

Edyta Strzelczyk
Biblioteka Główna Politechniki Warszawskiej
E.Strzelczyk@bg.pw.edu.pl

Otwarte dane badawcze – kolejny krok do otwierania nauki



Edyta Strzelczyk

Absolwentka studiów magisterskich w Instytucie Informacji Naukowej i Studiów Bibliologicznych oraz studiów licencjackich na kierunku socjologia w Uniwersytecie Warszawskim. Od 2011 r. pracuje jako bibliotekarz w Oddziale Udostępniania i Przechowywania Zbiorów w Bibliotece Głównej Politechniki Warszawskiej. Zainteresowania zawodowe skupiają się wokół statystyki bibliotecznej, analiz efektywności i jakości pracy bibliotek oraz udostępniania wyników prac naukowych w otwartych zasobach. Jest członkiem Zespołu SBP ds. badania efektywności bibliotek. Publikuje m.in. na temat tzw. wzmocnionych publikacji jako nowego modelu wsparcia komunikacji naukowej.

Streszczenie: Udostępnianie dorobku naukowego w otwartym dostępie z każdym rokiem staje się coraz powszechniejszą praktyką. Dotyczy to nie tylko publikacji, które są końcowym rezultatem działań naukowych, ale także produktów wytworzonych w trakcie całego procesu badawczego, w tym danych badawczych. Repozytoria służące do deponowania, przechowywania i udostępniania publikacji zaczęły być rozbudowywane o funkcje umożliwiające gromadzenie i magazynowanie danych badawczych lub też tworzone są repozytoria dedykowane dla tych zasobów. Specyfika danych ściśle związana jest zarówno z uprawianą dyscypliną wiedzy, jak i typem jednostki naukowej. Różnorodność rodzajów danych, niejednolite standardy ich zapisu i przechowywania oraz brak jednoznacznych regulacji prawnych stanowi utrudnienie dla ich umieszczania w otwartych zasobach. W celu zapewnienia efektywnego udostępniania danych badawczych należy zadbać o to, aby były zapisane w formatach umożliwiających ich otwarcie i ponowne wykorzystanie, aby były zachowane (najlepiej długoterminowo), dostępne (do pobrania), możliwe do odnalezienia (np. powiązane z publikacją), zrozumiałe (opisane) i zaufane (źródło pochodzenia). Zapotrzebowanie na dostęp do różnych produktów prac naukowych potwierdzają badania ankietowe zrealizowane w środowisku naukowym Politechniki Warszawskiej oraz innych polskich jednostkach naukowych i badawczo-rozwojowych. Kolejnym etapem do usprawnienia komunikacji naukowej może być tworzenie sieci powiązań między elementami działań naukowych, produktami prac badawczych oraz publikacjami, co prowadzi do tworzenia złożonych obiektów cyfrowych tzw. wzmocnionych publikacji (ang. *enhanced publications*).

Słowa kluczowe: otwarta nauka, otwarty dostęp, dane badawcze, otwarte dane, wzmocnione publikacje

Abstract: Publishing scientific achievements in open access becomes an increasingly common practice with every year. This applies not only to publications that are the end result of scientific activities, but also to the products of the entire research process, including research data. Repositories for depositing, storing and sharing publications have been expanded to include features that enable the collection and storage of research data or dedicated repositories for these resources are being created. The specificity of research data is related both to the given discipline of knowledge and to the type of scientific unit. The variety of the types of data, the non-uniform standards for their storage, and the lack of clear legal regulations make it difficult for them to be published in open resources. In order to ensure effective sharing of research data, care should be taken to allow them to be stored in formats that enable them to be re-opened and reused so that they are preserved (preferably long-term), accessible (downloadable), able to be found (e.g. related to a publication), understandable (described), and trusted (source of origin). The demand for access to various products of scientific work is confirmed by surveys conducted in the scientific community of the Warsaw University of Technology and other Polish scientific and research and development units. The next step in the improvement of scientific communication may be the creation of a network of links between the elements of scientific activities, research products and publications, leading to the creation of complex digital objects, so-called enhanced publications.

Keywords: open science, open access, research data, open data, enhanced publications

Prezentacja

Wprowadzenie

Idea otwartej nauki, czyli systemu, w którym zasadnicze etapy procesu produkcji i dystrybucji wiedzy realizowane są w ogólnie dostępnej sieci internetowej bez ograniczeń technicznych, finansowych i prawnych¹ obecnie jest zagadnieniem powszechnie znanym i stale rozwijającym się. Otwarta nauka to nie tylko swobodny dostęp do końcowego wytworu, czyli publikacji, ale dostęp do danych z całego toku prac naukowych, czyli otwarte dane badawcze, otwarte narzędzia i metodologia, tworzenie i stosowanie otwartych oprogramowań, otwarte modele pracy naukowej (badania prowadzone przy „otwartym notatniku”), otwarte debaty i dyskusje. Otwieranie nauki ma prowadzić do jej „usieciowienia”², czyli przeniesienia całokształtu komunikacji naukowej do internetu. Dokonywanie odkryć naukowych opierać się będzie wówczas na narzędziach umożliwiających przeszukiwanie już opisanej wiedzy i dostrzeganiu w niej nowych zależności. W artykule skoncentrowano się na kwestiach związanych z danymi badawczymi, ich udostępnianiem i stwarzaniem warunków do ich ponownego wykorzystania, zostaną one umiejscowione w szerszym kontekście pozostałych elementów procesu badawczego.

Znaczenie dostępności danych badawczych wzrosło w ostatnich latach, wraz ze stałym powiększeniem się liczby wytwarzanych danych oraz z pojawieniem się możliwości ich efektywnego udostępnienia z pomocą technologii cyfrowych³. Nastąpił rozwój w obszarze generowania, magazynowania, przetwarzania i szybkiego przesyłania danych nawet o ogromnych rozmiarach i wysokiej złożoności⁴. Stwarza to nowe możliwości dla nauki i naukowców, które to mogą być w pełni wykorzystane jedynie wówczas gdy powstające dane badawcze będą otwarte.

Dane badawcze – czym są?

Czym są dane badawcze (*research data*)? Odpowiedź na to pytanie nie jest prosta, gdyż występuje wiele różnych definicji. Często sami wytwórcy danych, czyli instytucje i organizacje naukowe, tworzą własne definicje w taki sposób, który najlepiej odpowiada ich potrzebom.

Definicja sformułowana w 1999 r. na potrzeby amerykańskiego rządu mówi, iż są to *zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane*

¹ STARCZEWSKI, M., STĘPIŃSKA-USTASIAK, L. *Otwarty dostęp czy otwarta nauka?* [online]. 2014, s. 162. [Dostęp 22.06.2017]. Dostępny w: <https://depot.ceon.pl/bitstream/handle/123456789/6317/Otwarty%20dost%C4%99p%20czy%20otwarta%20nauka.pdf?sequence=2&isAllowed=y>.

² Tamże, s. 162.

³ HOFMOKL, J., TARKOWSKI, A., BEDNAREK-MICHALSKA, B. i in. *Przewodnik po otwartej nauce* [online]. Warszawa: Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, 2009, s. 57. [Dostęp 22.06.2017]. ISBN 978-83-917150-4-8. Dostępny w: <https://otwartanauka.pl/images/PDFs/przewodnik-po-otwartej-nauce.pdf>.

⁴ *Otwarte dane w świecie wielkich danych. Porozumienie międzynarodowe. Wersja skrócona* [online]. ICSU-IAP-ISSC-TWAS, s. 2. [Dostęp 22.06.2017]. Dostępny w: http://www.science-international.org/sites/default/files/reports/otwarte-dane-wersja-skrocona_short_pl.pdf.

przez społeczność naukową za niezbędne do oceny wyników badań naukowych⁵. Zgodnie z dokumentami Komisji Europejskiej dane badawcze to *informacja, w szczególności zebrane fakty, liczby, które mogą posłużyć badaniom i być traktowane jako podstawa do dalszego wnioskowania, dyskusji lub obliczeń*⁶. Definicja ustalona na potrzeby Queensland University of Technology określa je jako zapisy faktów (w postaci liczbowej, tekstowej, graficznej czy dźwiękowej) będących rezultatem badań naukowych (np. obserwacji, pomiarów, doświadczeń, eksperymentów itp.), na których to opiera się wnioskowanie naukowe (tezy, argumentacja, wyniki)⁷. Do danych badawczych można zaliczyć zarówno surowe dane, czyli takie, które uzyskano bezpośrednio w wyniku zastosowania jakiegoś narzędzia badawczego (np. programu komputerowego, instrumentu pomiarowego, kwestionariusza ankiety) i które nie zostały jeszcze opracowane oraz takie dane, które poddane zostały pewnym operacjom służącym chociażby zmianie ich formy, oczyszczeniu, uporządkowaniu, przetworzeniu. W definicji stosowanej przez The University of Melbourne do danych badawczych zaliczane są także informacje na temat pochodzenia tych danych: jak, gdzie, kiedy zostały zebrane oraz za pomocą jakich narzędzi, czyli cała dokumentacja badań, która może obejmować takie elementy jak: notatki, raporty, opisy metod i narzędzi, korespondencja, dokumenty aplikacyjne, formularze zgody itp.⁸. Zatem w skład danych badawczych wchodzi także metadane (np. nazwa zbioru danych, autorzy itp.) stanowiące podstawowy opis umożliwiający identyfikację i zrozumienie struktury zbioru. Istotne jest używanie standardów zapisów metadanych (np. Dublin Core, OpenAIRE Guidelines) oraz wymiana metadanych za pomocą protokołu Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), co zapewnia danym interoperacyjność, czyli możliwość wymiany, łączenia i ponownego wykorzystania.

Ze wszystkich zaprezentowanych definicji można wywnioskować, iż do danych badawczych zaliczamy:

- dokumenty tekstowe, notatki,
- dane liczbowe (w formie nieprzetworzonej lub czytelnej dla komputera),
- kwestionariusze ankiet, wyniki badań ankietowych,
- nagrania audio i wideo, zdjęcia,
- modele matematyczne, algorytmy, schematy,
- oprogramowanie (skrypty, pliki wejściowe),
- wyniki symulacji komputerowych,
- protokoły laboratoryjne, opisy metodologiczne,
- próbki, artefakty, obiekty,
- metadane (opisy).

⁵ What is research data? W: *ANDS Guide* [online]. 2017, s. 1. [Dostęp 22.06.2017]. Dostępny w: http://www.ands.org.au/_data/assets/pdf_file/0006/731823/What-is-research-data.pdf.

⁶ Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 z dnia 11 grudnia 2013 r. Cyt. za: *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce* [online]. 2015, s. 8. [Dostęp 22.06.2017]. Dostępny w: http://www.nauka.gov.pl/g2/oryginal/2015_10/9f62cc350837b942e51ae23dd1f23df8.pdf.

⁷ *What is research...*, dz. cyt., s. 2.

⁸ Tamże, s. 2.

Otwarte dane badawcze – to znaczy jakie?

Otwarte dane badawcze to takie, które są dostępne w postaci cyfrowej w domenie publicznej, upowszechniane bez żadnych ograniczeń, czyli mogą być dowolnie używane, rozpowszechniane i przetwarzane przez kogokolwiek, gdziekolwiek w dowolnym celu⁹. Otwartość danych w odróżnieniu od otwartości publikacji związana jest nie tylko z ich dostępem, ale głównie z możliwością ich ponownego wykorzystania. Otwartość w odniesieniu do danych badawczych zachodzi wtedy, gdy mamy do czynienia z odpowiednikiem otwartego dostępu *libre* (obowiązek uznania autorstwa lub zakaz ograniczania wolności innych użytkowników)¹⁰.

Ograniczenia, na które możemy natrafić podczas korzystania z danych badawczych zamieszczonych w internecie mogą dotyczyć kwestii finansowych, technicznych lub prawnych. Bariery finansowe wiążą się z koniecznością uiszczenia opłaty podczas dostępu do danych, co wynika z postrzegania wiedzy w kategoriach towaru rynkowego. Ograniczenia techniczne pojawiają się wówczas, gdy dane zapisane są w formatach wymagających specjalistycznych oprogramowań lub też w formatach „zamkniętych”, uniemożliwiających ich ponowne wykorzystanie. Najbardziej złożone są zagadnienia prawnej ochrony danych badawczych. Mogą one podlegać ochronie prawa autorskiego, ale jedynie w obszarze związanym z kreatywnością i oryginalnością wykonywanej pracy (np. selekcja, rozmieszczenie, układ danych). Prawo autorskie nie chroni samych faktów oraz ich zapisów (czyli danych), ale odnosi się bardziej do oryginalnych form wyrażania idei. Prawo mające znaczne zastosowanie do danych badawczych to obowiązujące w Unii Europejskiej prawo tzw. *sui generis* do baz danych – związane z inwestycją włożoną w utworzenie (zdobycie, prezentację lub weryfikację) bazy danych. Prawo to zapewnia ochronę przed uzyskaniem (np. kopiowaniem) i wtórnym wykorzystaniem (np. redystrybucją) całej lub istotnej części bazy¹¹. Inne ograniczenia prawne mogą wiązać się z: prawami osób trzecich, np. osób zaangażowanych w powstanie danych (twórcy zawartości, respondenci), prawami pokrewnymi (np. dotyczącymi wydań krytycznych) czy też prawami przysługującymi pracodawcy wynikającymi np. z regulaminów nabywania, korzystania i ochrony własności intelektualnej uczelni¹².

Szereg trudności związanych z tradycyjnymi metodami uprawiania nauki i produkowania wiedzy oraz przyzwyczajeniami środowiska naukowego powoduje, iż zakres udostępniania danych badawczych jest nadal niewielki. Najbardziej istotne przeszkody to:

z perspektywy ogólnej oraz instytucji

- brak jednoznacznych uregulowań prawnych na poziomie krajowym (zalecenia, a nie obowiązek),

⁹ JAMES, L. Defining Open Data. W: *Open Knowledge International Blog* [online]. [Dostęp 22.06.2017]. Dostępny w: <http://blog.okfn.org/2013/10/03/defining-open-data/>.

¹⁰ SZPROT, J. (red.). *Otwarta nauka w Polsce 2014. Diagnoza* [online]. Warszawa: Wydawnictwo ICM, 2014, s. 28. [Dostęp 22.06.2017]. ISBN 978-83-63490-08-9. Dostępny w: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>.

¹¹ GUIBAULT, L., WIEBEGO, A. (red.). *Otwarte, czyli bezpieczne. Analiza prawnych form ochrony danych badawczych i zalecenia w kwestii dostępu i korzystania* [online]. Kraków: Biblioteka Jagiellońska, 2014, s. 21. [Dostęp 22.06.2017]. ISBN 978-83-934926-8-8. Dostępny w: <http://jbc.bj.uj.edu.pl/dlibra/doccontent?id=283369>.

¹² SZPROT, J., dz. cyt., s. 28.

- bariery stawiane przez prawa chroniące własność intelektualną,
- brak jednolitych standardów publikacji danych,
- obawa przed utratą potencjału dochodowego – wiedza traktowana jako towar rynkowy,
- ochrona danych osobowych, respektowanie prywatności osób, konieczność anonimizacji danych,
- tradycyjny system oceny dorobku naukowego, publikacja danych nie jest punktowana przy awansie zawodowym,
- koncentracja na wydaniu publikacji jako kluczowego wytworu prac naukowych, ignorowanie pozostałych elementów,
- ograniczenia nakładane przez wydawców komercyjnych (podobne do tych nakładanych podczas udostępniania publikacji),

z perspektywy naukowców

- obawa przed utratą autorstwa, wielokrotne wykorzystywanie danych,
- obawa przed utratą kontroli nad danymi, podział danych i utrata spójności, spadek jakości danych,
- ryzyko wykrycia błędów i nieprawidłowości – spadek wartości marki,
- efektywniejsze wykorzystanie danych przez innych badawczy,
- dodatkowa praca poświęcona na przygotowanie danych.

Pokonanie napotykanymi trudnościami zarówno tych formalnych, jak i tych związanych z przyzwyczajeniami i tradycyjnym podejściem, wymaga pewnego wysiłku, ale korzyści płynące z tych działań są niezaprzeczalne.

Dane należy otwierać we właściwy sposób, nie wystarczy samo ich zamieszczenie w internecie. Otwarte dane badawcze powinny być:

- Zapisane – w formatach łatwych do otworzenia i odczytania, najlepiej przy pomocy oprogramowania typu *open source* lub zalecane jest wskazanie (bądź udostępnienie) oprogramowania umożliwiającego te działania. Format powinien być „otwarty”, czyli dawać możliwość ich ponownego wykorzystania (modyfikacji, przetwarzania). Dane powinny być zapisane z wykorzystaniem standardowego kodowania (ASCII, Unicode).
- Przechowane (bezpieczna archiwizacja) – zdeponowane długoterminowo w repozytorium/archiwum godnym zaufania (jakość, marka), niezależnie od formatu, rozmiaru, aktualności itp.
- Dostępne – na określonych warunkach dostępu do danych i oprogramowania. Dane w repozytoriach/archiwach mogą być udostępniane bez licencji, na zasadach dozwolonego użytku lub poprzez wskazanie licencji. Niewskazanie licencji może zawęzić lub nawet ograniczyć zakres wykorzystania danych (np. ochrona prawno-autorska oraz ochrona *sui generis*). Zalecane licencje dla danych badawczych to *Creative Commons Zero* (CC 0) – oświadczenie o zrzeczeniu się praw lub *Creative Commons By* (CC BY) – uznanie autorstwa, choć ta ostatnia niesie ze sobą ryzyko nawarstwiania atrybucji (konieczność podawania autorów wszystkich danych, z których skorzystaliśmy). Pozostałe licencje w pewnych zakresach są niewłaściwe dla udostępniania danych, gdyż w ostatecznym rozrachunku ograniczają ich otwartość. *Creative Commons Share-alike* (CC SA) – na tych samych warunkach – może utrudnić łączenie danych z innymi zbiorami, które są

dostępne na odmiennych licencjach¹³. *Creative Commons Non-commercial* (CC NC) – użycie niekomercyjne – według niektórych interpretacji prawnych licencje niekomercyjne nie pozwalają nawet na wykorzystanie danych do wytworzenia artykułu naukowego, który jest potem sprzedawany (jako część czasopisma).

- Łatwe do znalezienia – np. poprzez zastosowanie DOI (*digital object identifier*) cyfrowego identyfikatora dokumentów elektronicznych, który w odróżnieniu od URL nie zależy od fizycznej lokalizacji dokumentu, lecz jest do niego na stałe przypisany. Dzięki temu dane są łatwe do zidentyfikowania za pomocą standardowego mechanizmu. Powiązanie danych z publikacją oraz innymi materiałami sprawia, że liczba ścieżek dotarcia do danych się zwiększa.
- Zrozumiałe – powinny być opisane w sposób kompletny i zrozumiały, tak aby osoby trzecie mogły je odczytać i ponownie wykorzystać. Zalecane jest stosowanie ujednoczonych standardów tworzenia metadanych (Dublin Core, OpenAIRE Guidelines). Należy dołączyć także dokumentację dotyczącą samego zbioru danych (nazwy i opis zmiennych, rekordów, wartości, opisy zastosowanych schematów klasyfikacyjnych, informacje o urządzeniach pomiarowych, dane przetworzone wraz z oprogramowaniem lub algorytmami zastosowanymi do ich uzyskania) jak i całego projektu (opis celu i kontekst badań, sposoby pozyskiwania danych, procedury, struktura plików z danymi, relacje między plikami, formularze zgody, informacje o różnych wersjach zbiorów danych itp.).
- Zaufane – z określonym autorstwem oraz afiliacją osób odpowiedzialnych za zebranie, przygotowanie czy przetworzenie danych, co pomaga w ocenie ich rzetelności, jakości i wiarygodności.
- Interoperacyjne – zapewniające możliwość ich swobodnej wymiany między naukowcami, instytucjami i innymi zainteresowanymi, możliwość ponownego wykorzystania w innych kontekstach niż ten, w którym zostały zebrane. Interoperacyjność oznacza także możliwość łączenia ich z innymi danymi (dzięki współpracy różnych systemów).
- Odtwarzalne – zapewniające możliwość ponownego wykorzystania w dowolnej formie oraz wykonania jakichkolwiek operacji: pobierania, zapisania, maszynowe przetwarzania, modyfikacji, zmiany, redystrybucji.

Idea powszechnej dostępności danych badawczych musi jednak dopuszczać pewne wyjątki. Nie wszystkie dane mogą być otwarte, w uzasadnionych przypadkach nieotwieranie danych jest jak najbardziej działaniem wskazanym. Kluczowym przykładem są tutaj dane osobowe, które umożliwiają identyfikację tożsamości konkretnych osób. Podlegają one innym zasadom i należy zachować wszelkie prawa związane z ochroną danych osobowych. Innym przykładem mogą być dane związane z bezpieczeństwem kraju, ochroną przemysłu czy też utrzymaniem przewagi strategicznej zarówno naukowej, jak i innowacyjnej¹⁴.

¹³ BEDNAREK-MICHALSKA, B. Repozytoria surowych danych – dlaczego biblioteki powinny je znać? *Biuletyn EBIB* [online]. 2012, nr 8 (135), s. 3. [Dostęp 22.06.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/152/292>.

¹⁴ SZCZĘSNY, P. *Otwarta nauka czyli dobre praktyki uczonych* [online]. Toruń: Stowarzyszenie EBIB, 2013, s. 23. [Dostęp 22.06.2017]. ISBN 978-83-63458-03-4. Dostępny w: http://repozytorium.uwb.edu.pl/jspui/download/Szczesny_Otwarta_nauka.pdf.

Regulacje prawne dotyczące danych badawczych

Ważne działania zmierzające do zwiększenia powszechnej dostępności danych badawczych prowadzi Komisja Europejska uruchamiając projekt Open Research Data (ORD) realizowany w ramach programu finansowania badań naukowych i innowacji Horyzont 2020. W latach 2014–2016 prowadzono działania testujące określane mianem pilotażu, którym zostały objęte wyznaczone dziedziny, np.: energia, środowisko, zdrowie, technologie informacyjne, infrastruktura badawcza, nauka w społeczeństwie, nauki społeczno-ekonomiczne i humanistyczne. Od 2017 r. projekt ORD został rozszerzony na wszystkie obszary tematyczne i dziedziny. Uczestnictwo w pilotażu w dużej mierze jest dobrowolne, w dokumencie *Guidelines on FAIR Data Management in Horizon 2020*¹⁵ znajdujemy informację, iż jest to „elastyczny pilotaż” a zaangażowanie w ORD nie jest elementem wpływającym na ocenę wniosku o dofinansowanie. Projekt ma na celu zachęcanie do udostępniania danych oraz propagowanie korzyści wypływających z tych działań. Dane badawcze uzyskane w procesie badawczym powinny być FAIR (*findable, accessible, interoperable, re-usable* – dające się znaleźć, dostępne, interoperacyjne, odtwarzane).

Niemalby nacisk położono na kwestie związane z właściwym zarządzaniem danymi wytworzonymi w procesie badawczym poprzez tworzenie planów zarządzania danymi (*Data Management Plan – DMP*). Plan taki ma określać: schemat zarządzania danymi badawczymi w trakcie programu i po jego zakończeniu, identyfikację i nazwy zestawów danych, opisy, standardy i metadane, sposoby wykorzystywania danych, wskazanie, które dane zostaną udostępnione, archiwizacja i zabezpieczenie (w tym przechowywanie i kopie zapasowe).

Programowi ORD przyświeca idea, iż nauka powinna być *tak otwarta jak tylko się da, zamknięta nie bardziej niż to konieczne*¹⁶. Nie jest wymagane udostępnianie wszystkich zebranych danych, ale takich, które są potrzebne do weryfikacji wyników badań zaprezentowanych w publikacji naukowej. Badacz ma możliwość samodzielnego zdecydowania, które dane przez niego zgromadzone znajdują się w publicznie dostępnych zasobach, jednak z zachowaniem zasady weryfikowalności wniosków. Zalecane jest udostępnianie także metadanych, które stanowią opis prezentowanego zbioru danych.

Możliwe jest wyłączenie się z pilotażu jednak należy uzasadnić swoją decyzję. Sytuacje, w których zebrane dane badawcze nie muszą zostać publicznie udostępnione mogą odnosić się do: planów ich komercyjnego lub przemysłowego wykorzystania, wymogów poufności związanych z bezpieczeństwem, zasad ochrony danych osobowych, niemożności osiągnięcia głównego celu działań lub jeśli nie zostaną wytworzone żadne dane¹⁷.

¹⁵ *Guidelines on FAIR Data Management in Horizon 2020* [online]. European Commission, Directorate-General for Research & Innovation, 2016. [Dostęp 22.06.2017]. Dostępny w: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

¹⁶ *Guidelines on FAIR...*, dz. cyt., s. 4.

¹⁷ JONES, S. Pilotaż otwartych danych badawczych w programie Horyzont 2020: jakie są wymagania? W: *Foster* [online]. S. 5. [Dostęp 22.06.2017]. Dostępny w: <https://www.fosteropenscience.eu/node/1234>.

Według wymagań ORD grantobiorcy powinni deponować dane w repozytoriach danych badawczych, a także podjąć starania umożliwiające innym badaczom i osobom zainteresowanym dotarcie do tych danych, ich maszynową analizę, ponowne wykorzystanie, kopiowanie i rozpowszechnianie bez konieczności ponoszenia dodatkowych opłat. Zalecane jest także zamieszczenie informacji o narzędziach niezbędnych do weryfikacji danych, a najlepszym rozwiązaniem będzie dostarczenie tego narzędzia¹⁸.

W polskim systemie prawodawstwa najbardziej aktualne zalecenia dotyczące udostępniania danych badawczych to dokument Ministerstwa Nauki i Szkolnictwa Wyższego (MNiSW) z października 2015 r. *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*. Czytamy w nim, że ministerstwo zaleca, aby krajowe podmioty finansujące badania naukowe ze środków publicznych (...) stosowały i upowszechniały zasady, zgodnie z którymi publikacje i dane badawcze powstające w wyniku finansowanych lub współfinansowanych przez nie badań znajdują się w otwartym dostępie¹⁹. Realizacja tych postanowień skierowana jest do instytucji dotujących badania, jednostek naukowych i szkół wyższych. Każda instytucja powinna określić własną politykę w zakresie dostępu do publikacji i danych badawczych zgodną z zasadami otwartości.

Ministerstwo Nauki i Szkolnictwa Wyższego w swoich wytycznych w dużej mierze odwołuje się do zasad przyjętych w programie Horyzont 2020²⁰. Sugeruje, żeby dane przed udostępnieniem były wyselekcjonowane w taki sposób, aby nie spowodowały naruszenia tajemnic ani interesów związanych z komercjalizacją. Należy mieć także na uwadze kwestie związane z ochroną danych osobowych, tajemnicą przedsiębiorstwa, bezpieczeństwem narodowym, uzasadnionymi interesami handlowymi oraz prawami własności intelektualnej. *Docelowo powszechny dostęp do wyników badań naukowych (zarówno publikacji, jak i danych badawczych) (...) powinien być ogólną zasadą, a stosowanie wyłączeń od tej zasady powinno wynikać z istnienia obiektywnych, uzasadnionych przesłanek*²¹. MNiSW powtarza za Komisją Europejską, iż udostępniane zbiory danych powinny być oznaczone w sposób umożliwiający ich łatwe zidentyfikowanie oraz zaopatrzone w dodatkowe informacje, które pozwolą na ich właściwą kategoryzację, ocenę i wykorzystanie²². W dokumencie *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce* wskazane zostały cele udostępniania danych: *podnoszenie jakości danych naukowych, ograniczenie konieczności powielania badań, przyspieszenie postępu naukowego, wspieranie innowacyjności przedsiębiorstw, walka z nadużyciami w dziedzinie nauki oraz umożliwienie ponownego wykorzystania danych*²³.

Wytyczne powyższego dokumentu zostały potwierdzone w lutym 2017 r. w liście Ministra Nauki i Szkolnictwa Wyższego, Jarosława Gowina: *otwieranie wyników badań*

¹⁸ JONES, S., dz. cyt., s. 8.

¹⁹ *Kierunki rozwoju otwartego...*, dz. cyt., s. 9.

²⁰ Tamże, s. 8.

²¹ Tamże.

²² Tamże, s. 17.

²³ Tamże.

naukowych (publikacje, dane badawcze) sprzyja rozwojowi nauki i innowacji oraz przyczynia się do lepszego rozumienia znaczenia nauki przez obywateli²⁴.

Otwarte dane badawcze w Polsce to jak na razie zalecenia a nie powszechna praktyka czy obowiązek. Brak jest ogólnych uregulowań oraz decyzji dotyczących stworzenia krajowej otwartej infrastruktury danych i publikacji. Nakaz udostępniania danych może zostać nałożony przez instytucje finansujące prace naukowe w umowach grantowych (zwłaszcza ze środków publicznych) – w pewnym zakresie praktyki są stosowane w programie Horyzont 2020. Spotykane są także przypadki wydawców naukowych, którzy zaostrzają swoją politykę wydawniczą i nie dopuszczają do pojawienia się w artykule stwierdzeń, które poparte są niedostępnymi danymi (schowanymi za sformułowaniem *data not shown*)²⁵. Pomimo przykładów pozytywnych działań propagujących ideę otwartych danych badawczych, skala tego zjawiska nadal jest niewystarczająca. Wydaje się, że najskuteczniejszym rozwiązaniem byłyby właściwe zarządzenia na szczeblu krajowym. Jednak wszystkie podmioty zainteresowane usprawnieniem produkcji i dystrybucji wiedzy na różnych szczeblach: naukowcy, instytucje, organy rządzące, jednostki międzynarodowe, powinny być zainteresowane tymi kwestiami. Choć nauka jest przedsięwzięciem międzynarodowym, to jej realizacja odbywa się w ramach odrębnych krajowych systemów odpowiedzialności, organizacji i zarządzania²⁶.

Repozytoria danych badawczych

Otwarte repozytoria stanowią jeden z ważnych elementów nowoczesnego i efektywnego systemu dystrybucji i zarządzania wiedzą, promocji instytucji naukowych, naukowców oraz ich osiągnięć²⁷. Dane badawcze to obiekty o zupełnie innej specyfice niż książki czy czasopisma. O ile deponowanie publikacji naukowych jest już praktyką od dawna znaną i realizowaną, o tyle lokowanie danych badawczych w otwartych repozytoriach to jeszcze sprawa świeża, zwłaszcza na gruncie polskim. Możemy odnaleźć już sporą liczbę europejskich i światowych repozytoriów umożliwiających zdeponowanie, zarchiwizowanie i udostępnienie danych badawczych. Międzynarodowa wyszukiwarka indeksująca repozytoria danych badawczych ze wszystkich dziedzin wiedzy – re3data.org – rejestruje obecnie ok. 1500 repozytoriów²⁸. Najwięcej jest repozytoriów z zakresu nauk przyrodniczych (biologii, medycyny, nauki o Ziemi) – 955. Dużą grupę stanowią repozytoria z obszaru nauk humanistycznych i społecznych (534). Część repozytoriów może być zaliczona do kilku dziedzin wiedzy, istnieją także repozytoria interdyscyplinarne czy też ogólne.

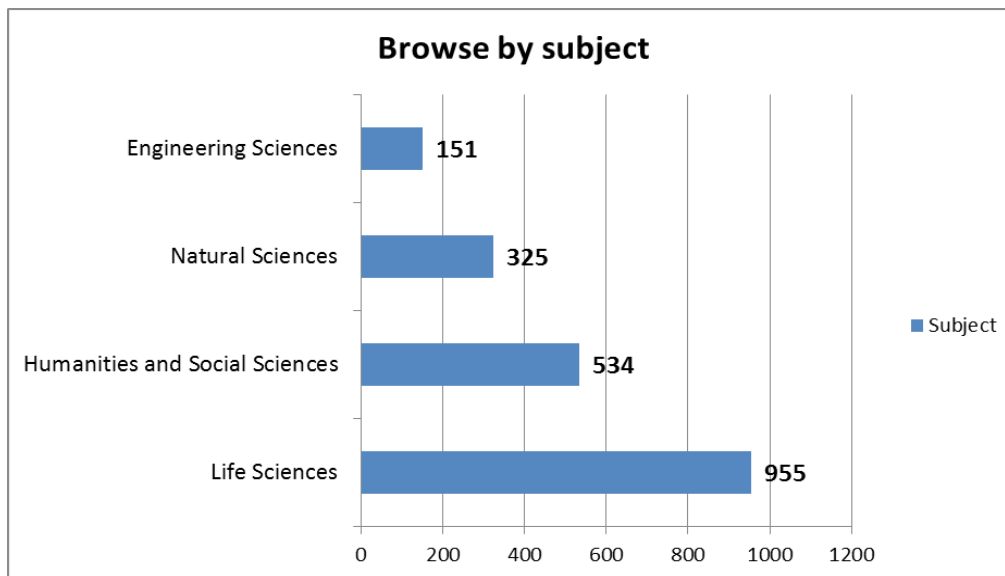
²⁴ List Ministra Nauki i Szkolnictwa Wyższego z dnia 10.02.2017 r. dotyczący: otwartego dostępu do publikacji naukowych [online]. [Dostęp 22.06.2017]. Dostępny w: http://www.nauka.gov.pl/g2/oryginal/2017_02/6b8c8c9c87a816fff3bd6e186b194b9e.pdf.

²⁵ SZCZĘSNY, P., dz. cyt., s. 15.

²⁶ *Otwarte dane w świecie...*, dz. cyt., s. 2.

²⁷ *Kierunki rozwoju otwartego ...*, dz. cyt., s. 11.

²⁸ re3data.org [online]. Stan w dn. 11.07.2017 r. [Dostęp 11.07.2017]. Dostępny w: <http://www.re3data.org/>.



Rys. 1. Liczba repozytoriów danych badawczych zewidencjonowanych w re3data.org według dziedzin wiedzy (stan na 12.07.2017)

Źródło: opracowanie własne na podstawie re3data.org.

Ze względu na specyficzne cechy możemy wyróżnić repozytoria:

- specjalistyczne (wąskie dziedzinowo): np. GenBank (amerykańska bioinformatyczna baza danych), Protein Data Bank (biologiczna baza danych gromadząca dane o strukturze przestrzennej białek i kwasów nukleinowych), British Atmospheric Data Centre (brytyjskie centrum informacji o atmosferze zawierające dane atmosferyczne),
- instytucjonalne/krajowe: British Oceanographic Data Centre (repozytorium tematyczne prowadzone przez brytyjską instytucję finansującą badania), Data Archiving and Networked Services (wielodziejzinowy portal naukowy Holandii tworzony przez koalicję naukowych organizacji rządowych tego kraju),
- tematyczne: UK Data Archive (repozytorium danych z nauk społecznych i humanistycznych), SeaDataNet (międzynarodowy projekt oceanograficzny) ADS Archiwum Danych Społecznych (tworzone od 2003 r. przez Uniwersytet Warszawski i Instytut Filozofii i Socjologii Polskiej Akademii Nauk),
- ogólne: Zenodo, FigShare, World Data System (otwarte dane z różnych dziedzin, stworzone w 2010 r. przez Międzynarodową Organizację Naukową International Council for Science), OpenDOAR, RepOD.

Komisja Europejska doradza korzystanie z re3data.org podczas poszukiwań właściwego repozytorium dla zdeponowania danych wytworzonych w ramach programu Horyzont 2020. Innym narzędziem, którym można się posłużyć jest OpenDOAR²⁹ – baza również o zasięgu międzynarodowym, która poza repozytoriami danych badawczych indeksuje także biblioteki cyfrowe i repozytoria instytucjonalne. Baza daje możliwość wyszukiwania zarówno samych repozytoriów, jak i przeszukiwania ich zawartości. Przykładowym repozytorium, w którym można zdeponować dane badawcze jest interdyscyplinarne repozytorium Zenodo stworzone przez OpenAIRE³⁰ i Europejską Organizację Badań Jądrowych CERN. Repozytorium to zostało uruchomione

²⁹ OpenDOAR [online]. [Dostęp 5.08.2017]. Dostępny w: <http://www.opendoar.org/>.

w celu zamieszczania danych z programów unijnych. Przyjmuje wiele rodzajów danych, publikacji oraz oprogramowania, nadaje identyfikator DOI, łączy informacje o źródłach finansowania z publikacjami, danymi i oprogramowaniem. Repozytorium gromadzącym otwarte dane z zakresu nauk humanistycznych jest FigShare. Inicjatywa ta zasługuje na uwagę ze względu na znaczne możliwości, jakie daje naukowcom publikującym dane badawcze w obrębie serwisu, m.in. możliwość dodawania różnych formatów, stały adres danych³¹. Polska inicjatywa to RepOD³² repozytorium otwartych danych stworzone przez ICM UW w ramach działań Platformy Otwartej Nauki. Archiwizuje i udostępnia wszystkie dane wytworzone, zebrane i opracowane na potrzeby badań naukowych. Przeznaczone jest dla tzw. małych danych. Obecnie w repozytorium zdeponowano 40 zbiorów danych.

Zapotrzebowanie na otwarte dane

Politechnika Warszawska, na potrzeby wniosku o dofinansowanie własnego projektu Sieć Instytucjonalnych Baz Wiedzy (SINBAW) w ramach Programu Operacyjnego Polska Cyfrowa, przeprowadziła poglądowe badania ankietowe skierowane do pracowników instytucji naukowych i badawczo-rozwojowych, szkół wyższych, reprezentantów grupy przedsiębiorców, pracowników administracji publicznej oraz społeczności akademickiej PW. Zebrano opinie od 424 osób (65% próby to naukowcy, 15% studenci, 6% doktoranci, 7% przedsiębiorcy, 7% pracownicy administracji publicznej). Celem badania było określenie popytu na różne typy informacji, sposób ich udostępniania, a także przeprowadzona przez użytkowników ocena funkcjonującego już w PW Systemu Bazy Wiedzy³³.

W ankiecie zapytano m.in. jakim typem informacji udostępnianych w otwartych zasobach użytkownicy są zainteresowani. *Wyniki badań naukowych – dane badawcze* były drugim wskazywanym zasobem (po publikacjach naukowych). W znaczący spo-

³⁰ Projekt OpenAIRE został zainaugurowany w 2009 r. jako narzędzie służące Komisji Europejskiej do monitorowania badań finansowanych przez Europejską Radę ds. Badań Naukowych (ERC) oraz siódmy program ramowy (7PR), a w szczególności do monitorowania stopnia realizacji wytycznych dotyczących otwartego dostępu zawartych w Pilotażu Otwartego Dostępu w 7PR i w ERC Guidelin W: MA-NOLA, N. *OpenAIRE – e-infrastruktura umożliwiająca publikowanie wyników badań naukowych w otwartym dostępie* [online]. [Dostęp 22.06.2017]. Dostępny w: <https://otwartanauka.pl/analysis/nauka-otwartosc-swiat/openaire-e-infrastruktura-umozliwiajaca-publikowanie-wynikow-badan-naukowych-w-otwartym-dostepie>.

³¹ KARWOWSKI, M. Otwarte dane – przyszłość humanistyki? W: *Tematy modne w humanistyce. Studia interdyscyplinarne* [online]. Toruń, 2015, s. 267. [Dostęp 22.06.2017]. Dostępny w: https://issuu.com/prolog_pismo/docs/tematy-modne-w-humanistyce__1.

³² RepOD. *Repozytorium Otwartych Danych* [online]. [Dostęp 5.08.2017]. Dostępny w: <https://repod.pon.edu.pl/pl>.

³³ Badania zostały wykonane przy użyciu kwestionariusza ankiety dystrybuowanego za pośrednictwem poczty elektronicznej. Dobór próby oparty był na dostępności badanych, starano się dotrzeć do wszystkich potencjalnych grup odbiorców projektu SINBAW. Badania miały charakter poglądowy. Ankieta zawierała 5 pytań głównych oraz metryczkę (status respondentów, typ instytucji i jej charakterystyka, wielkość miejscowości). W odpowiedziach na pytania główne poziom zainteresowania lub poparcia danej kwestii należało określić korzystając z 10-cio punktowej skali: 0 – w ogóle się nie zgadzam/ w ogóle nie jestem zainteresowany, 10 – całkowicie się zgadzam/ jestem zainteresowany w najwyższym stopniu.

sób zasób ten interesuje pracowników naukowych oraz studentów (mierniki poziomu zainteresowania to odpowiednio 8,44 i 8,33³⁴).

Tab. 1. Poziom zainteresowania otwartym dostępem do różnego typu informacji

	Rozprawy doktorskie	Publikacje naukowe	Inf. o patentach	Inf. o produktach (technologie, oprogramowanie)	Inf. o urządzeniach wysokospecjalistycznych	Inf. o laboratoriach do wykorzystania komercyjnego	Inf. o obiektach muzealnych	Inf. o zrealizowanych projektach	Wyniki badań naukowych – dane badawcze	Inf. o pracownikach naukowych i ich dorobku	Inf. o ekspertach i obszarze ich ekspertyzy
naukowcy	7,48	9,50	6,02	6,36	6,45	6,38	4,37	6,74	8,44	7,48	7,43
przedsiębiorcy	5,8	8,25	7,2	7,35	7	7,85	3,25	6,3	7,6	6,3	7,1
pracownicy administracji publicznej	4,45	6,59	3,77	5,91	4,68	4,05	4,59	5,68	5,82	5,00	6,18
doktoranci	7,74	9,37	6,00	6,37	6,68	6,32	3,84	5,63	7,95	7,32	7,53
studenci	6,49	8,89	6,47	7,13	6,47	6,27	4,20	6,00	8,33	6,89	6,93
ogólnie	6,75	8,82	5,80	6,41	6,30	6,20	4,34	6,43	8,02	7,00	7,13

Źródło: opracowanie własne.

Najistotniejszą kwestią dla respondentów podczas korzystania z otwartych zasobów nauki jest nieodpłatny dostęp do informacji (ogólny miernik stopnia zgodności³⁵ ze stwierdzeniem to 9,3). Ważne okazało się także to, aby dane były udostępniane w popularnych i łatwych do odczytu formatach oraz aby można je było pobrać i zapisać na własny komputer (ogólne mierniki stopnia zgodności to 8,48 i 8,36). Wysoki poziom akceptacji uzyskały także kwestie dotyczące zapotrzebowania na dane w postaci opracowanej (interpretacje oraz wyniki) a także obowiązek nieodpłatnego udostępniania przez publiczne jednostki naukowe wyników swoich prac.

³⁴ Miernik poziomu zainteresowania jest to średnia arytmetyczna z ocen wystawionych w skali od 0 do 10, gdzie 0 oznacza „w ogóle nie jestem zainteresowany”, a 10 „jestem zainteresowany w najwyższym stopniu”.

³⁵ Miernik stopnia zgodności to średnia arytmetyczna z ocen wystawionych w skali od 0 do 10, gdzie 0 oznacza „w ogóle się nie zgadzam”, a 10 „całkowicie się zgadzam”.

Tab. 2. Poziom zgodności ze stwierdzeniami związanymi z otwartymi zasobami nauki

	Interesują mnie dane w postaci surowej	Interesują mnie dane w postaci opracowanej	Interesują mnie dane udostępniane nieodpłatnie	Jestem gotowy zapłacić za dostęp do danych	Interesują mnie narzędzia analityczne na poziomie systemu udostępniającego dane	Możliwość pobrania danych na własny komputer jest dla mnie istotna	Przeglądanie danych online bez możliwości ich pobrania na własny komputer jest całkowicie wystarczające	Jestem zainteresowany dostępem do narzędzi wizualizujących dane	obiektów. Jestem zainteresowany zaawansowanymi możliwościami wielokryterialnego wyszukiwania	Jestem zainteresowany wyszukiwaniem obejmującym jednocześnie różne instytucje i obszary specjalizacji	Wysoka jakość wizualizacji danych jest dla mnie ważna	Udostępnianie danych w popularnych formatach (PDF, XML) jest bardzo ważne	Format udostępnianych danych jest bez znaczenia	To bardzo ważne, aby publiczne jednostki naukowe nieodpłatnie udostępniały wyniki swoich prac
naukowcy	6,51	8,48	9,54	3,81	6,34	8,52	4,16	7,10	7,87	7,93	7,10	8,76	3,15	8,47
przedsiębiorcy	6,40	8,65	9,40	5,50	6,05	8,70	4,70	6,70	7,50	7,60	6,35	8,40	3,80	8,05
pracownicy admin. publicznej	3,41	7,09	8,59	2,14	4,68	7,09	4,95	5,86	6,41	6,50	5,50	7,41	2,36	7,23
doktoranci	7,74	8,68	9,21	3,74	5,79	8,58	4,05	6,95	7,58	7,53	6,16	8,32	3,58	7,42
studenci	6,82	8,13	9,40	3,71	6,62	8,38	4,16	7,24	7,73	7,91	6,53	8,49	2,98	8,22
ogólnie	6,31	8,25	9,30	3,83	6,15	8,36	4,31	6,92	7,63	7,73	6,75	8,48	3,19	8,13

Źródło: opracowanie własne.

Przeprowadzone ankietowe badania poglądowe poparły hipotezę, iż zainteresowanie dostępem do otwartych danych jest duże i to w różnych grupach interesariuszy. Za-

sób ten może być przydatny nie tylko naukowcom i środowisku akademickiemu, ale także sektorowi przemysłowemu. Aspekty związane z otwieraniem danych – popularne i otwarte formaty zapisu, możliwość pobrania danych na własne nośniki – to kwestie, które znalazły znaczne poparcie wśród respondentów.

Ciekawych analiz stron internetowych polskich instytutów naukowo-badawczych, pod kątem zamieszczania i udostępniania na nich danych badawczych, dokonał w 2015 r. Adam Jachimczyk³⁶. Według jego analiz sytuacja w Polsce nie wygląda źle: 92% spośród zidentyfikowanych zbiorów danych jest umieszczona w internecie, z czego 75% jest dostępna online bez ograniczeń³⁷. Jednak według badań ankietowych zrealizowanych w 2014 r. przez Bibliotekę Główną PW (BG PW)³⁸ wśród, nie tylko instytucji naukowych i badawczo-rozwojowych, ale także instytucji PAN oraz publicznych uczelni wyższych, istnieje spora grupa instytucji, które nie zamieszczają wyprodukowanych przez siebie danych w internecie, a przechowują je wewnątrz instytucji (na stanowiskach pracy pracowników lub centralnym serwerze) – 79% respondentów³⁹. Te instytucje nie zostały uwzględnione w analizach A. Jachimczyka.

Z badań wynika, iż instytucje naukowe udostępniają głównie metadane (opisy bibliograficzne, opisy osób, organizacji, projektów), jedynie 40% danych zamieszczonych na stronach internetowych instytucji to dane surowe⁴⁰. Zauważalny jest także brak dbałości o otwarty format zapisu danych, w większości przypadków (w obydwu omawianych badaniach) wykazano, iż częstymi formatami zapisu plików są formaty (HTML, PDF) wymuszające dalszą, często pracochłonną konwersję plików do innych formatów, które umożliwią ich przetwarzanie i analizy. Praktyka zamieszczania informacji o warunkach dostępności danych jest jeszcze słabo stosowana. Według Jachimczyka 84% instytucji udostępniających dane online nie podaje informacji o warunkach ich udostępniania. Według badań BG PW 82% respondentów wskazało, iż w ich instytucjach nie stosuje się publicznych licencji dla udostępniania danych.

E-infrastruktura otwartej nauki

Idea otwartej nauki opiera się na założeniu, iż zbiór przydatnej wiedzy to nie tylko treść publikacji naukowych, ale wiedza produkowana na każdym etapie działalności badawczej. Zapewnienie otwartego dostępu nie tylko do publikacji, ale do całokształtu procesu naukowego, na wszystkich etapach pracy i komunikacji to działania, które powinny być podejmowane w dążeniu do otwierania nauki⁴¹. Poza umieszczeniem treści i danych w internecie ważne jest także stworzenie otwartej infrastruktury, która pozwoli na swobodne przeszukiwanie oraz łączenie ze sobą różnych źródeł, a także zapewni odpowiednią strukturę i opis zawartych w niej materiałów⁴². Tworzenie e-infrastruktur to konstruowanie całego zaplecza badawczego, niezależnego od barier

³⁶ JACHIMCZYK, A. Otwarte dane badawcze. Casus polskich instytutów badawczych. *Zagadnienia Naukoznawstwa* 2015, r. 206, nr 4, s. 409–424. ISSN 0044-1619.

³⁷ JACHIMCZYK, A., dz. cyt., s. 413.

³⁸ KĘDZIERSKA, E. Wzmocnione publikacje: nowy model wsparcia komunikacji naukowej. *Zagadnienia Informacji Naukowej* 2015, vol. 53, nr 1, s. 44–58. ISSN 0324-8494.

³⁹ Tamże, s. 50.

⁴⁰ JACHIMCZYK, A., dz. cyt., s. 418.

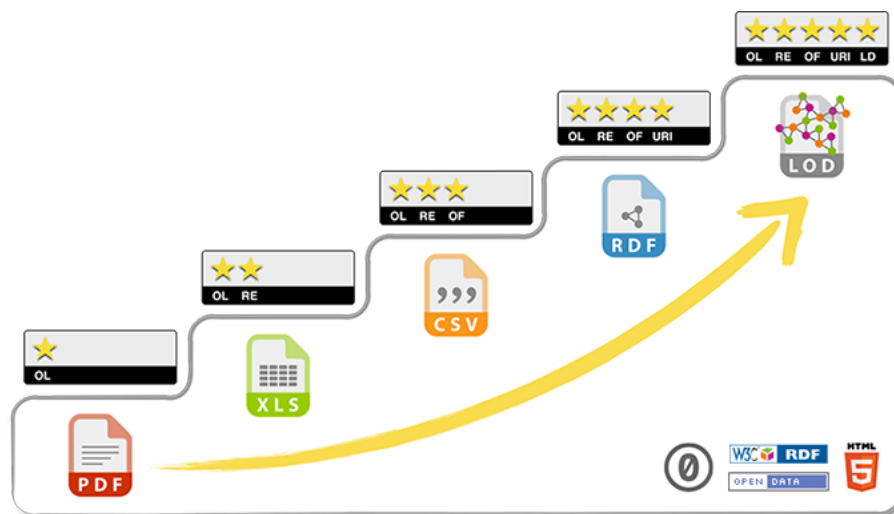
⁴¹ HOFMOKL, J., TARKOWSKI, A., BEDNAREK-MICHALSKA, B. i in., dz. cyt., s. 54.

⁴² Tamże, s. 55.

czasowych czy geograficznych. To nowy sposób dokonywania odkryć naukowych, który opiera się na narzędziach umożliwiających eksplorowanie już opisanej wiedzy i odkrywanie nowych aspektów i wymiarów.

Aby w pełni korzystać z możliwości, jakie dają otwarte dane badawcze nie powinny pozostać one „osamotnione”. Należy je osadzić w odpowiednim kontekście, czyli połączyć z pozostałymi wytworami procesu naukowego: z publikacjami, dokumentacją z badań, materiałami pomocniczymi, ale także z innymi zasobami danych i publikacjami o podobnej tematyce. Dane takie zostały określone jako *Linked Open Data* (otwarte połączone dane), czyli dane nie tylko opublikowane i udostępnione, ale także skomunikowane z innymi materiałami. Termin ten odnosi się do praktyk związanych z publikowaniem i łączeniem strukturalnych danych w sieci⁴³. Tim Berners-Lee wyznaczył ścieżkę rozwoju otwartych danych w stronę otwartych połączonych danych w postaci modelu pięciu gwiazdek (*5-star Open Data*)⁴⁴:

- ★ OL (Open Licenses) – otwarte licencje
- ★★ RE (Reusable) – gotowe do ponownego przetwarzania
- ★★★ OF (Open Formats) – otwarte formaty
- ★★★★ URI (Uniform Resource Identifiers) – stałe identyfikatory
- ★★★★★ LD (Linked Data) – kontekstualizacja



Rys. 2. Model rozwoju *Linked Open Data*

Źródło: 5 * Open Data [online]. [Dostęp 22.06.2017]. Dostępny w: <http://5stardata.info/en/>.

Końcowym rezultatem tego modelu ma być otrzymanie danych, które można odszukać, pobrać, przetworzyć za pomocą powszechnie dostępnego oprogramowania, ponownie używać, łączyć z innymi zbiorami. Te cechy *Linked Open Data* mają ułatwić odnajdywanie danych, zwiększyć ich wartość, umożliwić odkrywanie nowych danych. Podobny model tworzenia powiązań i sieci relacyjnej między zasobami nauki proponuje pojęcie *enhanced publications* – wzmocnione publikacje. Są to złożone obiekty cyfrowe, w skład których wchodzi poza typowymi publikacjami naukowymi (książka-

⁴³ PAWEŁOSZEK, I. Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji. *Roczniki KAE* [online]. 2014, z. 33, s. 462. [Dostęp 22.06.2017]. ISSN 1232-4671. Dostępny w: http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z33_26.pdf.

⁴⁴ Tamże, s. 466.

mi, artykułami, raportami) połączone z nimi dodatkowe elementy, które zostały wytworzone w trakcie procesu badawczego np. dane badawcze (surowe i przetworzone), nagrania filmowe i dźwiękowe, algorytmy, plany, modele i inne materiały wizualne, ale także notatki czy materiały niepublikowane (formularze zgody, opisy pomiarów) oraz wszelkiego rodzaju recenzje, komentarze czy też dyskusje toczące się wokół badań⁴⁵. Owe dodatkowe elementy dołączone do publikacji mają na celu pokazanie pełnego kontekstu badań, uwiarygadniają uzyskane wyniki i wyciągnięte wnioski, zwiększają czytelność, pomagają zrozumieć badania, pozwalają na weryfikację i wielokrotne wykorzystanie danych i materiałów badawczych⁴⁶. Ważnym elementem przy tworzeniu wzmocnionych publikacji nie jest samo zamieszczenie materiałów w internecie, ale ich wzajemne powiązanie i zbudowanie właściwej struktury relacyjnej. Dennis Adriaansen i Jürgen Hooft⁴⁷ wymienili typowe elementy, którymi powinny odznaczać się wzmocnione publikacje: załączone dane badawcze lub informacja o miejscu ich zdeponowania, dodatkowe materiały, interaktywne treści (tabele, mapy), linki do cytowanych publikacji, wskazanie publikacji nas cytujących, komentarze, dyskusje, recenzje, relacje z innymi publikacjami, plikami, nawigacja po publikacji, metadane, otwarte formaty, możliwość rozwoju. Pojęcie wzmocnionych publikacji na arenie międzynarodowej pojawia się już od jakiegoś czasu. W wyszukiwarce repozytoriów danych badawczych re3data.org jednym z filtrów zawężających wyniki wyszukiwania jest kategoria *enhanced publication*, która wskazuje czy dane repozytorium zawiera publikacje tego typu, czy nie (lub też wartość tego parametru może być nieznana). Według tej wyszukiwarki występuje 636 repozytoriów⁴⁸ zawierających wzmocnione publikacje.

Podsumowanie

W ostatnich latach nastąpił rozwój w dziedzinie generowania, magazynowania, przetwarzania i natychmiastowego przesyłania danych o ogromnych rozmiarach i dużej złożoności, co ma znaczne następstwa dla nauki⁴⁹. Jednak zagadnienie otwartego dostępu do danych badawczych jest złożone zarówno ze względów technicznych, organizacyjnych, jak i prawnych. Mimo różnych przeszkód i ograniczeń korzyści płynące z otwierania danych badawczych są niezaprzeczalne:

- obniżenie kosztów pracy naukowej: niepowielanie wysiłków, zmniejszenie kosztów finansowych, redukcja konieczności ponownego gromadzenia materiału badawczego (np. względy etyczne – badania na zwierzętach),
- tworzenie nowego materiału badawczego, wzbogacenie danych (łączenie danych),
- samokorygowanie nauki – weryfikacja danych, wychwytywanie nierzetelności w nauce, poprawa jakości,
- nawiązywanie i rozwój współpracy między badaczami, instytucjami, zespołami (współpraca interdyscyplinarna),

⁴⁵ KĘDZIERSKA, E., dz. cyt., s. 46.

⁴⁶ Tamże.

⁴⁷ ADRIAANSEN, D., HOOFT, J. *Properties of enhanced publications and the supporting tools* [online]. S. 2. [Dostęp 22.06.2017]. Dostępny w: <http://www.cs.uu.nl/docs/vakken/mdic/papers/Adriaansen-fin.pdf>.

⁴⁸ *re3data.org* [online]. Stan w dn. 7.07.2017 r. [Dostęp 11.07.2017]. Dostępny w: <http://www.re3data.org/>.

⁴⁹ *Otwarte dane w świecie ...*, dz. cyt., s. 2.

- większa liczba kanałów dostępu dla użytkownika – wzrost docieralności, zwiększenie wyszukiwalności i trafności,
- szerszy krąg odbiorców wyników pracy badawczej (przedsiębiorcy, administracja, obywatele),
- wzrost wartości marki,
- realizacja misji publicznej – dorobek wyprodukowany ze środków publicznych dostępny dla obywateli,
- wykorzystanie danych przez podmioty gospodarcze – rozwój innowacji,
- szybszy przepływ informacji (ważny w niektórych dziedzinach np. medycyna), usprawnienie komunikacji, pobudzenie dyskusji, wymiana opinii,
- bezpieczna archiwizacja (repozytoria/archiwa), dane są opisane i udokumentowane.

Do rozwiązania jest jeszcze wiele kwestii związanych z przechowywaniem, archiwizowaniem i udostępnianiem danych badawczych, biorąc pod uwagę fakt, iż ich specyfika jest odmienna od typowych publikacji a procedury i praktyki zarządzania danymi badawczymi nie są jeszcze unormowane. Udostępnianie w otwarty sposób 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 krąg odbiorców, komentatorów, uczestników badań, także podmiotów badań⁵⁰.

Obecnie otwartość polskich zasobów nauki jest fragmentaryczna, ponieważ nie ma spójnej strategii otwierania nauki na poziomie krajowym czy instytucjonalnym. Dostępne otwarte zasoby są wynikiem starań pojedynczych jednostek lub grup działających na rzecz otwartości⁵¹. Zachodzi jednak wiele pozytywnych zmian. Koncepcję otwartych danych można będzie skutecznie zrealizować tylko wtedy, gdy podejmowane będą systemowe działania na poziomach indywidualnym, dyscyplinarnym, krajowym i międzynarodowym⁵².

Działania na przyszłość, które mogą poprawić obecną sytuację i doprowadzić do zwiększenia otwartości polskich zasobów wiedzy to:

- uregulowania prawne na poziomie krajowym,
- wypracowanie odpowiedniego modelu prawnego, niwelowanie barier stawianych przez system własności intelektualnej,
- usankcjonowanie wiedzy jako dobra publicznego,
- jasne wytyczne dotyczące przechowywania i udostępniania danych (licencje, standardy metadanych, trwałe identyfikatory DOI),
- włączenie publikacji danych badawczych do systemu oceny dorobku naukowego,

⁵⁰ BEDNAREK-MICHALSKA, B., *Otwarta nauka. Open Access* [online]. Warszawa: Koalicja Otwartej Edukacji, 2010, s. 16. [Dostęp 22.06.2017]. Dostępny w: <http://kpbk.umk.pl/dlibra/doccontent?id=52395>.

⁵¹ SZCZĘSNY, P., dz. cyt., s. 24.

⁵² *Otwarte dane w świecie ...*, dz. cyt., s. 2.

- pokonywanie barier technicznych (automatyczne metody agregacji, wyszukiwania i opracowania danych),
- wspieranie i zachęcanie naukowców do znajdowania oraz wielokrotnego użycia danych do ich dalszych badań.

Bibliografia:

1. ADRIAANSEN, D., HOOFT, J. *Properties of enhanced publications and the supporting tools* [online]. [Dostęp 22.06.2017]. Dostępny w: <http://www.cs.uu.nl/docs/vakken/mdic/papers/Adriaansen-fin.pdf>.
2. BEDNAREK-MICHALSKA, B. *Otwarta nauka. Open Access* [online]. Warszawa: Koalicja Otwartej Edukacji, 2010. [Dostęp 22.06.2017]. Dostępny w: <http://kpbpc.umk.pl/dlibra/doccontent?id=52395>.
3. BEDNAREK-MICHALSKA, B. Repozytoria surowych danych – dlaczego biblioteki powinny je znać? *Biuletyn EBIB* [online]. 2012, nr 8 (135). [Dostęp 22.06.2017]. ISSN 1507-7187. Dostępny w: <http://open.ebib.pl/ojs/index.php/ebib/article/view/152/292>.
4. GUIBAULT, L., WIEBEGO, A. (red.). *Otwarte, czyli bezpieczne. Analiza prawnych form ochrony danych badawczych i zalecenia w kwestii dostępu i korzystania* [online]. Kraków: Biblioteka Jagiellońska, 2014. [Dostęp 22.06.2017]. ISBN 978-83-934926-8-8. Dostępny w: <http://jbc.bj.uj.edu.pl/dlibra/doccontent?id=283369>.
5. *Guidelines on FAIR Data Management in Horizon 2020* [online]. European Commission. Directorate-General for Research & Innovation, 2016. [Dostęp 22.06.2017]. Dostępny w: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
6. HOFMOKL, J., TARKOWSKI, A., BEDNAREK-MICHALSKA, B. i in. *Przewodnik po otwartej nauce* [online]. Warszawa: Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego, 2009. [Dostęp 22.06.2017]. ISBN 978-83-917150-4-8. Dostępny w: <https://otwartanauka.pl/images/PDFs/przewodnik-po-otwartej-nauce.pdf>.
7. JACHIMCZYK, A. Otwarte dane badawcze. Casus polskich instytutów badawczych. *Zagadnienia Naukoznawstwa* 2015, r. 206, nr 4, s. 409–424. ISSN 0044-1619.
8. JAMES, L. Defining Open Data. W: *Open Knowledge International Blog* [online]. [Dostęp 22.06.2017]. Dostępny w: <http://blog.okfn.org/2013/10/03/defining-open-data/>.
9. JONES, S. Pilotaż otwartych danych badawczych w programie Horyzont 2020: jakie są wymagania? W: *Foster* [online]. [Dostęp 22.06.2017]. Dostępny w: <https://www.fosteropenscience.eu/node/1234>.
10. KARWOWSKI, M. Otwarte dane – przyszłość humanistyki? W: *Tematy modne w humanistyce. Studia interdyscyplinarne* [online]. Toruń, 2015. [Dostęp 22.06.2017]. Dostępny w: https://issuu.com/prolog_pismo/docs/tematy-modne-w-humanistyce__1.
11. KĘDZIERSKA, E. Wzmocnione publikacje: nowy model wsparcia komunikacji naukowej. *Zagadnienia Informacji Naukowej* 2015, vol. 53, nr 1, s. 44–58. ISSN 0324-8494.
12. *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce* [online]. 2015. [Dostęp 22.06.2017]. Dostępny w: http://www.nauka.gov.pl/g2/oryginal/2015_10/9f62cc350837b942e51ae23dd1f23df8.pdf.
13. *List Ministra Nauki i Szkolnictwa Wyższego z dnia 10.02.2017 r. dotyczący: otwartego dostępu do publikacji naukowych* [online]. [Dostęp 22.06.2017]. Dostępny w: http://www.nauka.gov.pl/g2/oryginal/2017_02/6b8c8c9c87a816fff3bd6e186b194b9e.pdf.
14. MANOLA, N. *OpenAIRE – e-infrastruktura umożliwiająca publikowanie wyników badań naukowych w otwartym dostępie* [online]. [Dostęp 22.06.2017]. Dostępny w: <https://otwartanauka.pl/analysis/nauka-otwartosc-swiat/openaire-e-infrastruktura-umozliwiajaca-publikowanie-wynikow-badan-naukowych-w-otwartym-dostepie>.
15. *OpenDOAR* [online]. [Dostęp 5.08.2017]. Dostępny w: <http://www.opendoar.org/>.
16. *Otwarte dane w świecie wielkich danych. Porozumienie międzynarodowe. Wersja skrócona* [online]. ICSU-IAP-ISSC-TWAS. [Dostęp 22.06.2017]. Dostępny w: http://www.science-international.org/sites/default/files/reports/otwarte-dane-wersja-skrocona_short_pl.pdf.

17. PAWEŁOSZEK, I. Wybrane problemy wdrożenia koncepcji otwartych danych w e-administracji. *Roczniki KAE* [online]. 2014, z. 33, s. 455–470. [Dostęp 22.06.2017]. ISSN 1232-4671. Dostępny w: http://rocznikikae.sgh.waw.pl/p/roczniki_kae_z33_26.pdf.
18. *re3data.org* [online]. [Dostęp 11.07.2017]. Dostępny w: <http://www.re3data.org/>
19. *RepOD. Repozytorium Otwartych Danych* [online]. [Dostęp 5.08.2017]. Dostępny w: <https://re-pod.pon.edu.pl/pl>
20. STARCZEWSKI, M., STĘPIŃSKA-USTASIAK, L. *Otwarty dostęp czy otwarta nauka?* [online]. 2014. [Dostęp 22.06.2017]. Dostępny w: <https://depot.ceon.pl/bitstream/handle/123456789/6317/Otwarty%20dost%C4%99p%20czy%20otwarta%20nauka.pdf?sequence=2&isAllowed=y>.
21. SZCZĘSNY, P. *Otwarta nauka, czyli dobre praktyki uczonych* [online]. Toruń: Stowarzyszenie EBIB, 2013. [Dostęp 22.06.2017]. ISBN 978-83-63458-03-4. Dostępny w: http://repozytorium.uwb.edu.pl/jspui/download/Szczesny_Otwarta_nauka.pdf
22. SZPROT, J. (red.). *Otwarta nauka w Polsce 2014. Diagnoza* [online]. Warszawa: Wydawnictwo ICM, 2014. [Dostęp 22.06.2017]. ISBN 978-83-63490-08-9. Dostępny w: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>.
23. What is research data? W: *ANDS Guide* [online]. 2017. [Dostęp 22.06.2017]. Dostępny w: http://www.ands.org.au/_data/assets/pdf_file/0006/731823/What-is-research-data.pdf.

STRZELCZYK, E. Otwarte dane badawcze – kolejny krok do otwierania nauki. W: Sójkowska, I., Derfert Wolf, L. (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