



# An Interoperable Data Economy to Enable GeoAI via Spatial Tokenizers



BUSINESS  
INCUBATION  
CENTRE

*Michael Jendryke<sup>1,2</sup>\*, João Manuel<sup>1,2</sup>, Ludovic Augé<sup>3</sup>, Emmanuel Mondon<sup>4</sup>, Gino Caspari<sup>1,2</sup>*

<sup>1</sup>GeoInsight, Kaiserstr. 25, 59505 Bad Sassendorf, DE; <sup>2</sup>GeoInsight, Forstweg 65A, 3012 Bern, CH;

<sup>3</sup>Dataionics, 1 rue Jean-Paul Laurens, 31400 Toulouse, FR, <sup>4</sup>Space Cooperative Europe

# Outline

- Introduction
- Interoperable Data Spaces
- Connecting Data Spaces and AI
- Spatial Tokenization - DGGS
- Use Cases
- Implementations
- Outlook



# Introduction



# Obstacles to GeoAI

Big Geodata are:



- costly to store
- difficult to integrate
- siloed and messy

GeoAI is the foundation for distilling information from Big Geodata.  
But applying complex ML algorithms requires overcoming  
the major challenge of training data standardization and integration.



# Interoperable Data Spaces



# Data Spaces for AI

“Data spaces are **decentralised** infrastructures, where diverse **actors** can share and use data in a **secure**, reliable and **trustworthy** manner” - *EC*

“These include the potential of data spaces to realize efficiencies in data **exchange**, to promote data **sovereignty**, and to enable new business.” - *Gaia-X Hub Germany*

“A dataspace is defined as a set of “**participants**”, or data **sources**, and the **relations** between them” - *Wikipedia*

“**Interoperable** framework, based on common **governance** principles, **standards**, practices and enabling services, that **enables** trusted data transactions between **participants**.” - *Data Spaces Support Centre*

Highlights are mine



# Copernicus Data Space Ecosystem



CDSE COPERNICUS DATA SPACE ECOSYSTEM

DATA ANALYSIS SERVICES ECOSYSTEM Explore Data SUPPORT LOGIN

## Welcome to Copernicus Data Space Ecosystem

Welcome to the Copernicus Data Space Ecosystem, an open ecosystem that provides free instant access to a wide range of data and services from the Copernicus Sentinel missions and more on our planet's land, oceans and atmosphere.

The Copernicus Data Space Ecosystem not only ensures the continuity of the open and free access to Copernicus data but also extends the portfolio for data processing and data access possibilities. Delve into the data immediately via the Copernicus Browser or register to create an account and have an even better comprehensive exploration experience.



Get started

TRUE-COLOUR FALSE-COLOUR NDVI NDWI

SENTINEL-2 QUARTERLY MOSAIC

# Interoperable Europe

“Citizens and businesses across the EU expect high-quality public services from their administrations. Public service delivery rarely functions in isolation. Before the final delivery to the end-user, data must often be **exchanged** between different public authorities. This exchange happens primarily at the local level but also substantially at regional, national, and European levels - **requiring seamless interoperability.**”



# Simpl Program

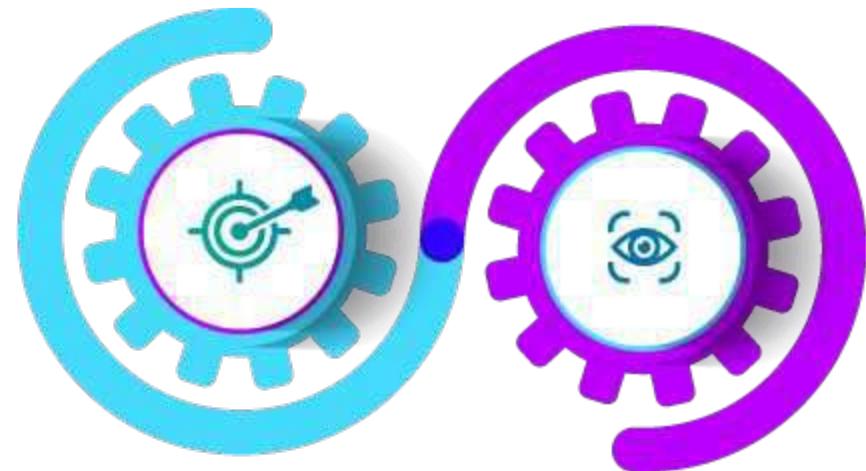
“Simpl is a middleware platform ... that supports **secure** data sharing .... It enables **interoperability** by promoting common data **standards** and policies, and by providing compatible **open source components** ....

Simpl’s mission is to unleash the power of **existing data** by enabling the creation of **data chains** across multiple industries while maintaining the ownership and control of data at its source.“



# Gaia-X

“It enables a **federated** and secure data infrastructure, whereby data are shared, with **users retaining control** over their data access and usage. It enables the creation of links between many cloud service providers in a wider, transparent and fair **ecosystem** to drive the European Data economy of tomorrow.”



# Connecting Data Spaces and AI



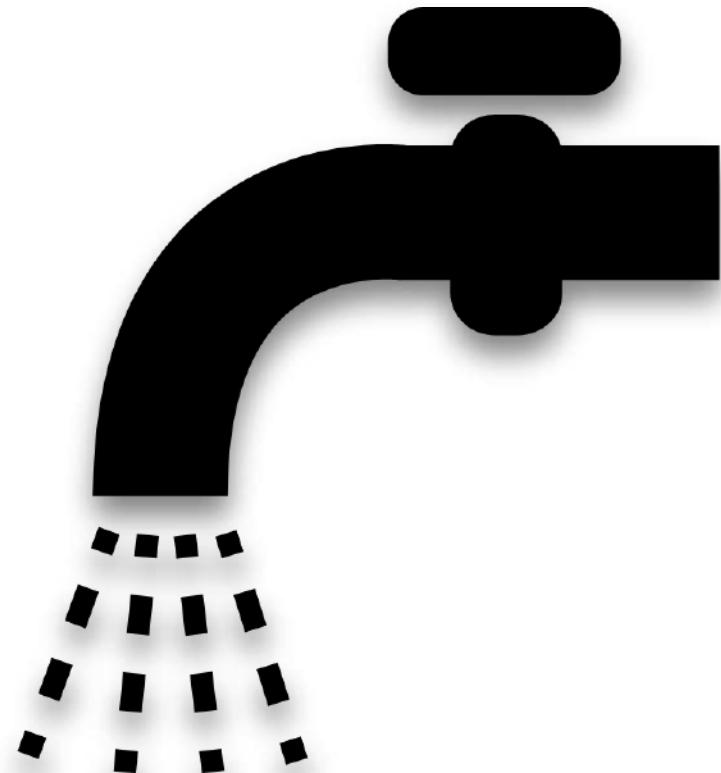
## Serving the (Geo)AI community

One Data Space is great.

Multiple Data Spaces are (currently) problematic for AI.

The GeoAI community needs a mechanism to make geographic data (not only imagery) machine readable, they call this mechanism **Spatial Tokenization**.

How to open the **data tap**?

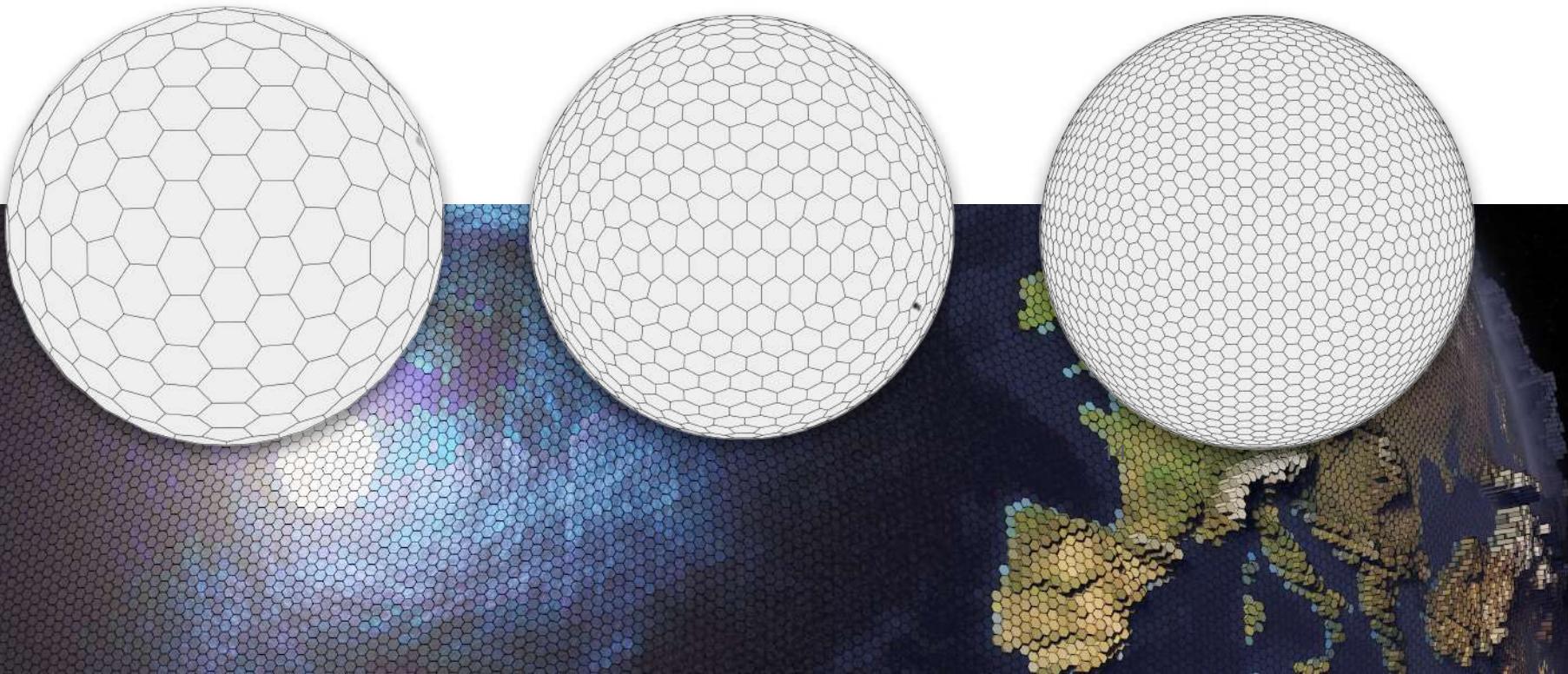


# Spatial Tokenization



# Spatial Tokenization

A Discrete Global Grid System with equal area size cells consistent across the entirety of the globe ideal for training GeoAI



# Spatial Tokenization

A Discrete Global Grid System with equal area size cells consistent across the entirety of the globe ideal for training GeoAI



# CEOS: Grid and Cells

## Grid

Network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way.

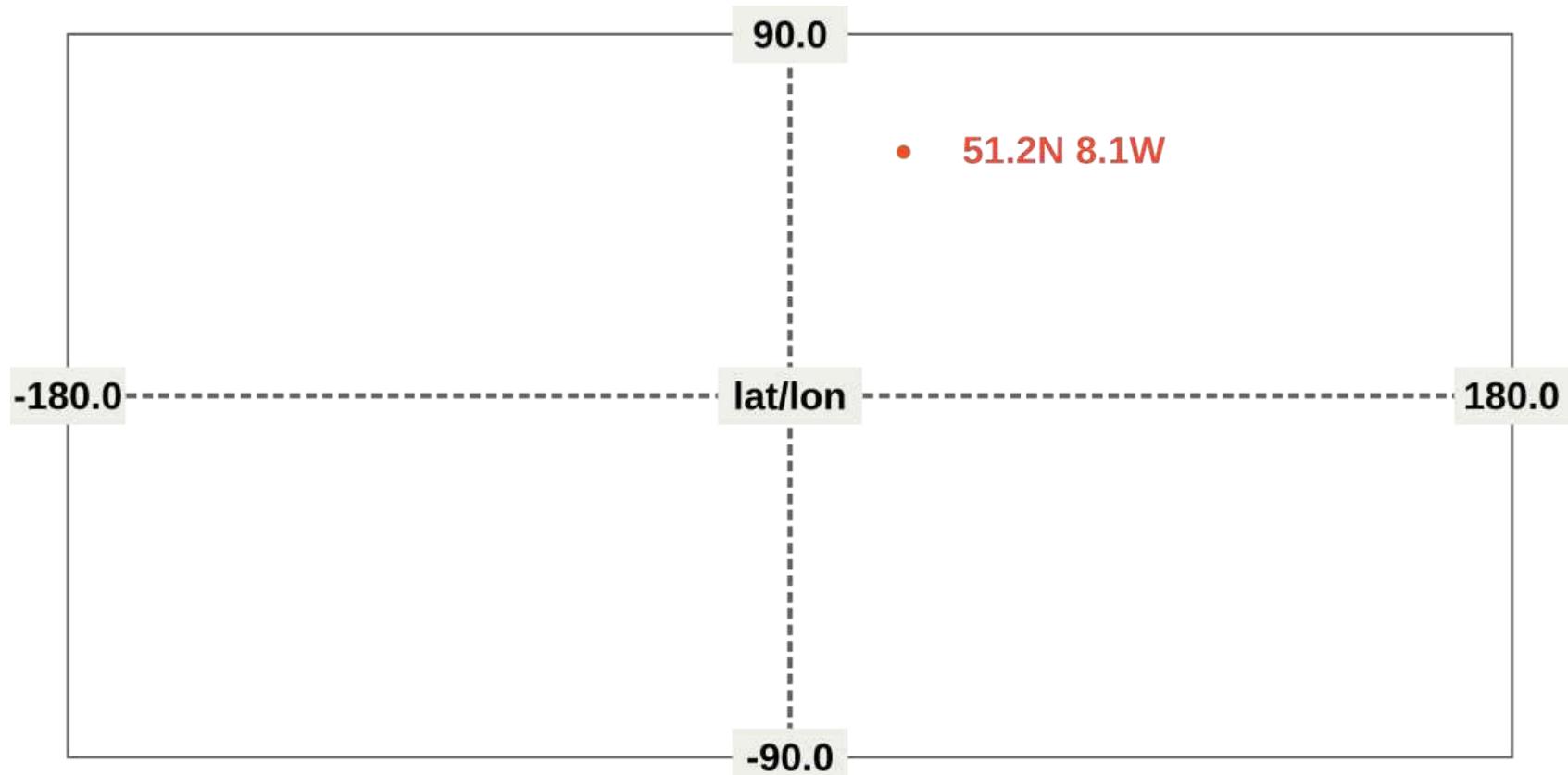
## Cell

Spatial, spatio-temporal zone or temporal unit of geometry with dimensionality greater than 0, associated with a zone.

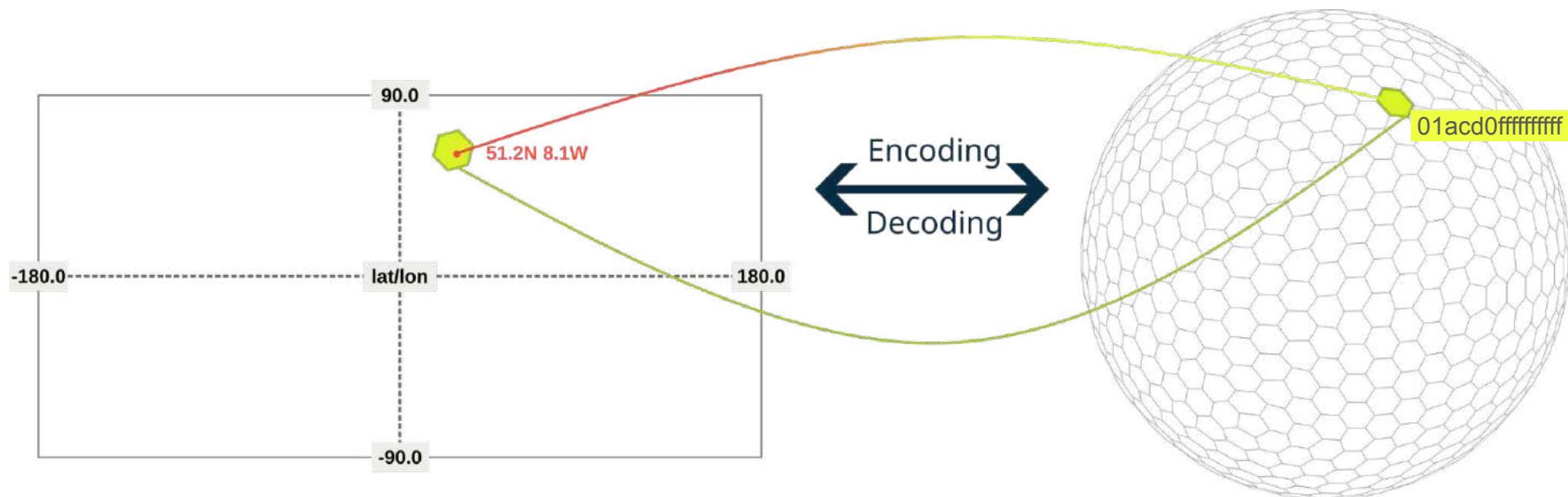
<https://ceos-org.github.io/eo-glossary/>



# Lat/Lon space: Analog or Digital/Discrete?



# Tokenization of Lat/Lon space into zone IDs

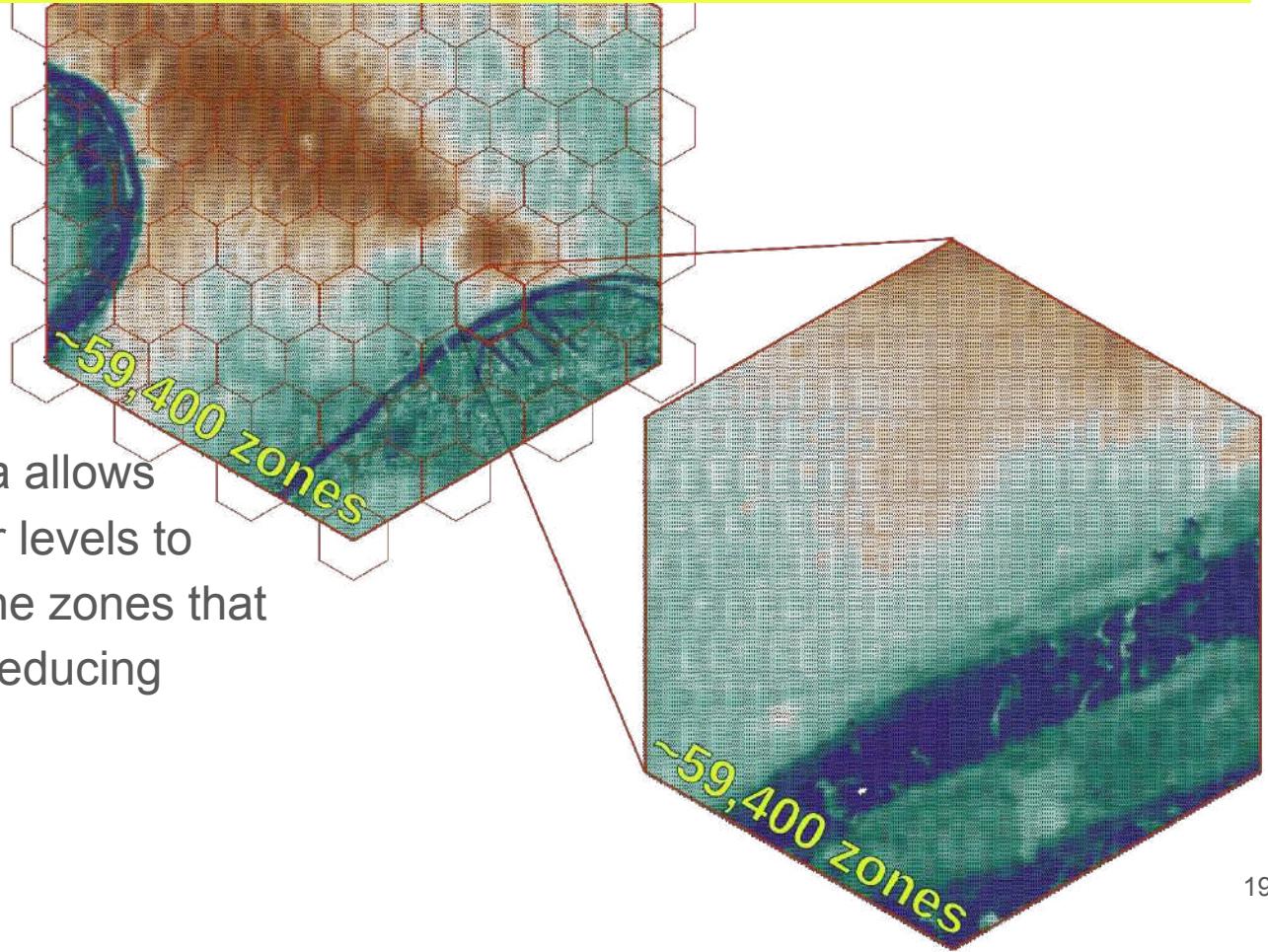


Compact, hexagonal  
zones of equal size <sub>18</sub>



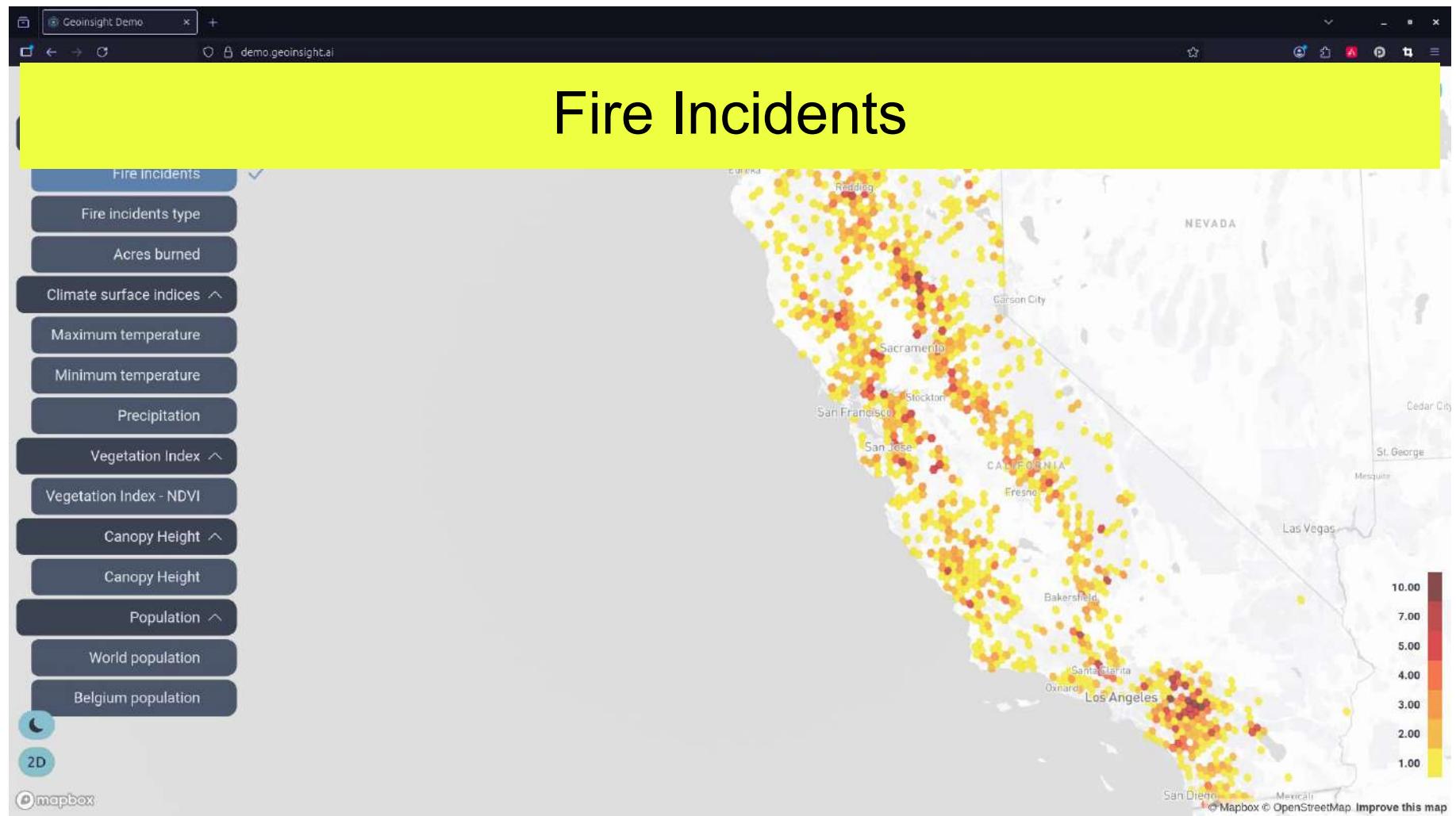
# Hierarchy of zones

- Inherent in the DGGS concept
- Partitioning of data into parent zones and sub-zones
- Nested
- Aggregation of data allows analysis on coarser levels to then 'zoom-in' on the zones that are significant = reducing resources.

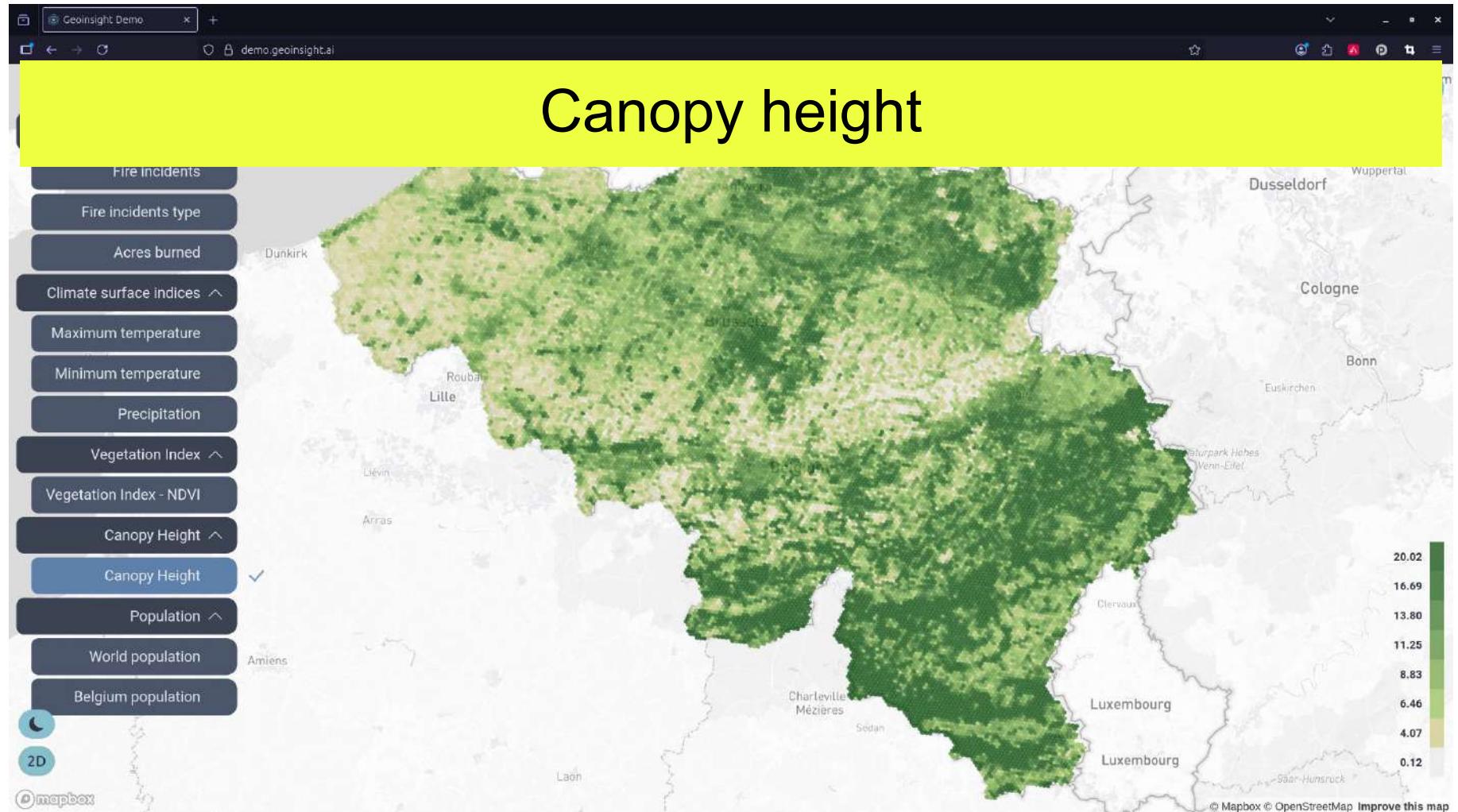


# Use Cases





GeoInsight



GeoInsight

# Global population

Fire Incidents

Fire incidents type

Acres burned

Climate surface indices ^

Maximum temperature

Minimum temperature

Precipitation

Vegetation Index ^

Vegetation Index - NDVI

Canopy Height ^

Canopy Height

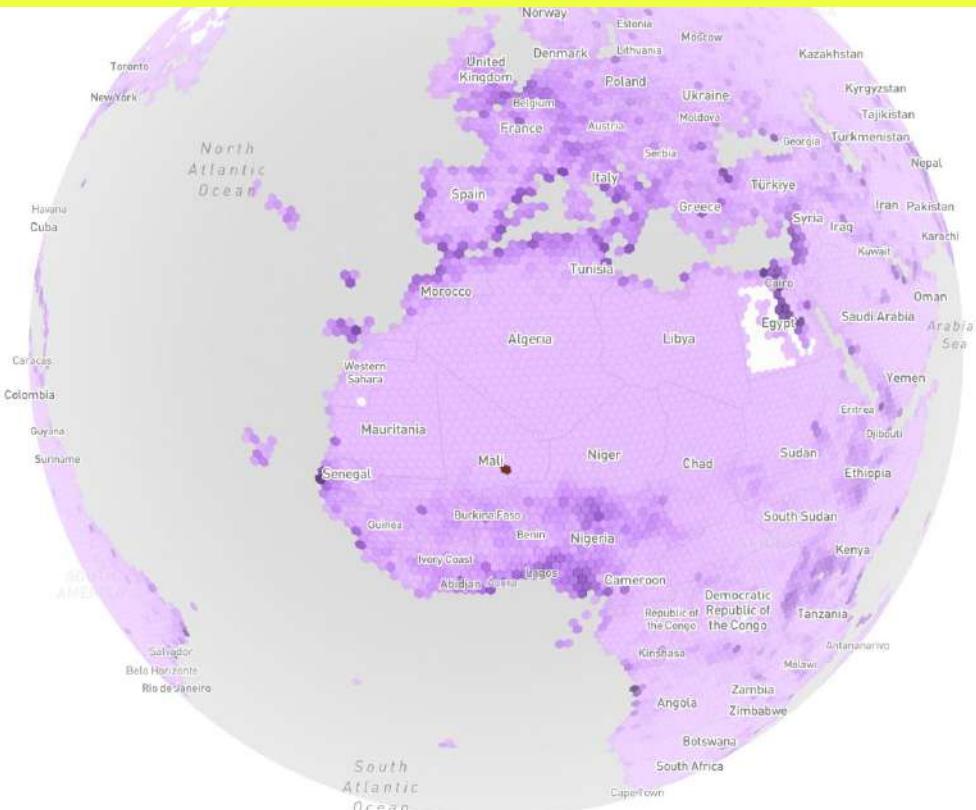
Population ^

World population

Belgium population



mapbox

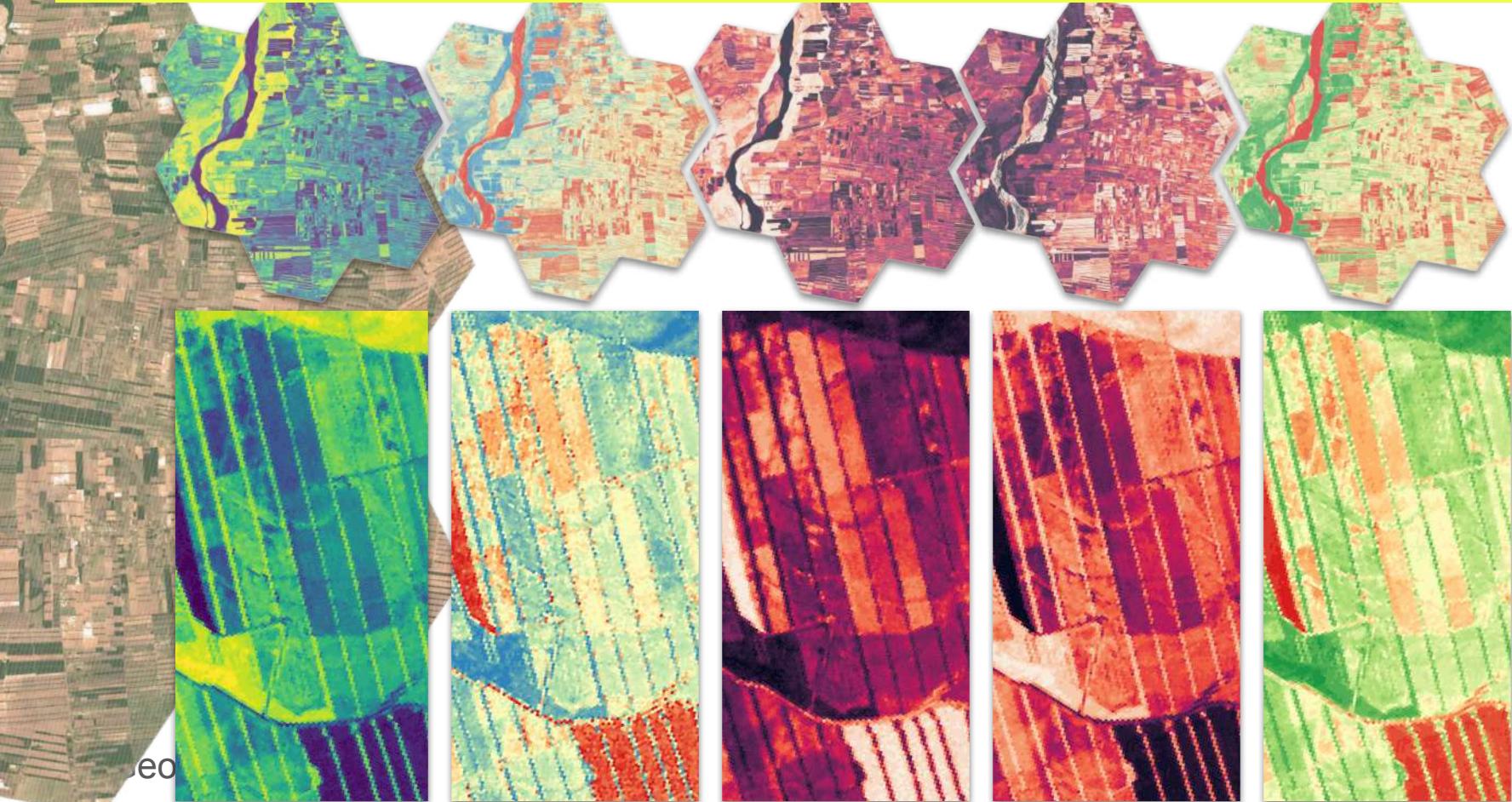


© Mapbox © OpenStreetMap Improve this map

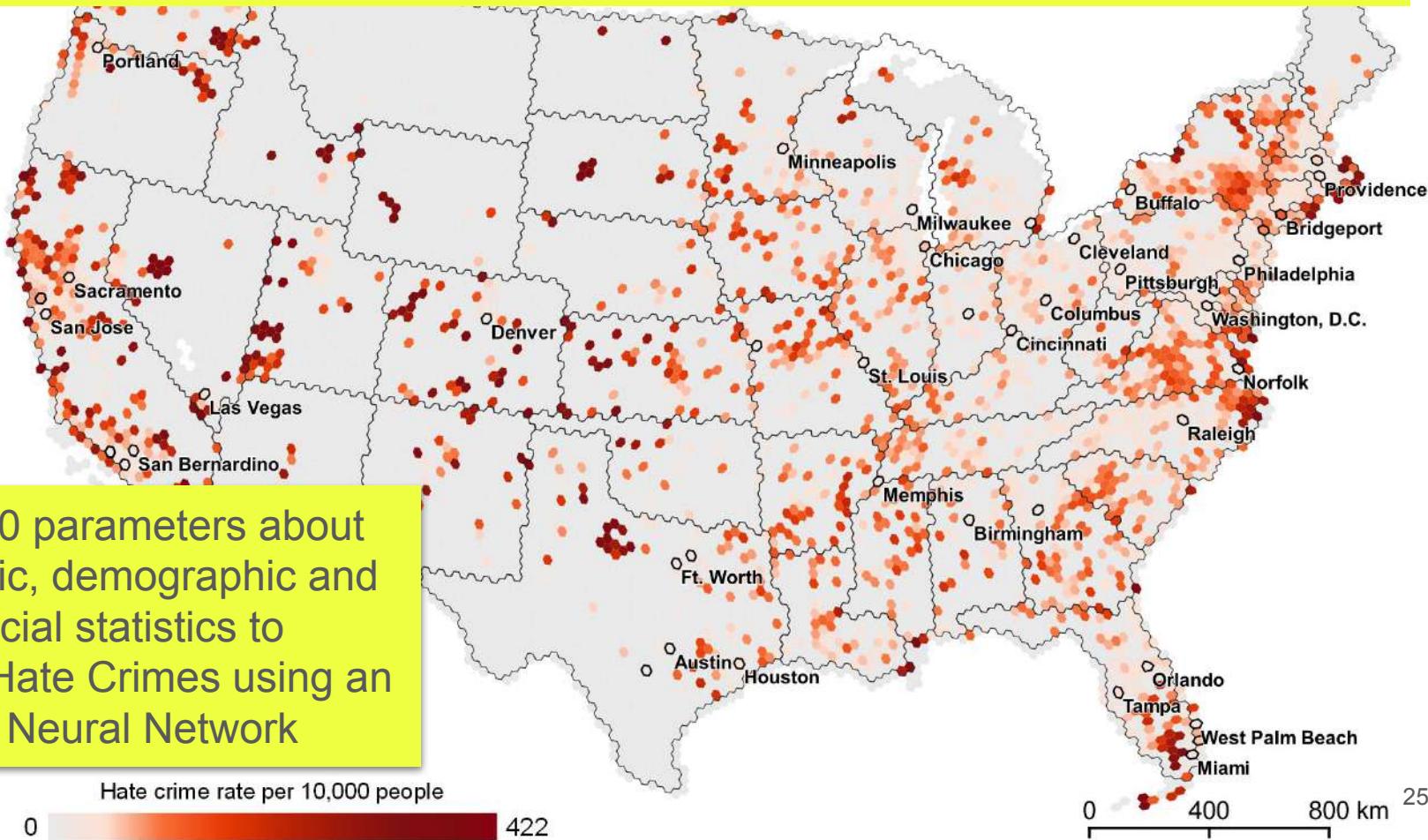


GeoInsight

# High resolution agricultural indices



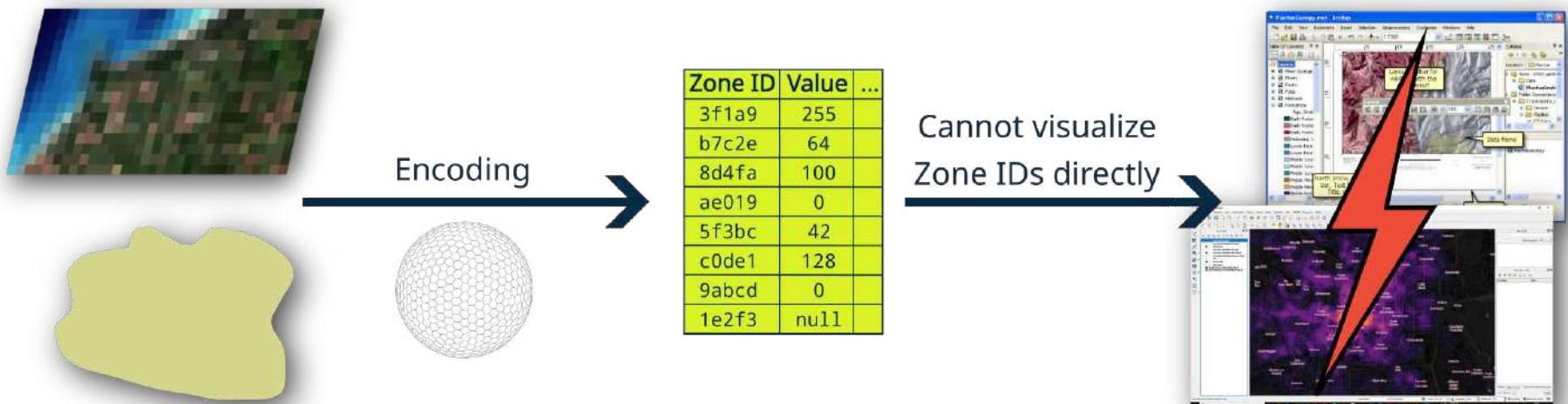
# Hate crimes in the USA



# Implementations



# DGGS aware and unaware clients



# GeoPlegma - an open-source DGGS library



# OGC: DGGS API and AI+DGGS Pilot

## CONTENTS

- I. ABSTRACT
- II. KEYWORDS
- III. PREFACE
- IV. SECURITY CONSIDERATIONS
- V. SUBMITTING ORGANIZATIONS
- VI. SUBMITTERS
- VII. ACKNOWLEDGEMENTS
- 1. SCOPE
- 2. CONFORMANCE
  - 2.1. Requirements classes defining resources
  - 2.2. Requirements classes for integration within the OGC API framework
  - 2.3. Requirements classes defining query parameters
  - 2.4. Requirements classes defining



OGC API - Discrete Global Grid Systems - Part 1: Conformance Requirements

Open Geospatial Consortium

Dr. Matthew Brian John Purss  
Jérôme Jacovella-St-Louis

Responses due by Jun 20, 2025

CALL FOR  
PARTICIPATION

Organizing  
the World in  
Grid Cells:  
The AI-DGGS  
Pilot



<https://dggs.io>



CELL ID	SO2	CO	NO	NO2	NOX
4.568740092774877	249.18369581341184	8.928823833748101	26.401432636687523	49.993857	
4.5811653514569	249.70894273534796	8.969992749392837	26.4950402337883	50.22143	
4.595961197454199	249.6267838439203	8.928747620970408	26.515087993563206	50.09633528615073	
4.547195842265073	250.63383434343974	10.340026837645325	26.147188039057884	50.5847550460639	
4.5702683696117165	249.05745115420637	8.935244765931735	26.416095894178346	49.163404126512294	

Choose a resolution

8

Choose a field

NOX

Add datasets (.geojson, .json)

Datasets

air\_quality\_hexagons

Developing a portal similar to [geojson.io](https://geojson.io) to deal with DGGS data, APIs and use cases

# Outlook



# Cloudferro

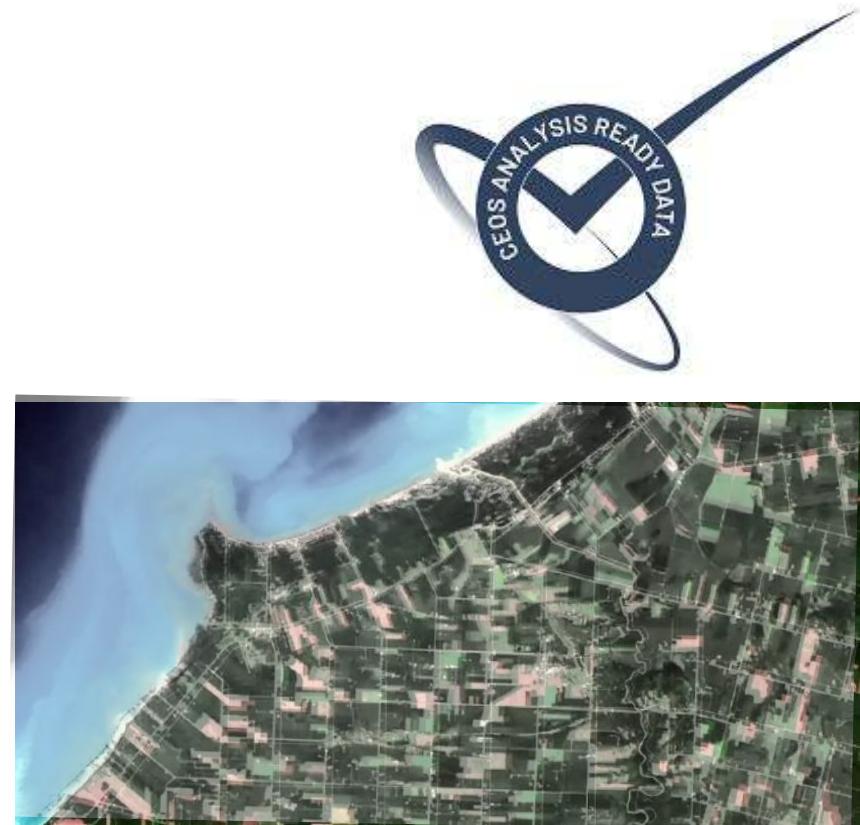
We are working on attaching GeoInsights **Spatial Tokenizer**, DGGS, to the entire Copernicus data archive in the Cloudferro infrastructure via the **OGC DGGS API**

```
euuser@webinar-vm:~$ ls /
bin          boot  eodata  home  lib usr-is-merged  lost+found  mnt  proc  run  sbin usr-is-merged  sys  usr
bin.var-is-merged  dev  etc    lib   lib64          media      opt  root  sbin  sry  tmp  var
euuser@webinar-vm:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
tmpfs          739M  692K  739M  1% /run
/dev/sda3       38G   1.5G   27G   6% /
tmpfs          3.7G   0   3.7G   0% /dev/shm
tmpfs          5.0M   0   5.0M   0% /run/lock
/dev/sda1       549M  184K  549M  1% /boot/efi
s3fs           4.0G   0   4.0G   0% /eodata
tmpfs          739M  8.0K  739M  1% /run/user/1001
euuser@webinar-vm:~$ ls /eodata/
CJIS  CEMS  CLMS_archive  Envisat-ASAR  Jason-3  Landsat-7  Landsat-8-ESA  SMOS  Sentinel-1-RTC  Sentinel-3  Sentinel-6  Terra
CAMS  CLMS  Envisat      Global-Mosaics  Landsat-5  Landsat-8  Landsat-9  Sentinel-1  Sentinel-2  Sentinel-5P  Sentinel-NPP  Suomi-NPP  auxdata
euuser@webinar-vm:~$
```

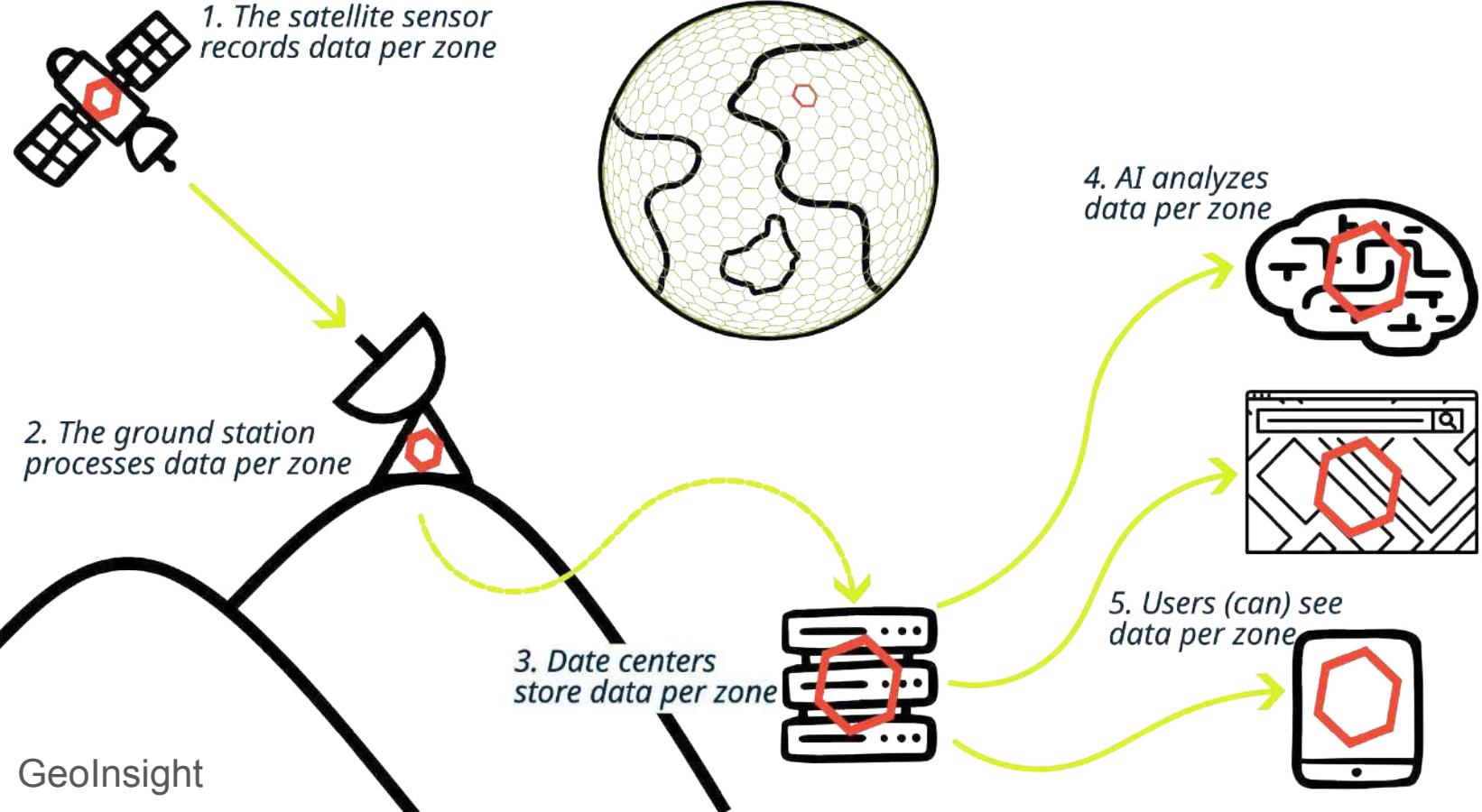


# CEOS Analysis Ready Data

- Each and every satellite image that is being taken has basically its own local grid
- The idea presented by CEOS during LPS was to take a global grid as the mechanism to geometrically align the grids to a common DGGS reference.



# Spatial Tokens everywhere





# Thank you!

[info@geoinsight.ai](mailto:info@geoinsight.ai)

