

## Роль бібліотек у Data Curation

---

НАЗАРОВЕЦЬ С. А.

Львів, Україна

---

### Анотація

*Розглянуто новий напрямок діяльності сучасних бібліотек, що скерований на зберігання та обслуговування науковими даними. Надано рекомендації щодо впровадження технологій і практик Data Curation у роботу бібліотек вчз.*

**Ключові слова:** *e-наука, відкритий доступ, інформаційне обслуговування, наукові дані, citizen science, Data Curation, FigShare.*

### Abstract

*A new direction in the work of modern libraries, aiming at the storage and maintenance of scientific data is considered. Recommendations on the introduction of technologies and practices Data Curation in the work of university libraries are provided.*

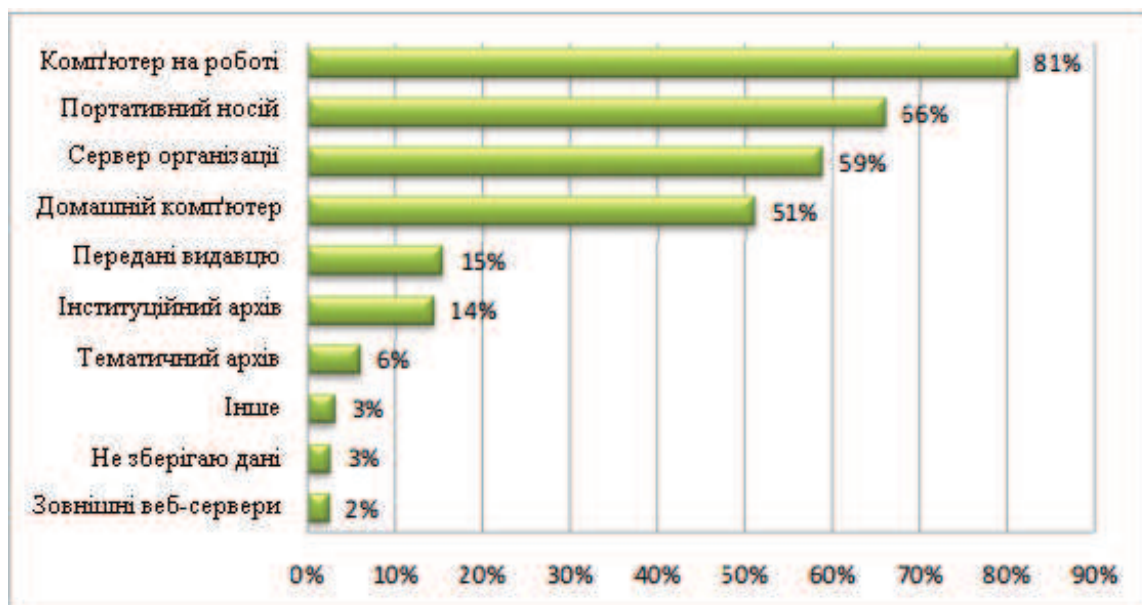
**Keywords:** *e-science, open access, information services, scientific data, citizen science, Data Curation, FigShare.*

У результаті розвитку Інтернет-технологій за останні роки процес наукової комунікації зазнав значних трансформацій. Нині говорячи про нові відкриті моделі передачі знань все частіше згадують про феномен е-науки – науки, яка відбувається в Інтернет-середовищі з використанням великих наборів даних, що вимагають опрацювання на різних серверах. Як правило, такий процес наукової комунікації відбувається відкрито, що гарантує повну прозорість та легкість приєднання до дослідницького процесу усіх зацікавлених вчених, чи організацій.

Одним з яскравих прикладів таких змін у процесі наукової комунікації стала поява науково-дослідних Інтернет-ініціатив – «громадянин науки» (citizen science) [4, с. 467]. На веб-ресурсах, котрі підтримують ідеологію «громадян науки» (таких як Galaxy Zoo, SETI@home, EyeWire, Cell Slider, SciStarter), значну частку робіт виконують добровольці-непрофесіонали. Передусім, це роботи, що не потребують спеціальної освіти і пов'язані з опрацюванням величезних масивів даних, як, наприклад, класифікація галактик, чи укладання прогнозів клімату Землі до 2300 року.

Безперешкодний доступ та можливість використання існуючих даних дозволяють вченим, що володіють різними стартовими можливостями, зменшити економічні витрати на проведення досліджень, відкривають нові шляхи для науково-технічного співробітництва та прискорюють науковий прогрес. Відтак, з ростом важливості відкритого доступу до вихідних наукових даних та збільшенням масиву зібраних даних, перед науковим співтовариством все гостріше повстає потреба у розроблені та впроваджені оптимальних управлінських рішень і технологій для їх ефективної передачі та зберігання.

Так, згідно результатів масштабного дослідницького проекту Survey Report PARSE.Insight, що відбувався за фінансової підтримки Європейської комісії, близько 60% європейських учених хотіли б використовувати наукові дані, отримані іншими дослідниками [3]. Водночас, понад 40% вчених зіштовхуються з серйозними труднощами при спробах поділитися власними науковими даними, а тому тотальна більшість наукових даних продовжує зберігатися виключно на приватних електронних носіях інформації (Діагр 1.).



Діагр. 1. Де вчені зберігають свої наукові дані згідно Survey Report PARSE.Insight (було дозволено кілька варіантів відповіді)

Управлінську діяльність, що скерована на підтримку зберігання та повторне використання наукових даних, у наш час прийнято позначати терміном Data Curation [1, с. 371]. Оскільки збереження та надання доступу до інформації є однією з основних місій бібліотек, новий напрямок роботи, пов'язаний з Data Curation, швидко з'явився серед послуг багатьох з них. Нині університетські бібліотеки у всьому світі прагнуть стати активнішими партнерами в науково-дослідній роботі працівників університету, пропонуючи допомогу на різних етапах наукових досліджень і поступово частково відмовляючись від сервісів, побудованих на основі сталих фондів, на користь обслуговування користувачів з використання динамічних наукових даних.

Одне з найоригінальніших рішень щодо можливостей використання наукових даних у процесі обслуговування користувачів реалізоване у бібліотеці університету Джона Хопкінса (Балтімор, США). Дослідники Грег Хагер та Саїд Чоудхурі розмістили на стіні бібліотеки прототип пристрою, що складається з кількох моніторів, процесорів та систему Microsoft Kinect, яка здатна розпізнавати рухи і жести людського тіла.

Прототип дозволяє обробляти великі масиви даних, отримувати їх візуалізацію на систему моніторів і за допомогою жестів проводити складні наукові дослідження [2]. У майбутньому даний пристрій планується використовувати для візуалізації та маніпулювання молекулами, або у навчанні студентів класифікувати галактики.

Подібні вдалі проекти поки залишаються рідкісним явищем, адже впровадження Data Curation у роботу бібліотеки внз вимагає не лише переосмислення пріоритетів, а й вироблення відповідних технічних умінь персоналу. За результатами анкетування персоналу бібліотек–учасниць Європейської мережі наукових бібліотек (LIBER), понад 80% наукових бібліотек відчують високу затребуваність з боку вчених в організації роботи з даними, однак лише 12% анкетованих вважають, що бібліотекарі володіють достатнім рівнем умінь, щоб задовольняти нові потреби користувачів [6]. Попри успішний досвід створення та підтримки інституційних репозитаріїв бібліотеками внз, робота з науковими даними потребує вироблення свого власного підходу до питань селекції, збереження, доступу, правових аспектів та стандартів метаданих.

Окрім того, як слушно зауважив Андрій Зємсков – масив наукових даних сам по собі не має великого значення. Винятки становлять лише заздалегідь сформовані колективні проекти – такі, як геном людини, пошук бозона Гітса, проекти в астрономії, спільні дослідження клімату Землі. Якщо ж між потенційними користувачами не було встановлено попередньої домовленості, то важко зрозуміти, навіщо все це потрібно [5, с. 100]. Сучасна трансформація бібліотек та невизначеність їх ролі у суспільстві створюють перешкоди для налагодження тісного взаємозв'язку зі науковою спільнотою потенційних користувачів сервісів Data Curation, тим самим значно ускладнюючи запуск подібних проектів.

В умовах обмеженого фінансування, навіть за умов подолання економічних, правових та технічних бар'єрів,

підтримка усіх процесів пов'язаних з Data Curation може стати непосильною ношею для більшості вітчизняних установ. Оптимальним шляхом вирішення цієї проблеми може стати активне використання бібліотеками вnz хмарних сервісів, що дозволяють безкоштовно зберігати дані на віддалених Інтернет-серверах. За використання такого підходу, персонал та бюджет бібліотеки буде звільнений від таких витратних процесів, як зберігання та надання доступу, що дасть змогу зосередитись на створенні оптимальних стандартів опису даних.

Для задоволення інформаційних потреб, що пов'язані зі зберігання та управлінням великими масивами наукових даних, варто рекомендувати хмарний сервіс FigShare (<http://figshare.com>). Сервіс дозволяє приватно та безпечно керувати результатами власних наукових досліджень та зробити їх видимими, доступними і цитованими. Для цього усім користувачам FigShare безкоштовно надається необмежений простір для публічного поширення документів та 1 Гб простору для приватного збереження даних.

FigShare гарантує постійний доступ до наукових даних, забезпечує легку фільтрацію та пошук документів (за автором, науковими дисциплінами, тегами), підтримує різні формати файлів, дозволяє автоматично отримувати альтиметричні показники використання завантажених документів. Також, значною перевагою FigShare, порівняно з іншими подібними хмарними сервісами, є те, що усі розміщені документи отримують унікальний ідентифікатор DOI, а відтак вони можуть бути легко знайдені та коректно процитовані іншими науковцями. Таким чином, використання бібліотеками вnz хмарного сервісу FigShare здатний не лише суттєво зменшити витрати на впровадження нових послуг пов'язаних з Data Curation, а й значно покращити видимість наукових досліджень, що проводяться в установі.

Поки що бібліотеки тільки розпочинають запроваджувати сервіси з обслуговування користувачів науковими даними. Однак, подібні попередні експериментальні бібліотечні веб-проекти, що будувалися навколо ідеї відкритості наукових документів, як правило, отримували достатньо велику кількість прихильників. Тому, можемо прогнозувати, що впровадження сервісів Data Curation у роботу університетських бібліотек активно відбудуватиметься у наступні роки, що в свою чергу зміцнить позиції бібліотек внаслідок дослідницької та освітньої діяльності своїх установ.

### **Список використаних джерел:**

1. From data deluge to data curation / Philip Lord, Alison Macdonald, Liz Lyon, David Giaretta // Proc 3th UK e-Science All Hands Meeting. – 2004. – P. 371–375.
2. Monastersky R. Publishing frontiers: The library reboot / Richard Monastersky // Nature. – 2013. – № 495. – P. 430–432.
3. PARSE.Insight Survey Report // PARSE.Insight : Website. – 2009. – Mode of access: URL : [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf). – Title from screen.
4. Silvertown J. A new dawn for citizen science / Jonathan Silvertown // Trends in Ecology & Evolution. – 2009. – Vol. 24, № 9. – P. 467–471.
5. Земсков А. И. Data Curation хранение научных данных и обслуживание ими – новое направление деятельности библиотек / А. И. Земсков // Научные и технические библиотеки. – 2013. – № 2. – С. 85–101.
6. Рейли С. Роль библиотек в поддержке обмена научными данными / Сюзан Рейли // Научные и технические библиотеки. – 2013. – № 2. – С. 102–107.