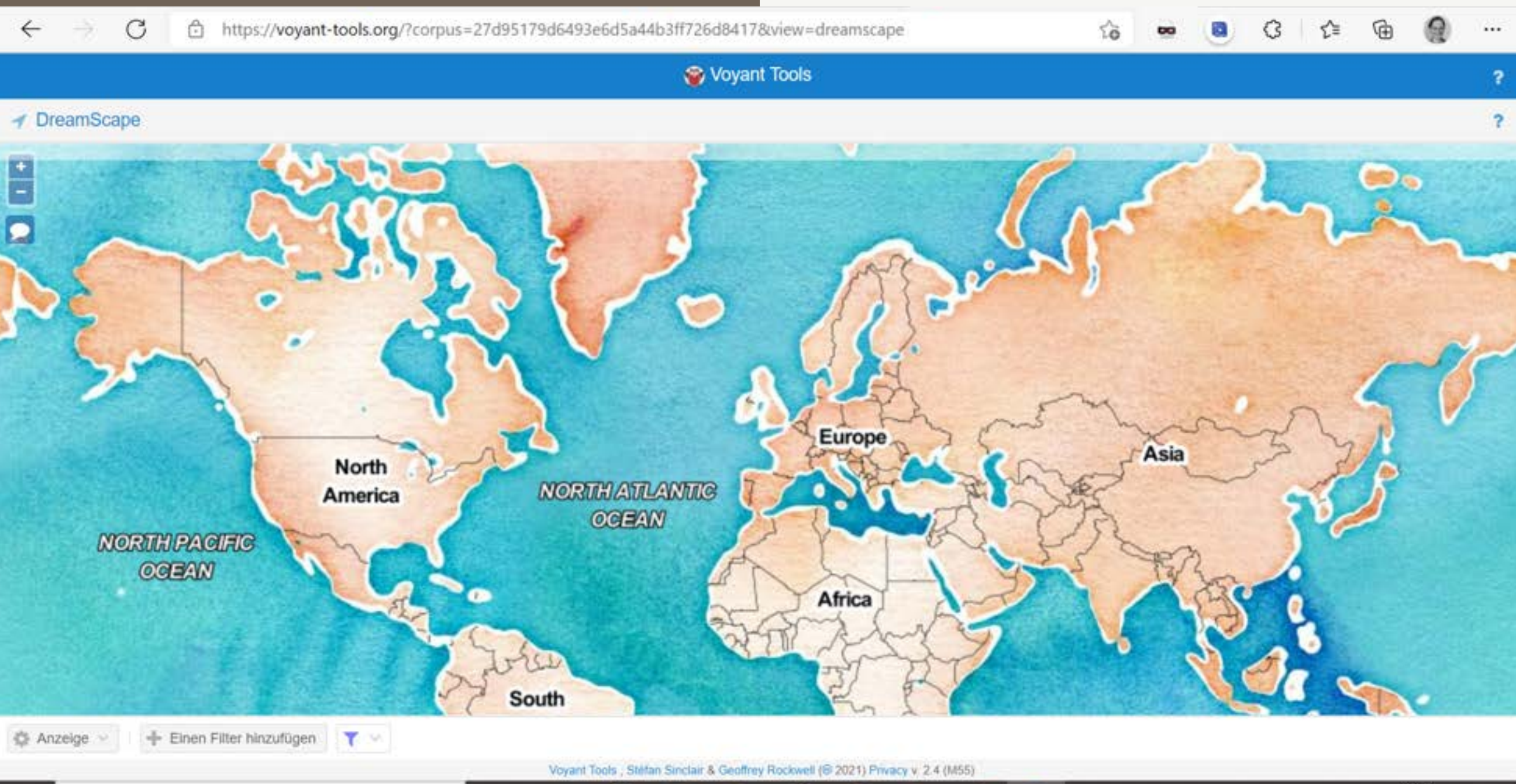
Tab-delimited-text_Palladio.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe
Places "Coordinates"
Mainz "50.0012314, 8.2762513"
Meißen "51.1630871, 13.4704939"
Wiesbaden "50.0820384, 8.2416556"
Köln "50.938361, 6.959974"
Paris "48.8588897, 2.320041"
Bonn "50.735851, 7.10066"
Kopenhagen "55.6867243, 12.5700724"
Eltville "50.0559826, 8.0891474"
Maastricht "50.8579855, 5.6969882"
Lanaken "50.8892784, 5.6513208"
Heerlen "50.8775239, 5.9815066"
Aachen "50.776351, 6.083862"

```

PALLADIO (BROWSER)

Palladio (Stanford University) allows the creation of simple maps and network diagrams. The data should ideally be entered in a "tab-delimited" text format. The possibilities for visualisation are limited. A high-quality export of static maps is not possible.





Voyant Tools is a constantly growing collection of text analysis tools. Data can be uploaded as TXT, XML or even PDF (with embedded OCR).

The "Dreamscapes" tool promises to automatically recognise place names in texts and display them on a map.

The results are very mixed. Often no places are recognised.

VOYANT
DREAMSCAPES
(BROWSER)



VOYANT

see through your text

GEOJSON.IO (BROWSER)

Semi-manual editing of geodata in GeoJSON format on geojson.io

"geojson.io is a tool for editing GeoJSON data on the internet. It enables editing through a map interface, raw GeoJSON, and exporting and importing a large number of formats. geojson.io is a tool for editing GeoJSON data on the internet."

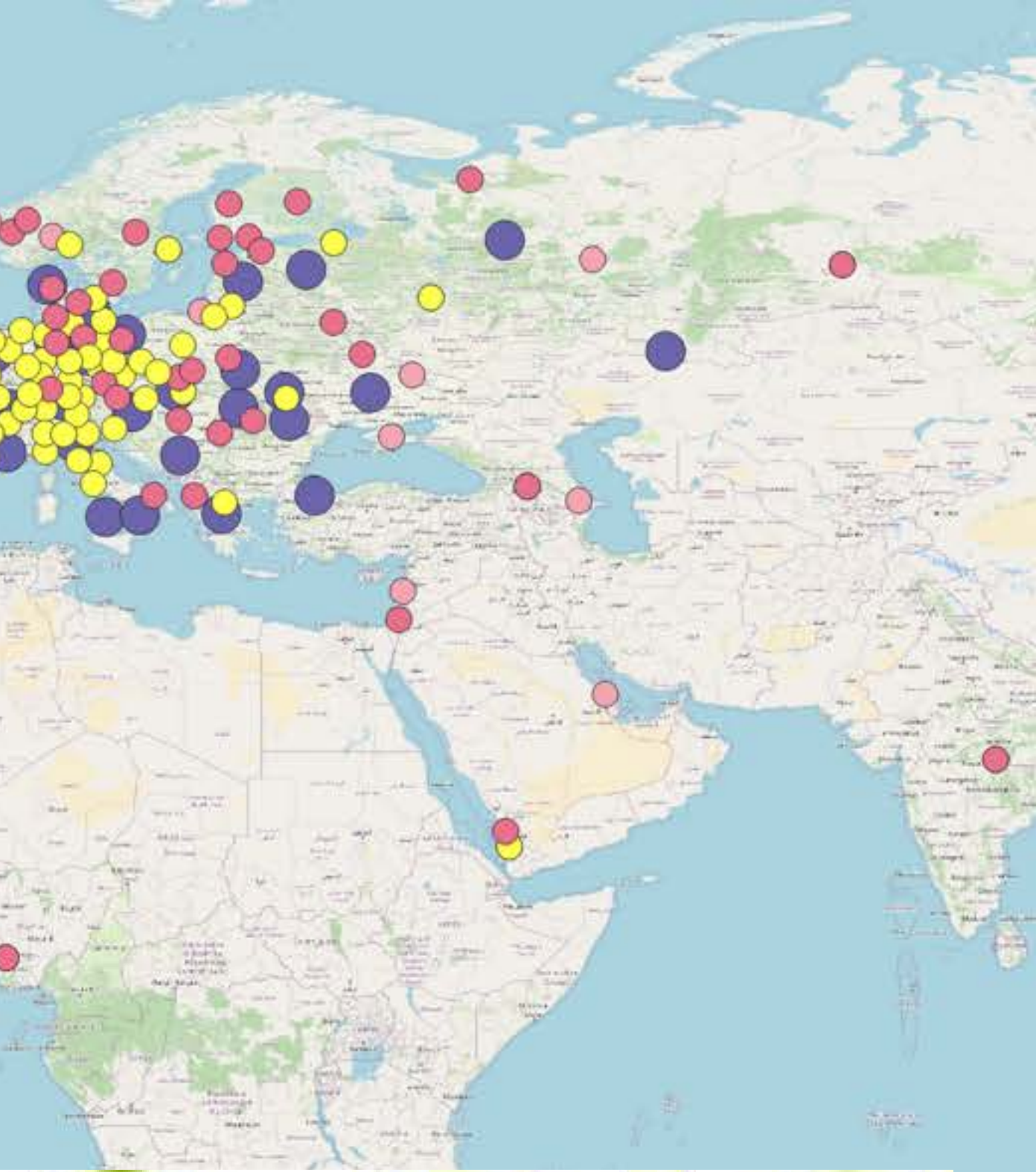
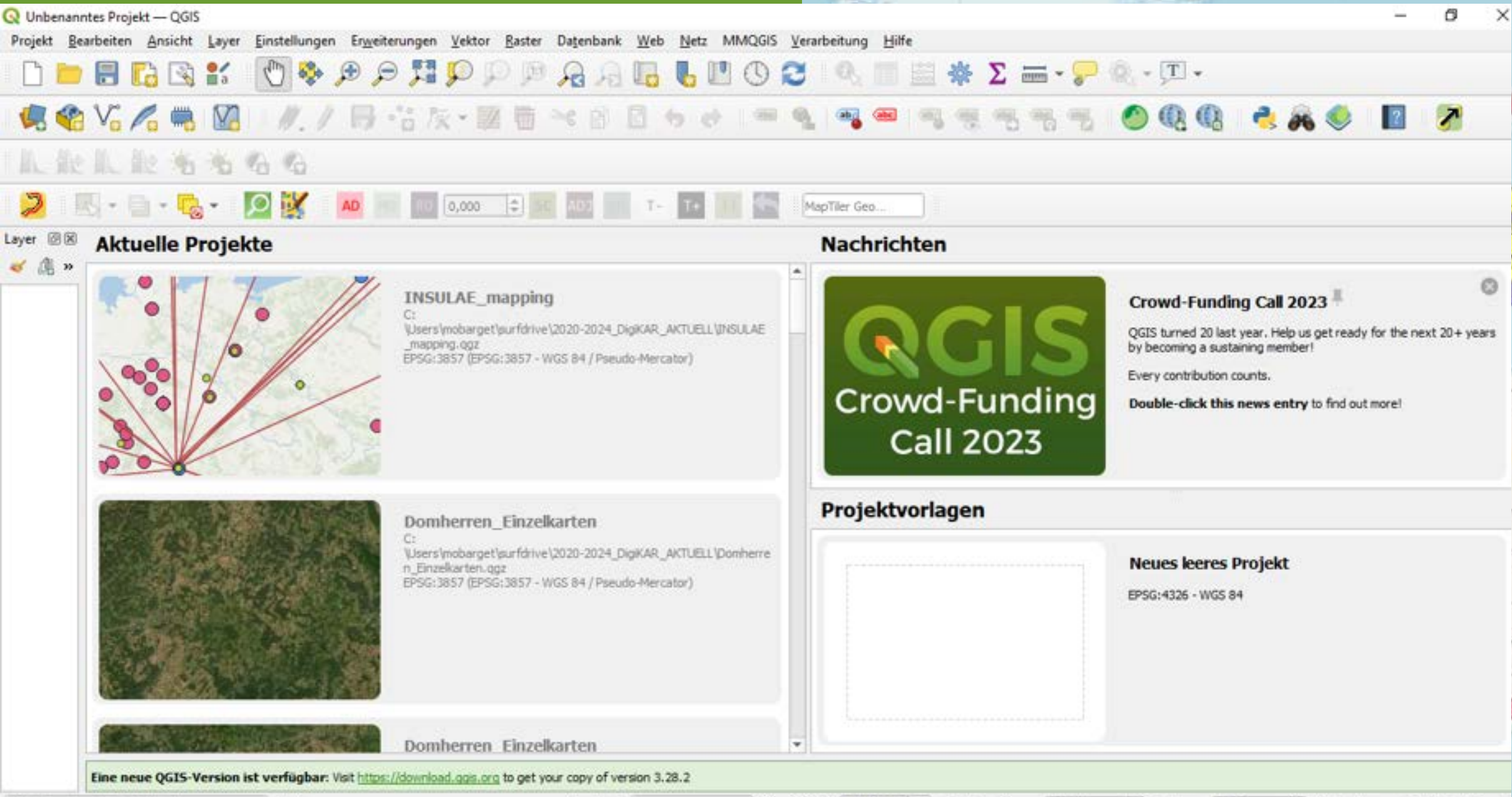
(GeoJSON, 12. October 2020)

Tutorial:

https://monikabarget.github.io/GeoHumTutorials/Tutorial_GeoJSON

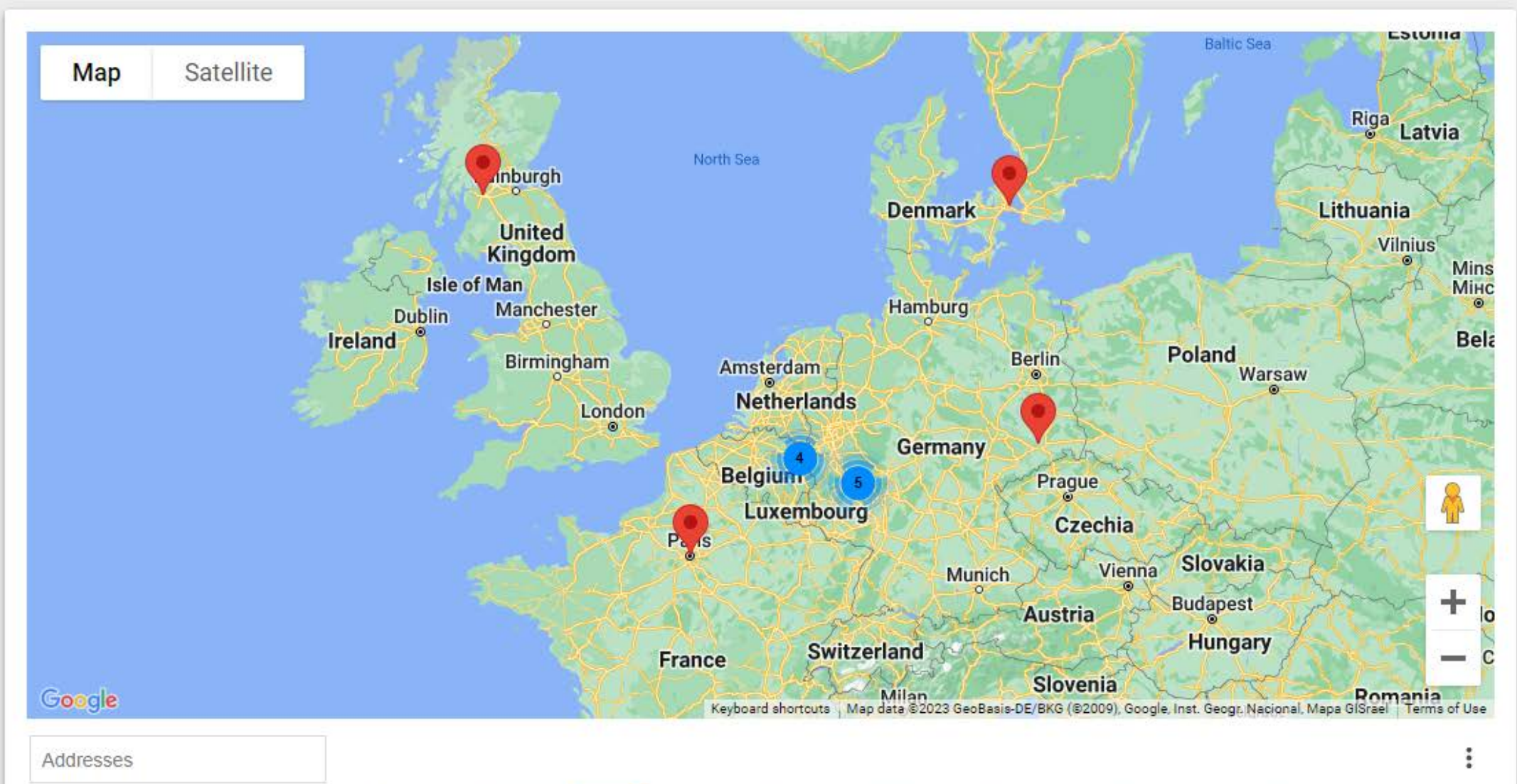
The screenshot displays the geojson.io web application. The browser's address bar shows the URL `https://geojson.io/#map=11/42.6908/18.0114`. The interface is divided into two main sections: a map on the left and a JSON editor on the right. The map shows a coastal region with a red pin at Dubrovnik. The JSON editor shows a GeoJSON FeatureCollection with two features: one for Dubrovnik (red pin) and another for Cavtat (blue pin). The JSON editor shows the following code:

```
1 {
2   "type": "FeatureCollection",
3   "features": [
4     {
5       "type": "Feature",
6       "properties": {
7         "marker-color": "#f30a6d",
8         "marker-size": "medium",
9         "marker-symbol": "",
10        "NAME": "Dubrovnik"
11      },
12      "geometry": {
13        "type": "Point",
14        "coordinates": [
15          18.0889892578125,
16          42.65416193033991
17        ]
18      }
19    },
20    {
21      "type": "Feature",
22      "properties": {
23        "marker-color": "#2440d9",
24        "marker-size": "medium",
25        "marker-symbol": ""
```



GIS SOFTWARE
(E.G. QGIS)





GOOGLE AWESOME



Geocode is a tool that helps you get latitudes & longitudes from addresses in a Google Sheet to display them on a map you can share.

By: Talarian

Listing updated: January 10, 2023



GEOCODIERUNG & PLOTTING WITH SCRIPTS

```
github.com/MonikaBarget/GeoHumTutorials/blob/master/Colab_Geocoding/Geocode_Plot_Geonames_interactiveMAP.ipynb
```

Geocoder for Python supports the following ones:

- (geocoding) retrieve GeoNames's geocoded data from a query string, and various filters
- (details) retrieve all geonames data for a given geonames_id
- (children) retrieve the hierarchy of a given geonames_id
- (hierarchy) retrieve all children for a given geonames_id

Full documentation: <https://geocoder.readthedocs.io/providers/GeoNames.html>

The first step is to get COLAB working:

```
In [ ]: ## mount drive
from google.colab import drive
drive.mount("/content/drive")
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

A file path needs to be defined for storing input or output files linked with this script:

```
In [ ]: directory="/content/drive/My Drive/Colab_Geohumanities/"
```

Now we can install packages that are not part of Python's standard distribution but are necessary for geocoding and plotting maps. There will most likely be a dependency error for NumPy, but the script should still work.

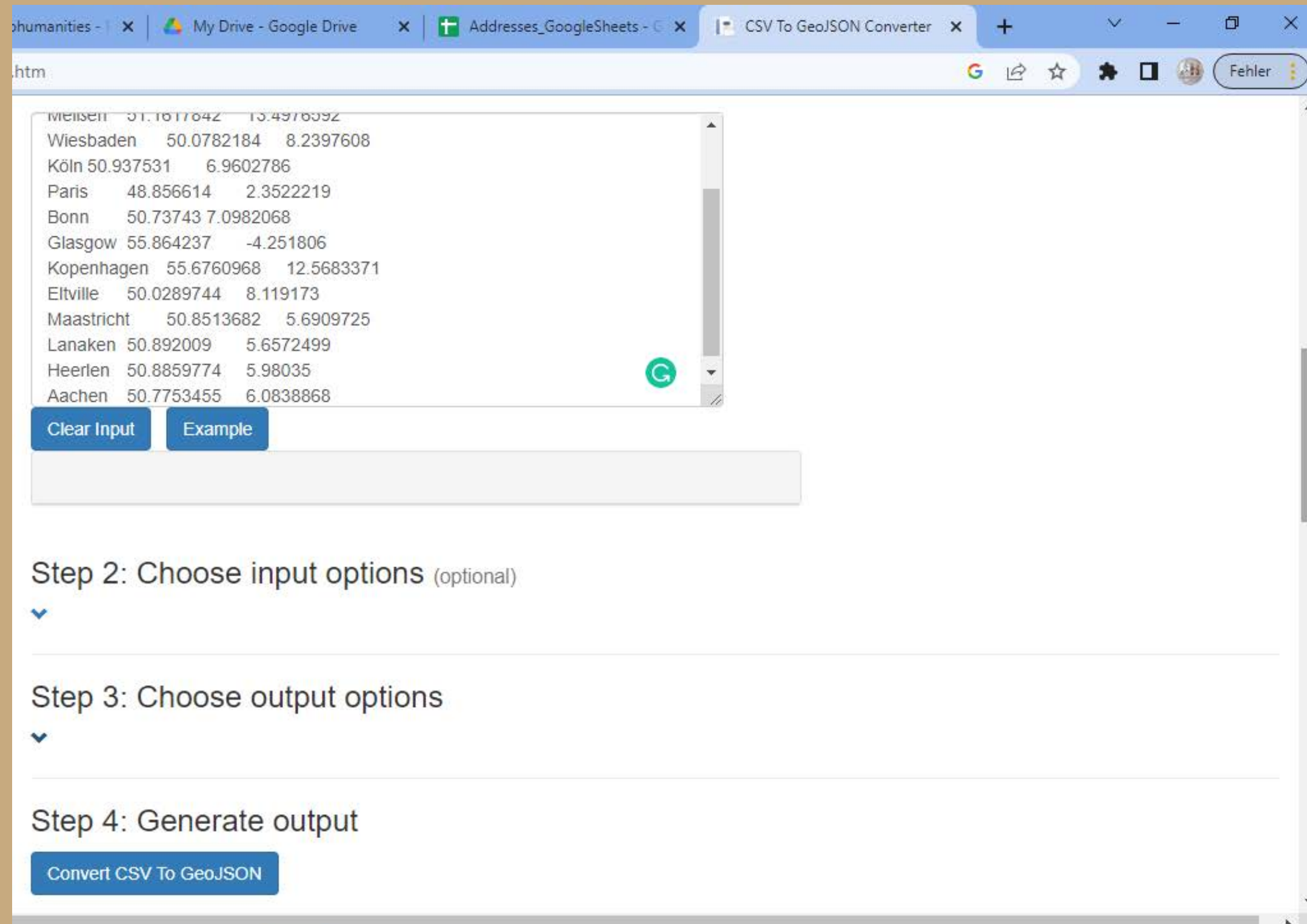
```
In [ ]: ## install packages that are not part of Python's standard distribution

!pip install geocoder
!pip install basemap
!pip install ipyleaflet
!pip install geojson
```

Suchbegriff hier eingeben

CSV TO GEOJSON:

[HTTPS://WWW.CONVERTCSV.COM/CSV-TO-GEOJSON.HTM](https://www.convertcsv.com/csv-to-geojson.htm)



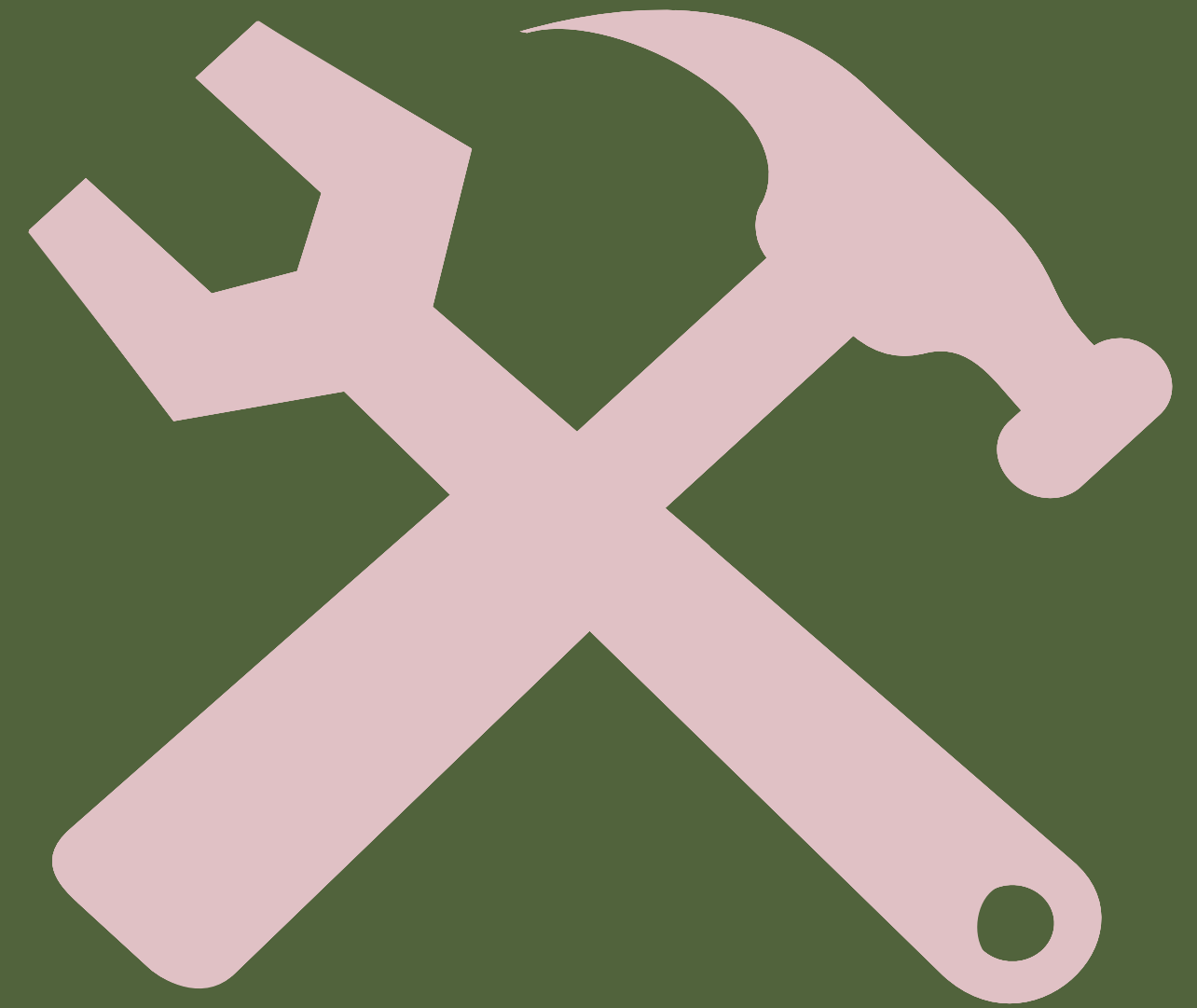
The screenshot shows a web browser window with the URL <https://www.convertcsv.com/csv-to-geojson.htm>. The page contains a text input field with a list of city names and their coordinates. Below the input field are two buttons: "Clear Input" and "Example". The page is divided into four steps: "Step 2: Choose input options (optional)", "Step 3: Choose output options", and "Step 4: Generate output". A large blue button labeled "Convert CSV To GeoJSON" is located at the bottom of the page.

Meisen	51.1617642	13.4976392
Wiesbaden	50.0782184	8.2397608
Köln	50.937531	6.9602786
Paris	48.856614	2.3522219
Bonn	50.73743	7.0982068
Glasgow	55.864237	-4.251806
Kopenhagen	55.6760968	12.5683371
Eltville	50.0289744	8.119173
Maastricht	50.8513682	5.6909725
Lanaken	50.892009	5.6572499
Heerlen	50.8859774	5.98035
Aachen	50.7753455	6.0838868

This online tool allows the creation of GeoJSON files from tables that cannot be read by GIS software.

However, the creation in the graphical interface is time-consuming. Scripts do this task faster.

4) Interactive session



TO TRY AT HOME: GOOGLE AWESOME

The screenshot shows a Google Sheets spreadsheet titled "Addresses_GoogleSheets" with a list of cities in column A. The "Extensions" menu is open, and the "Geocode by Awesome Table" option is selected, showing a sub-menu with "Start Geocoding", "Geocode on form submit", and "Help". On the right, the "Geocode" sidebar is open, displaying instructions and configuration options for geocoding the addresses.

	A	B
1	Addresses	
2	MapsTooltip - Hidden - StringFilter	
3	Mainz	
4	Meißen	
5	Wiesbaden	
6	Köln	
7	Paris	
8	Bonn	
9	Glasgow	
10	Kopenhagen	
11	Eltville	
12	Maastricht	
13	Lanaken	
14	Heerlen	
15	Aachen	
16		
17		
18		
19		
20		
21		
22		
23		

Geocode

Geocode gets latitudes and longitudes from **full addresses** and creates a customizable map.

Current sheet
Tabelle1

Address column
Addresses

[Are your addresses in multiple columns?](#)

Try wider results

Map

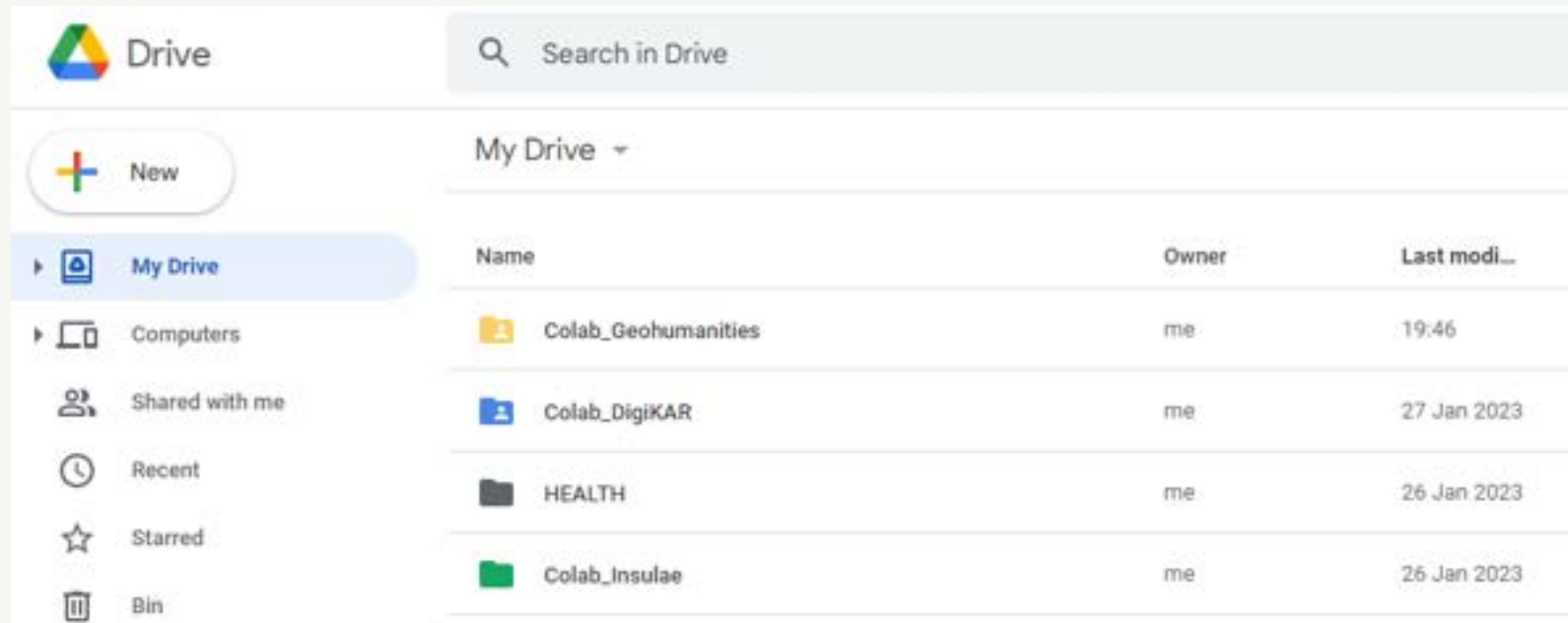
The map has been created. If you click on a marker in the map, the tooltip will simply display the address. But you can do much better and display any info you want in this tooltip.

[Customize Tooltip](#) [Open map](#)

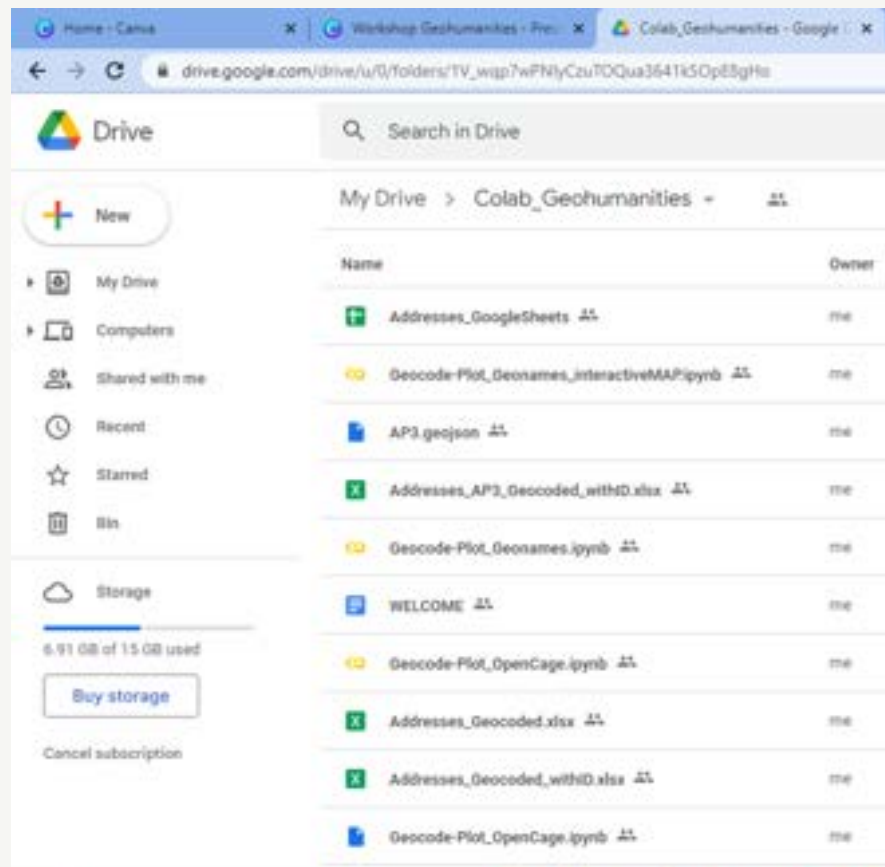
[Geocode!](#)

Proudly brought to you by [Awesome Table](#)

GEOCODING WITH PYTHON & API



Copy the synchronised folder or download the virtual folder and add it to the drive as your own folder with a new name (!).



Open scripts in Jupyter Notebook format.

Add the path to your own folder in the script and execute the code:

```
DIRECTORY="/CONTENT/  
DRIVE/MY DRIVE/  
COLAB_GEOHUMANITIES/"
```



RUN



OK

The screenshot shows a Google Colab notebook interface. The browser address bar displays the URL: `colab.research.google.com/drive/13nCyb2p2KU2YZUBr7oN9jcylyX4HcpBR`. The notebook title is "Geocode-Plot_Geonames_interactiveMAP.ipynb" and it was last edited on 26 January. The notebook content includes:

- A section titled "GeoNames." with the text: "GeoNames is mainly using REST APIs. It offers 40 different webservices."
- A section titled "Geocoder for Python supports the following ones:" followed by a bulleted list:
 - (geocoding) retrieve GeoNames's geocoded data from a query string, and various filters
 - (details) retrieve all geonames data for a given geonames_id
 - (children) retrieve the hierarchy of a given geonames_id
 - (hierarchy) retrieve all children for a given geonames_id
- A link to "Full documentation: <https://geocoder.readthedocs.io/providers/GeoNames.html>"
- A text prompt: "The first step is to get COLAB working:"
- A code cell with a play button icon and the following code:

```
## mount drive
from google.colab import drive
drive.mount("/content/drive")
```
- A message: "Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount(\"/content/drive\", force_remount=True)." with a copy icon.
- A text prompt: "A file path needs to be defined for storing input or output files linked with this script:"
- A code cell with the following code:

```
[ ] directory="/content/drive/My Drive/Colab_Geohumanities/"
```

THANKYOU

DigiKAR Website: <https://digikar.eu/>

(inklusive Links zu Vorträgen, Daten, Blogposts und anderen Veröffentlichungen)

Geohumanities Repository:

<https://monikabarget.github.io/GeoHumTutorials>

Monika's research blog:

<https://insulae.hypotheses.org/>

M

m.barget@maastrichtuniversity.nl

