

CZU: 001.891

DOI: [10.5281/zenodo.3886720](https://doi.org/10.5281/zenodo.3886720)**MODELE ALE CICLULUI DE VIAȚĂ AL DATELOR DE CERCETARE***Viorica LUPU, Nelly ȚURCAN**Universitatea de Stat din Moldova*

În lucrare sunt analizate și prezentate diferite modele ale ciclului de viață al datelor de cercetare. Un model de ciclu de viață al datelor ar putea include crearea de date (conceptualizare), colectarea și descrierea datelor, stocarea datelor, arhivarea și conservarea, accesul la date, descoperirea, analiza, reutilizarea și transformarea datelor. Modelele pot ajuta instituțiile să articuleze și să perfecționeze fluxurile activității de cercetare, precum și să dezvolte servicii de asistență care să permită o gamă largă de părți interesate implicate pentru a răspunde cerințelor de gestionare și custodie a datelor.

Cuvinte-cheie: *modele, ciclu de viață, date de cercetare.*

RESEARCH DATA LIFE CYCLE MODELS

In the article, different models of the life cycle of the research data are analyzed and presented. A data lifecycle model could include data creation (conceptualization), data collection and description, data storage, archiving and preservation, data access, discovery, analysis, reuse and data transformation. These models can help institutions articulate and refine the flows of research, as well as develop support services that allow a wide range of stakeholders to respond to data management and custody requirements.

Keywords: *models, life cycle, research data.*

Introducere

În condițiile creșterii schimbului de date la nivel global, ciclul de viață al datelor devine un factor din ce în ce mai important în activitatea științifică a cercetătorului, extinzând parcursul tipic de cercetare. Creșterea continuă a volumului și diversității de date adaugă inevitabil complexitate ridicată și dificultăți severe în toate etapele de existență a datelor.

Un bun management al datelor este fundamental pentru rezultate de cercetare de înaltă calitate. Acesta presupune, în principal, facilitarea partajării datelor, asigurarea durabilității și accesibilității datelor pe termen lung, adică posibilitatea reutilizării lor de către viitorii cercetători.

Dacă datele de cercetare sunt bine organizate, documentate, păstrate și accesibile, iar exactitatea și validitatea lor poate fi controlată în orice moment, rezultatele pe termen lung se materializează în date de înaltă calitate, cercetări eficiente, constatări bazate pe dovezi solide și economie de timp și resurse. Iată de ce este necesar ca instituțiile academice să se concentreze pe dezvoltarea infrastructurilor de partajare a datelor, pe sprijinirea cercetătorilor în gestionarea și diseminarea datelor prin instrumente, îndrumări practice și instruire, pe conectarea datelor de cercetare cu publicațiile pentru a crește vizibilitatea, accesibilitatea datelor și a cercetării în sine.

Bibliotecile au oferit dintotdeauna acces la publicațiile tipărite, însă astăzi ele au devenit mai interesate în a ajuta savanții și cercetătorii la transformarea comunicării științifice, oferind alternative pentru publicarea, diseminarea și păstrarea rezultatelor cu scopul de a încuraja dezvoltarea ideilor prin schimburi de informații deschise și gratuite. Ne-am propus studierea unor modele după care se derulează, în general, procesul de cercetare și, în special, ciclul de viață al datelor de cercetare, ca mai apoi, în baza acestor modele, să identificăm rolul bibliotecilor academice în gestionarea datelor de cercetare.

Analiza modelelor ciclului de viață al datelor de cercetare

În vederea realizării studiului au fost colectate informații privind modelele publicate în articole de specialitate recenzate, dar și privind unele modele puse în discuție de practicieni. Peisajul datelor de cercetare poate fi deconstruit în mai multe moduri, fiecare dintre acestea producând o perspectivă valoroasă [1].

„Modelele ciclului de viață conturează felul în care studiem procesele informaționale digitale. Aceste modele reprezintă cursul vieții unui sistem mai mare, cum ar fi procesul de cercetare, printr-o serie de etape sau faze legate în mod secvențial în care informațiile sunt produse sau manipulate” [2]. Modelele „ajută la definierea și ilustrarea vizuală a acestor procese complexe, făcând mai ușoară identificarea părților componente sau a etapelor distincte ale datelor de cercetare” și a persoanelor sau entităților responsabile [3].

Într-o gamă largă de domenii de date, mulți cercetători din mediul academic și din industrie propun diferite modele pentru a gestiona viața datelor ca active valorizate [4].

Modelul ciclului de viață al datelor definește un cadru la nivel înalt și reprezintă o imagine de ansamblu a vieții datelor de la stadiul de producție până la stadiul de consum. Scopul unui astfel de model este de a optimiza gestionarea datelor, de la o organizare eficientă la eliminarea oricărui tip de deșeurii, pentru a oferi produse de date adecvate pentru utilizare de către utilizatorii finali care corespund cerințelor de calitate preconizate. Un model simplu de ciclu de viață al datelor ar putea fi reprezentat astfel [4]:

- Achiziție: Procesul de colectare, înregistrare sau generare a unor date;
- Custodie: Activitatea de gestionare a utilizării datelor din punctul de creare cu scopul de a se asigura că acestea sunt disponibile pentru descoperire și reutilizare în viitor;
- Conservare: Procesul de păstrare a datelor într-o formă fizică pentru utilizarea viitoare.

Pornind de la acest model de bază au fost elaborate un șir de modele ale ciclului de viață al datelor, fiecare cu diferită focalizare sau perspectivă. Acestea pot fi clasificate în funcție de forma de vizualizare ori în funcție de creatorul sau utilizatorul modelului. După forma de vizualizare, deosebim modele liniare, circulare și neliniare. Un exemplu de model de tip liniar este modelul pentru datele științifice USGS (United States Geological Survey) (Fig.1).

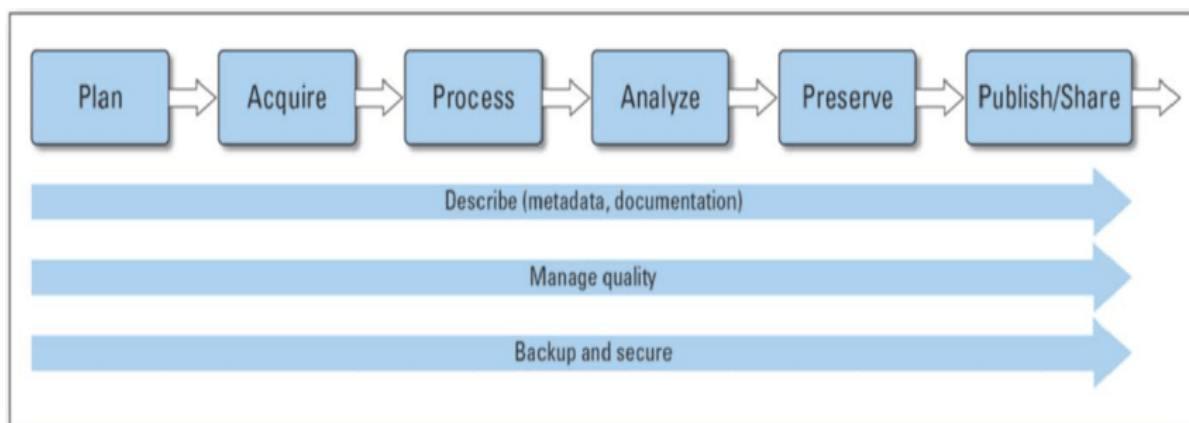


Fig.1. Modelul USGS pentru datele științifice.

Sursa: FAUNDEEN, J.L. ș.a. [5].

Acest model descrie ciclul de viață al datelor din perspectiva proiectelor de cercetare și a activităților care sunt efectuate în faze, de exemplu planificarea, colectarea, prelucrarea, analiza, conservarea, publicarea și diseminarea datelor pentru reutilizare [5, 6]. Pe lângă acestea, există și alte activități care trebuie să fie efectuate continuu în toate fazele ciclului de viață, cum ar fi documentarea procesului de lucru și furnizarea de metadata, asigurarea calității datelor, precum și backup-ul datelor pentru a împiedica pierderea lor fizică [7, 8].

Există, de asemenea, modele circulare, care își propun să reflecte natura iterativă a procesului de cercetare. Modelele circulare par mai potrivite pentru a descrie practicile actuale de cercetare, bazate tot mai mult pe partajarea și reutilizarea datelor.

Pornind de la criteriul de clasificare în funcție de creatorul sau utilizatorul datelor de cercetare, Jake Carlson [3] descrie trei tipuri diferite de modele ale ciclului de viață al datelor bazate pe individ, organizație și comunitate. Modelele bazate pe individ se constituie pe proiecte specifice și, adesea, nu sunt la un nivel abstract, ci conțin informații detaliate legate de proiectul de cercetare. Astfel de modele individuale pot fi utile în elaborarea unui plan de gestionare a datelor pentru un proiect specific. Modelele bazate pe organizație sunt produse de organizațiile care oferă servicii sau asistență cercetătorilor în gestionarea datelor. Aceste organizații includ universități, biblioteci, depozite de date, editori etc. În comparație cu modelele individuale, cele bazate pe organizație mai mult generalizează diferitele faze ale ciclului de viață al datelor, întrucât acestea nu sunt concentrate pe un proiect specific. Un exemplu în acest sens servește modelul Universității din Oxford [9, 10], care nu este organizat doar ținându-se cont de procesul de cercetare, ci și în baza serviciilor pe care organizația le poate oferi cercetătorilor în diferitele etape ale ciclului de viață al datelor (Fig.2).

Modelul include furnizarea accesului la servicii și facilități pentru stocarea, securitatea și păstrarea datelor și înregistrărilor, cea ce permite cercetătorilor să-și îndeplinească cerințele față de finanțatorii de cercetare, de asemenea presupune accesul la instruire, sprijin și consultanță în gestionarea datelor și a documentelor de cercetare, precum și oferirea resurselor necesare unităților operaționale însărcinate cu furnizarea acestor servicii, facilități și instruirii. Sprijinul oferit include și o listă de verificare a planificării gestionării datelor.



Fig.2. Modelul de management al datelor de cercetare al Universității din Oxford.

Sursa: WILSON, James A.J., JEFFREYS, Paul W. [9].

Modelele de tipul al treilea sunt cele bazate pe comunitate. Ele au fost dezvoltate pentru a sprijini nevoile unei anumite comunități de cercetare și pentru a transmite cele mai bune recomandări și practici într-un mod care să ducă la o înțelegere și adoptare cât mai eficientă a acestora în comunitatea interesată [3]. Un exemplu de model de ciclu de viață al datelor bazat pe comunitate este Digital Curation Centre Lifecycle Model [2, 11-14], ce oferă o imagine grafică de ansamblu a etapelor ciclului de viață [14] pentru gestionarea cu succes a datelor în formă digitală. Distribuția etapelor din acest model este destul de sofisticată, ele fiind plasate într-o structură ciclică cu mai multe straturi [15] (Fig.3).

În ceea ce privește domeniul de aplicare, spre deosebire de modelele prezentate anterior, modelul propus de DCC este preocupat în mod exclusiv de custodia datelor [16]. Nu cuprinde toate etapele procesului de cercetare, cum ar fi generarea de idei. Conceptualizarea, ca punct de plecare pentru crearea de date, este stabilită simbolic în afara ciclului principal [17]. Modelul identifică, în schimb, acțiuni de custodie a datelor aplicabile pe întreg ciclul lor de viață, acțiuni secvențiale și acțiuni ocazionale impuse de circumstanțe. Acțiunile complete ale ciclului de viață al datelor sunt împărțite în patru etape: 1) descrierea și reprezentarea informațiilor; 2) planificarea conservării; 3) monitorizarea și participarea comunitară; 4) custodia și conservarea datelor. Acțiunile secvențiale sunt conceptualizarea, crearea sau recepționarea, evaluarea, selectarea, ingerarea, conservarea, stocarea, accesarea, utilizarea, reutilizarea și transformarea. Acțiunile ocazionale includ reevaluarea și migrarea datelor pentru păstrare și securitate. Modelul poate fi utilizat în calitate de instrument de planificare organizațională, este adaptabil la diferite domenii și extensibil pentru a permite planificarea activităților de custodie și conservare a datelor la diferite niveluri [13].

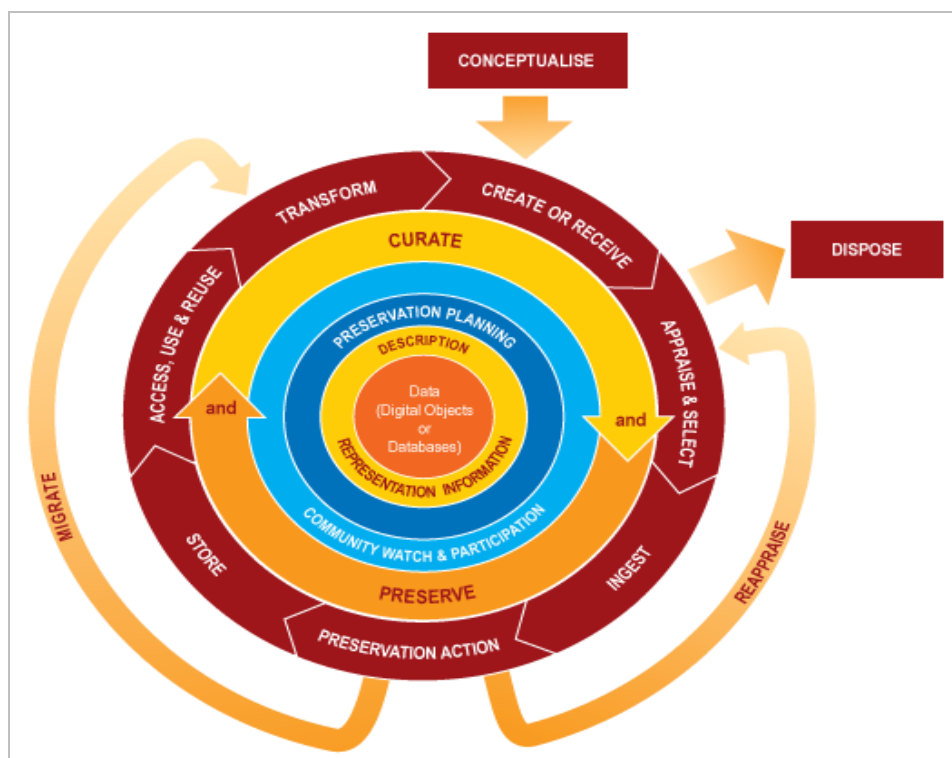


Fig.3. Modelul ciclului de viață propus de DCC Curation.

Sursa: LEWIS, M.J. [15].

Un alt model de acest gen îl reprezintă modelul cadru de capacitate comunitară (CCM), dezvoltat de Biroul Regatului Unit pentru Rețeaua de Biblioteci și Informare (UKOLN) cu scopul de a ajuta finanțatorii, instituțiile și cercetătorii la creșterea capacității comunităților de efectuare a cercetărilor intensive de date [18]. Modelul în discuție cuprinde opt factori de capacitate care includ colaborarea, abilitățile și instruirea, deschiderea, infrastructura tehnică, practicile comune, aspectele economice și de afaceri, juridice și etice, umane, tehnice și de mediu.

Modelul de capacitate comunitară se concentrează pe dezvoltarea capacităților instituționale în abordarea managementului datelor de cercetare, dar nu ține cont în detaliu de custodia datelor. La acest nivel, modelul dat poate fi completat de cel elaborat de Centrul de Custodie a Datelor (DCC), care se axează anume pe custodia datelor, cu toate elementele constituente – colectare, evaluare, descriere, conservare, acces și reutilizare a datelor.

Există mai multe modele ale ciclului de viață al datelor care diferă din perspectiva practicilor din domenii sau comunități. Unul dintre ele este modelul DataONE (Fig.4), dezvoltat de către echipa DataONE Leadership în colaborare cu comunitatea mai largă DataONE și bazat pe modelul prezentat de Fundația Națională a Științei din SUA [19].

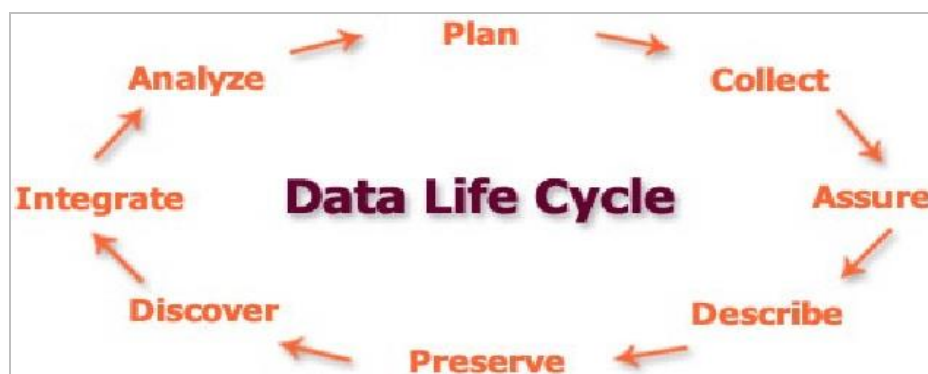


Fig.4. Modelul ciclului de viață al datelor DataONE.

Sursa: STRASSER, C. ș.a. [19].

DataOne a adoptat un model de ciclu de viață al datelor ce permite identificarea fluxurilor de date și a proceselor de lucru pentru oamenii de știință [20]. DataONE definește opt faze din ciclul de viață al datelor științifice: planificarea, colectarea, asigurarea, descrierea, conservarea, descoperirea, integrarea și analiza datelor. Numărul etapelor selectate depinde de tipul de proiect sau de cerințele proiectului [6]. Modelul poate fi utilizat pentru stocarea și preluarea informațiilor spre utilizare pe termen lung, dar nu poate fi privit ca un model complet, deoarece nu se concentrează asupra securității datelor. În schimb, poate servi drept cadru de bază pentru dezvoltarea de instrumente, servicii și materiale educaționale.

Un alt model al ciclului de viață al datelor de cercetare a fost dezvoltat de Institutul Transporturilor din Texas (Fig.5).

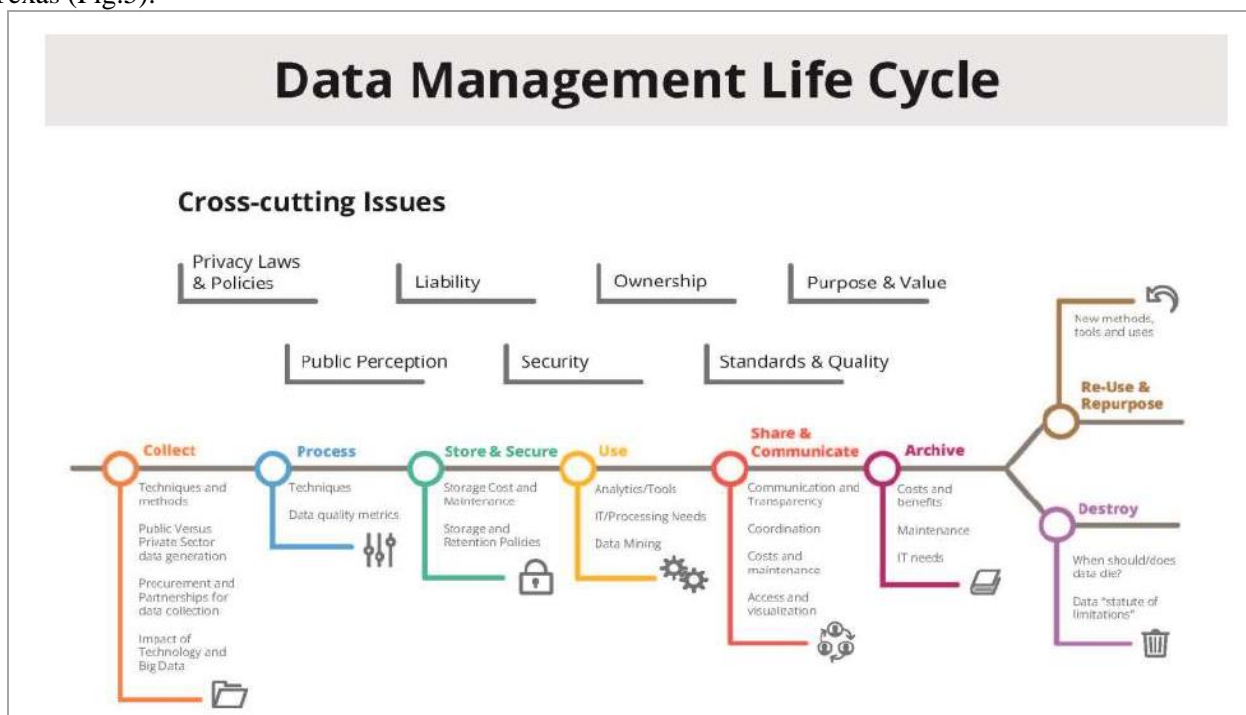


Fig.5. Ciclul de viață al datelor, Institutul Transporturilor din Texas.

Sursa: MILLER, Kristi ș.a. [21].

Acest model prezintă o modalitate de a organiza datele, de a caracteriza natura și valoarea acestora în timp și de a identifica implicațiile politice în problemele transversale de gestionare a datelor [21]. Modelul descrie ciclic și iterativ aspectele-cheie ale datelor pe parcursul tuturor celor șapte etape ale ciclului lor de viață: colectarea, procesarea, stocarea și securizarea, utilizarea, diseminarea și comunicarea, arhivarea, reutilizarea și distrugerea datelor. În același timp, modelul conține și aspecte transversale ce afectează etapele ciclului de viață al datelor: scopul și valoarea, confidențialitatea, proprietatea datelor, răspunderea, percepția publică, securitatea, standardele și calitatea.

Punctele de tranziție ale ciclului de viață sunt de asemenea joncțiuni cruciale pentru alte activități importante în cercetare. Cadrul ciclului de viață poate cuprinde alte evoluții semnificative în managementul informațiilor și cunoștințelor prin fluxurile de activitate care se desfășoară pe etapele cercetării.

Multe fluxuri de activități curg de-a lungul ciclului de viață al cercetării, inclusiv administrarea proiectelor, achiziționarea granturilor, gestionarea datelor, crearea de cunoștințe, judecăți etice, supravegherea proprietății intelectuale și managementul tehnologiei. Repartizarea activităților pe etape necesită coordonare și un sentiment de continuitate în procesul general [2]. Predarea produselor de cercetare între etape implică deseori o mutare a responsabilităților de la o persoană la alta sau de la o organizație la alta. Aceste schimbări ale responsabilității sunt esențiale pentru conceptul de custodie digitală, care reprezintă procesul prin care se negociază acorduri formale între parteneri cu privire la responsabilitatea privată de custodie pentru obiectele digitale de-a lungul întregului ciclu de viață al lor.

Concluzii

Modelele ciclului de viață al datelor oferă o gamă largă de abordări în reprezentarea cercetării și/sau a datelor: de la reprezentări abstracte ale ciclului de viață al cercetării, în ansamblu, până la modele mult mai detaliate, cum ar fi de creare și reutilizare a datelor sau chiar numai de custodie a datelor. Fiecare are atuurile sale relevante pentru scopul particular în care a fost creat. Unele sunt specifice unui anumit domeniu, proiect de cercetare sau unei instituții academice, altele încearcă să generalizeze orice formă de cercetare.

Modelele ciclurilor de viață ale datelor sunt menite să ajute instituțiile să articuleze și să perfecționeze fluxurile de lucru de cercetare și să dezvolte servicii de asistență care să permită o gamă largă de părți interesate implicate pentru a răspunde cerințelor de gestionare și custodie a datelor. De asemenea, aceste modele sunt eficiente la implementarea și susținerea serviciilor de management al datelor de cercetare, precum și la identificarea posibilelor servicii de date și a parteneriatelor construite în sprijinul acestor servicii.

Referințe:

1. LYON, L. The informatics transform: Re-engineering libraries for the data decade. In: *International Journal of Digital Curation*. 2012, vol.7, pp.126-138. eISSN 1746-8256. Disponibil: <http://ijdc.net/index.php/ijdc/article/view/210>
2. HUMPHREY, Ch. *E-Science and the Life Cycle of Research*, 2008. Disponibil: <https://era.library.ualberta.ca/items/3334684b-fa6a-4c9d-a74b-559fec42f9f/view/79b064d6-7b51-4d18-8e4e-3d42b9faa81f/Lifecycle-science060308.pdf>
3. CARLSON, J. The Use of Life Cycle Models in Developing and Supporting Data Services. In: Joyce M Ray (ed). *Research Data Management. Practical Strategies for Information Professionals*. West Lafayette, In: Purdue University Press, 2014, pp.63-86. ISBN 9781557536648.
4. FOX, P. *Data Management Considerations for the Data Life Cycle* [online]. Tetherless World Constellation, 2011 [Accesat: 28.11.2019]. Disponibil: http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_065993.pdf
5. FAUNDEEN, J.L., BURLEY, T.E., CARLINO, J.A., GOVONI, D.L., HENKEL, H.S., HOLL, S.L., HUTCHISON, V.B., MARTÍN, E., MONTGOMERY, E.T., LADINO, C.C., Tessler, Steven, and Zolly, L.S. *The United States Geological Survey Science Data Lifecycle Model: U.S. Geological Survey Open-File Report 2013–1265*, 2013. 4 p. Disponibil: <http://dx.doi.org/10.3133/ofr20131265>
6. MAINDZE, A., SKAF, Z., JENNIONS, I. Towards an Enhanced Data – and Knowledge Management Capability: A Data Life Cycle Model Proposition for Integrated Vehicle Health Management. Annual Conference of the PHMS, Scottsdale, AZ, USA, 21-26 September 2019. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society*. 2019, vol.11, no.1, pp.1-14. ISSN 0000-2015. Disponibil: <https://www.phmpapers.org/index.php/phmconf/article/view/842>
7. ARASS, M. EL, TIKITO, I., SOUSSI, N. Data lifecycles analysis: towards intelligent cycle [online]. In: *IEEE ISCV*. 2017, Apr 2017, [hal-01593851], Fez, Morocco [Accesat: 28.11.2019]. Disponibil: <https://hal.archives-ouvertes.fr/hal-01593851>
8. FAUNDEEN, J.L., HUTCHISON, V.B. The Evolution, Approval and Implementation of the U.S. Geological Survey Science Data Lifecycle Model. In: *Journal of eScience Librarianship*. 2017, vol.6(2), e1117. ISSN 2161-3974. Disponibil: <https://doi.org/10.7191/jeslib.2017.1117>
9. WILSON, J.A.J., JEFFREYS, P.W. Towards a Unified University Infrastructure: The Data Management Roll-Out at the University of Oxford. In: *The International Journal of Digital Curation*. 2013, vol.2(8), pp.235-246. eISSN 1746-8256. Disponibil: <https://doi.org/10.2218/ijdc.v2i8.287>
10. WILSON, J.A.J., MARTINEZ-URIBE, L., FRASER, M.A., JEFFREYS, P. An institutional approach to developing research data management infrastructure. In: *International Journal of Digital Curation*. 2012, vol.6(2), pp.274-287. eISSN 1746-8256. Disponibil: <https://doi.org/10.2218/ijdc.v6i2.203>
11. BALL, A. *Review of Data Management Lifecycle Models* [online]. Bath: University of Bath, 2012 [Accesat: 26.11.2019]. Disponibil: <https://purehost.bath.ac.uk/ws/files/206543/redm1rep120110ab10.pdf>
12. BURGESS, L., JEFFERIES, N., RUMSEY, S., SOUTHALL, J., TOMKINS, D., WILSON, J. A. From compliance to curation. In: *Alexandria: The Journal of National and International Library and Information Issues*. 2016, vol.26(2), pp.107-135. eISSN 2050-4551. Disponibil: <https://doi.org/10.1177/0955749016657482>
13. HIGGINS, S., HARVEY, R., WHYTE, A. *The DCC Curation Lifecycle Model* [online]. Centre of expertise in data curation and preservation, 2008 [Accesat: 26.11.2019]. Disponibil: <https://pdfs.semanticscholar.org/49eb/905fc13f033ebae6b368f220ee3152fd654f.pdf>
14. HIGGINS, S. The DCC curation lifecycle model. In: *The International Journal of Digital Curation*. 2008, vol. 3(1), pp. 134–140. eISSN 1746-8256. Disponibil: <http://www.ijdc.net/article/download/69/48/0>

15. LEWIS, M.J. Libraries and the management of research data. In: McKnight, S. (ed.). *Envisioning future academic library services*. London: Facet Publishing, 2010, pp. 145-168. ISBN 9781856046916. Disponibil: http://eprints.whiterose.ac.uk/11171/1/LEWIS_Chapter_v10.pdf
16. PINFIELD, S, COX, A.M., SMITH, J. Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. In: *PLoS ONE*. 2014, vol.9(12). e-ISSN 114734. Disponibil: <https://doi.org/10.1371/journal.pone.0114734>
17. COX, A.M., TAM, W. A critical analysis of lifecycle models of the research process and research data management. In: *Aslib Journal of Information Management*. 2018, 18 March. ISSN 2050-3806. Disponibil: <https://doi.org/10.1108/AJIM-11-2017-0251>
18. LYON, L. *Exploring the roles and responsibilities of data centres and institutions in curating research data – a preliminary briefing* [online]. UKOLN, 2012 [Accesat: 29.11.2019]. Disponibil: <http://www.ukoln.ac.uk/projects/data-cluster-consultancy/briefing-paper/briefing-final.pdf>
19. STRASSER, C., COOK, R., MICHENER, W., BUDDEN, A. *Primer on Data Management: What you always wanted to know* [online]. CDL, 2012 [Accesat: 24.11.2019]. Disponibil: <https://doi.org/10.5060/D2251G48>
20. POUCHARD, L. Revisiting the Data Lifecycle with Big Data Curation. In: *International Journal of Digital Curation*. 2015, vol. 10, no.2, pp.176-192. eISSN 1746-8256. Disponibil: <https://doi.org/10.102218/ijdc.v10i2.342>
21. MILLER, K., MILLER, M., MORAN, M., DAI, B. *Data Management Life Cycle: Final report* [online]. No. PRC 17-84 F. Texas A&M Transportation Institute, 2018 [Accesat: 24.11.2019]. Disponibil: <https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-84-F.pdf>

Date despre autori:

Viorica LUPU, doctorandă, Școala doctorală Științe Sociale, Universitatea de Stat din Moldova.

E-mail: v.lupu@uasm.md

ORCID: 0000-0002-3905-7499

Nelly ȚURCAN, doctor habilitat, profesor universitar, Facultatea de Jurnalism și Științe ale Comunicării, Universitatea de Stat din Moldova.

E-mail: tsurcannelly@gmail.com

ORCID: 0000-0001-8475-0770

Prezentat la 10.10.2019