GROUP ON EARTH OBSERVATIONS

INSIGHT FOR A CHANGING WORLD

GEO WEEK 2017

23-27 OCTOBER 2017

GROUP ON EARTH OBSERVATIONS

INSIGHT FOR A CHANGING WORLD

GEO WEEK 2017

23-27 OCTOBER 2017
WASHINGTON, D.C., USA

www.earthobservations.org/geoweek2017

GROUP ON EARTH OBSERVATIONS

GEO-CRADLE REGIONAL DATA HUB USING GEO DAB APIs

http://datahub.geocradle.eu/

Haris KONTOES, Geo-Cradle Coordinator
NOA Research Director

Coordinating and integRating state-of-the-art Earth Observation Activities in the regions of North Africa, Middle East and Balkans and Developing Links with GEO related intiatives toward GEOSS

http://geocradle.eu/

GEO-CRADLE REGIONAL DATA HUB

The GEO-CRADLE Project

GEO-CRADLE

Funded under H2020 - Climate action, environment, resource efficiency and raw materials
ACTIVITY: Developing Comprehensive and Sustained Global Environmental Observation and Information Systems
CALL IDENTIFIER: H2020 SC5-18b-2015 Integrating North African, Middle East and Balkan Earth Observation capacities in GEOSS
Project GA number: 690133
Total Budget: 2,910,800.00 €

… is a unique EU funded Coordination Action running at regional level;
… is looking at the territories of North Africa, Middle East and Balkans;

It seeks to identify common needs, create synergies, and integrate capacities;

Fosters the regional cooperation and integration of monitoring capabilities and networks, as well as scientific skills;

Define and communicate goals that are clear and beneficial from societal and market wise point of view, and also realistic and in line with the domestic priorities and user needs;

Proposes/sets up large scale regional initiatives based on the Earth Observation (space based and in-situ) for capacity building and also addressing societal priorities in the thematic areas of the project such as Adaptation to Climate Change, Access to Raw Materials, better exploitation of the renewable Energy resources, and Food Security.

Objectives

- ✓ **Promote** the uptake of EO services and data in response to regional needs.
- ✓ **Support** the effective integration of existing Earth Observation Capacities in the region.
- ✓ **Facilitate** the engagement of the complete ecosystem of EO stakeholders in the region.
- ✓ **Enhance** the participation in and contribution to the implementation of **GEOSS** and **Copernicus** in **North Africa**, **Middle East** and the **Balkans.**

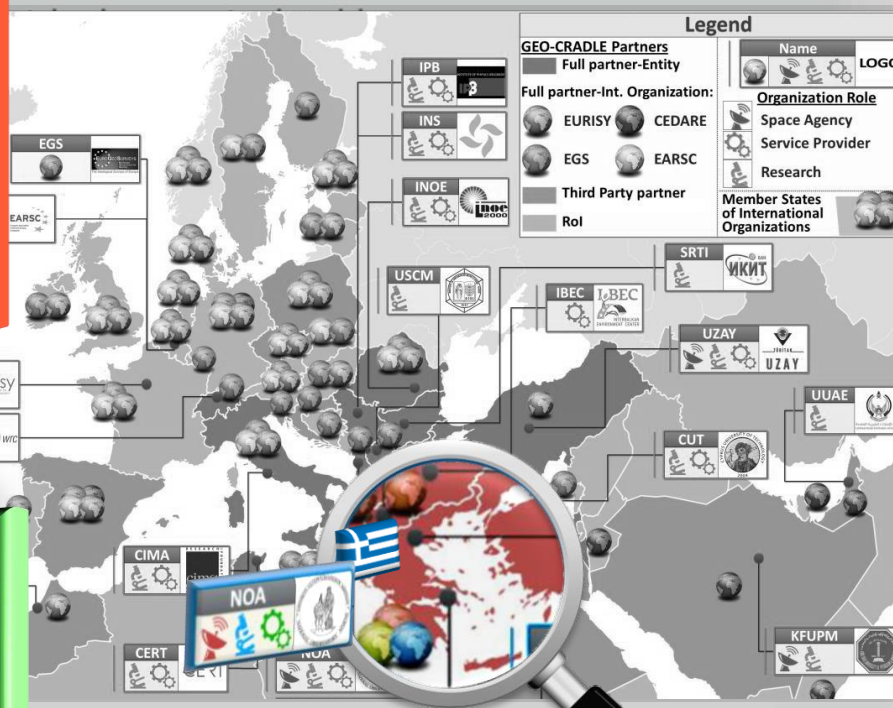
GEO-CRADLE REGIONAL DATA HUB

The GEO-CRADLE Project

25 Partners, 3 Continents, 1 Team

25 Partners, 3 Continents, 1 Team

- **GEO-CRADLE** brings together a highly-complementary team combining a strong background in GEO-related coordination activities with proven excellence in the field of Earth Observation:



Adaptation to Climate Change (ACC)

Improved Food Security – Water Management (IFS-WEM)

Access to Raw Materials (ARM)

Access to Energy (SENSE)

@GE C2025

www.earthobservat rg

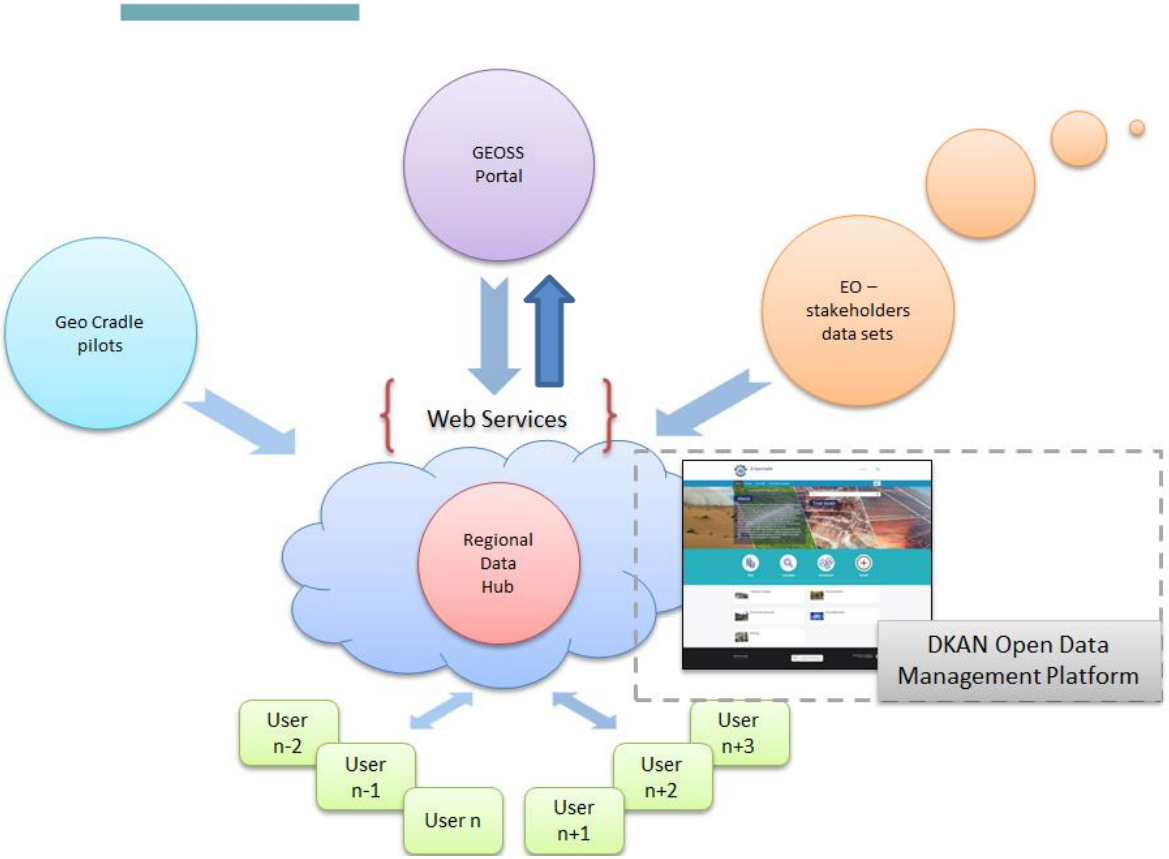
GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Concept

What question are you answering?

How to provide access to region-related (N. Africa, Middle East, Balkans) datasets and services, directly fed from the GEOSS-portal, and local portals, while at the same time being the centralised gateway for regional data providers to contribute easily and timely their products to GEOSS

GEO-CRADLE has designed and put in operation the Regional Data Hub which is the focal node for data sharing in the regions of N. Africa, Middle East, Balkans in the context of GEOSS and Copernicus implementation



GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Concept

- The concept is simple, but not easy to implement.
- The RDH aspires to become a de facto “one-stop-shop” for the Region of Interest (RoI) specific data/information/knowledge access for EO players, service providers, and end users.
- In order for this to be accomplished, the RDH has to satisfy the following conditions:
  i. Be online
  ii. Connect with GEOSS
  iii. Connect with multiple remote data sources (big variety in data formats, data types, and data distribution system architectures)
  iv. Act as a gateway to the users
  v. Be efficient, and user friendly to both users and administrators
  vi. Act as an integrated open data management platform

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Solution

THE SOLUTION:

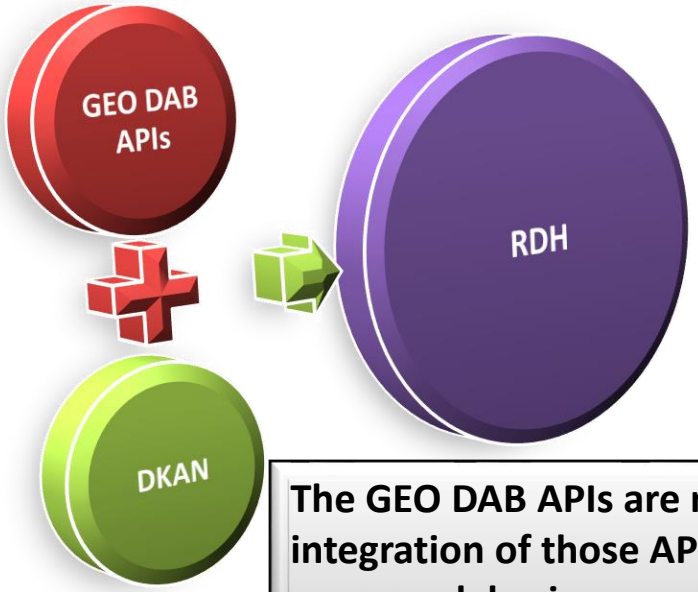
- RDH supports two types of users: (a) unauthorized users and (b) authorized users (i.e. administrators, content authors, group owners, etc.)
- The 1st category is interested more to the search, view/preview and download
- The 2nd category is related with an admin panel (i.e. crud functions, content organization, user management, publishing options, etc.)
- DKAN to the rescue!
- Why DKAN ?
    i. Ultimately, DKAN is a complimentary offering to CKAN in the effort to make data more open and accessible (source)
    ii. Integrates open data catalog features into Drupal CMS, which is build upon PHP. PHP powers a significant percentage of Web, while Drupal powers ~2% of the Internet as a whole
    iii. Has a wide community of active users/developers.

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Connection with GEOSS

The connection of GEOSS to RDH was a challenging objective of the GEO CRADLE project

For this to be achieved the GEO DAB APIs were used, and in particular the OpenSearch API



The GEO DAB APIs are really plug and play. However the integration of those APIs in DKAN in an efficient way, has proven a laborious work that demanded novel software architectural design from the development team

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Highlights

- **GEO DAB APIs are really plug and play, with very good response time**
- **DKAN on the other hand provides several modules, working as add-ons, which support all the necessary functionalities used by an open data management platform**

DKAN which is built on top of Drupal assumes that its content resides in a local database

However, a basic idea for the RDH is not to replicate the GEOSS data

What's the point if the GEOSS data (or data from other remote sources) are replicated in DKAN's local database? DKAN's performance will degrade significantly with that volumes of data!

On the other hand is it feasible to online fetching GEOSS data (and remote data in general) in every user request?

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Highlights (2)

The RDH's uses a smart data caching mechanism which contribute to the efficient rendering of remote data. It is a novel solution to the tradeoff between the storage of the remote data and the online fetch of the data

The logic behind this mechanism is very simple and is inspired by the way the World Wide Web caching proxies and the TCP/IP protocol are working

So, how is this mechanism working?

Whenever a user searches for data through the RDH search machine, an API call is dispatched to GEOSS using the GEO DAB APIs

The GEOSS returns "reports" (i.e. metadata describing datasets). If a "report" is already cached as a DKAN dataset is immediately returned to the user, otherwise a corresponding DKAN dataset will be created and cached

This functionality is taking place behind the scenes, and in the end, the user will get the search results in a uniform way (the DKAN way), regardless of the remote data source (and the source's data model)

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Highlights (3)

- **With the caching mechanism, the RDH avoids to create every time a DKAN dataset**
- **Fetching a dataset from cache is faster, than creating it on the fly**
- **Avoids to store data in the local database. The cache resides entirely in the server's memory (e.g. 4gb). Thus, it is faster than the database (replication is also avoided in that way)**
- **If the cache is full, the cache replacement policy (e.g. LRU) will assure that only those datasets that are not frequently used will be replaced be the new ones**
- **If the dataset is already cached , then it is served from the cache, and there is no any more need for communication with the remote source (e.g. a GEOSS dataset) for viewing the data**

But if in every user's search there is communication with the remote source of data for the creation of the corresponding DKAN dataset and the eventual caching, is finally any benefit from this mechanism?

What happens if the remote data are modified in the source? How does the caching mechanism ensures the delivery of "fresh" remote data?

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Highlights (4)

Well, there is another caching mechanism too, that is the HTTP caching. This caches the HTTP responses of external API calls through the GEO DAB APIs, avoiding the communication with the remote source every time a user searches for data. Another benefit of this mechanism is that it offloads the remote servers and the network bandwidth in general

The rendering cache mechanism provides also a validation policy based on the Adaptive-Time-To-Live (ATTL) scheme, also inspired by WWW caching proxies. So, whenever a DKAN dataset is created/cached for remote data, an adaptive TTL is assigned to this dataset. When this TTL expires, the mechanism will check its validation against the remote data and if the remote data has changed, the mechanism will re-create/re-cache the dataset

GEO-CRADLE REGIONAL DATA HUB

USING GEO DAB APIs – The Highlights (5)

Other enhancements, improvements, novelties are introduced in RDH such as:

- Preview capabilities for GEOSS data (WMS/WFS preview, PDF, CSV preview, etc.)
- Cleaning of GEOSS data; removing/merging duplicate resources reported by GEOSS
- Assigning titles to resources without titles
- Detecting through HTTP HEAD requests formats for resources that have not format information
- URL cleaning; discover URLs that are not working anymore, discarding data with invalid URL schemes
- Transforming a GEOSS report in RDF and Project Open Data Schema formats

Even more the most important enhancement stems from the fact that the RDH provides an architecture that is highly modular and replicable

The software could be used for the creation of similar hubs. Even in its current version the SW is easy to install and used by other sites/projects in just two steps:

Step 1: install DKAN
Step 2: install the RDH module (which actually is a family of modules)

GEO-CRADLE REGIONAL DATA HUB

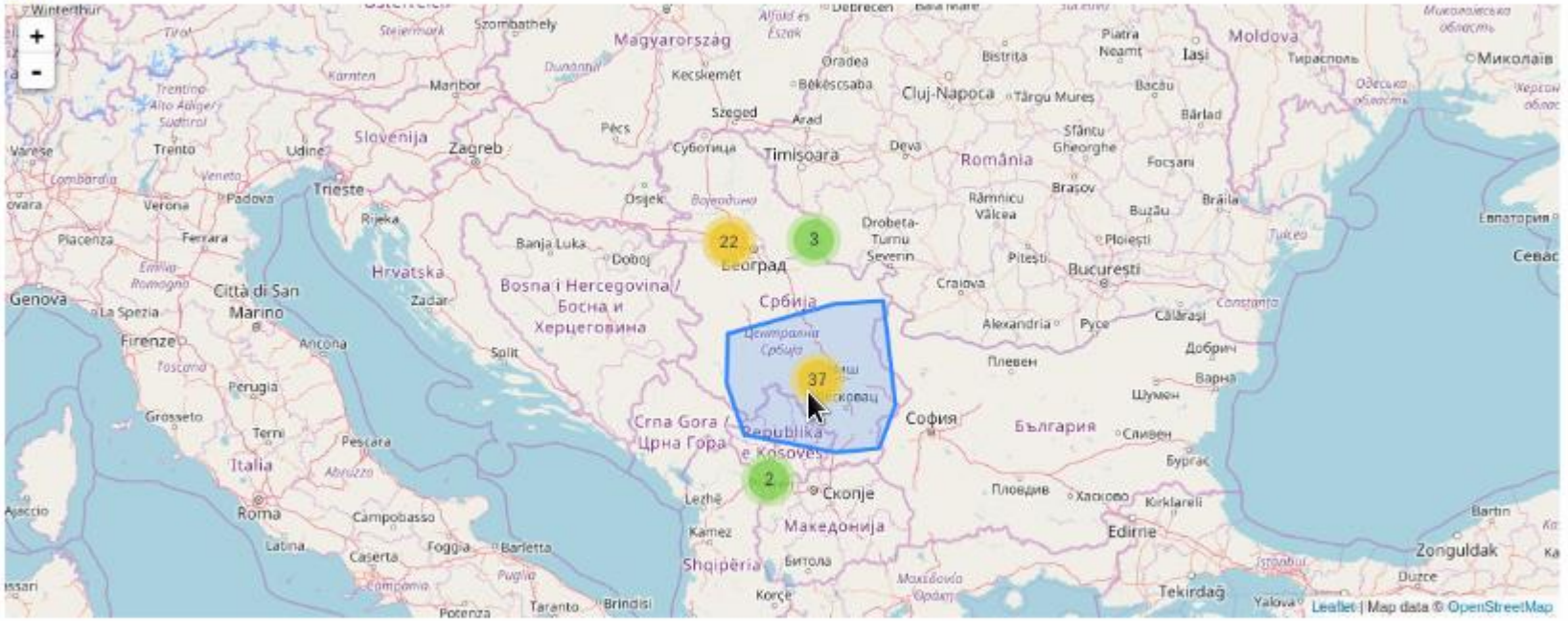
FEEL and LOOK

Home / Datasets / Summer temperature trend 1970-1990 after change year in Serbia

View | Add Resource | Edit | Revisions | Devel

osgl_2:tac_summer2

Preview of a resource that is WFS layer!!!



Resources

osgl_2:tac_summer2

via the DKAN API

Harvested	Yes

changes in the
h the end of the 1990s.

Social
Twitter
LinkedIn

Related Content:

GEO DAB API (OpenSearch)

irect
ataset's

Datas

Other
which
access
POD formats.

Facets to further filter results according to da (e.g. GEOSS), format

Social widget to share dataset in social media.

Link to dataset's source data (e.g. GEO DAB).

GEO-CRADLE REGIONAL DATA HUB

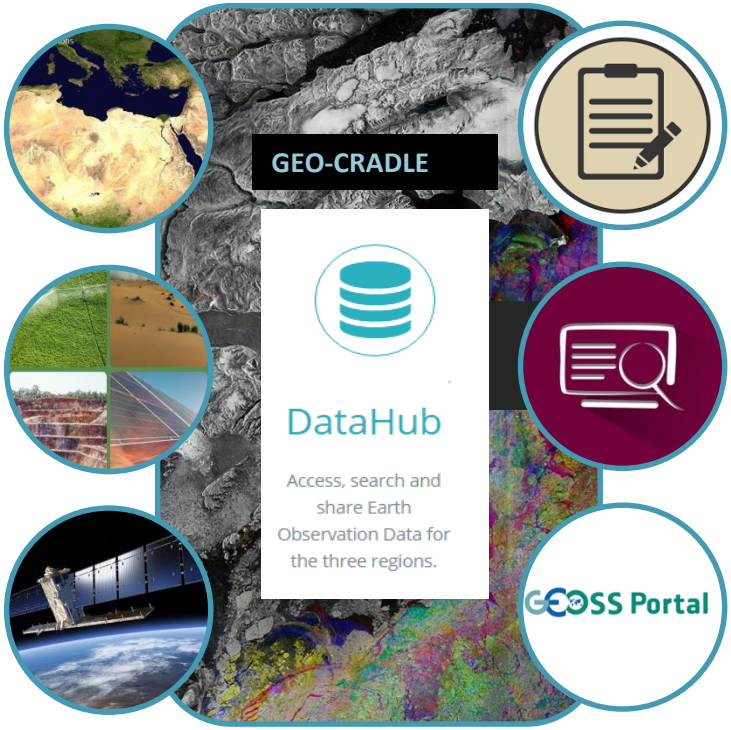
VIEWS OF DATA/DATA SOURCES IN RDH

DATA

Regions of Interest
- Balkans
- Middle East
- North Africa

Thematic Areas
- Climate Change
- Food Security & Water Extremes
- Raw Materials
- Energy

EO Data Categories
- Space borne
- Ground based
- Modelling
- Data exploitation



DATA PROVIDERS

GEO-CRADLE Survey
- 10 national portals and sites

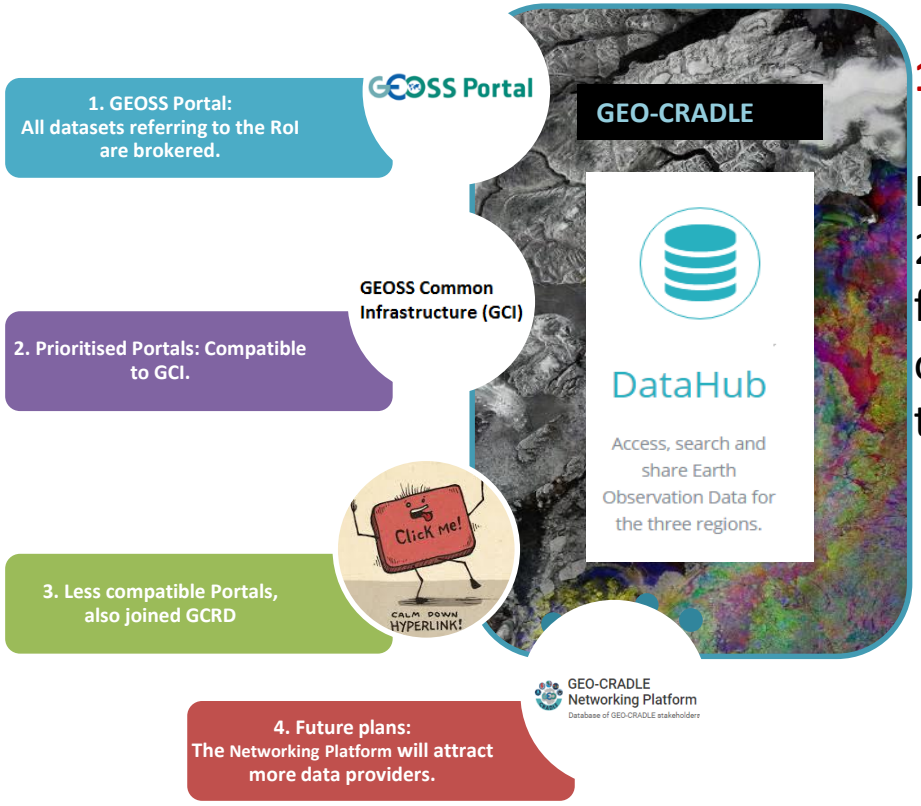
Portals **in total**:
42 national
54 national & international

Desk Research
- 32 national portals and sites
- 12 continental and global portals and sites

GEOSS Portal
- 25.131.225 datasets

GEO-CRADLE REGIONAL DATA HUB

VIEWS OF DATA/DATA SOURCES IN RDH (2)

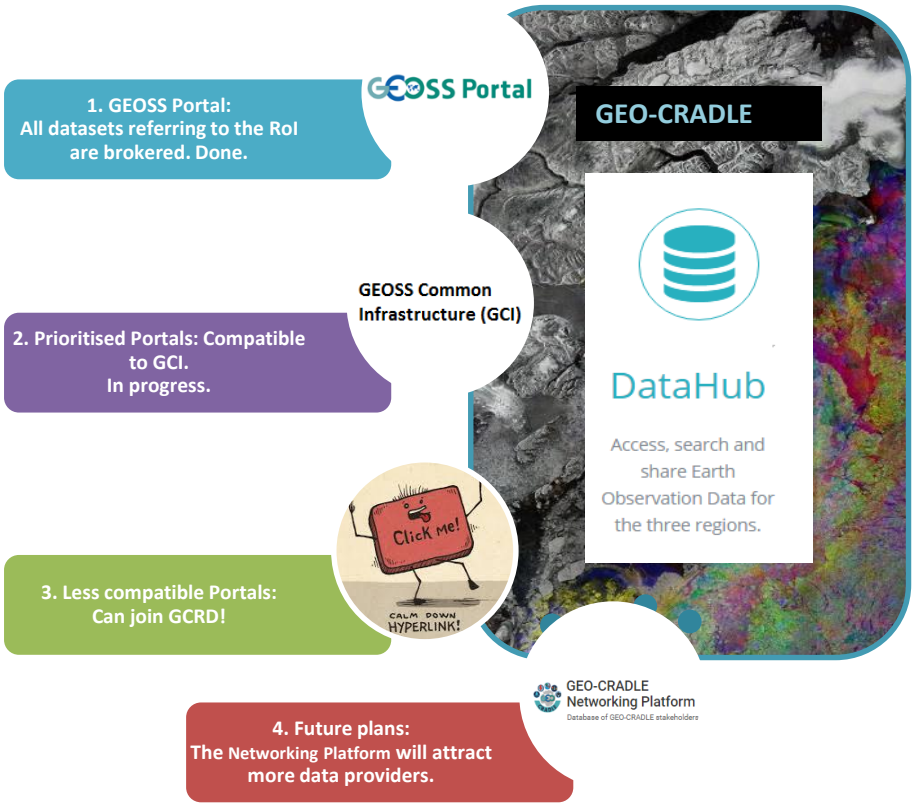


1. Step

Regional GEOSS Portal:
25.131.225 datasets are brokered from GEOSS Portal,
only the areas which belong at the Region of Interest of GEO-CRADLE

GEO-CRADLE REGIONAL DATA HUB

VIEWS OF DATA/DATA SOURCES IN RDH (3)



2. Step

According to GCI team:
a Top-14 (out of 54) national portals were more compatible to GCI standards & offer free and open data.

6 out of 14 portals agreed to be brokered by GEOSS

GEO-CRADLE REGIONAL DATA HUB

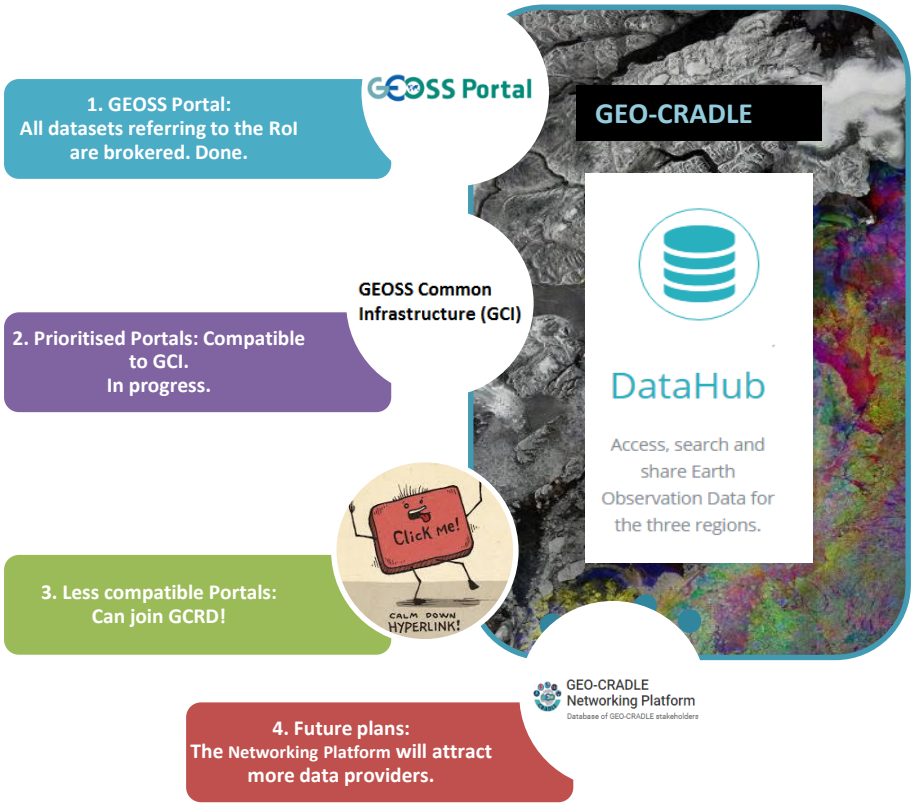
VIEWS OF DATA/DATA SOURCES IN RDH (5)

- The List of the 6 new added portals in GCI

a/a	Portals	Brokered by GEOSS
1.	Danube Reference Data and Services Infrastructure (DRDSI)	Brokered
2.	Eusoils	In progress
3.	Moldova - National geospatial data of Moldova	Brokered
4.	FYROM – Soil information system	Pending
5.	Cyprus - Geoportal	Brokered
6.	Cyprus - Air quality	Pending

GEO-CRADLE REGIONAL DATA HUB

VIEWS OF DATA/DATA SOURCES IN RDH (4)



3. Step

The rest (national and international) portals & sites have:

1. less compatible infrastructure
   or/and
2. restricted policies on data sharing

Have been added to RDH:

- tagged with keywords and
- returned as links.

User can search/access data manually & according the policy of each portal.

Thank You

Communicate and Collaborate with GEO:

GROUP ON EARTH OBSERVATIONS

@GEOSEC2025
www.earthobservations.org