**Proiectul:** *Organizarea bazei de date a INCD GeoEcoMar pentru operaționalizarea*

*conceptului de Fair Data*

**Etapa3:** *Instalarea portalului de date (termen predare:* **13.09.2024)**

**Activitati derulate Etapa 3**:

* *i****nstalare portal si extensii***

Au instalate si configurate librăriile adiționale necesare funcționarii acestuia: python 3, PHP 8.2, PostgreSQL, Solr, Nginx, redis.

Portalul CKAN poate fi instalat atât prin compilarea sursei cit si folosit pachete precompilate. Pentru instalarea portalului de date a fost aleasa soluția instalării unui pachet precompilat (versiunea 2.10.4), instalarea prin compilarea sursei fiind mai dificil de efectuat si mai inclinata spre erori la instalare. De asemenea, aceasta metoda de instalare este aleasa in special pentru teste si dezvoltare.

După finalizarea instalării a fost testata funcționarea cu succes a portalului prin vizitarea adresei web: <https://geodb.geoecomar.ro>:



 In etapa următoare au fost instalate extensiile necesare îmbunătățirii funcționalităților portalului.

* ***configurare portal***

Instalarea de baza a portalului a fost configurata inițial in următoarele etape:

- Personalizare prin folosirea logo-ului INCD GEOECOMAR

- Descriere principala a portalului

- Crearea organizației „Geoecomar”

- Editarea fișierului de configurare „ckan.ini” prin modificarea unor parametri de funcționare adaptați la necesitățile portalului de date

Extensiile au fost activate in fișierul de configurare si in funcție de particularitățile la instalare au fost făcute configurările necesare pentru adaptare la nevoile portalului de date GeoEcoMar.

* ***Testare functionare corecta portal***

Toate funcționalitățile portalului si extensiile instalate au fost testate folosind seturi de date de proba.

Funcționarea generala a portalului a fost analizata si nu au fost descoperite probleme tehnice sau de alta natura (securitate, performanta, etc).

Portalul a fost repornit de câteva ori prin repornirea serviciului web nginx pentru a simula căderi de sistem si nu au fost descoperite erori la repornirea acestuia serviciile de baza de autentificare si afișare funcționând in parametrii normali.

In cadrul organizației „Geoecomar” a fost creat un set de date de referința numit „test-00”. Acesta a fost folosit pentru testarea funcționalităților extensiilor instalate odată cu portalul.

**Rezultate Etapa3:**

* ***Instalare inițiala a portalului desemnat in etapa anterioara si extensii pe un server fizic.***
* ***Configurare inițiala portal si extensii.***
* ***Testare bug-uri si funcționalități ansamblu portal extensii***

In cadrul acestei etape a fost instalata soluția open source CKAN desemnata in urma analizei făcute in etapa anterioara pentru crearea portalului de date al INCD Geoecomar. A fost făcută o configurare inițiala a acestuia si o personalizare de baza, urmând ca ajustări ulterioare sa fie făcute pe toata durata proiectului in funcție de specificul situațiilor apărute.

Au fost instalate si configurate extensii pentru îmbunătățirea funcționalităților de baza oferite de către portal, a fost încărcat un set de date cu fișiere de referința pentru testarea extensiilor si funcționalităților de baza. De asemeni a fost testata funcționalitatea de ingestie a datelor/metadatelor printr-un harvest de test făcut de pe platforma data.gov.ro

In cadrul acestor activități a fost analizata de asemenea si funcționarea corecta a acestora nefiind necesare intervenții majore, bug-urile apărute fiind de mica importanta si rezolvate cu succes.

Portalul de date, in configurația inițiala împreuna cu extensiile instalate, funcționează fără probleme majore si va fi folosit pe toata durata proiectului in cadrul activităților prevăzute in etapele viitoare pentru stocarea si punerea la dispoziție a seturilor de date rezultate in urma activităților de cercetare științifica din INCD GEOECOMAR.

**Diseminare Etapa 3:**

Nu este cazul

**Echipamente Etapa 3:**

* Server hostare+ DB, laptop procesare date

**Etapa4:** *Elaborarea unui model de structura si acces pentru stocarea datelor științifice (termen predare:* **13.11.2024)**

**Activitati derulate Etapa 4**:

* ***Obtinerea scehmei catalog***

In aceasta etapa au fost analizate seturi de date sample din fiecare domeniu si subdomeniu asociat de cercetare.

In urma analizei efectuate a fost elaborata schema model de catalog care sa fie reprezentativa pentru toate tipurile de date existente. A fost făcută o analiza a capetelor de tabel pentru seturile de date existente si a fost ales un set de parametrii comuni care sa fie relevanți in descrierea seturilor de date indiferent de domeniul/subdomeniul științific de care aparțin.

* ***Analiza metadate descriptive***

Schema model catalog obținuta in etapa anterioara a fost creata in așa fel incit sa ofere un set de metadate descriptive comune pentru toate seturile de date existente in cadrul INCD GEOECOMAR indiferent de domeniul științific din care provin. In definirea termenilor implicați in aceasta s-a avut in vedere obținerea unui grad de FAIRness cit mai ridicat.

In urma analizei efectuate se va explora adăugarea unui termen de tip <DOI></DOI> generat de Zenodo si încărcarea seturilor de date intr-un data repository „trusted” cum ar fi Zenodo pe lingă baza de date a GeoEcoMar.

De asemenea se va introduce in schema de catalog si termenul „cuvinte cheie” pentru a optimiza căutarea după termeni relevanți pentru utilizator.

* ***Analiza licente de acces***

Desemnarea unor/unei licențe de acces este un pas extrem de important in expunerea seturilor de date pe internet. Pe baza acestora se decide cine si cum poate utiliza datele științifice expuse. De asemenea o licența de acces definita este un pas importat in respectarea criteriului de reutilizare (Reusable) FAIR.

Dintre tipurile de licențe open existente ne-am concentrat pe analiza celor „creative commons (CC)” întrucât acestea sunt cele mai des întâlnite in mediul științific. Licența de acces de tip Digital Peer Publishing (DPPL) și Licența pentru documentație gratuită a Inițiativei GNU (GNU-FDL) sunt utilizate mai rar. În timp ce licențele CC și DPPL sunt compatibile la nivel internațional, GNU-FDL este adaptat domeniului juridic anglo-american.

 A fost făcută o analiza a licențelor de tip „Creative Commons CC” existente si pe baza acesteia a fost aleasa licența de tip CC-BY (acum ajunsa la versiunea 4). Licențele de tip CC BY-NC si CC By NC-ND-SA sunt restrictive si nu sunt in acord cu obiectivul de a avea o baza de date open acces.

**Rezultate Etapa4:**

* ***Crearea schema model catalog.***
* ***Definire termeni implicați.***
* ***Desemnare licențe de acces***

In cadrul acestei etape a fost elaborata schema model de metadate descriptive pentru seturile de date ce vor fi incarcate in baza de date a INCD GEOECOMAr. Au fost analizate seturi de date reprezentative perntru fiecare domeniu/subdomeniu stiintific, a fost creat un model uniform pentru acestea.

De asemenea au fost definiți termenii implicați in schema model si a fost urmărit obținerea unui grad FAIRness cit mai ridicat prin alegerea unor termeni reprezentativi pentru principiile FAIR.

A fost făcută o analiza a tipurilor de licențe existente si au fost desemnate licențele de acces pentru date sub forma licenței CC BY4.0 pentru majoritatea datelor îndeplinind-se astfel obiectivul de open acces la datele din cercetarea științifica. Pentru cazuri particulare care necesita restricții (date încă in prelucrare, date sub embargo legate de cerinte proiect, etc) poate da acces liber doar la metadatele aferente setului de date.

**Diseminare Etapa 4:**

**Nu este cazul**

**Echipamente Etapa 2:**

* **Server testare si procesare primara date.**