



An Interoperable Data Economy to Enable GeoAI via Spatial Tokenizers



• esa

BUSINESS
INCUBATION
CENTRE

Michael Jendryke^{1,2}, João Manuel^{1,2}, Ludovic Augé³, Emmanuel Mondon⁴, Gino Caspari^{1,2}*

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

³Dataionics, 1 rue Jean-Paul Laurens, 31400 Toulouse, FR, ⁴Space 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

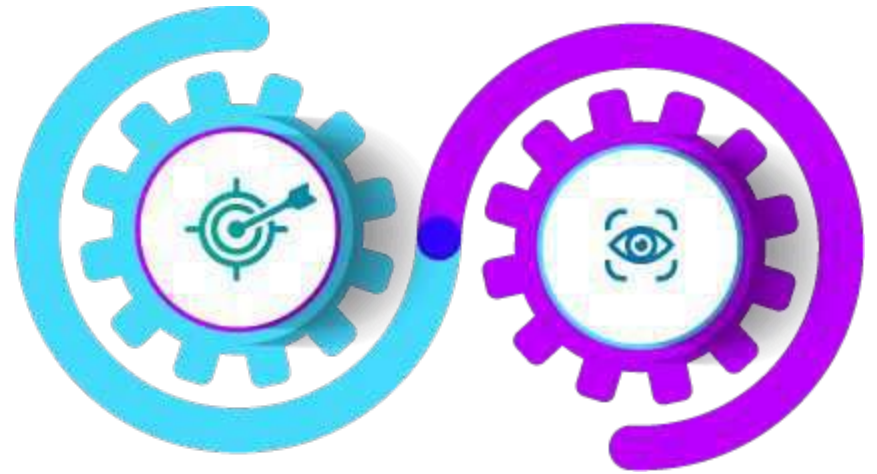
“Simpl is a middleware platform ... that supports **secure** data sharing It enables **interoperability** by promoting common data **standards** and policies, and by providing compatible **open source components**

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



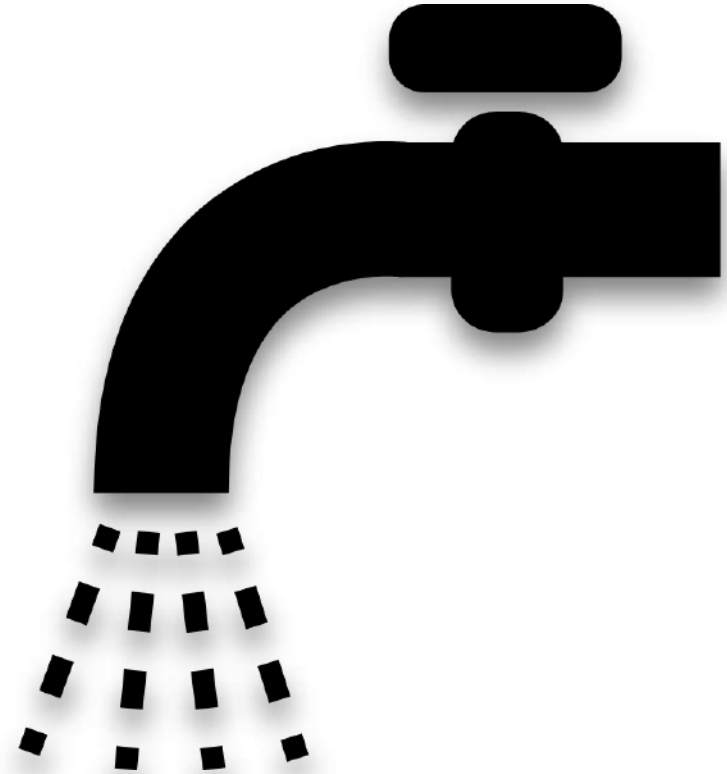
Serving the (Geo)AI community

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**?

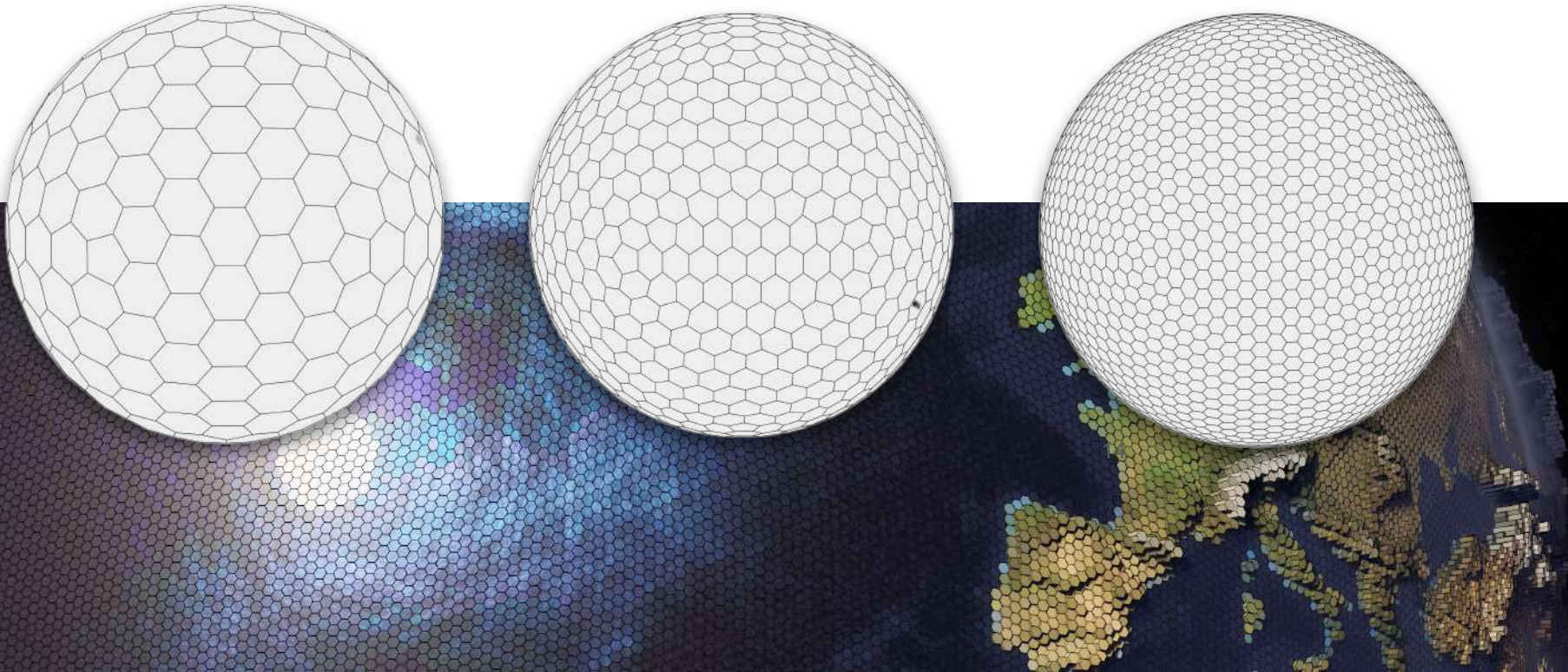


Spatial Tokenization



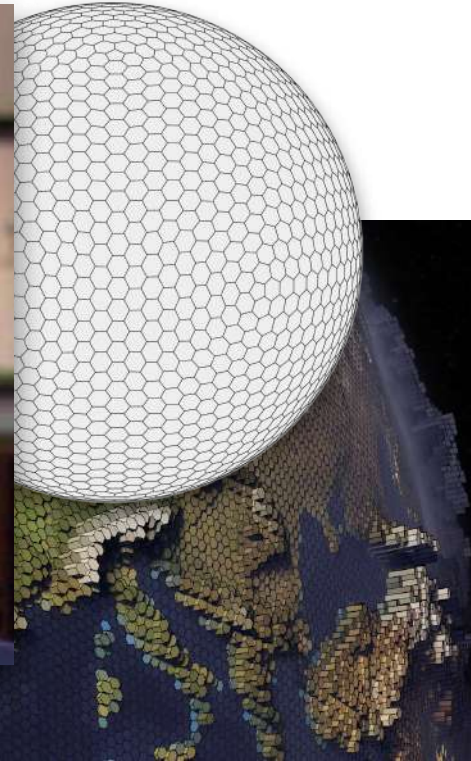
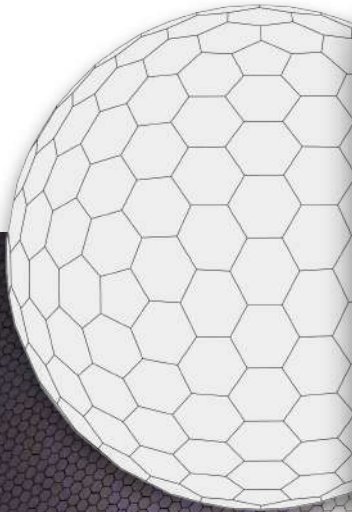
Spatial Tokenization

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.

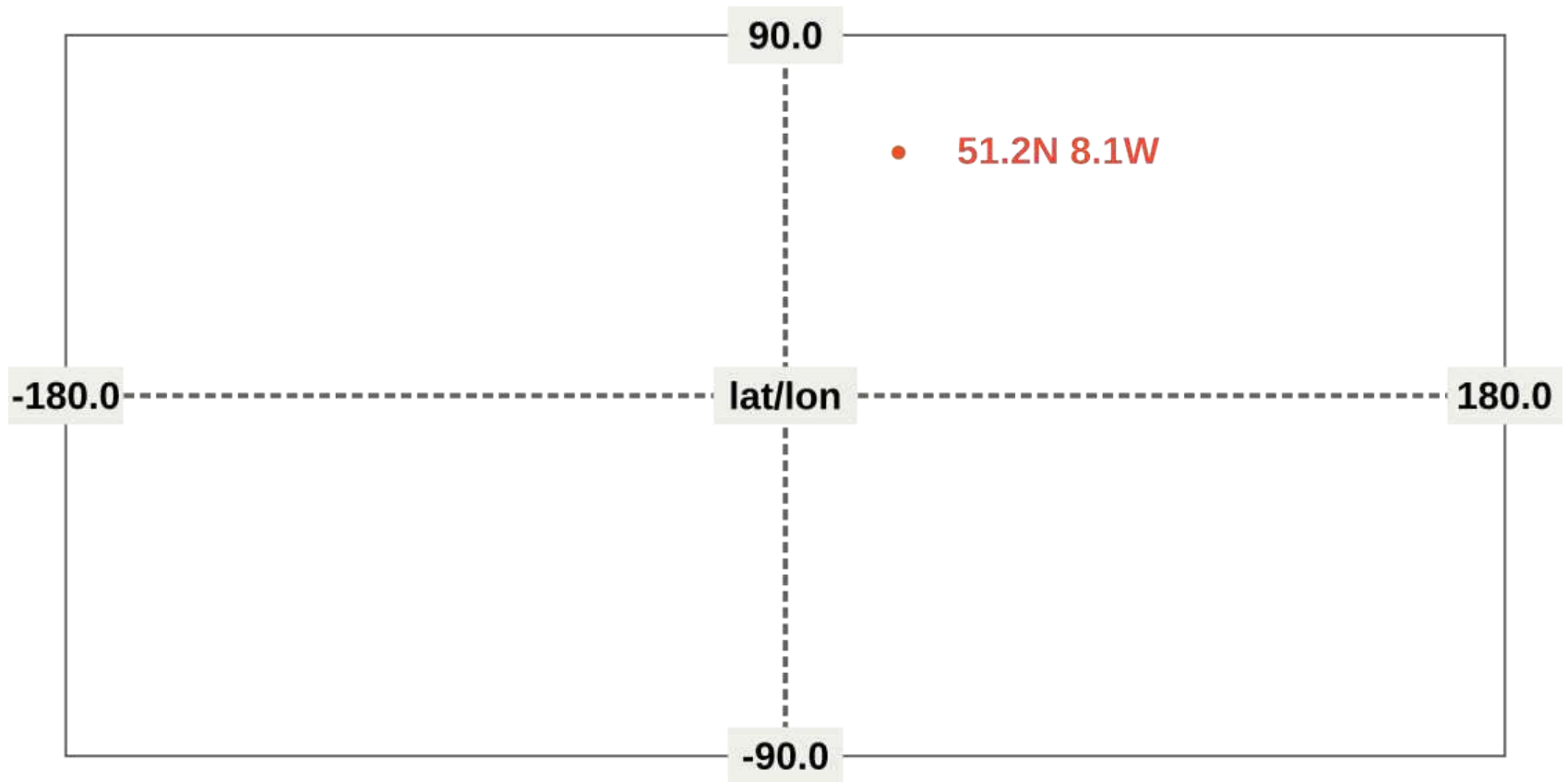
Cell

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

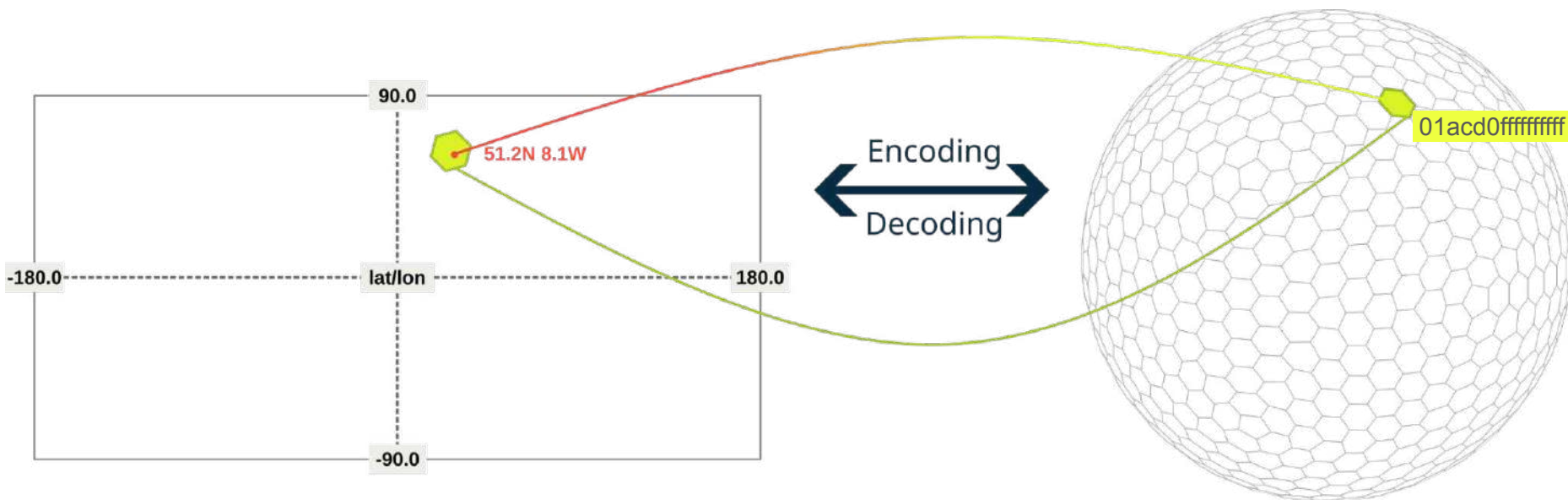
<https://ceos-org.github.io/eo-glossary/>



Lat/Lon space: Analog or Digital/Discrete?



Tokenization of Lat/Lon space into zone IDs

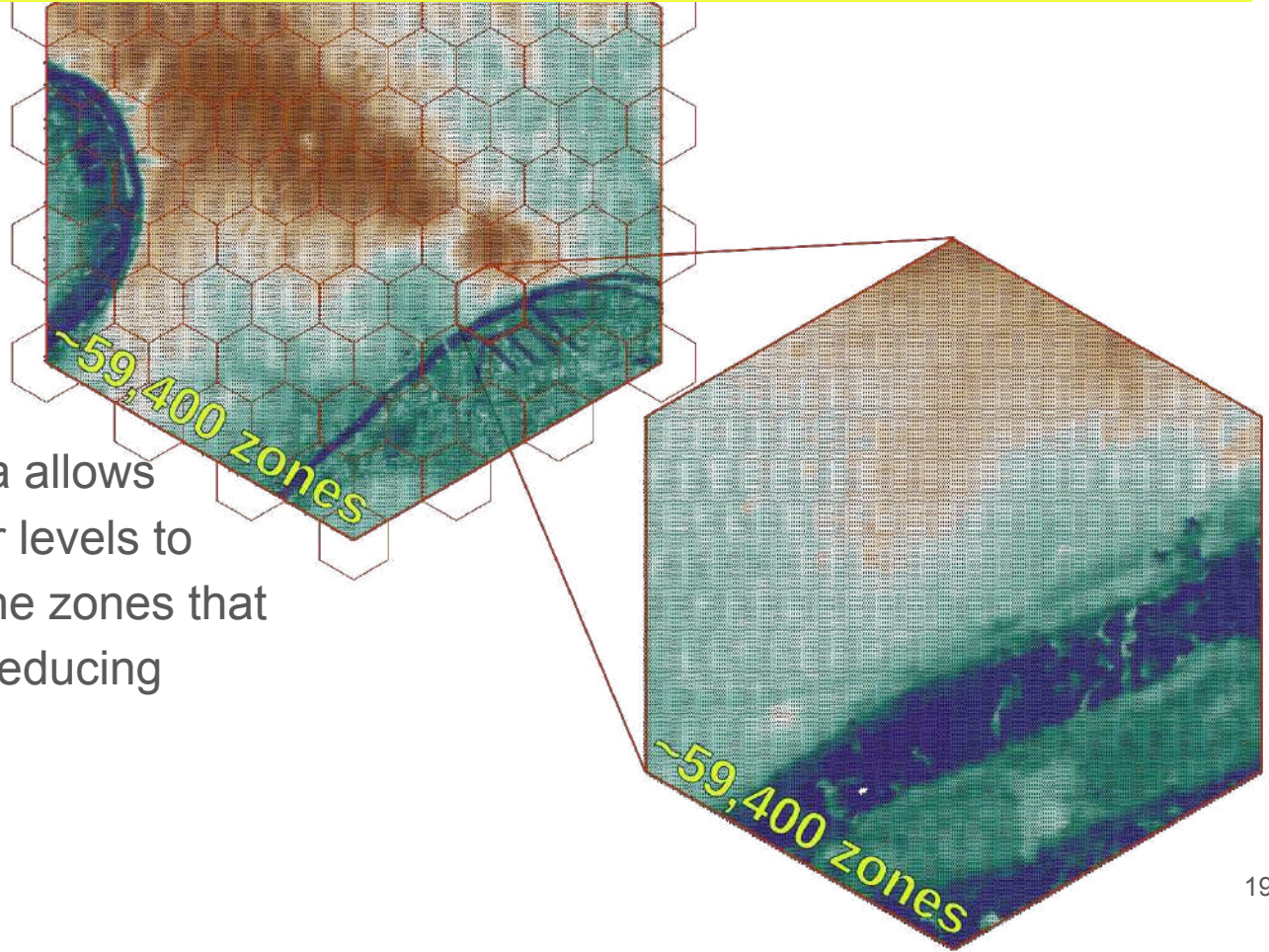


Compact, hexagonal
zones of equal size ₁₈



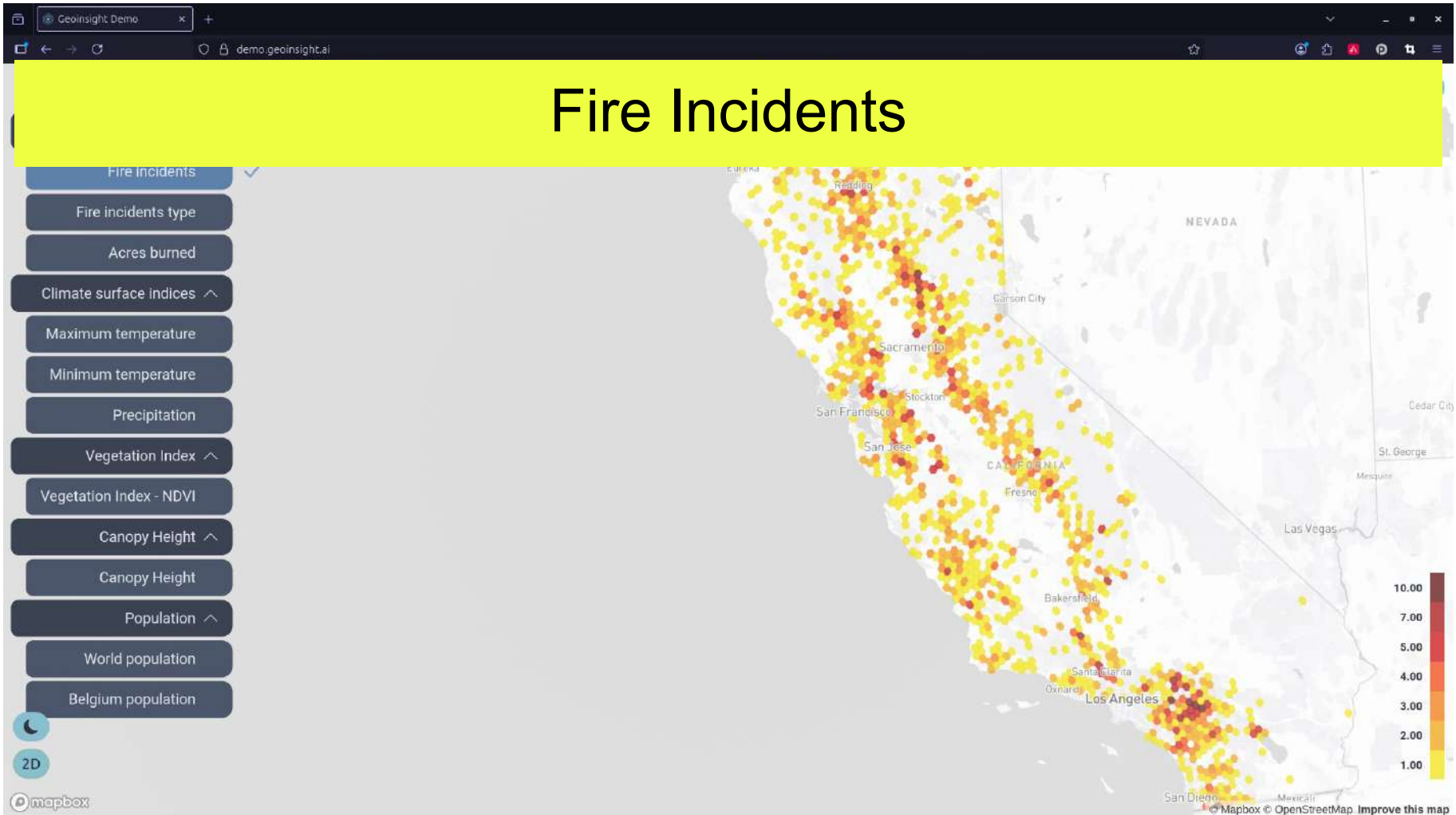
Hierarchy of zones

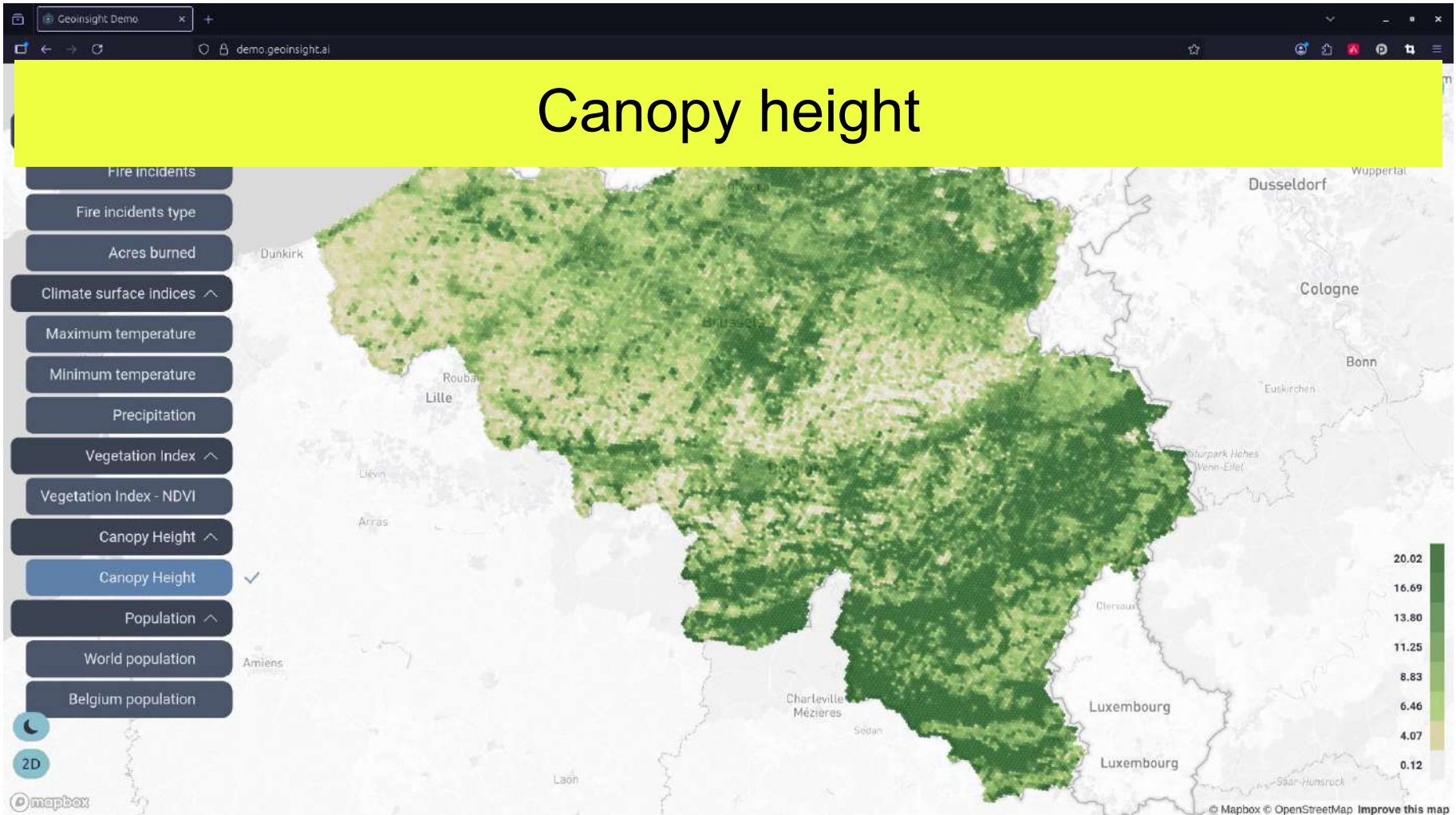
- Inherent in the DGGs concept
- Partitioning of data into parent zones and sub-zones
- Nested
- Aggregation of data allows analysis on coarser levels to then 'zoom-in' on the zones that are significant = reducing resources.



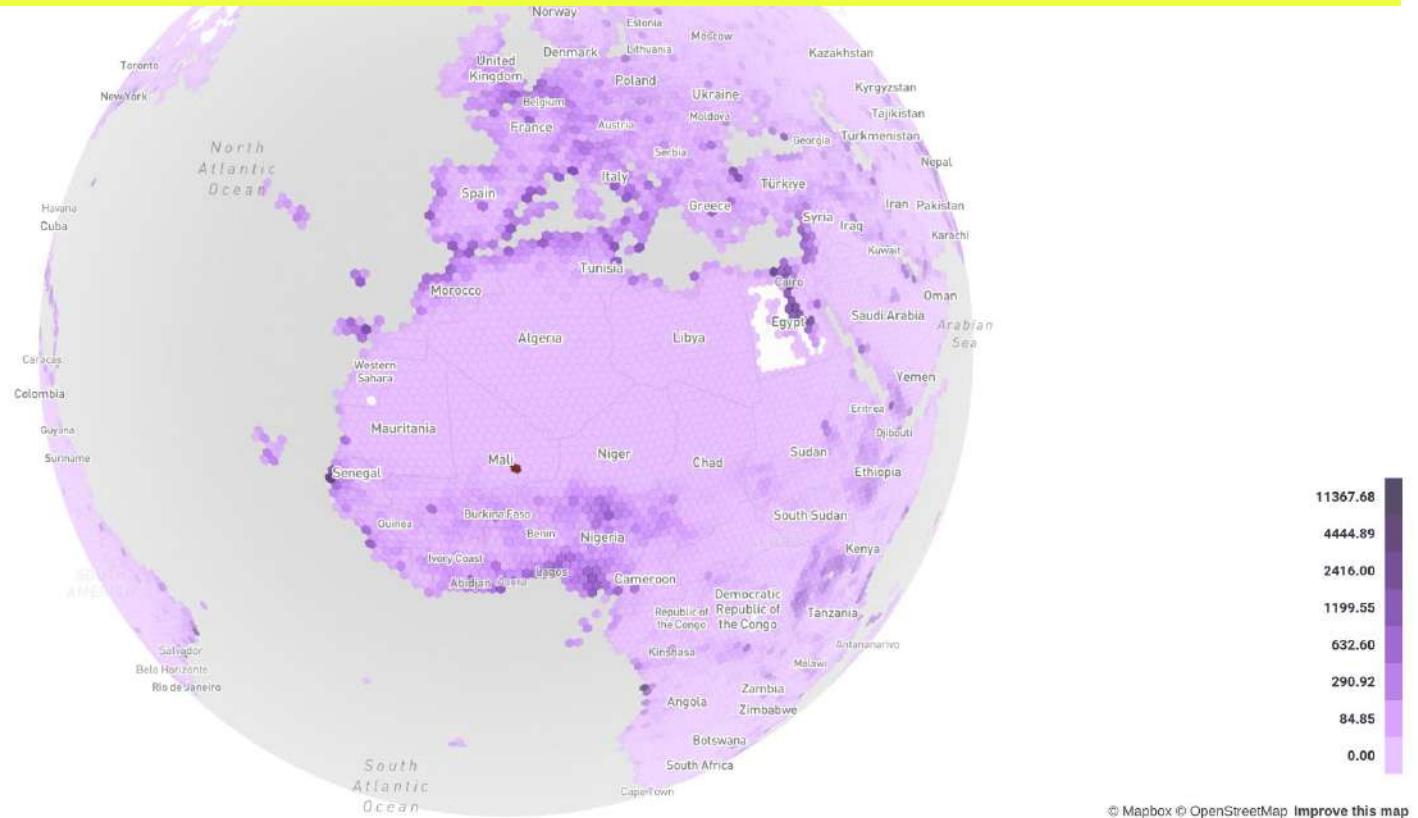
Use Cases







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

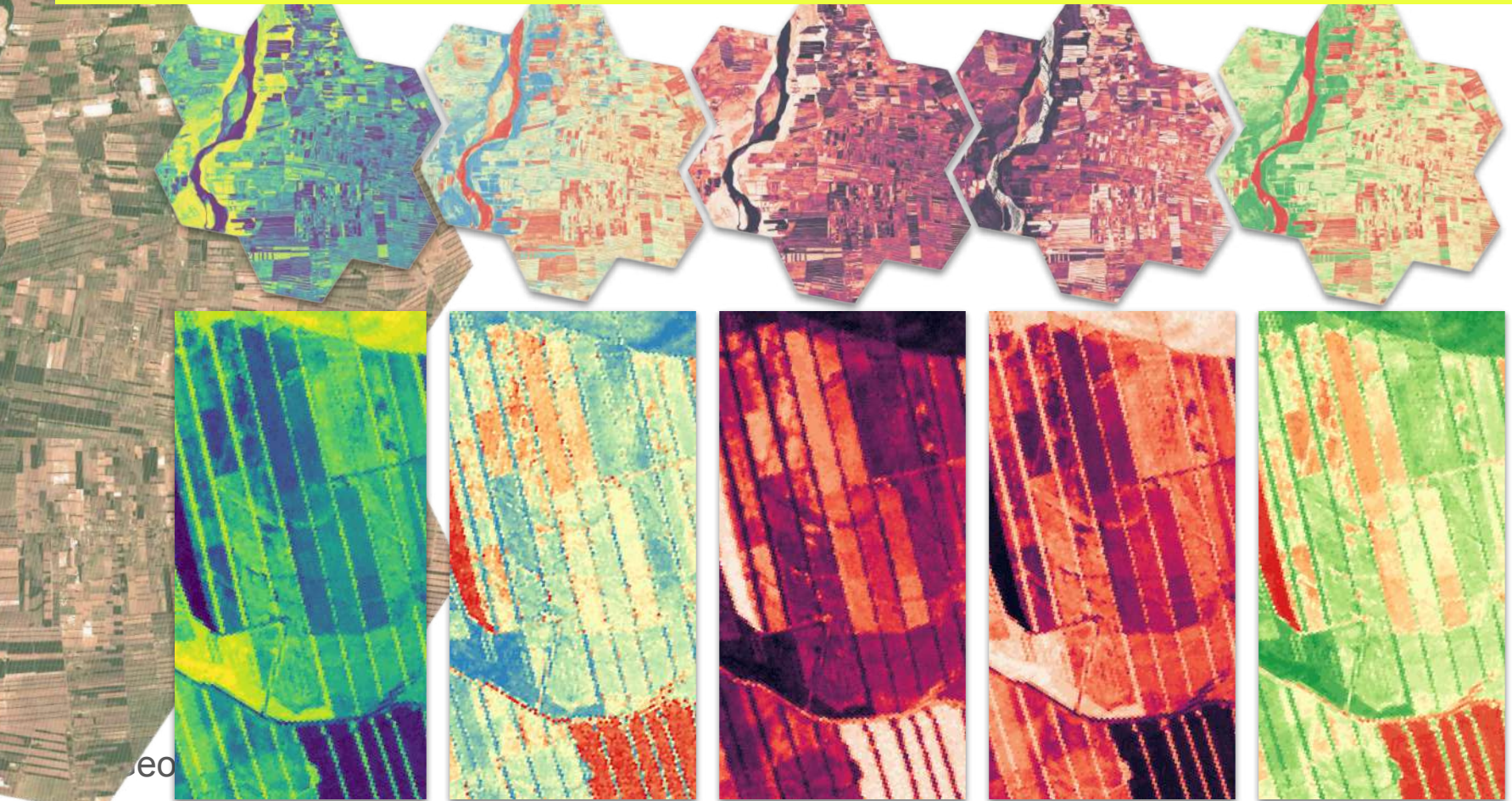


mapbox

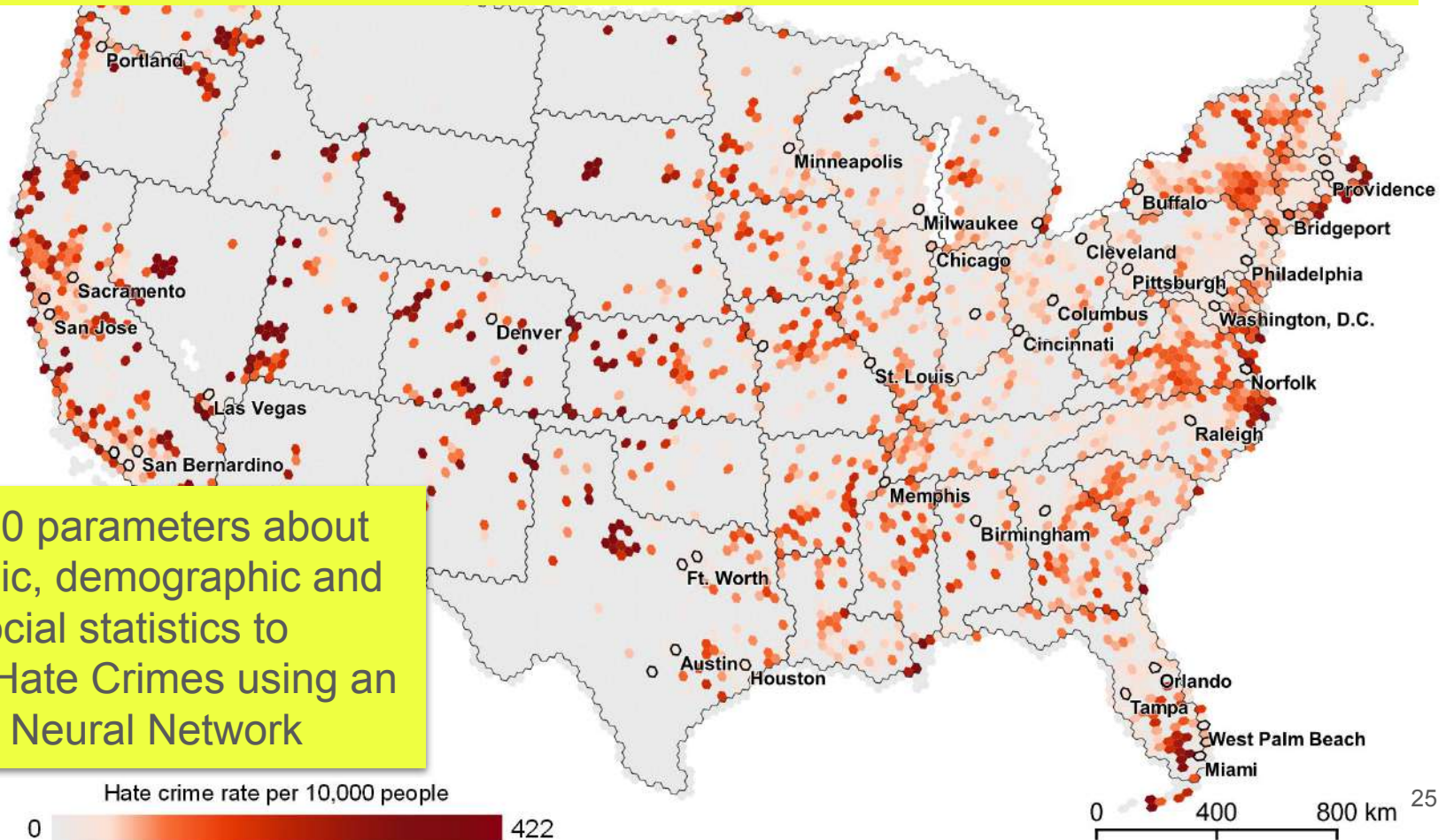


GeoInsight

High resolution agricultural indices



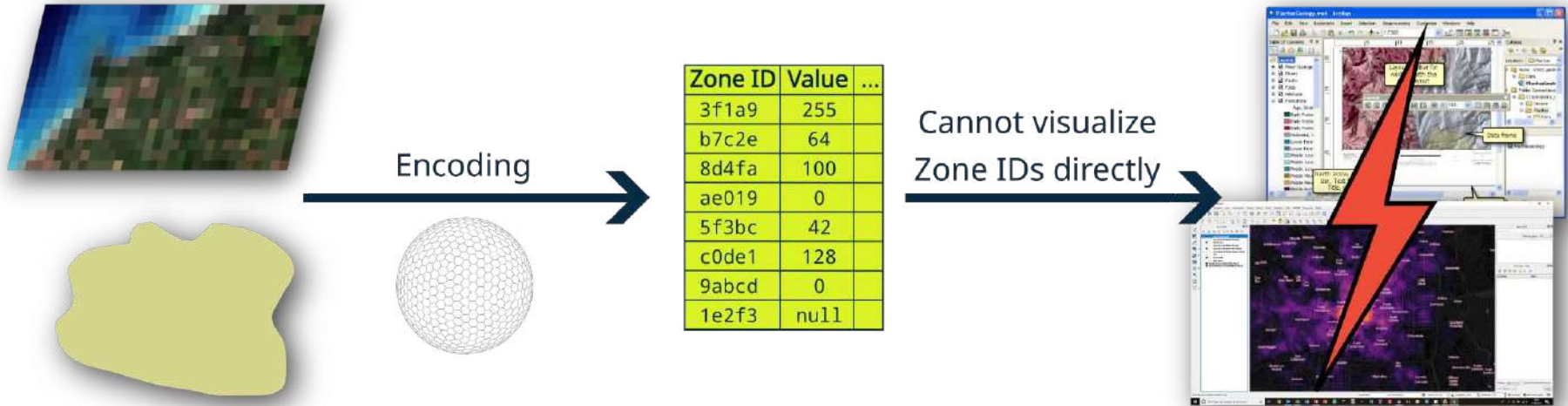
Hate crimes in the USA



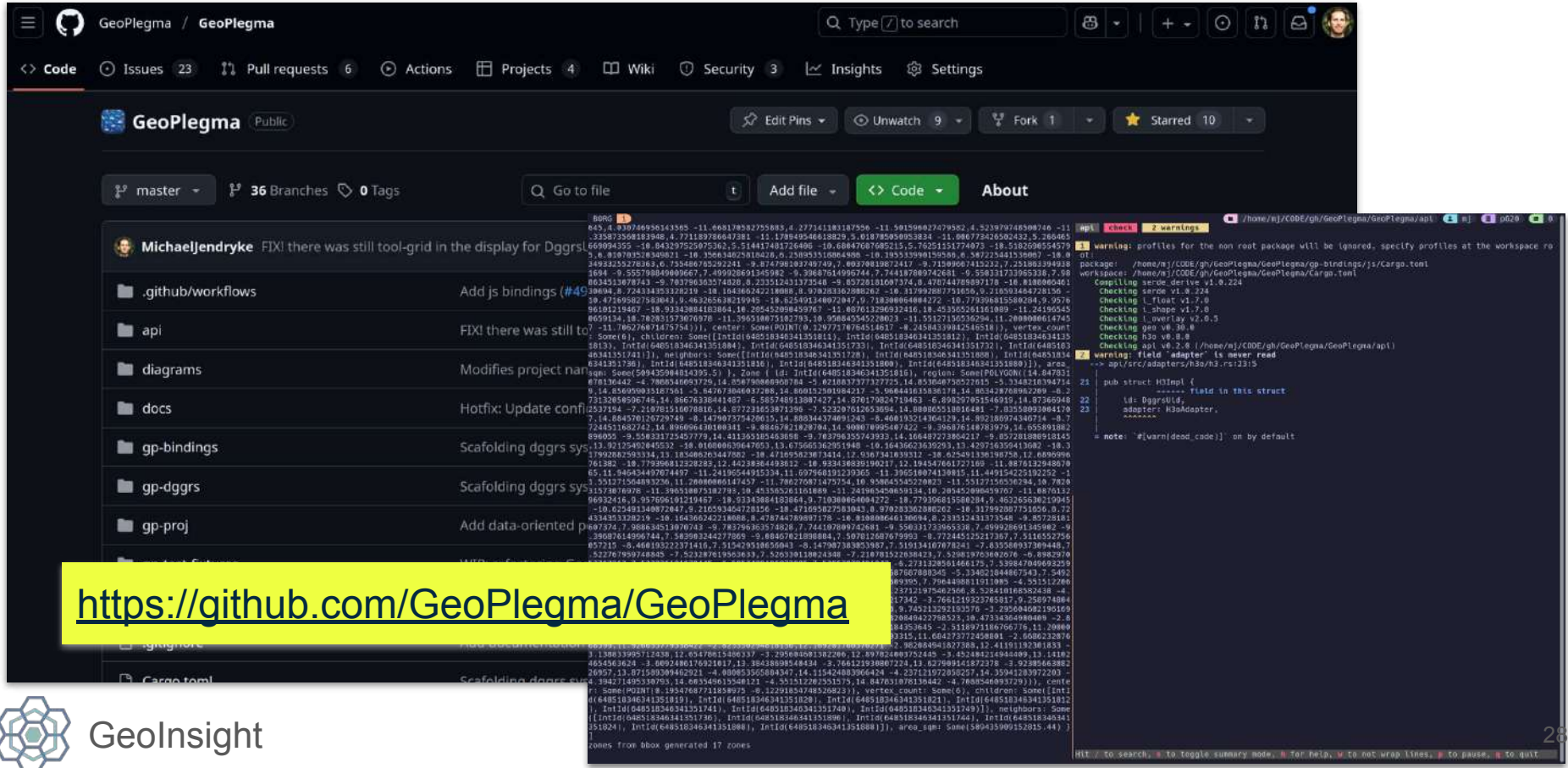
Implementations



DGGS aware and unaware clients



GeoPlegma - an open-source DGGS library



The screenshot displays the GitHub repository for GeoPlegma, an open-source DGGS library. The repository is owned by MichaelJendryke and is public. It has 36 branches and 0 tags. The README file is selected, showing a commit by MichaelJendryke with the message "FIX! there was still tool-grid in the display for Dggrs". The README content includes a list of features and a list of files. The files listed are .github/workflows, api, diagrams, docs, gp-bindings, gp-dggrs, and gp-proj. The commit history shows a list of commits with their messages and the files they modified. The commit messages include "FIX! there was still tool-grid in the display for Dggrs", "Modifies project name", "Hotfix: Update config", "Scaffolding dggrs sys", "Scaffolding dggrs sys", and "Add data-oriented project". The files modified in these commits include .github/workflows, api, diagrams, docs, gp-bindings, gp-dggrs, and gp-proj.

<https://github.com/GeoPlegma/GeoPlegma>

GeoInsight

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



Open
Geospatial
Consortium

Dr. Matthew Brian John Purss

Jérôme Jacovella-St-Louis



CALL FOR
PARTICIPATION

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

**RESPONSES DUE BY
JUN 20, 2025**



GeoInsight

<https://dggs.io>



Choose a resolution

8

Choose a field

NOX



Add datasets (.geojson, .json)

Datasets

air_quality.geojson



CELL ID	SO2	CO	NO	NO2	NOX
	4.568740092774877	249.18369581341184	8.928823833748101	26.401432636887623	49.99385
	4.5811653514569	249.70854273534796	8.969992749392837	26.49504023357883	50.22143
	4.595961197454199	249.6267838439203	8.928747628973408	26.515087993563206	50.06633528619073
N	4.647195882265073	250.63389434343974	10.340926837645323	26.147185039057884	50.58475560460639
	4.5702663698117185	249.05745115420837	8.095744765931735	26.416093604178348	49.863404146512294

Develop geojson data, A

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

Calculator

30

Outlook



Cloudferry

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

```
ecouser@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
ecouser@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
ecouser@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
```

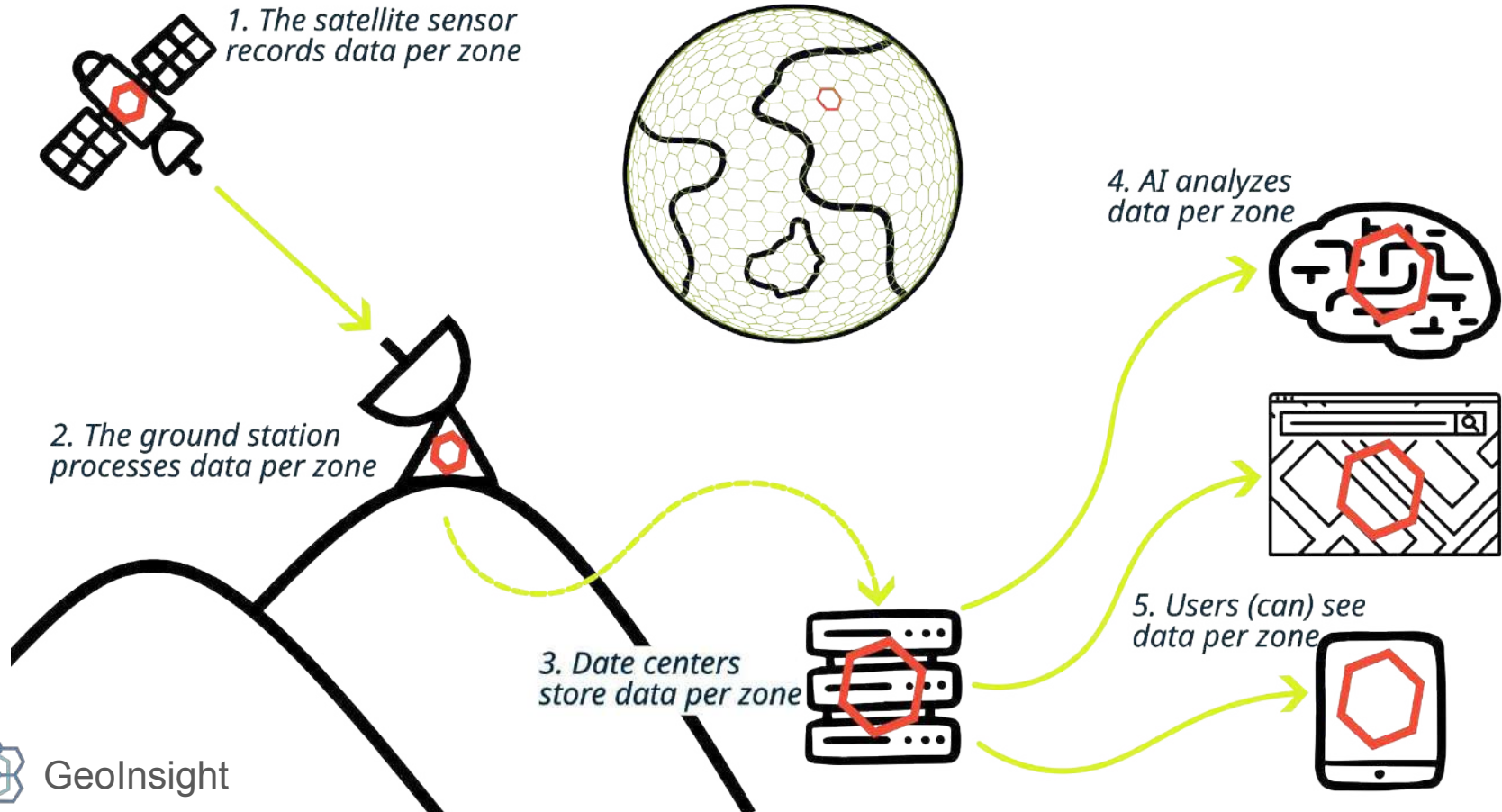


CEOS Analysis Ready Data

- 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

