

### **Anexa 3**

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 transpunerea principiilor FAIR în cadrul e-Infrastructurii datelor de cercetare e-IDSM**

(Proiect)

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

Chișinău 2019

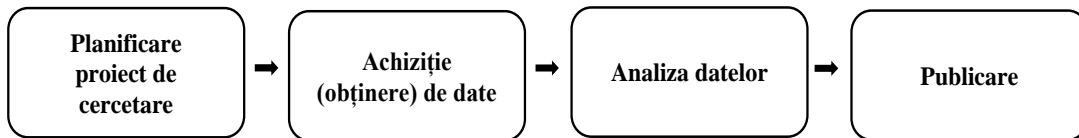
# Cuprins

Introducere.....	3
1. De ce accesul deschis la publicații și date? .....	4
2. Necesitatea planificării managementului pentru datele de cercetare .....	5
3. Principiile directoare FAIR .....	6
4. Ce trebuie să se întreprindă pentru ca datele de cercetare să devină FAIR?.....	7
5. Date regășibile .....	7
6. Date accesibile .....	8
7. Interoperabilitatea datelor.....	9
8. Date Reutilizabile.....	9
Referințe: .....	11
Anexa .....	12

## Introducere

Un ecosistem modern în spațiul cercetării științifice în condițiile actuale trebuie să facă față unei provocări majore privind utilizarea optimă a datelor și a metodelor de cercetare. Provocarea se referă la un management corespunzător al datelor provenite din cercetările științifice care să asigure integrarea și reutilizarea datelor și a cunoștințelor ca o condiție prioritară pentru valorificarea rezultatelor cercetărilor, dar și pentru a facilita noi descoperiri și inovații. Accentul este pus pe datele de cercetare disponibile în formă digitală.

Starea actuală privind ecosistemele datelor de cercetare nu asigură încă condiții suficiente pentru a putea beneficia pe deplin de datele rezultate din activitățile de cercetare. Cercetătorii, finanțatorii proiectelor de cercetare, agențiile guvernamentale, businessul ș.a. conștientizează tot mai mult necesitatea ca datele cercetărilor să fie gestionate într-un mod care înseamnă nu doar colectarea și păstrarea lor în condiții de laborator (Figura 1) pentru necesitățile proiectului în care au fost generate, ci și administrarea lor pe parcursul unui ciclu de viață extins (Figura 2) care cuprinde, de asemenea, partajarea, preservarea și reutilizarea datelor pentru cercetări de viitor.



**Figura 1. Ciclul de viață tradițional al datelor științifice**

Activitățile de cercetare în condițiile actuale au nevoie stringentă de o îmbunătățire esențială a infrastructurii care să asigure managementul și (re)utilizarea datelor de cercetare. Conștientizarea acestei necesități a condus la elaborarea și adoptarea unui set concis și măsurabil de principii directoare numite Principiile FAIR pentru date [1]. Ideea care stă la baza acestor principii este că furnizorii și consumatorii de date trebuie să poată să descopere, să acceseze și să reutilizeze mai ușor și mai eficient informații din volumele mari de date, generate de cercetarea modernă, precum și să fie instituite comportamente integrative și exploaratorii care să facă posibilă crearea unei game largi de instrumente pentru furnizarea, vizualizarea și consumul datelor științifice pe întreg spațiul de cercetare.

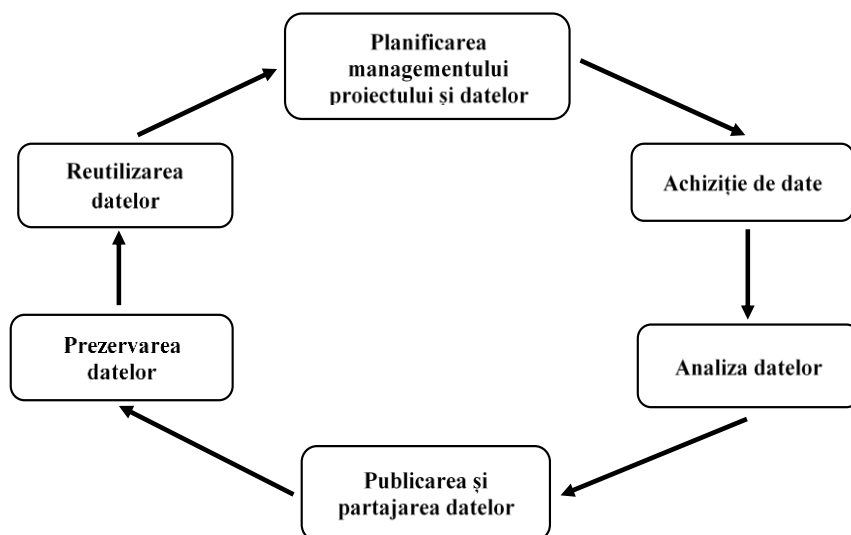


Figura 2. Ciclul de viață al datelor științifice în infrastructură digitală

## 1. De ce accesul deschis la publicații și date?

Datele de cercetare se referă la informații, în special fapte sau numere, colectate pentru a fi examinate și considerate ca bază pentru raționamente științifice, discuții sau calcule. Într-un context larg al cercetării, exemplele de date pot include date statistice, rezultate ale experimentelor, măsurători, observații rezultate din lucrările de teren/laborator, rezultate ale sondajelor, înregistrări foto / video ș.a.

Utilizatorii de date trebuie să poată, în mod normal, să acceseze, să exploateze, să reproducă și să difuzeze datele de cercetare accesibile în mod gratuit.

Accesul deschis la datele cercetării se referă la dreptul de acces și reutilizare a datelor științifice în format digital în termenii și condițiile prevăzute.

Cercetarea modernă se bazează pe un dialog științific extins și avansează prin îmbunătățirea cercetărilor efectuate anterior. Accesul mai larg la publicații și la date științifice ajută la:

- Îmbunătățirea calității rezultatelor pe baza datelor cercetărilor anterioare;
- Creșterea eficienței cercetărilor prin încurajarea colaborării și evitarea duplicării în activitatea de cercetare;
- Accelerarea inovației și a transferului tehnologic;
- Implicarea mai largă a cetățenilor și a societății prin îmbunătățirea transparenței procesului de cercetare științifică.

Din aceste motive, este necesar să se îmbunătățească accesul la informațiile științifice stimulând astfel beneficiile investițiilor publice în cercetare.

Datele cercetărilor obținute în proiectele de cercetare finanțate de la bugetul public trebuie să fie puse la dispoziție online, fără costuri suplimentare, informațiile științifice obținute din proiecte cu finanțare de la bugetul public/ finanțate public pentru cercetători, industriile inovatoare și publicul larg, asigurând-se în același timp că acestea sunt păstrate pe termen lung.

Nu toate datele pot fi deschise. Pot exista excepții de la obligațiile asociate condițiilor de deschidere a datelor, dacă acestea:

- sunt incompatibile cu obligația de a proteja rezultatele care pot fi exploatate în mod rezonabil comercial sau industrial;
- sunt incompatibile cu nevoia de confidențialitate în legătură cu problemele de securitate;
- sunt incompatibile cu regulile privind protecția datelor cu caracter personal;
- ar putea însemna că scopul principal al proiectului nu poate fi atins;
- proiectul nu va genera / colecta date de cercetare;
- există orice alte motive legitime.

Proiectele de cercetare pot alege să păstreze seturi de date selectate sau chiar toate datele închise pentru oricare dintre motivele de mai sus, incluzând aceste specificații în Planul de management al datelor cercetărilor.

Un plan de management al datelor va include prevederi pentru următoarele tipuri de date:

- Date de bază - date necesare pentru validarea rezultatelor prezentate în publicațiile științifice, inclusiv metadatele asociate - metadate care descriu datele de cercetare depuse,
- Orice alte date, de exemplu, date care nu pot fi atribuite în mod direct unei publicații sau date brute, inclusiv metadate asociate.

## **2. Necesitatea planificării managementului pentru datele de cercetare**

Știința se bazează pe descoperirile precedente. Progresul științific este intrinsec conectat la informațiile cercetărilor anterioare pe care le poate reutiliza. Odată ce cercetarea științifică a intrat în era digitală, cantitatea de date produse a început să ajungă la dimensiuni foarte mari.

În acest context, proiectele de cercetare ar trebui să îndeplinească un șir de cerințe privind disponibilitatea datelor pentru cercetări ulterioare:

- Datele de cercetare trebuie depuse, de preferință, în depozite de date de cercetare autorizate. Acestea pot fi arhive de date de cercetare online, constituite pe domenii/ discipline, instituționale sau centralizate. Depozitele de date vor putea oferi informații și asistență suplimentară privind conectarea publicațiilor la datele de cercetare de bază. Unele depozite vor putea permite cercetătorilor să depună atât publicații cât și date de cercetare,

oferind, totodată, și instrumente pentru conectarea acestora. Conectarea publicațiilor și datele care stau la baza lor la depozitele de date se face prin intermediul identificatorilor și citărilor de date persistente.

- Pe cât va fi posibil, proiectele de cercetare vor trebui, în timp, să ia măsuri pentru a permite terților părți interesate să acceseze, să exploateze, să reproducă și să difuzeze (gratuit, pentru orice utilizator) aceste date de cercetare. Un mod simplu și eficient de a face acest lucru este atașarea licențelor Creative Commons (CC BY sau CC0) la datele depuse.
- Proiectele de cercetare vor trebui să ofere informații prin intermediul depozitului ales despre instrumentele care vor fi puse la dispoziția beneficiarilor și care vor fi necesare pentru validarea rezultatelor, precum, de ex., software specializat sau cod software, algoritmi și protocoale de analiză. Atunci când va fi posibil, proiectele de cercetare vor trebui să furnizeze singure aceste instrumente pentru eventualii utilizatori.
- Costurile legate de accesul deschis la datele de cercetare trebuie să fie eligibile, reprezentând o parte componentă a propunerilor de proiect.

### **3. Principiile directoare FAIR**

Principiile de date FAIR sunt principii directoare cu privire la modul în care datele pot fi găsite, accesibile, interoperabile și reutilizabile (a se vedea Anexa 1).

Tendința spre digitalizare a condus la dezvoltarea de platforme digitale unde datele pot fi stocate și transmise. Cum pot fi folosite aceste platforme digitale într-o manieră organizată? Principiile FAIR (Regăsibile, Accesibile, Interoperabile, Reutilizabile) reprezintă un cadru util și eficient pentru a concepe schimbul de date într-un mod care să permită utilizarea și reutilizarea acestor date cu eficiență maximă.

Adoptarea principiilor FAIR reprezintă un pas important în a obține noi beneficii pentru cercetare. Organizarea datelor de cercetare conform principiilor FAIR oferă o serie de beneficii pentru cercetători, organizații și comunități de cercetare, servicii de infrastructură de cercetare, inclusiv:

- Obținerea unui impact maxim din cercetare.
- Creșterea vizibilității și citărilor cercetărilor.
- Îmbunătățirea reproductibilității și fiabilității cercetării.
- Atragerea de noi parteneriate cu cercetători, afaceri, politici și comunități mai largi.
- Permitearea răspunsurilor la noile întrebări de cercetare.

## **4. Ce trebuie să se întreprindă pentru ca datele de cercetare să devină FAIR?**

Există niveluri diferite de a face ca datele să corespundă principiilor FAIR. Vom menționa că, în virtutea anumitor circumstanțe, nu întotdeauna este posibil ca datele de cercetare să fie conforme FAIR, dar aplicarea unor principii asupra datelor va ajuta la regăsirea, accesibilitatea, interoperabilitatea și reutilizarea datelor de cercetare.

La crearea datelor FAIR, un rol important le revine metadatelor. Datele trebuie să fie însoțite de metadata. Metadatale sunt esențiale pentru a regăsi, a reutiliza și a gestiona datele, dar și pentru a înțelege mai bine contextul datelor și al seturilor de date. Cu ajutorul metadatelor poate fi descris cine este cercetătorul responsabil de date, când, unde, cine și de ce a colectat datele respective, care este modul de citare a datelor cercetării etc. Conținutul și formatul metadatelor trebuie să fie ghidate de specificul unei anumite discipline și / sau un depozit specializat care utilizează standarde de metadata adecvate acestei discipline.

Cadrul normativ al e-Infrastructurii de date de cercetare va trebui să ofere directoare de standarde pentru metadata din diverse domenii și discipline științifice. Acestea vor trebui, de asemenea, să ofere instrumente cu ajutorul cărora să poată fi căutate standarde specifice disciplinei precum și instrumente asociate acestora.

## **5. Date regășibile**

Primul pas în (re)utilizarea datelor este regăsirea lor. Metadatale și datele trebuie să fie ușor de regăsit atât pentru oameni, cât și pentru mașină. Metadatale care pot fi citite de mașini sunt esențiale pentru descoperirea automată a seturilor de date și a serviciilor de date.

Pentru ca datele să fie regășibile trebuie să se asigure că:

- Datele sunt descrise cu metadata îmbogățite;
- (Meta)datelor le sunt atribuite anumiți identificatori unici și persistenți la nivel global (de exemplu, un DOI).
- (Meta)datele sunt înregistrate sau indexate într-o resursă căutabilă.

Depozitele (pe discipline științifice) vor atribui un identificator persistent la arhivarea unui set de date.

Se va utiliza un depozit de încredere și adecvat și autorizat pentru publicarea datelor.

Se recomandă înregistrarea ORCID pentru obținerea unui identificator personal persistent care să poată fi folosit pentru publicații și pentru date.

Se vor adăuga metadate bogate (descrierea contextului, calitatea, starea și caracteristicile setului de date).

Cu cât sunt mai elaborate informațiile despre context, conținut și caracteristici ale datelor, cu atât acestea vor putea fi regăsite mai ușor.

## 6. Date accesibile

Trebuie să fie posibil atât pentru oameni cât și pentru mașină să poată să obțină acces la datele cercetărilor în condiții specifice sau de restricții, după caz. FAIR nu înseamnă neapărat că datele trebuie să fie deschise. În cazurile în care datele nu pot fi făcute accesibile în mod deschis, trebuie să se asigure ca metadatele să fie disponibile publicului.

Pentru ca datele să poată fi regăsite trebuie să fi asigurate condiții precum:

- Depozitul pe care îl utilizăm pentru a partaja datele cercetărilor trebuie să atribuie datelor identificatori persistenti prin care acestea să poată fi preluate;
- Procedura de acces la date trebuie să includă autentificarea și autorizarea, dacă este necesar;
- Metadatele trebuie să fie accesibile, oricând este posibil, chiar dacă datele nu sunt în acces liber.

*Recomandări practice:*

- Depozitele de date trebuie să poată fi utilizate atât pentru publicarea și acordarea accesului la datele cercetărilor cât și pentru descărcare publică, ca date deschise, sau pentru ca datele să fie disponibile la cerere sau cu acces restricționat;
- Atunci când cercetarea este în desfășurare, depozitele de date vor permite să fie acordat accesul la datele părților externe, astfel, încât în proiectele multinstituționale sau multidisciplinare să fie posibil schimbul de date și colaborarea;
- În cazul în care accesul la date este restricționat, se va asigura că sunt furnizate suficiente informații de contact pentru alți cercetători, dacă aceștia doresc să acceseze datele (e-mail personal, contacte ale managerului de proiect sau al administratorului de date etc.);

Formatele în care vor fi stocate datele de cercetare vor avea un rol important în asigurarea accesibilității. Se recomandă să fie utilizate formate deschise, neproprietate sau comune pentru a contribui la sporirea accesibilității.

De asemenea, vor fi importante instrumentele software care asigură accesul la date. Dacă este necesar, acestea trebuie să fie documentate în modul corespunzător.



## 7. Interoperabilitatea datelor

Pentru a îmbunătăți procesul de cercetare și a face posibilă descoperirea de noi informații, datele de cercetare trebuie să fie ușor combinate cu alte seturi de date, cu aplicații și cu fluxuri de lucru atât de către oameni cât și de către sisteme informatice care funcționează în cadrul e-Infrastructurii de date de cercetare.

Pentru ca datele să devină interoperabile pentru utilizare, se recomandă să se întreprindă următoarele:

- Atunci când este posibil, să fie utilizate formate și software binecunoscute și, preferabil, deschise;
- Să fie utilizate standarde relevante pentru metadate;
- În măsura în care este posibil, să fie utilizate scheme convenite de comunitate, vocabulare controlate, cuvinte cheie, tezaure sau ontologii.

*Recomandări practice.* În cazurile când nu sunt utilizate standarde cunoscute pentru organizarea datelor, se recomandă să fie create fișiere README (.txt sau .pdf) pentru a ne asigura că datele cercetărilor pot fi interpretate și reanalizate corect de către alții. De obicei, un fișier README conține următoarele informații:

- pentru fiecare nume de fișier, o scurtă descriere a datelor pe care le include, descriind opțional relația cu tabelele, cifrele sau secțiunile din publicația însoțitoare;
- pentru date tabulare: definiții ale rubricilor de coloană și a etichetelor de rând; coduri de date (inclusiv date care lipsesc); și unități de măsură;
- orice etapă de prelucrare a datelor, în special dacă nu este descrisă în publicație, care poate afecta interpretarea rezultatelor;
- o descriere a seturilor de date asociate stocate în altă parte, dacă este cazul;
- pe cine să contactăm atunci când sunt întrebări.
- Se atașează scripturile de programe care s-au folosit pentru a analiza sau a aduna datele.
- Se adnotează corect scripturile de programare pentru ca alții să-l înțeleagă.
- Se folosește consecvența în numele fișierelor, a variabilelor de date, scripturilor, variabilelor scripturilor și a adnotărilor similare.

## 8. Date Reutilizabile

Datele cercetărilor ar trebui să fie organizate astfel încât să poată fi utilizate în cercetările viitoare și în procesarea datelor acelor cercetări, asigurând condițiile necesare ca descoperirile (rezultatele

cercetărilor) să poată fi reproduse și că noile cercetări se vor putea baza efectiv pe rezultatele deja obținute.

Pentru ca datele să fie reutilizabile, trebuie să se aigure că:

- Datele sunt bine documentate pentru a sprijini interpretarea lor corectă;
- Există o licență clară și accesibilă de utilizare a datelor, astfel încât alții să poată să știe ce tipuri de reutilizări sunt posibile;
- Datele sunt însoțite de informații de proveniență pentru a clarifica cum, de ce și de cine au fost create și procesate;
- Datele și metadatele respectă standardele relevante pentru domeniul de cercetare în cauză.

*Recomandări practice.* Documentarea trebuie furnizată pe trei niveluri:

- Documentația la nivel de proiect explică obiectivele studiului, ipoteza din spatele acestuia, instrumentele și metodologia;
- Documentația la nivel de fișier explică modul în care toate fișierele care alcătuiesc un set de date se raportează între ele;
- Documentația la nivel de element explică numele variabilelor și semnificațiile acestor variabile;
- Datele se vor furniza cu o licență clară pentru a putea stabili condițiile de reutilizare a acestora. Licențele utilizate în mod obișnuit, cum ar fi Creative Commons (CC) sau MIT, pot fi conectate la date sau la software-ul cu ajutorul căruia acestea au fost produse sau obținute.

Liniile directoare din cadrul Programului Orizont 2020 recomandă CC-0 sau CC-BY ca o modalitate simplă și eficientă de a face posibilă exploatarea și reproducerea datelor pentru alții.

De obicei, în cadrul depozitelor de date, o licență de utilizare a datelor face parte din metadate.

Aceste indicații trebuie urmate atunci când se procedează la elaborarea Planului de management al datelor.

## Referințe:

1. H2020 Guidelines on FAIR Data Management: Version 3.0 [online]. EUROPEAN COMMISSION, Directorate-General for Research & Innovation, 26 July 2016 [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).
2. CUCIUREANU, Gheorghe, COJOCARU, Igor, COJOCARU, Irina, ȚURCAN, Nelly, GRECU, Mihai, ROȘCA, Alfreda, RUSU, Andrei, CUJBA, Rodica, COȘULEANU, Ion, UNGUREANU, Elena, MINCIUNĂ, Vitalie. Știința deschisă în Republica Moldova: Studiu = Open science in the Republic of Moldova: Study. Chișinău: Institutul de Dezvoltare a Societății Informaționale, 2018, 264 p. ISBN 978-9975-3220-3-4, eISBN 978-9975-3220-4-1. [https://zenodo.org/record/1468418#.W\\_7EmegzbIV](https://zenodo.org/record/1468418#.W_7EmegzbIV).
3. WILKINSON, Mark D. et al. The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data* [online]. 2016, nr. 3, article nr. 160018 [citat 04.03.2018]. Disponibil: <https://www.nature.com/articles/sdata201618>.

### Principiile FAIR

(După *Știința deschisă în Republica Moldova: Studiu* = Open science in the Republic of Moldova: Study. Chișinău: Institutul de Dezvoltare a Societății Informaționale, 2018 [2])

În ecosistemul eScience (eȘtiință), provocarea de a facilita o utilizare optimă a datelor și a metodelor de cercetare este una complexă, existând multiple părți interesate:

- *Cercetătorii* care doresc să partajeze datele și interpretările lor;
- *Editorii de date profesionale* care oferă servicii, creatorii de software și instrumente care oferă servicii de analiză și prelucrare a datelor;
- *Agențiile finanțatoare* (private și publice) care se ocupă tot mai mult de gestionarea corectă a datelor (Data Stewardship); și comunitatea (pentru) date științifice care se ocupă de căutarea (mining), integrarea și analizarea rezultatelor pentru a promova descoperirile științifice. Analiza computațională pentru a descoperi modele semnificative în seturile de date masive interconectate devine rapid o activitate de cercetare de rutină.

În 2014, la inițiativa Netherlands eScience Center și Dutch Techcentre for the Life Sciences (DTL), în cadrul unei reuniuni a mai multor părți interesate privind consolidarea ecosistemului eScience, desfășurate la Lorentz Center din Leiden, Țările de Jos, s-a constatat că, prin definirea și promovarea la scară largă a unui set minim de principii și practici directe convenite de comunitatea furnizorilor și a consumatorilor de date – atât mașini cât și oameni – ar putea fi mai ușor descoperite, accesate, interoperabile și sensibil reutilizate, cu o citare corectă, cantități mari de informații generate de știința contemporană bazată pe date intensive. Aceste principii și practici simple trebuie să permită o gamă largă de comportamente integrative și exploratorii și să sprijine o gamă largă de opțiuni și implementări tehnologice, la fel cum Internet Protocol-ul (IP) a oferit un nivel minim – facilități pentru crearea unei game largi de instrumente de furnizare, consum și vizualizare a datelor pe Internet.

Principiile FAIR (FAIR Principles) sunt menite să ghideze implementatorii de medii de date FAIR pentru a verifica dacă alegerile lor specifice de punere în aplicare dau efectiv datele rezultate FAIR.

S-a convenit asupra a patru deziderate (FAIR Data Guiding Principles) pe care un mediu modern de publicare a datelor ar trebui să le ofere pentru a sprijini stocarea, explorarea, partajarea și utilizarea atât automată, cât și manuală a datelor:

- Datele trebuie să fie regăsibile (Findable),

- Datele trebuie să fie accesibile (Accessible),
- Datele trebuie să fie interoperabile (Interoperable),
  - Datele trebuie să fie reutilizabile (Re-usable).

Aceste aspecte ale FAIR sunt în mod evident legate între ele, dar, din punct de vedere tehnic, sunt oarecum independente una de alta și pot fi implementate incremental în orice combinație, sporind gradul de acceptare a principiilor FAIR.

Prin adoptarea principiilor FAIR, obiectele de date devin pe deplin regășibile, accesibile, interoperabile și reutilizabile [3].

Descrierea principiile FAIR privind datele:

1. Pentru a fi regășibil (a putea fi găsit - Findable) un obiect de date trebuie să fie identificabil în mod unic și persistent.
  - 1.1. Același obiect de date trebuie să fie regășibil în orice moment, precum și să fie persistent, accentul punându-se pe metadatele acestuia.
  - 1.2. Un obiect de date trebuie să conțină un minimum de metadate de bază prin care poate fi distins de alte obiecte de date.
  - 1.3. Identificatorii pentru orice concept utilizat în obiectul de date trebuie să fie unici și persistenți.
2. Datele sunt accesibile prin faptul că pot fi obținute întotdeauna de mașină sau/și de oameni:
  - 2.1. În urma autorizării corespunzătoare.
  - 2.2. Prin intermediul unui protocol bine definit.
  - 2.3. Mașinile și oamenii trebuie să poată să aprecieze accesibilitatea reală a fiecărui obiect de date.
3. Obiectele de date pot fi interoperabile doar dacă:
  - 1.1. (Meta)datele pot fi acționate de către mașină.
  - 1.2. Formatele de (meta)date utilizează vocabulare și/sau ontologii comune.
  - 1.3. (Meta)datele din obiectul de date ar trebui să fie atât lizibile sintactic, cât și accesibile semantic mașinilor.
4. Pentru ca obiectele de date să fie reutilizabile, criteriile suplimentare sunt:
  - 4.1. Obiectele de date ar trebui să respecte principiile 1-3 – să fie regășibile, accesibile și interoperabile.
  - 4.2. (Meta)datele trebuie să fie suficient de bine descrise și bogate astfel încât să poată fi conectate sau integrate automat (sau cu un efort uman minim), la fel ca și alte surse de date.
  - 4.3. Obiectele de date publicate trebuie să facă referință la sursele lor cu metadate suficient de bogate și la proveniența lor pentru a permite o citare corectă.