

Anexa 6

La Raportul științific final privind executarea proiectului de cercetări științifice aplicative 18.50.07.10A/PS „Elaborarea cadrului conceptual și metodologic pentru e-Infrastructura de date în sfera CDI din Republica Moldova” (e-IDSM)”

Recomandări privind Seturile și Serviciile de date din sfera CDI

(Proiect)

(Document recomandat spre validare. Proces-verbal nr.3 din 18.12.2019 al ședinței Consiliului științific al Institutului de Dezvoltare a Societății Informaționale)

Chișinău 2019

Cuprins

1. Introducere	3
2. Date de cercetare	3
3. Seturi de date	4
4. Formate ale seturilor de date	4
5. Metadate	5
<i>Conținutul metadatelor</i>	6
6. Servicii de date	7
7. Managementul datelor de cercetare	8
<i>Exemplu de descriere a unui set de date</i>	8

1. Introducere

Organizarea datelor de cercetare în sfera CDI din Republica Moldova are la bază responsabilitatea asumată de instituțiile și proiectele de cercetare privind asigurarea accesului la date și conservarea datelor.

Prezentul document conține recomandări privind modul de organizare a seturilor de date și a serviciilor de date din sfera cercetării, dezvoltării și inovării din Republica Moldova în conformitate cu prevederile Concepției e-Infrastructurii de date științifice din Republica Moldova - e-IDSM.

Recomandările privind organizarea seturilor de date și a serviciilor de date în cadrul infrastructurii de date CDI sunt elaborate în baza proiectului 18.50.07.10A/PS ”Elaborarea cadrului conceptual și metodologic pentru e-infrastructura de date din sfera CDI din Republica Moldova (e-IDSM)”.

Aceste Recomandări se adresează entităților publice care participă la activitatea de dezvoltare și suport al Infrastructurii de date din domeniul cercetării științifice.

Entitățile publice, participante la dezvoltarea și suportul e-IDSM, vor urma Recomandările la crearea seturilor de date și a metadatelor din domeniul CDI conform specificului domeniilor și disciplinei de cercetare științifică în care activează.

2. Date de cercetare

Datele în cadrul e-IDSM vor reprezenta fapte, măsurători, înregistrări sau observări despre lumea înconjurătoare, colectate de oamenii din diverse domenii de activitate, cu un minim de interpretare contextuală.

Datele pot fi în orice format sau sub formă de scriere, note, numere, simboluri, text, imagini, filme, video, înregistrări audio, reproduceri picturale, schițe, desene sau alte reprezentări grafice, manuale procedurale, formulare, diagrame de flux, descrieri ale echipamentelor, fișiere de date, algoritmi de prelucrare a datelor sau înregistrări statistice.

Datele de cercetare se referă la toate informațiile (independent de formă sau prezentare) necesare pentru a susține sau valida evoluția, rezultatele, observările sau constatările unui proiect de cercetare, inclusiv informațiile contextuale.

Datele de cercetare includ toate materialele create în cursul activității academice, inclusiv înregistrările digitizate, sursele de cercetare, experimentele, măsurătorile, sondajele și interviurile, software-ul și codul sursă.

Infrastructura națională de date în domeniul CDI trebuie să asigure accesul deschis la publicațiile și datele cercetărilor științifice finanțate de la bugetul public, implementarea principiilor FAIR, conformitatea cu prevederile standardelor și a bunelor practici internaționale în domeniu.

Modul de organizare a datelor cercetărilor trebuie să aibă drept scop sprijinul eforturilor privind îmbunătățirea schimbului și a accesului la datele de cercetare sub cele mai diferite aspecte prin dezvoltarea de servicii de date de înaltă performanță care să garanteze interoperabilitatea diferitelor infrastructuri în domeniul cercetării, să depășească izolarea și fragmentarea informațională și să asigure o bună funcționare a activităților de cercetare științifică colaborativă.

3. Seturi de date

Seturile de date reprezintă colecții de statistici și informații brute generate de studiile de cercetare. Seturile de date produse în cadrul proiectelor de cercetare finanțate din bugetul public trebuie să poată fi, de regulă, descărcate gratuit. Seturile de date create de instituții și companii comerciale vor putea fi disponibile și contra cost.

Seturile de date trebuie să poată fi localizate fiind posibilă identificarea instituției sau organizației care le-a creat. De regulă, acestea vor fi sistematizate în funcție de domeniul și disciplina de cercetare.

Conceptul „date deschise” presupune că datele ar trebui să fie accesibile în mod liber. Infrastructura de date ale cercetării din Republica Moldova trebuie să încurajeze și să susțină eforturile în această direcție creând disponibilitate și accesibilitatea seturilor de date din domeniul cercetării științifice.

4. Formate ale seturilor de date

Cercetătorii se vor asigura că datele cercetărilor lor vor fi în deplină siguranță și vor fi recuperabile pentru a putea fi utilizate pe termen lung.

La etapa de stocare a datelor, se va specifica termenul de valabilitate a formatelor de date ale cercetărilor precum și a instrumentelor software care asigură servicii în baza acestor formate de date.

Pentru a asigura păstrarea pe termen lung a funcționalității complete și originale a formatelor de date astfel ca conținutul datelor cercetărilor să poată fi disponibil, se recomandă ca, la selectarea formatelor de date, să fie luate în considerare următoarele caracteristici:

- să fie asigurată o documentație completă și deschisă pentru formatele utilizate pentru ca datele organizate astfel să poată fi preluate și utilizate și în alte cercetări viitoare;
- să fie folosită o platformă independentă pentru utilizarea și managementul acestor formate, neproprietară și independentă de furnizorul ei nefiind pusă astfel în pericol utilizarea acestor formate în alte aplicații și cercetări;
- la completarea seturilor de date să nu fie utilizate compresii ale datelor, deoarece în procesul de comprimare au loc pierderi ale anumitor proprietăți ale acestora, fapt care ar crea anumite dificultăți în utilizarea și interpretarea datelor respective;
- în cadrul seturilor de date să nu fie utilizate programe/scripturi incorporate care ar putea pune în dificultate reutilizarea datelor;

- de asemenea, nu se recomandă să se facă criptarea datelor sau protecția lor prin parolă ceea ce ar putea constitui un impediment în procesul de preluare și reutilizare a lor în cadrul unor cercetări colaborative.

Formatele datelor de cercetare care vor fi operate în cadrul e-Infrastructurii vor trebui să poată contribui la dezvoltarea de servicii de date pentru cercetători într-un context larg. În acest sens, vor fi luate în considerație următoarele recomandări:

- producătorii și furnizorii de date ale cercetărilor vor specifica metodele utilizate la obținerea și analiza datelor;
- se vor specifica standardele utilizate pentru datele din domeniul/disciplina în cauză; conformitatea cu standardele internaționale sau posibilitatea conversiei la aceste standarde, atunci când este cazul;
- se va asigura compatibilitatea și aplicabilitatea mijloacelor informatice și a echipamentelor hardware în perioada de păstrare a datelor;
- se vor lua toate măsurile privind respectarea drepturilor, dacă există, de proprietate intelectuală asupra aplicațiilor software utilizate la analiza și managementul datelor.

Cercetătorii, odată cu formatele utilizate la organizarea și stocarea datelor cercetărilor, vor documenta, în modul în care este specificat în actele normative în vigoare, și mijloacele software utilizate la obținerea și analiza acestor date.

Odată cu stocarea, în modul corespunzător, a datelor în depozite autorizate, se vor atașa și copii ale aplicațiilor software pentru a face posibilă prelucrarea acestor date la reutilizarea lor în viitor.

Formatele seturilor de date utilizate în cursul colectării și a lucrărilor legate de analiza și interpretarea datelor proiectelor de cercetare nu sunt întotdeauna conforme cerințelor privind reutilizarea sau conservarea lor pe termen lung și pot să nu îndeplinească cerințele arhivelor sau depozitelor de date sau să nu satisfacă așteptările beneficiarilor cercetării.

Dacă nu există cerințe specifice ale finanțatorilor sau ale depozitelor de date, recomandările generale pentru selectarea formatelor de seturilor de date pentru ca acestea să poată fi prezervate și reutilizate este să fie utilizate cu preponderență, formate deschise și non-proprietare. Formatele deschise, care nu sunt proprietare, sunt mult mai probabil să rămână utilizabile chiar dacă software-ul prin care au fost create nu mai este disponibil sau nu este funcțional. Formatele de date pentru care există o documentare completă și disponibilă liber au, de asemenea, o probabilitate mai mare de conservare pe termen lung.

Dacă programele software utilizate la crearea seturilor de date sunt singura opțiune pentru citirea sau accesarea datelor, este probabil să fie un format proprietar, care nu este deschis. Formatele deschise sunt, de obicei, mai convenabile pentru reutilizare și prezervare pe termen lung.

5. Metadate

Metadatele sunt documentații care descriu date. Ele reprezintă informații descriptive sau contextuale care sunt asociate publicațiilor și datelor pentru a ajuta la identificarea, localizarea și recuperarea acestora de către utilizatori, facilitând în același timp gestionarea conținutului și a accesului. Metadatele au de obicei forma unui set structurat de elemente.

Existența metadatelor despre date facilitează căutarea și regăsirea datelor atunci când acestea sunt depuse într-un depozit de date. Descrierea și documentarea corectă a datelor cu ajutorul metadatelor permite utilizatorilor de date științifice să înțeleagă și să urmărească detalii importante ale cercetărilor.

Instituțiile antrenate în procesul de creare a metadatelor CDI vor actualiza metadate deținute în conformitate cu prevederile actelor normative valabile în cadrul e-IDSM

Deținătorii /furnizorii de metadate vor încarcă metadatele pentru seturile și serviciile de date științifice pe care le dețin/furnizează și le vor actualiza pe conturile de utilizator proprii pe care le dețin pe portalul infrastructurii naționale de date CDI.

Deținătorii /furnizorii de date și metadate CDI vor fi responsabili de asigurarea veridicității descrierii seturilor și serviciilor de date ale cercetărilor, pe care le dețin/ furnizează, prin metadate suficient de bogate cu privire la conținutul datelor în acord cu cerințele cadrului normativ al infrastructurii naționale de date ale cercetărilor – e-IDSM.

Furnizorii de date și de servicii de date ale cercetărilor științifice vor asigura calitatea și fiabilitatea infrastructurilor de date, atât a celor instituționale cât și a celor naționale sau pe anumite domenii și discipline de cercetare aparte, aplicând realizările în materie de metodologii și instrumente de management, utilizare și partajare a datelor cercetărilor.

Metadatele vor putea include conținut precum informații de contact, locații geografice, detalii despre unitățile de măsură, prescurtări sau coduri utilizate în setul de date, informații despre instrumente și protocoale, informații privind instrumentele de sondaj, informații despre proveniență și versiune și multe altele.

Se recomandă ca informațiile utilizate în procesul de descriere a datelor să fie structurate folosind formate standard de metadate.

Atunci când nu există un standard adecvat pentru metadatele utilizate se poate proceda la elaborarea unui document de metadate în format liber (în stil „readme”).

Conținutul metadatelor

Metadatele pentru seturile și serviciile de date ale cercetărilor științifice vor conține informații prevăzute în documentele normative și recomandările cu privire la infrastructura națională de date e-IDSM.

Instituțiile de cercetare vor crea și vor actualiza metadatele în conformitate cu prevederile cadrului normativ al e-IDSM.

Autoritățile coordonatoare din cadrul e-IDSM verifică conformitatea metadatelor pentru seturile și serviciile de date ale cercetărilor științifice în corespundere cu standardele adoptate, documente normative și recomandări adoptate de organele competente ale e-IDSM, validează și publică metadatele prezentate de către entitățile care le furnizează.

Instituțiile de cercetare deținătoare și furnizoare de metadate vor completa metadatele pentru domeniile de cercetare în conformitate cu specificațiile adoptate de forurile de decizie din cadrul e-IDSM pe domeniile și disciplinele de cercetare pentru care au fost acreditate (în care activează).

Valorile pentru fiecare element al metadatelor referitoare la seturile și serviciile de date ale cercetărilor sunt incluse în catalogul de metadate, o singură dată sau de mai multe ori.

Valorile fiecărui element de metadate pentru seturile și serviciile de date de cercetare sunt necesare pentru a asigura interoperabilitatea cu alte seturi și servicii de date și pot avea forma de texte libere, date, coduri derivate din standardele ISO, cum ar fi coduri lingvistice, cuvinte-cheie provenite din liste sau tezaure controlate ori șiruri de caractere.

6. Servicii de date

Serviciile de date de cercetare se referă la instrumente, suport și instruire pentru a ajuta personalul și cercetătorii care lucrează cu datele de cercetare. Serviciile de date oferă soluții pentru organizarea, managementul și (re)utilizarea datelor științifice și răspund cerințelor cercetătorilor și ale altor categorii de utilizatori referitoare la date.

Serviciile privind datele de cercetare trebuie să ofere instrumente și suport care să ajute cercetătorii ca datele lor de cercetare să poată fi utilizate eficient atât pe durata desfășurării proiectului cât și după ce proiectul a luat sfârșit, iar datele au fost stocate în modul corespunzător, partajate și disponibile pentru alte cercetări.

Serviciile de date sunt de un folos tot mai mare pentru cercetători. Articolele științifice cu date deschise sunt citate cu până la 50 % mai mult. Factorii de decizie din sfera cercetării, dezvoltării și inovării, agențiile de finanțare a cercetărilor trebuie să încurajeze sau să mandateze elaborarea planurilor de management al datelor și partajarea datelor cercetărilor.

Stocarea datelor de cercetare în depozite specializate va contribui la o mai bună asigurare a integrității, securității și disponibilității acestora pentru noi cercetări.

Serviciile de date trebuie să contribuie la maximizarea utilizării datelor de cercetare.

Partajarea datelor și utilizarea bunelor practici privind datele de cercetare trebuie să devină un lucru normal.

În condițiile actuale, trebuie să fie conștientizată nevoia de instruire și abilități privind datele de cercetare. Serviciile de instruire pentru cercetători privind datele de cercetare trebuie să ofere o gamă largă de programe de instruire, inclusiv ateliere programate în mod regulat, cursuri online și evenimente personalizate cu privire la cele mai diferite aspecte ale producerii, analizei, stocării, partajării și alte acțiuni care privesc managementul și utilizarea datelor având drept scop eficientizarea cercetărilor bazate pe date și creșterea potențialului de cercetare colaborativă.

Necesitatea de suport privind managementul datelor de cercetare fiind în creștere, serviciile de date trebuie să se concentreze, în principal, pe următoarele activități:

- Servicii de suport pentru managementul datelor instituționale: cercetătorii trebuie ajutați în a face ca datele lor să fie ușor de găsit, de accesat, de înțeles (atât de oameni cât și de mașină) și de citat.
- Servicii de instruire în domeniul datelor de cercetare: se vor oferi soluții de instruire pentru ca echipele de cercetători și instituțiile de cercetare să poată să aibă mai multe beneficii din utilizarea datelor de cercetare.
- Servicii de raportare privind disponibilitatea datelor: ajutor pentru instituțiile de cercetare de a înțelege modul în care cercetătorii pot partaja datele.
- Servicii de suport în crearea unui plan de management al datelor
- Instrumente pentru a asista cercetătorii în stocarea, managementul, protecția și colaborarea privind datele de cercetare pe parcursul executării proiectului de cercetare.
- Servicii de arhivare, păstrare și partajare a datelor cercetărilor pe termen lung – atât până la sfârșitul proiectului de cercetare cât și după finalizarea lui.

7. Managementul datelor de cercetare

Managementul datelor de cercetare (Research data management - RDM) se referă la practicile privind planificarea, colectarea, stocarea, utilizarea, schimbul și conservarea datelor generate în cadrul proiectelor de cercetare.

De ce este important managementul datelor de cercetare? Partajarea și managementul responsabil al datelor produse sau reutilizate în cadrul activităților într-un proiect de cercetare sunt necesare cercetătorilor, finanțatorilor, editorilor și altor părți interesate și sunt acțiuni care sprijină cercetarea deschisă.

Este din ce în ce mai acceptat faptul că bunele practici de management al datelor oferă o serie de beneficii cercetătorilor și că angajamentul instituțiilor de cercetare de a oferi sprijin și infrastructură pentru datele cercetătorilor săi este un lucru benefic.

Agențiile de finanțare a proiectelor de cercetare și editorii au nevoie din ce în ce mai mult ca datele rezultate din proiectele de cercetare să fie puse la dispoziție pentru evaluare și reutilizare, ori de câte ori cerințele legale și etice permit acest lucru. În cazul în care acest lucru nu este posibil.

Datele de cercetare trebuie să fie făcute publice fiind înregistrate într-un depozit public de date sau în orice altă locație cu condiții de acces. De regulă, o versiune a datelor trebuie să fie păstrată în siguranță fiind regășibilă o anumită perioadă de timp după încheierea proiectului de cercetare.

Crearea unui plan de management al datelor (DMP) va ajuta la clarificarea a ce trebuie să se întâmple cu datele de cercetare, dar și în ceea ce privește modul în care acestea vor fi gestionate pe parcursul întregului lor ciclu de viață.

Exemplu de descriere a unui set de date

Descrierea unui set de date va conține caracteristici cheie precum cele ce urmează:

- DOI și citări de date sugerate,
- Un titlu util și expresiv,
- Numele autorului exact/corect,

- Descrierea completă, inclusiv contextul datelor din studiu și metoda de colectare a datelor,
- Informații privind entitatea finanțatoare,
- Indicatori altmetrici - informații despre descărcări și referințe,
- Link către articolul evaluat peer-review cu acest set de date,
- Categoriile și cuvinte cheie relevante care să permită altor cercetători să regăsească setul de date,
- Informații privind licența, pentru clarificări privind condițiile de reutilizare.

Se recomandă, de asemenea, ca descrierea setului de date să fie însoțită de o declarație privind disponibilitatea datelor și citarea acestora, inclusiv citarea datelor din publicațiile pe hârtie.