

BENEFICIILE DATELOR DE CERCETARE DESCHISE

Viorica LUPU

Pe măsură ce societatea devine mai digitizată, cercetarea științifică utilizează și produce în mod progresiv mai multe date. Această rată de creștere nu este de așteptat să încetinească, deoarece se estimează că numărul de date generate va crește de 29 de ori până în 2020. Acest lucru a adus necesitatea unei concentrări mai accentuate asupra gestionării datelor și, în special, pentru ca toate datele să fie mai deschise și mai ușor accesibile. În acest context, procesele eficiente de gestionare a datelor din cercetare sunt esențiale pentru a permite rezultate mai transparente și mai accesibile ale cercetării.

Gestionarea datelor de cercetare este un subiect larg discutat, care câștigă interesul cercetătorilor, instituțiilor și finanțatorilor deopotrivă. Este considerat a fi un aspect important al cercetării academice și central al reproductibilității și integrității științifice. Importanța și rolul datelor de cercetare crește în contextul proiectului Orizont 2020 finanțat de Uniunea Europeană, care recomandă ca datele să fie „deschise atât cât este posibil, închise atât cât este necesar”, acest principiu bazându-se pe principiile datelor FAIR, care afirmă că datele de cercetare ar trebui să fie detectabile, accesibile, interoperabile și reutilizabile [6, p. 3].

Similar cu alte mărfuri globale, datele de cercetare au un potențial semnificativ de a oferi beneficii. În prezent, se spune că „datele sunt noul petrol” [1, p. 1], deoarece atât datele cât și petrolul au o valoare intrinsecă, ambele trebuie „rafinat” sau transformate în alt mod pentru a-și realiza potențialul.

Din ce în ce mai mulți cercetători, instituții și organisme internaționale menționează beneficiile datelor de cercetare deschise.

Datele deschise încurajează transparența și controlul democratic prin încurajarea participării, facilitând autonomia de sine și oferind dispoziții pentru îmbunătățirea eficienței și eficacității serviciilor guvernamentale și crearea de noi cunoștințe din seturile de date existente [7, p. 142].

Datele de mult timp au fost piatra de temelie a științei [11, p. 1]. Prin urmare, abilitatea de a împărtăși, reutiliza și combina datele oferă oamenilor de știință o multitudine de oportunități: reanalizarea dovezilor, verificarea rezultatelor, dublarea minimă a efortului și inovarea accelerată [10, p. 704].

Beneficiile datelor deschise sunt diverse și variază de la îmbunătățirea eficienței administrației publice și creșterea economică în sectorul privat până la bunăstarea socială în sens mai larg [3, p. 11]. Valoarea economică a datelor de cercetare devine din ce în ce mai evidentă. Potrivit portalului european de date, numai în cadrul Uniunii Europene, mărimea pieței deschise de date în 2016 se ridică la 55,3 mld. de euro. Se estimează că mărimea pieței va crește cu până la 37% în 2020 – ajungând la 325 mld. de euro (cu corecția pentru inflație) [5, p. 8].

Datele deschise sunt valoroase și pentru o colaborare eficientă între cetățeni și administrația guvernamentală. Ele facilitează descoperirea și accesul ministerelor guvernamentale atât la datele proprii, cât și la datele altor ministere și organizații, reduc costurile de achiziție, redundanță și cheltuieli generale. Datele deschise facilitează monitorizarea activităților guvernamentale, cum ar fi monitorizarea cheltuielilor bugetare publice și a impactului, încurajează participarea mai mare a cetățenilor la afacerile guvernamentale și sprijină societățile democratice.

Comunicarea deschisă a rezultatelor cercetării aduce beneficii societății prin obținerea mai rapidă și mai largă a informațiilor de către cercetători, practicieni, studenți și factori de decizie politică. Acest lucru accelerează ritmul descoperirii și reduce decalajele de diseminare a informațiilor [2].

În mod crucial, atunci când datele sunt disponibile în mod deschis, acestea pot fi utilizate în mai multe moduri: ele pot fi combinate cu alte seturi de date, utilizate pentru a aborda întrebări care nu au fost gândite de autorii studiilor inițiale, analizate prin metode statistice noi

care nu erau disponibile în momentul publicării sau utilizate ca set de date independente de replicare.

Accesul la date în multe domenii stimulează, de asemenea, noi tipuri de gândire, deoarece cercetătorii dezvoltă noi înțelegeri prin reunirea datelor dintr-o varietate de surse. Acest lucru permite noi perspective asupra problemelor multidisciplinare într-o mare varietate de domenii, permite obținerea celor mai interesante cunoștințe.

Avantaje importante aduc datele deschise și finanțatorilor de cercetare. Un număr din ce în ce mai mare de finanțatori de cercetare solicită sau recomandă ca atât rezultatele cercetării, cât și datele cercetării produse ca parte a unui proiect finanțat să fie făcute publice.

Multe reviste au recunoscut ineficiența cercetărilor publicate care nu sunt susținute de datele care stau la baza acestora și au făcut obligatorie transmiterea datelor de cercetare de susținere împreună cu manuscrisele [8, p.1]. Astfel, ei consideră că publicarea datelor de cercetare este un mijloc de îmbogățire a unei publicații de text interpretativ (de obicei, un articol în revistă), contribuie la o mai bună înțelegere a cercetării și oferă mai multă eficiență în utilizare a resurselor științifice. Unele reviste, inclusiv „Nature Research”, obligă autorii să pună la dispoziția cititorilor fără restricții materiale, date, coduri și protocoale calificate, astfel încât autorii să poată reproduce și să se bazeze pe afirmațiile publicate [9, p. 1].

Un beneficiu al datelor de cercetare deschise, ce pare la prima vedere foarte simplu, este buna păstrare a datelor, menținerea integrității datelor, adică a exactității și coerenței (valabilității) datelor în timpul ciclului lor de viață. De fiecare dată când datele sunt replicate sau transferate, acestea trebuie să rămână intacte și neschimbate între actualizări. Integritatea datelor poate fi compromisă în mai multe moduri: erori umane, amenințări cibernetice (viruși, hacking etc.), erori de transfer, inclusiv modificări neintenționate sau compromiterea datelor în timpul transferului de la un dispozitiv la altul etc. Utilizarea unor facilități robuste și adecvate de stocare a datelor contribuie la reducerea pierderii datelor de cercetare prin accidentare sau prin neglijență. Locul potrivit pentru depozitarea datelor de cercetare este depozitul instituțional digital de date sau,

eventual, un depozit disciplinar, care poate fi identificat prin a apela la ajutorul bibliotecii sau utilizând o listă extinsă în registrul re3data.org al depozitelor de date de cercetare. Majoritatea depozitelor digitale utilizează servere profesionale care oferă backup-uri continue ale datelor stocate și nu există niciun motiv pentru pierderea datelor.

Un aspect important în partajarea datelor îl reprezintă citarea datelor. Cunoașterea impactului datelor de cercetare este destul de importantă pentru cercetători, în special pentru avansarea în carieră, pentru aplicațiile, proiectele de finanțare și pentru a demonstra amploarea și importanța activității sale. Referințele sunt o măsură bine stabilită a impactului cercetării; o citare poate însemna recunoașterea sau validarea datelor de cercetare de către alți specialiști.

Totodată, tot mai mulți editori de reviste solicită citări de date, departamentele specializate impun citarea datelor ca o practică fundamentală a unei bune cercetări, agențiile acordă premii pentru citările de date și recunosc citările de date în evaluările anuale ale performanței cercetătorilor și asumarea responsabilității de către cercetători pentru citarea surselor de date. Citarea datelor este acum recunoscută drept una dintre practicile-cheie care duce la recunoașterea datelor ca rezultat principal al cercetării și reprezintă aceeași practică în care cercetătorii oferă în mod obișnuit o referință bibliografică la articole de revistă, cărți, rapoarte etc. Citările de date oferă o modalitate de a urmări, înregistra și raporta asupra transmierii și reutilizării datelor și asigurarea că cercetătorii obțin recunoaștere științifică pentru munca lor.

În anul 2014 au fost lansate Principiile de citare a datelor FORCE11 [4, p. 1] cu scopul de a face ca datele de cercetare să devină o parte integrantă a practicii academice. Principiile au recunoscut că un factor esențial pentru creșterea disponibilității datelor de cercetare a fost să asigure autorii să primească recunoaștere științifică pentru partajare prin citarea corespunzătoare a datelor de cercetare.

În concluzie, menționăm că datele de cercetare sunt o resursă valoroasă care necesită adesea o cantitate mare de timp și bani pentru a le crea, organiza, depozita și păstra. Există însă o serie de motive foarte bune pentru care datele de cercetare ar trebui gestionate în timp

util și adecvat: pentru a asigura integritatea cercetării și validarea rezultatelor, a facilita securitatea datelor și a minimiza riscul de pierdere a datelor, a asigura o diseminare mai largă și un impact crescut, a asigura respectarea cerințelor agențiilor de finanțare etc.

Referințe:

1. AISHAH, Selamat (2018). How can Open Data benefit business?. [Accesat 26.08.2019]. Disponibil: <http://blog.ukdataservice.ac.uk/how-can-open-data-benefit-business/>
2. BEAGRIE, Neil, HOUGHTON, John (2014). The Value and Impact of Data Curation and Sharing: A synthesis of three recent studies of UK research data centers. JISC (March 2014). Available: http://repository.jisc.ac.uk/5568/1/iDF308_Digital_Infrastructure_Directions_Report-%2C_Jan14_v1-04.pdf.
3. CODATA (2015). The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations. [Accesat 09.09.2019]. Disponibil: <http://doi.org/10.5281/zenodo.33830>
4. DATA CITATION SYNTHESIS GROUP. Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014. [Accesat 09.09.2019]. Disponibil: <https://www.force11.org/group/joint-declaration-data-citation-principles-final>.
5. EUROPEAN COMMISSION (2015). Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources. Luxemburg, 2015. 112 p. <https://www.europeandataportal.eu/en/highlights/creating-value-through-open-data>. ISBN 978-92-79-52791-3. DOI 10.2759/328101.
6. EUROPEAN COMMISSION (2016). H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. 12 p. [Accesat 25.08.2019]. Available: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
7. KHAN, N.A., AHANGAR, H. (2017). Emerging trends in open research data," *2017 9th International Conference on Information and Knowledge Technology (IKT)*, Tehran, 2017, pp. 141-146. doi: 10.1109/IKT.-2017.8258631
8. KULKARNI, Sneha (2017). To share or not to share: What motivates researchers to share their data? <https://www.editage.com/insights/to-share-or-not-to-share-what-motivates-researchers-to-share-their-data>
9. NATURE RESEARCH (2019). Reporting standards and availability of data, materials, code and protocols. [Accesat 09.09.2019]. Disponibil:

<https://www.nature.com/nature-research/editorial-policies/reporting-standards#top>.

10. REICHMAN, O.J., JONES, M.B., SCHILDHAUER, M.P. (2011). Challenges and Opportunities of Open Data in Ecology. In: *Science*, vol. 331, pp. 703-705.
11. WALLIS, J.C., ROLANDO, E., BORGMAN, C.L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. In: *PLoS ONE*, vol. 8(7): e67332. [Accesat 25.08.2019]. Disponibil: <https://doi.org/10.1371/journal.pone.0067332>.