**Project:** *Organization of INCD GeoEcoMar database for operationalization*

 *the Fair Data concept*

**Stage3:** *Installation of the data portal (deadline: 13.09.2024)*

**Stage 3 activities**:

* ***installation of portal and extensions***

There have been installed and configured the additional libraries necessary for portal operation: python 3, PHP 8.2, PostgreSQL, Solr, Nginx, redis.

The CKAN portal can be installed both by compiling the source and using precompiled packages. For the installation of the data portal, the solution of installing a precompiled package (version 2.10.4) was chosen, the installation by compiling the source being more difficult to perform and more prone to errors during installation. Also, this installation method is chosen especially for tests and development.

After completing the installation, the successful operation of the portal was tested by visiting the web address: https://geodb.geoecomar.ro:



 In the next stage, the necessary extensions to improve the functionality of the portal were installed.

* ***Portal initial setup***

The basic installation of the portal was initially configured in the following steps:

* Personalization by using the INCD GEOECOMAR logo
* Main description of the portal
* Creation of the "Geoecomar" organization
* Editing the "ckan.ini" configuration file by changing some operating parameters adapted to the needs of the data portal
* The extensions were activated in the configuration file and depending on the particularities of the installation, the necessary configurations were made to adapt to the needs of the GeoEcoMar data portal.

***- Testing the correct functioning of the portal***

All portal functionalities and installed extensions have been tested using sample data sets.

The general functioning of the portal was analyzed and no technical or other problems were discovered (security, performance, etc.).

The portal was restarted several times by restarting the nginx web service to simulate system crashes and no errors were discovered when it was restarted, with the basic authentication and display services operating within normal parameters.

Within the "Geoecomar" organization, a reference data set called "test-00" was created. This was used to test the functionality of the extensions installed with the portal.

**Stage 3 results:**

***• Initial installation of the portal designated in the previous stage and extensions on a physical server.***

***• Initial configuration of the portal and extensions.***

***• Testing bugs and functionality as a whole portal extensions***

During this stage, the CKAN open source solution designated following the analysis made in the previous stage for the creation of the INCD Geoecomar data portal was installed. An initial configuration of it and a basic customization were made, with subsequent adjustments to be made throughout the duration of the project depending on the specifics of the situations that arose.

Extensions were installed and configured to improve the basic functionalities offered by the portal, a data set with reference files was uploaded for testing the extensions and basic functionalities. The data/metadata ingestion functionality was also tested through a test harvest made from the data.gov.ro platform

As part of these activities, their correct functioning was also analyzed, no major interventions being necessary, the bugs that appeared being of minor importance and successfully resolved.

The data portal, in the initial configuration together with the installed extensions, works without major problems and will be used throughout the project in the activities provided in the future stages for the storage and provision of data sets resulting from scientific research activities in INCD GEOECOMAR.

**Stage 3 dissemination:**

* none

**Stage 3 equipment**

* Hosting + DB server+ DB, data processing notebook

**Stage4:** *Development of a structure and access model for scientific data storage (deadline: 13.11.2024)*

**Stage 4 activities**:

- ***Obtaining the catalog scheme***

In this stage, sample data sets from each field and associated subfield of research were analyzed.

Following the analysis, a model catalog scheme was developed to be representative of all types of existing data. An analysis of the table headers for the existing data sets was made and a set of common parameters was chosen to be relevant in the description of the data sets regardless of the scientific field/subfield they belong to.

***Analyzing descriptive metadata***

The catalog model scheme obtained in the previous stage was created in such a way as to provide a set of common descriptive metadata for all existing data sets within INCD GEOECOMAR regardless of the scientific field from which they come. In the definition of the terms involved in this, it was taken into account to obtain a degree of FAIRness as high as possible.

Following the analysis carried out, the addition of a <DOI></DOI> type term generated by Zenodo and the loading of data sets into a "trusted" data repository such as Zenodo will be explored next to the GeoEcoMar database.

The term "keywords" will also be introduced in the catalog scheme to optimize the search for relevant terms for the user.

***Analysis of access licenses***

Assigning access licenses is an extremely important step in exposing data sets on the Internet. Based on them, it is decided who and how can use the exposed scientific data. Also, a defined access license is an important step in complying with the FAIR Reusable criterion.

Among the types of existing open licenses, we focused on the analysis of the "creative commons (CC)" ones, since they are the most common in the scientific environment. The Digital Peer Publishing License (DPPL) and the GNU Initiative Free Documentation License (GNU-FDL) are less commonly used. While the CC and DPPL licenses are internationally compatible, the GNU-FDL is tailored to the Anglo-American legal domain.

An analysis of the existing "Creative Commons CC" type licenses was made and based on this the CC-BY type license (now reached version 4) was chosen. CC BY-NC and CC By NC-ND-SA licenses are restrictive and do not agree with the objective of having an open access database.

**Stage 4 results:**

***• Creation of catalog model scheme.***

***• Definition of terms involved.***

***• Designation of access licenses***

During this stage, the descriptive metadata model scheme was developed for the data sets that will be loaded into the INCD GEOECOMAr database. Representative data sets were analyzed for each scientific field/subfield, a uniform model was created for them.

Also, the terms involved in the model scheme were defined and it was aimed to obtain the highest degree of FAIRness by choosing terms representative of the FAIR principles.

An analysis of the existing license types was made and data access licenses were designated in the form of the CC BY4.0 license for most of the data, thus fulfilling the objective of open access to scientific research data. For particular cases that require restrictions (data still in processing, data under embargo related to project requirements, etc.) it can give free access only to the metadata related to the data set.

**Stage 4 dissemination:**

* **none**

**Stage 4 equipment:**

* **Testing and basic data processing server.**