Otwarte dane badawcze

Powszechny i bezpłatny dostęp do aktualnej wiedzy stanowi jeden z fundamentów funkcjonowania oraz rozwoju demokratycznego społeczeństwa. Koncepcja otwartego dostępu (*open access*) dotyczy udostępniania w postaci cyfrowej w Internecie publikacji naukowych i wyników badań, które finansowane są ze środków publicznych w celu umożliwienia bezpłatnego ich wykorzystania przez naukowców, studentów, przedsiębiorców i całe społeczeństwo.

Znaczenie dostępności danych badawczych wzrosło w ostatnich latach i ma to związek z przyrostem liczby wytwarzanych danych oraz z pojawieniem się możliwości ich efektywnego udostępnienia za pomocą technologii cyfrowych. Nastąpił rozwój w obszarze generowania, magazynowania, przetwarzania i szybkiego przesyłania danych o ogromnych rozmiarach i wysokiej złożoności. Dla nauki i naukowców kreuje to nowe możliwości, które mogą być w pełni wykorzystane wyłącznie wówczas, gdy powstające dane badawcze będą ogólnodostępne. W artykule skoncentrowano się na zagadnieniu otwartych baz danych badawczych, sposobach ich udostępniania oraz kwestii zarządzania danymi.

Co to są dane badawcze (research data)?

W literaturze naukowej występuje wiele definicji wytworzonych przez różne instytucje oraz organizacje. Pierwsza z nich, która powstała w roku 1999 na potrzeby amerykańskiego rządu, określa, że *dane badawcze są to zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.*

Kolejna definicja została stworzona przez Komisję Europejską i stwierdza, iż *dane badawcze to informacja, w szczególności zebrane fakty, liczby, które mogą posłużyć badaniom i być traktowane jako podstawa do dalszego wnioskowania, dyskusji lub obliczeń.*

Trzecia definicja – ustalona na potrzeby Queensland University of Technology – określa dane badawcze jako *zapisy faktów (w postaci liczbowej, tekstowej, graficznej bądź dźwiękowej) będących rezultatem badań naukowych (np. obserwacji, pomiarów, doświadczeń, eksperymentów itp.), na których to opiera się wnioskowanie naukowe (tezy, argumentacja, wyniki)*.

Na postawie powyższych definicji można przyjąć, że dane badawcze to jest wszystko, co zostało wyprodukowane w ramach prowadzonych badań. Do danych badawczych zalicza się więc: notatki, dane liczbowe, ankiety, kwestionariusze, nagrania audio i wideo, modele matematyczne, algorytmy, oprogramowanie, skrypty, wyniki symulacji komputerowych. W skład danych badawczych wchodzą także metadane (np. nazwa zbioru danych, autorzy itp.), które stanowią podstawowy opis umożliwiający identyfikację oraz zrozumienie struktury zbioru. Dane badawcze to zarówno surowe dane (*row data*), które uzyskano bezpośrednio w wyniku zastosowania jakiegoś narzędzia badawczego i które nie zostały jeszcze opracowane, a także takie, które zostały podane analizie.

Otwarte dane badawcze (*open research data***)** są to dane pozyskane w procesie badawczym, które są dostępne w postaci cyfrowej i upowszechniane bez żadnych ograniczeń, co oznacza, że mogą być dowolnie używane oraz rozpowszechniane i przetwarzane przez kogokolwiek, gdziekolwiek i w dowolnym celu.

Otwarte dane badawcze powinny być udostępnianie w bezpiecznych otwartych repozytoriach cyfrowych spełniających zasady FAIR Data, czyli powinny być:

- ***F****indable* – łatwo znajdowane i wyszukiwane.
- ***A****ccessible* – dostępne dla wszystkich.
- ***I****nteroperable* – interoperacyjne, tak aby można było je połączyć z innymi danymi.
- ***R****eusable* – wielokrotnego użytku.

Otwartość danych, w odróżnieniu od otwartości publikacji, związana jest nie tylko z ich dostępem, lecz głównie z możliwością ich ponownego wykorzystania. Otwartość w odniesieniu do danych badawczych zachodzi wtedy, kiedy mamy do czynienia z odpowiednikiem otwartego dostępu libre (obowiązek uznania autorstwa bądź zakaz ograniczania wolności innych użytkowników).

Ważnym elementem otwartych danych badawczych jest właściwe zarządzanie danymi (*research data management*) poprzez tworzenie planów zarządzania danymi (*Data Management Plan* – DMP). Jest to formalny dokument określający sposób zarządzania danymi badawczymi podczas projektu badawczego oraz po jego zakończeniu. Powinien zawierać informacje: jakie dane będą pozyskane lub wytworzone, w jakiej formie, w jaki sposób będą opisane, jak będą przechowywane oraz jak zostaną udostępnione. Coraz więcej instytucji finansujących badania naukowe (w Polsce Narodowe Centrum Nauki, Program Horizon 2020) wymaga od naukowców i jednostek naukowych przedstawienia we wnioskach grantowych Planu Zarządzania Danymi.

Najważniejszym krokiem służącym do podjęcia konkretnych działań, związanych z udostępnianiem publikacji naukowych oraz danych badawczych, jest zrozumienie korzyści, jakie z tego płyną, a są to:

1. w ramach zastosowania otwartości następuje ograniczenie kosztów powielania badań, zwiększenie ich wydajności, co przekłada się na potencjalnie szybsze odkrycia, a to z kolei przyspiesza postęp naukowy,
2. poprzez dostęp do bezpłatnej literatury ulegnie zwiększeniu widoczność dorobku naukowego, co wpłynie na stopień oddziaływania badań i wzrost liczby cytowań,
3. dzięki możliwości weryfikacji oraz odtworzenia danych badawczych (*reproducibility*) oraz ich ponownego wykorzystania (*re-use*), w tym również komercyjnie, podniesiona zostanie jakość późniejszych badań,
4. otwieranie dostępu do danych badawczych wpływa na zwiększenie prawdopodobieństwa nawiązywania kontaktów z innymi grantodawcami oraz naukowcami, również spoza bazowej dziedziny, działających na przykład na styku różnych dyscyplin.

Podsumowując, swobodne udostępnianie zarówno danych, jak i pozostałych elementów procesu badawczego, ma służyć budowaniu na nich podstawy wiedzy – dzięki temu, że mogą być ponownie wykorzystane do analiz, badań, testów oraz publikowania nowych wyników. Otwarty dostęp wydobywa naukę z jej zamkniętego kręgu, pozwala poszerzać grono odbiorców i komentatorów oraz uczestników badań, jak i podmiotów badań.

Magdalena Wiederek-McRobb
Katarzyna Rybicka

Bibliografia:

1. https://www.gov.pl/web/nauka/otwarty-dostep-do-publikacji-naukowych.
2. E. Strzelczyk, *Otwarte dane badawcze – kolejny krok do otwierania nauki* [w]: I. Sójkowska, L. Derfert Wolf (red.), *Bibliograficzne bazy danych: perspektywy i problemy rozwoju*. III Konferencja Naukowa Konsorcjum BazTech, Kraków, 26-27 czerwca 2017 [online]. Stowarzyszenie EBIB, 2017. [Dostęp 20.11.2017]. Materiały Konferencyjne EBIB, nr 25. ISBN 978-83-63458-08-9. Dostępny w: http://open.ebib.pl/ojs/index.php/Mat_konf/article/view/599.
3. M. Rożniakowska-Kłosińska, *Otwarte dane badawcze w warsztacie pracy naukowca*, Biuletyn EBIB [online] 2018, nr 6 (183), Współpraca bibliotek z naukowcami. [Dostęp 18.12.2018]. ISSN 1507-7187. Dostępny w: http://open.ebib.pl/ojs/index.php/ebib/article/view/668.
4. https://www.liverpool.ac.uk/media/livacuk/computingservices/research-data-management/OpenResearch.jpg.

Fot. Źródło: https://www.liverpool.ac.uk/media/livacuk/computingservices/research-data-management/OpenResearch.jpg

Wojskowa Akademia Techniczna

OTWARTE DANE BADAWCZE

www.bg.wat.edu.pl