An Interoperable Data Economy to Enable GeoAI via Spatial Tokenizers

Michael Jendryke[1,2], João Manuel[1,2], Ludovic Augé[3], Emmanuel Mondon[4], Gino Caspari[1,2]*

[1]GeoInsight, Kaiserstr. 25, 59505 Bad Sassendorf, DE; [2]GeoInsight, Forstweg 65A, 3012 Bern, CH;
[3]Dataionics, 1 rue Jean-Paul Laurens, 31400 Toulouse, FR, [4]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 destilling 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*

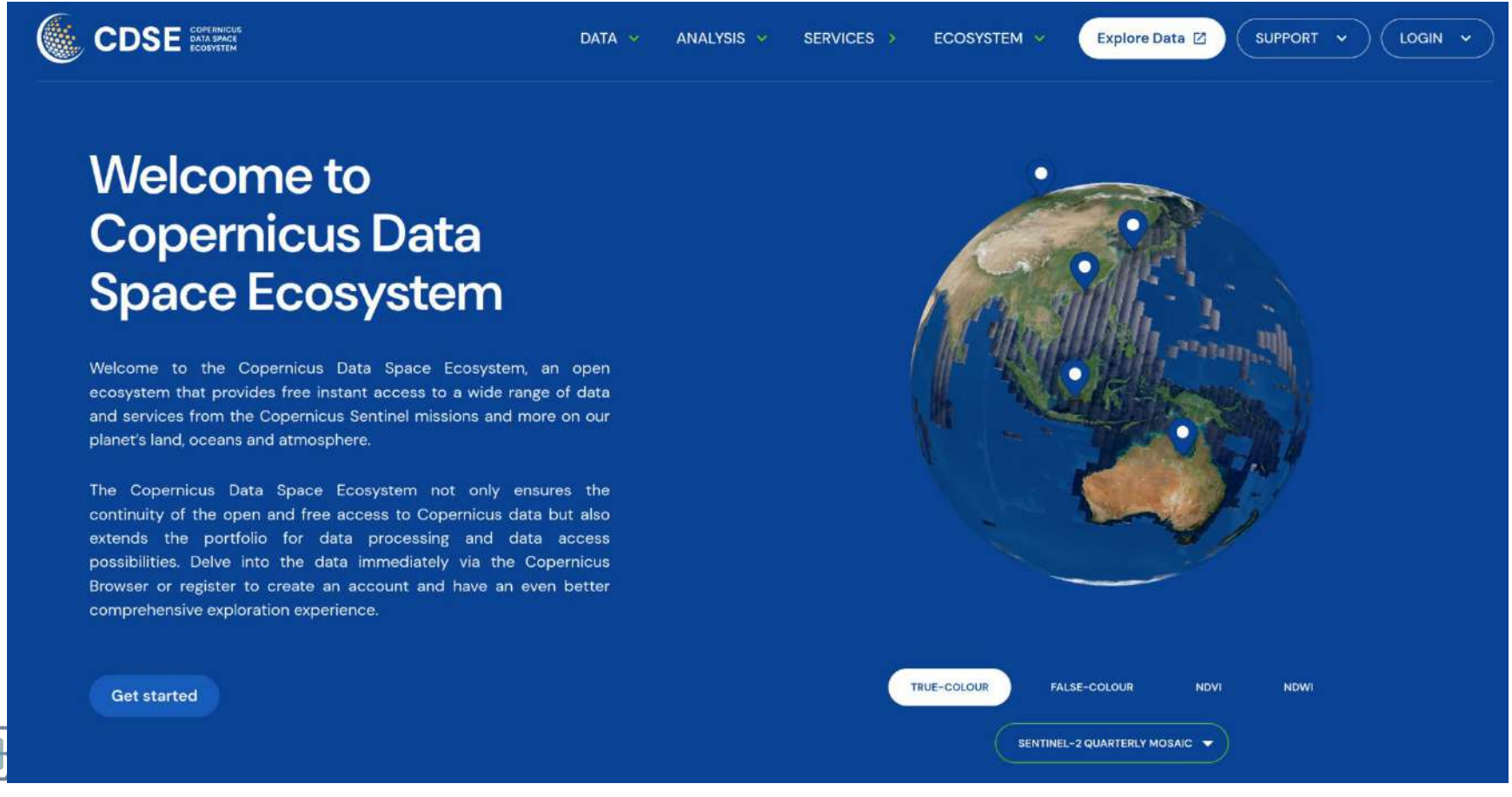
“These include the potential of data spaces to realize efficiencies in data **exchange**, to promote data **sovereignty**, and to enable new business.” - *Gaia-X Hub Germany*

“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



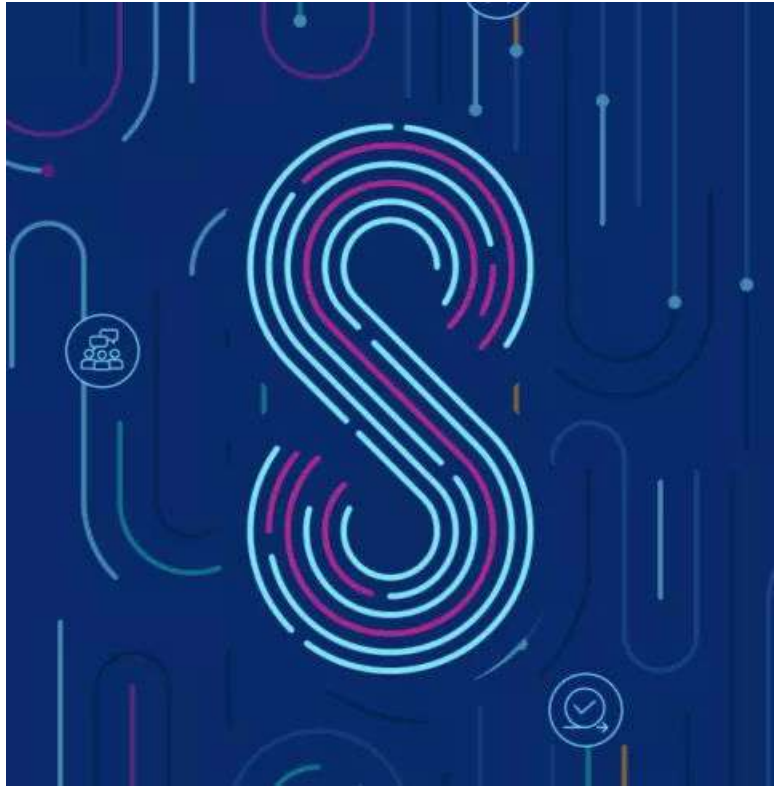
Interoperable Europe

“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

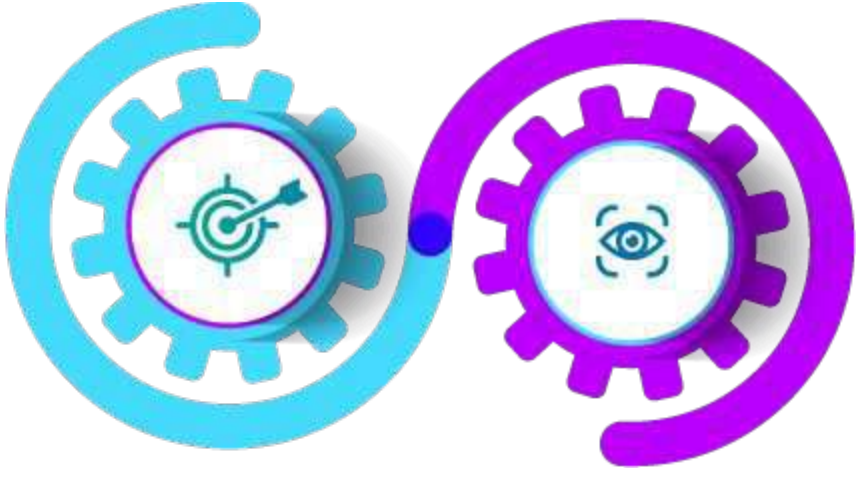
“Simpl is a middleware platform … that supports **secure** data sharing …. It enables **interoperability** by promoting common data **standards** and policies, and by providing compatible **open source components** ….

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

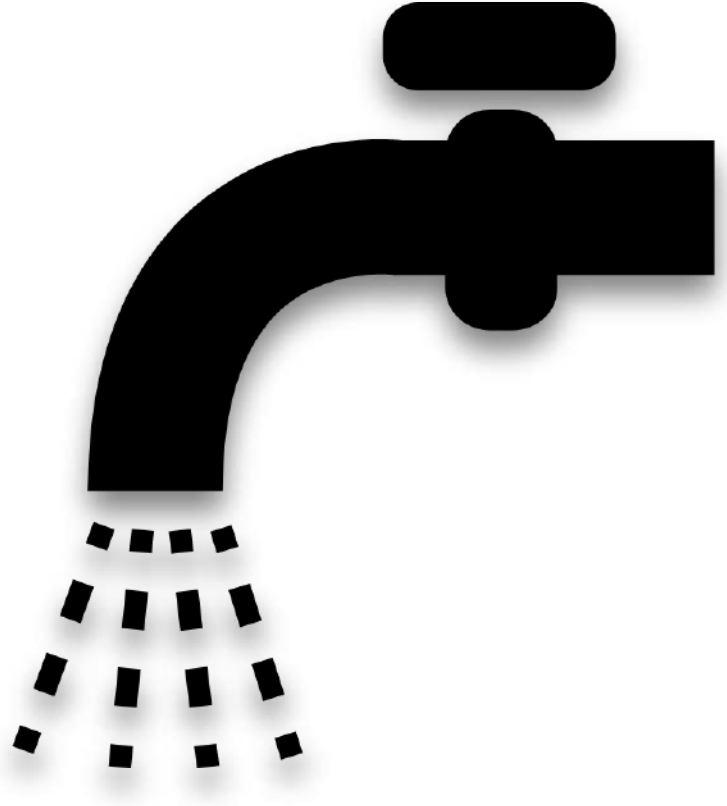
Serving the (Geo)AI community

One Data Space is great.

Multiple Data Space**s** 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

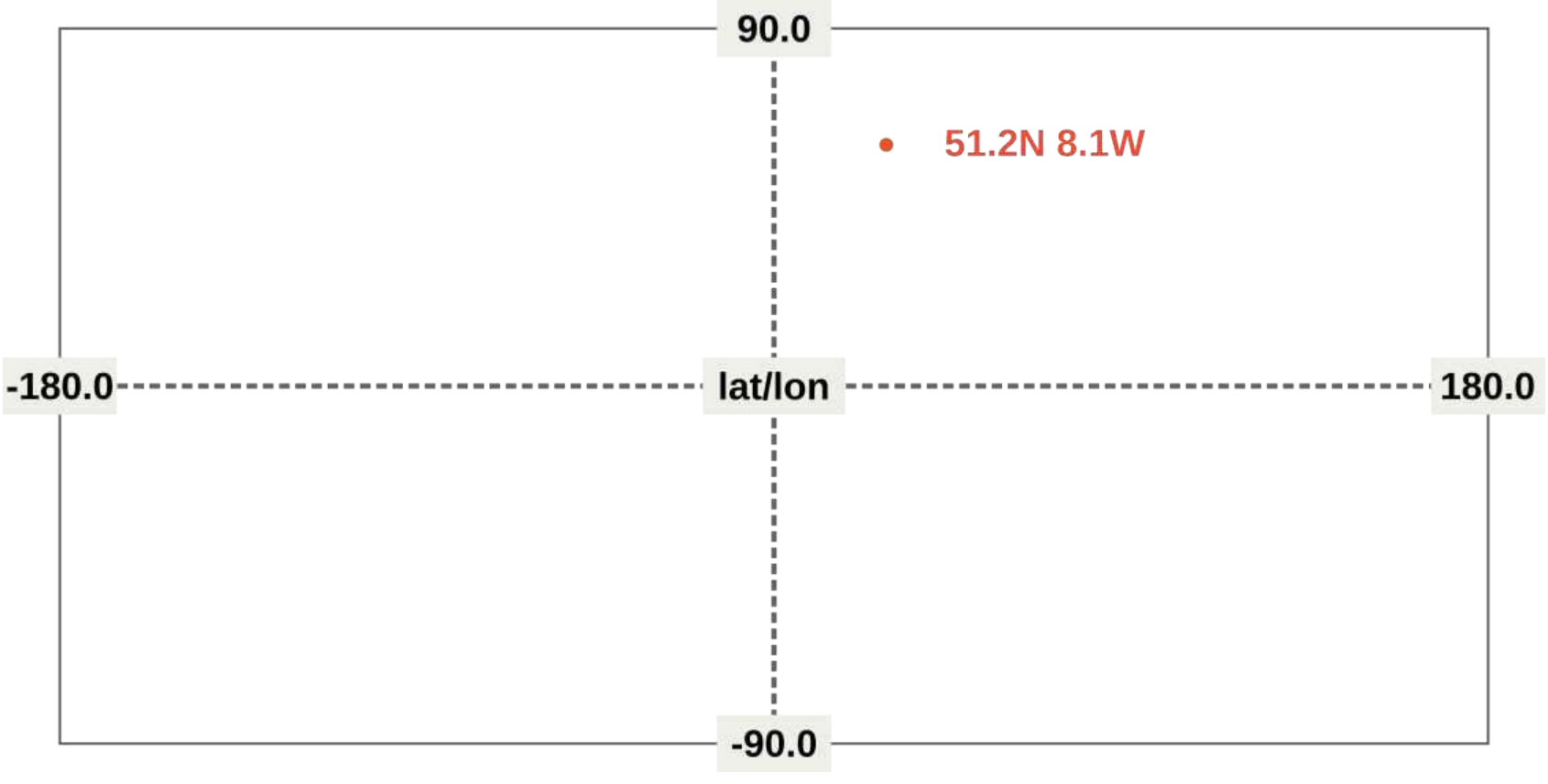
Network composed of two or more sets of curves in which the members of each set intersect the members of the other sets in an algorithmic way.

Cell

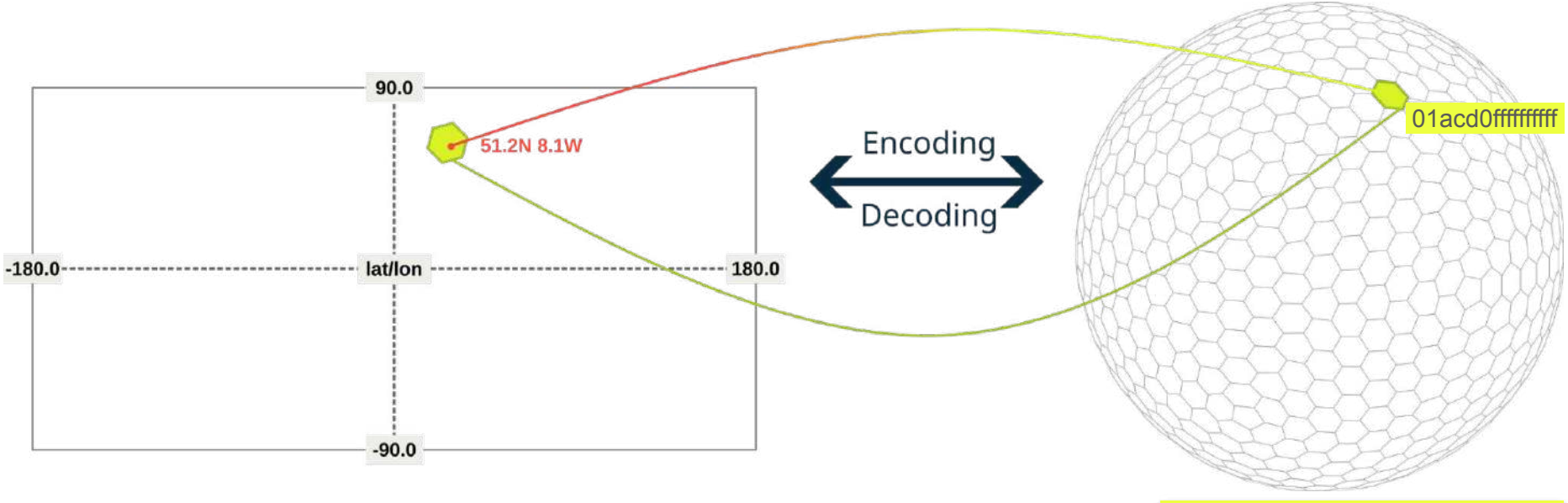
Spatial, spatio-temporal zone or temporal unit of geometry with dimensionality greater than 0, associated with a zone.

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

Lat/Lon space: Analog or Digital/Discrete?



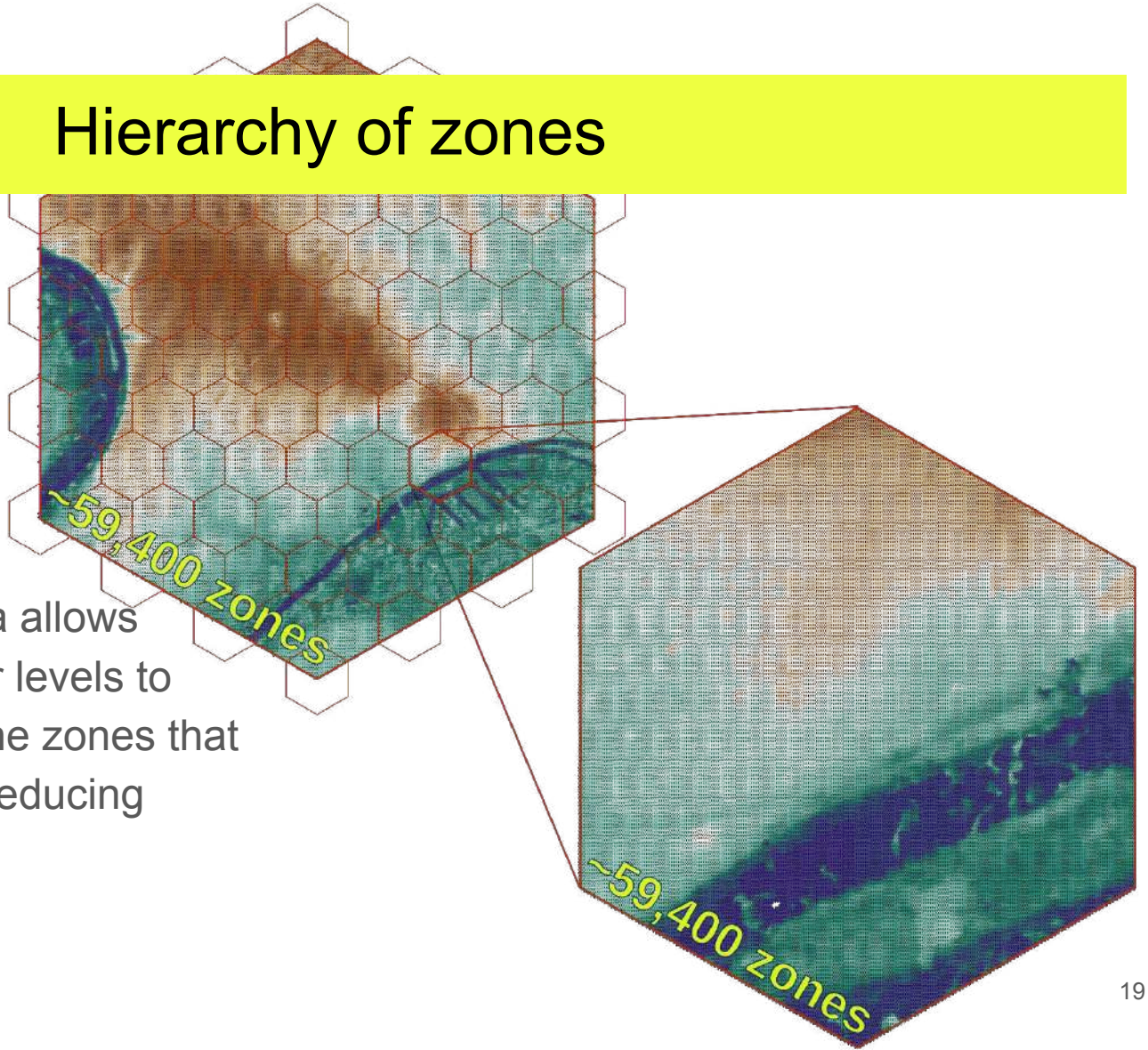
Tokenization of Lat/Lon space into zone IDs



Compact, hexagonal zones of equal size

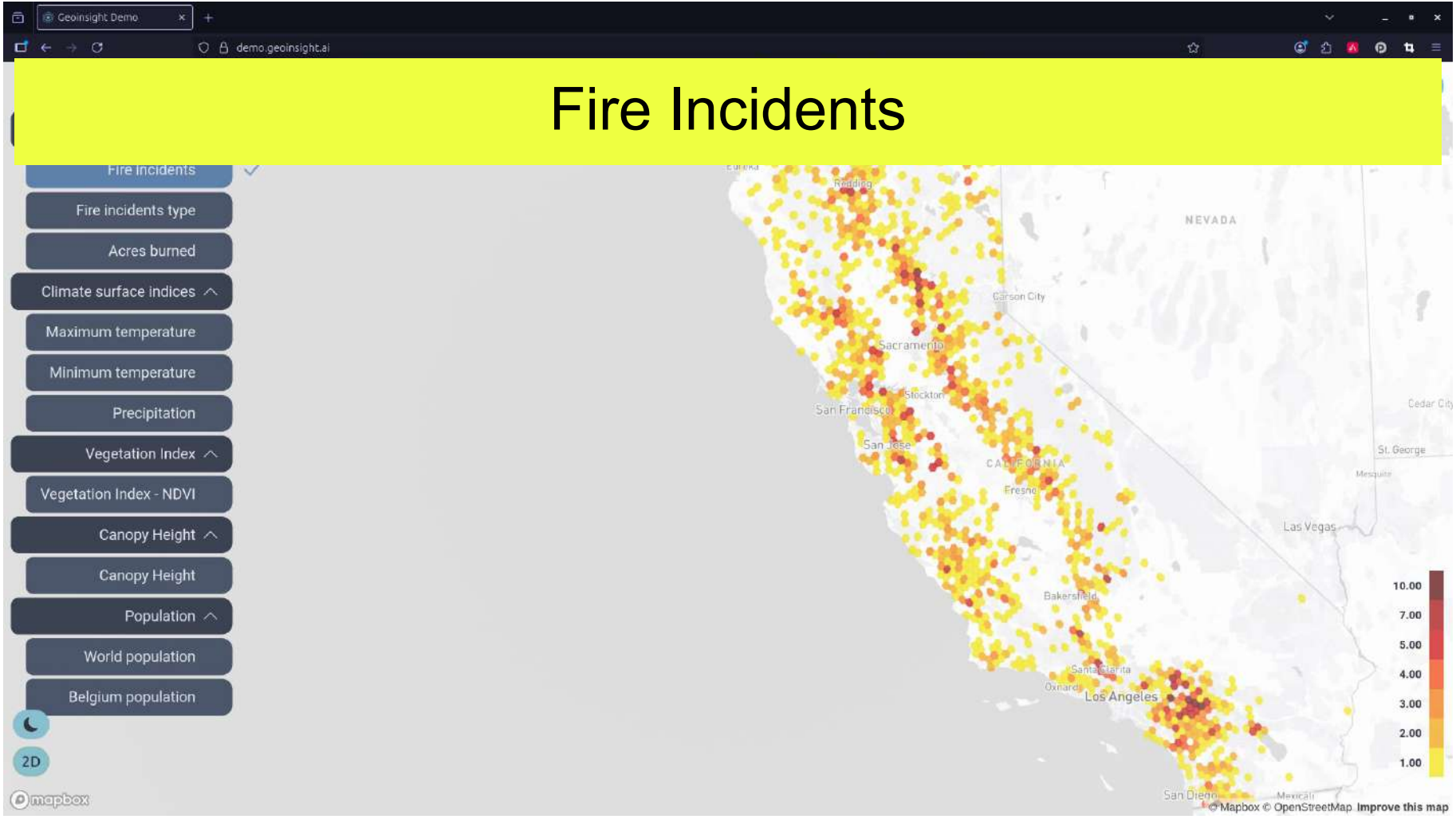
Hierarchy of zones

- 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 signification = reducing ressources.



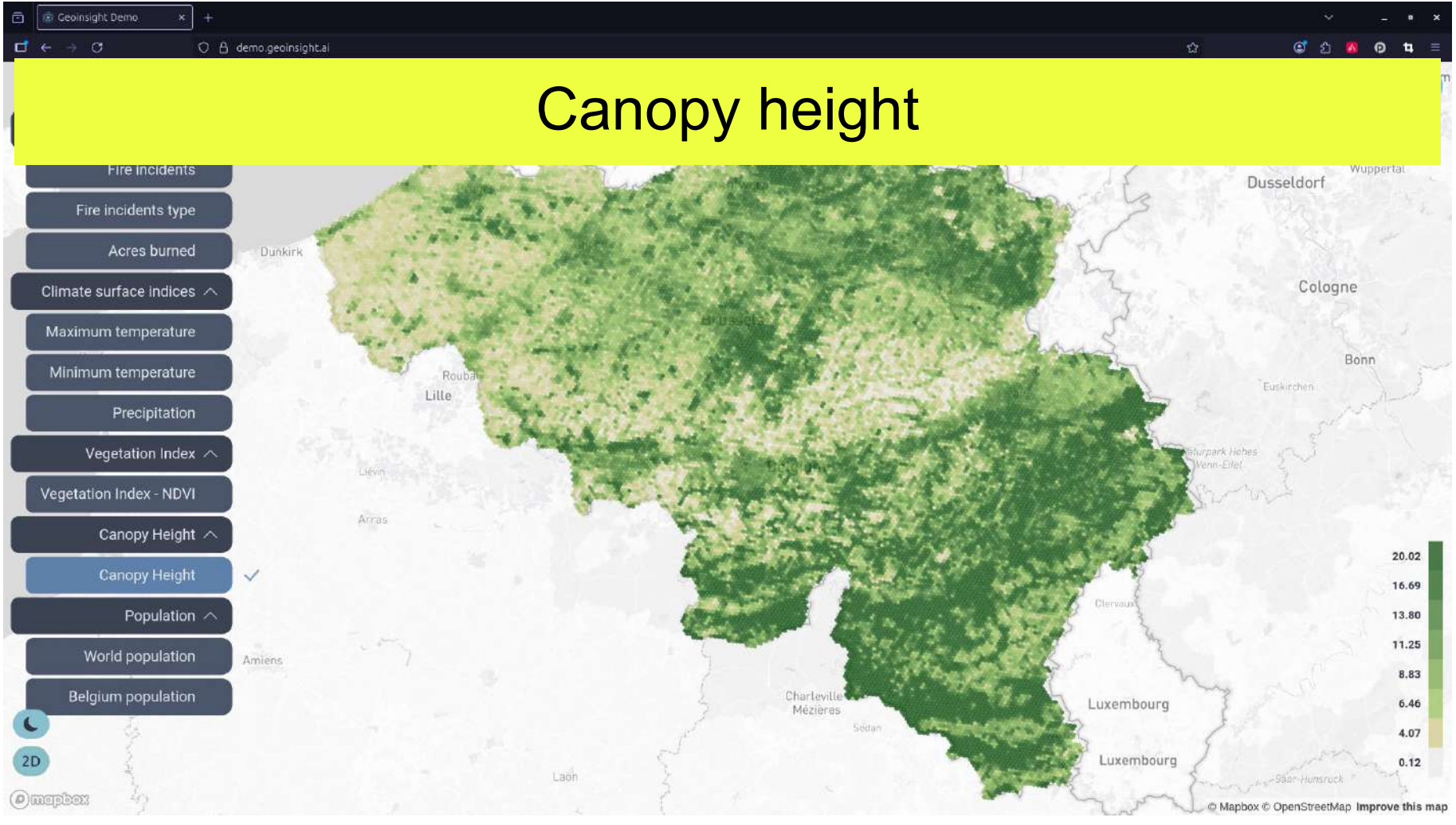
Use Cases

Fire Incidents



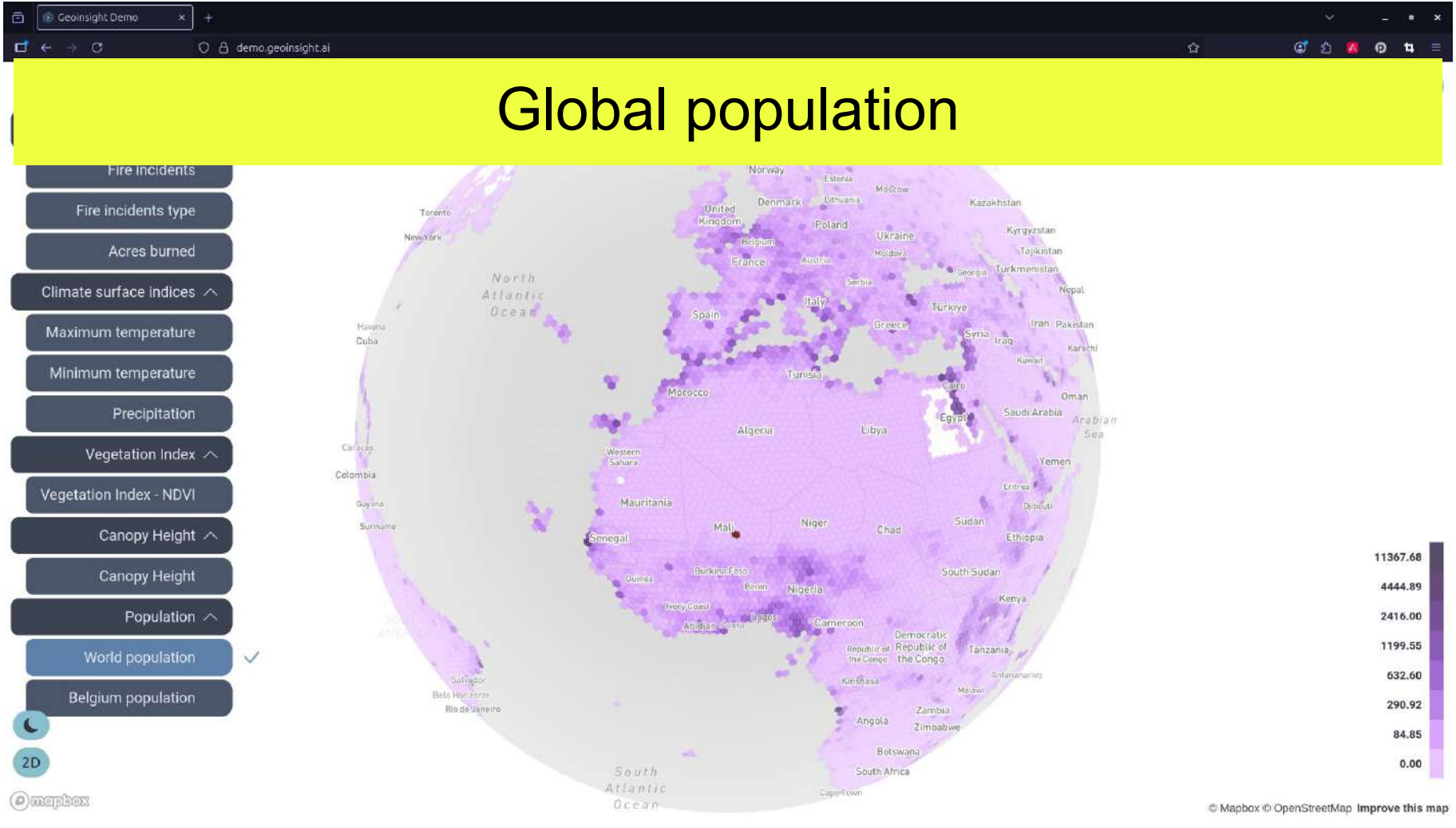
GeoInsight

Canopy height



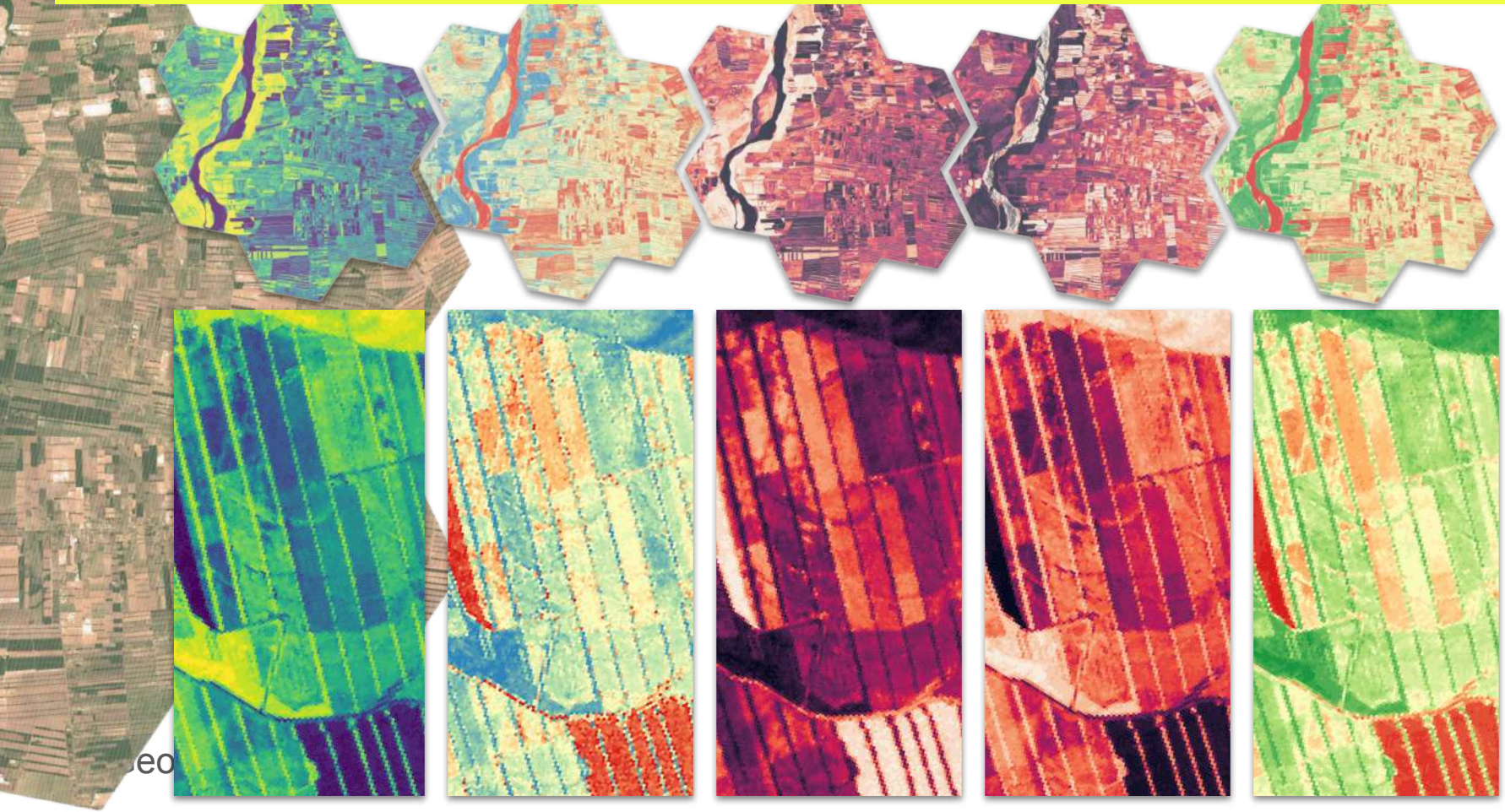
GeoInsight

Global population

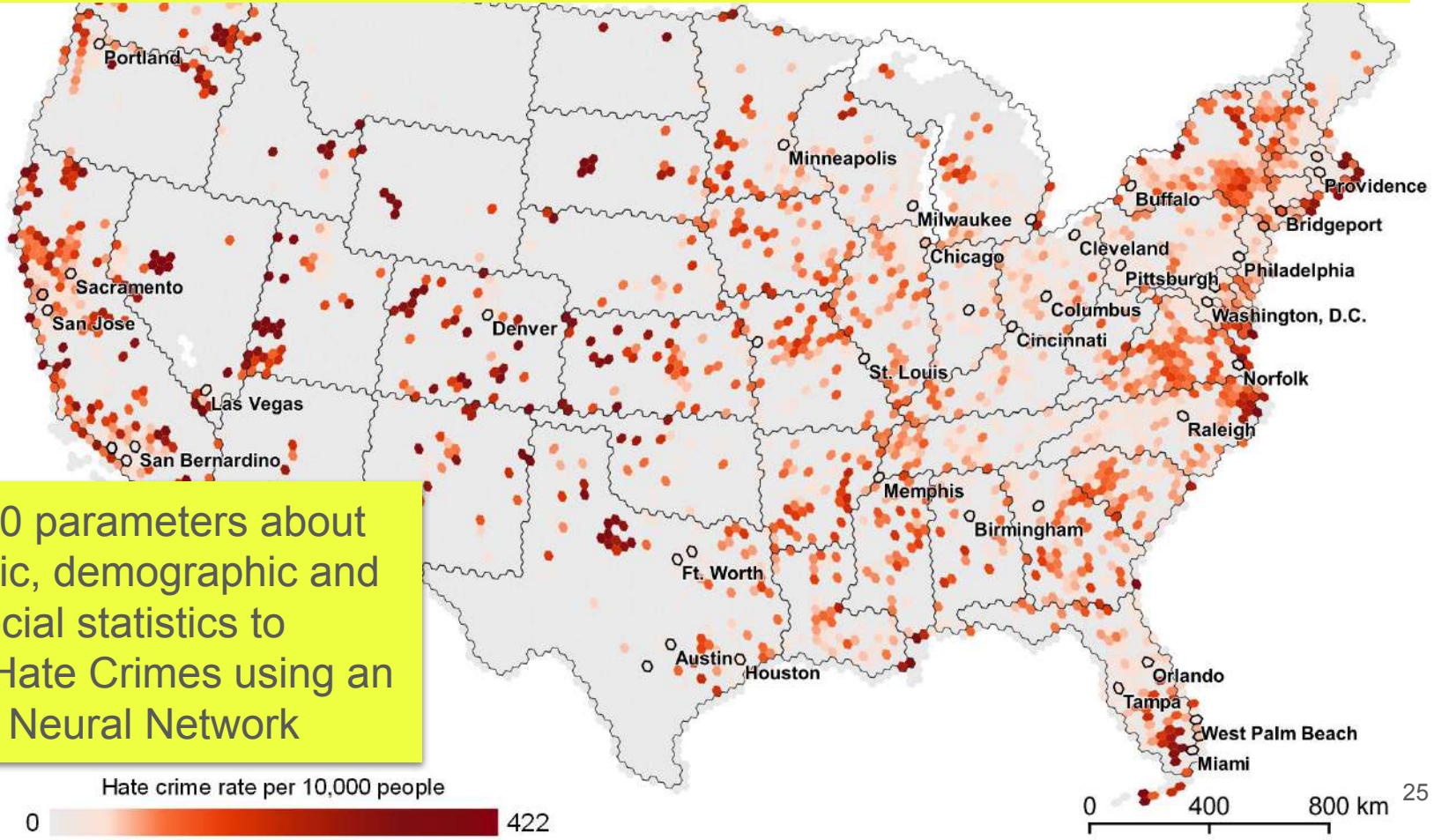


GeoInsight

High resolution agricultural indicies



Hate crimes in the USA



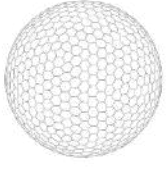
Over 400 parameters about economic, demographic and other social statistics to predict Hate Crimes using an Artificial Neural Network

Geo

Implementations

DGGS aware and unaware clients

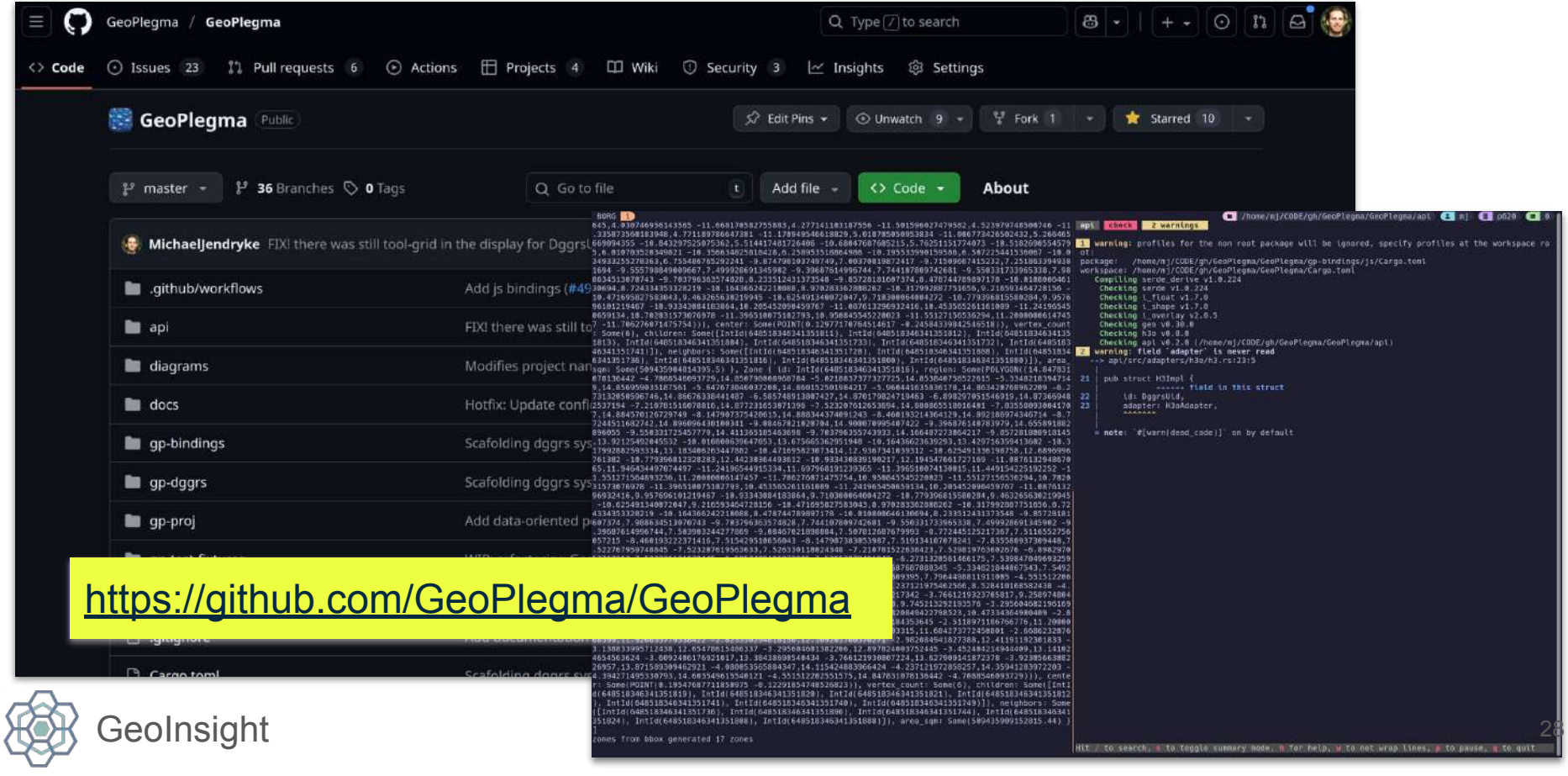
Encoding



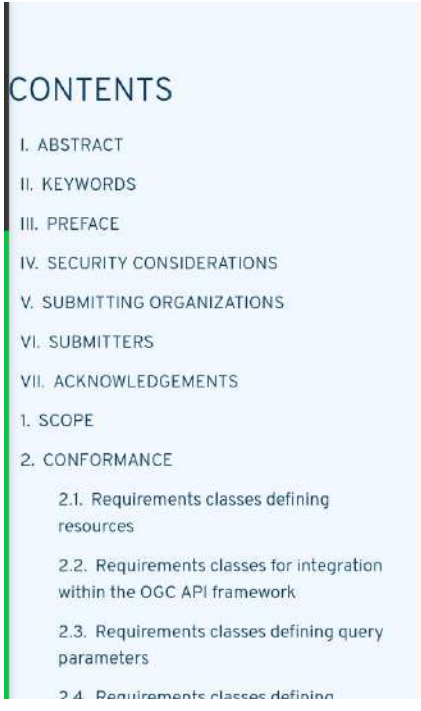
Zone ID	Value	...
3f1a9	255	
b7c2e	64	
8d4fa	100	
ae019	0	
5f3bc	42	
c0de1	128	
9abcd	0	
1e2f3	null	

Cannot visualize
Zone IDs directly

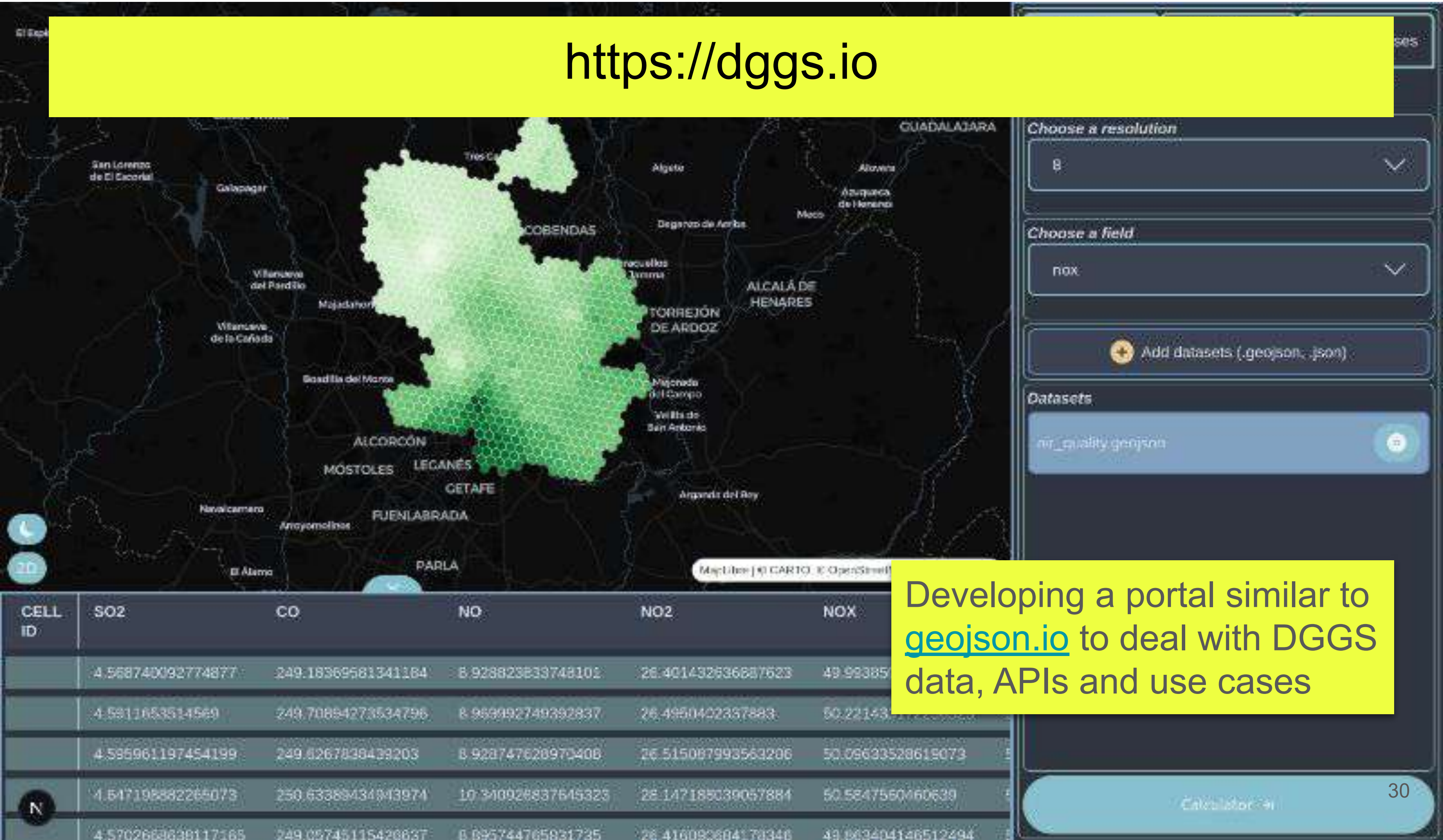
GeoPlegma - an open-source DGGS library



OGC: DGGS API and AI+DGGS Pilot



https://dggs.io



Developing a portal similar to geojson.io to deal with DGGS data, APIs and use cases

Outlook

Cloudferro

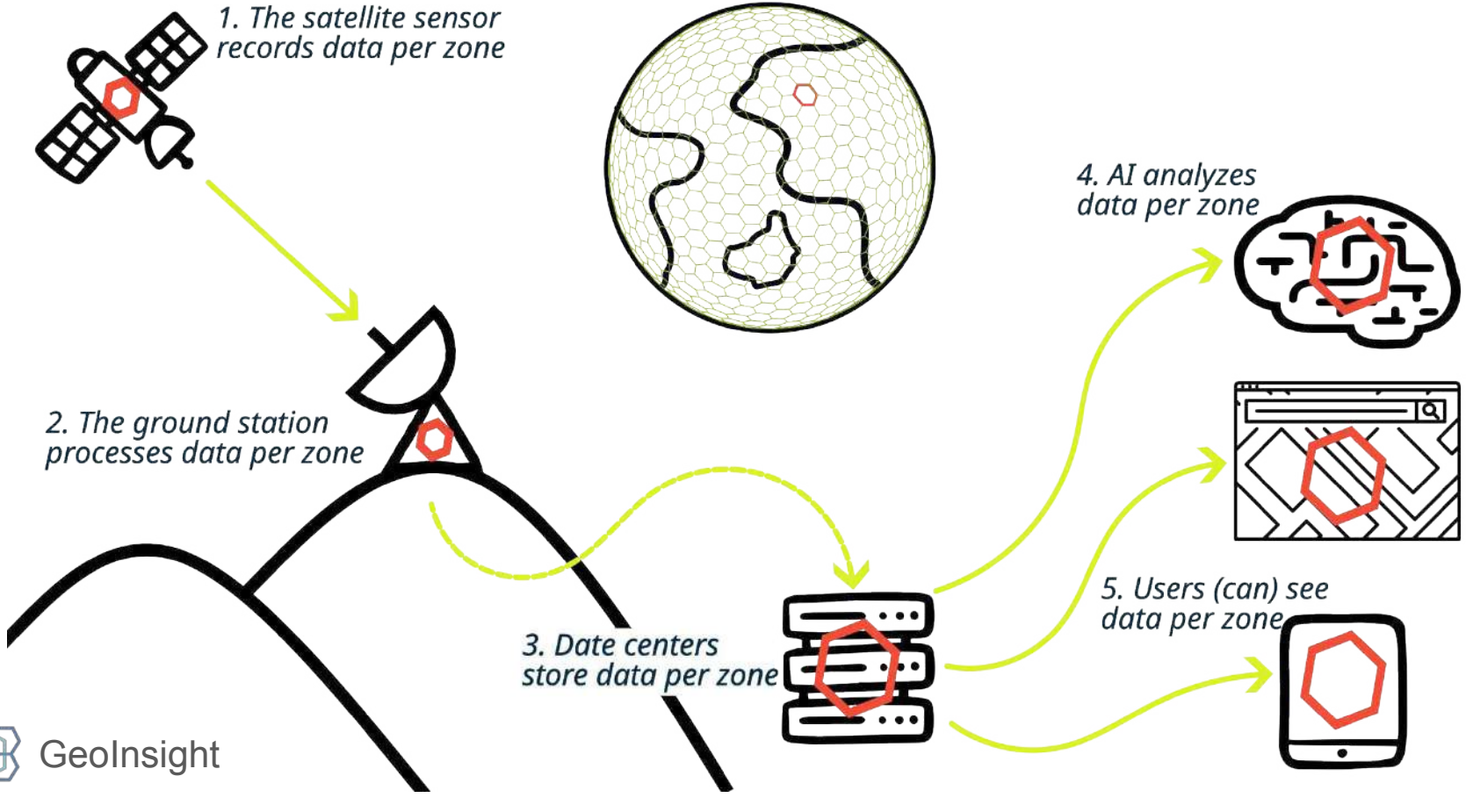
We are working on attaching GeoInsights **Spatial Tokenizer**, DGGS, to the entire Copernicus data archive in the Cloudferro infrastructure via the **OGC DGGS API**

```
eouser@webinar-vm:~$ ls /
bin                boot  eodata  home  lib.usr-is-merged  lost+found  mnt  proc  run   sbin.usr-is-merged  sys  usr
bin.usr-is-merged  dev   etc     lib   lib64              media       opt  root  sbin  srv                 tmp  var
eouser@webinar-vm:~$ df -h
Filesystem      Size  Used Avail Use% Mounted on
tmpfs           739M  692K  739M   1% /run
/dev/sda3        30G  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
eouser@webinar-vm:~$ ls /eodata/
C3S   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  Suomi-NPP   auxdata
eouser@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