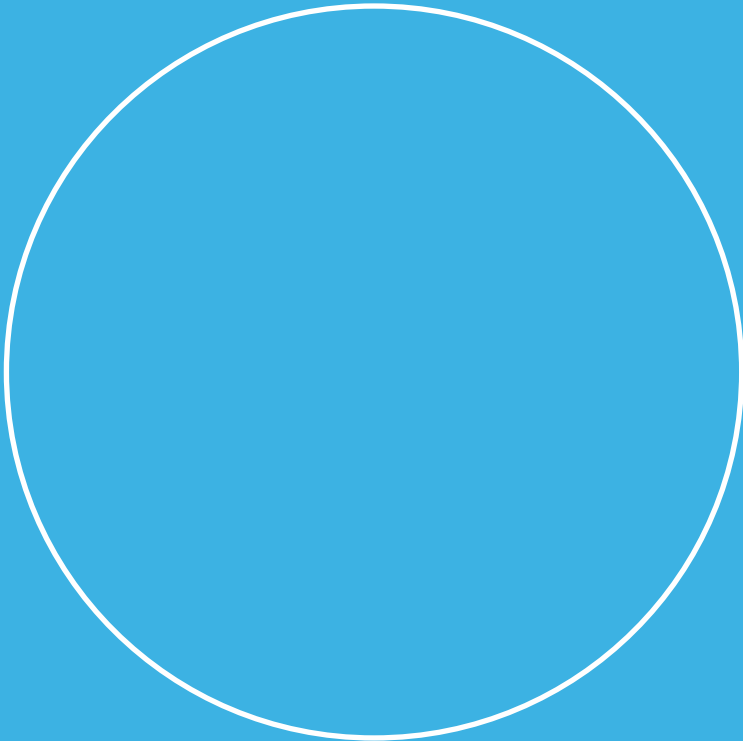


Jak korzystać z zasobów w repozytoriach danych

Wersja 1.0



Opracowanie Natalia Gruenpeter

Konsultacja Wojciech Fenrich, dr Krzysztof Siewicz, Jakub Szprot

Broszura powstała w ramach projektu Dziedziny Repozytoria Otwartych Danych Badawczych, finansowanego ze środków Programu Operacyjnego Polska Cyfrowa.

Strona internetowa projektu: drodb.icm.edu.pl

© Copyright by Uniwersytet Warszawski, Warszawa 2019

Publikacja dostępna na licencji Creative Commons – Uznanie Autorstwa 4.0.
Postanowienia licencji dostępne są pod adresem

<https://creativecommons.org/licenses/by/4.0/pl/legalcode>

Uniwersytet Warszawski

Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego

ul. Tyniecka 15/17

02-630 Warszawa

www.icm.edu.pl



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Jak korzystać z zasobów w repozytoriach danych

Wersja 1.0

Spis treści

Słowniczek	5
Wstęp	7
Część 1. Sposoby wyszukiwania danych badawczych	8
<i>Data journals</i>	8
Repozytoria danych badawczych.	10
Repozytoria ogólnego przeznaczenia	11
Rejestry repozytoriów	11
Wyszukiwarki danych badawczych	14
Część 2. Najpopularniejsze funkcjonalności repozytoriów.	16
Wyszukiwanie i pobieranie	16
Udostępnianie i cytowanie	17
Kontrola wersji	17
Część 3. Dostęp do danych	18
Dostęp do repozytorium a dostęp do danych badawczych	18
Część 4. Korzystanie z danych	20
Prawne aspekty korzystania z danych badawczych	20
Wolne licencje	20
Podsumowanie.	23

Słowniczek

Dane badawcze – dane zebrane lub wytworzone jako materiał do analizy w ramach badań naukowych.

Data access committee – składająca się z ekspertów grupa, do której należy decyzja o udostępnieniu zbioru danych.

Data journal – czasopismo naukowe, które publikuje artykuły opisujące zbiory danych badawczych, udostępnione w repozytoriach danych lub (rzadko) w formie suplementu do samego artykułu.

Data management plan – zob. plan zarządzania danymi.

DMP – zob. plan zarządzania danymi.

DOI – ang. „Digital Object Identifier”, jeden z trwałych identyfikatorów obiektów cyfrowych, pozwalający na ich odnalezienie w internecie niezależnie od wiodącego do nich adresu URL. Posiadający DOI zbiór danych można za jego pomocą zidentyfikować nawet wtedy, gdy zostanie on przeniesiony na inny serwer czy do innego repozytorium.

Embargo – okres, przez który dane badawcze nie mogą zostać udostępnione publicznie. Jest on zwykle wykorzystywany po to, aby uzyskać związane z nimi patenty i/lub inne prawa własności intelektualnej oraz przygotować oparte na nich publikacje naukowe. Po jego upływie opublikowanie danych badawczych staje się możliwe.

FAIR – akronim słów „findable” (możliwy do znalezienia), „accessible” (dostępny), „interoperable” (interoperacyjny) i „reusable” (możliwy do ponownego wykorzystania), określający wymogi, jakie powinny spełniać udostępnione dane badawcze.

Interoperacyjność – cecha tych danych, które można łączyć z innymi danymi, wykorzystywać w wielu różnych systemach komputerowych i analizować przy użyciu różnorodnego oprogramowania.

Licencja – upoważnienie do korzystania w określony sposób z utworu lub bazy danych. Przedmiotem licencji może być na przykład zbiór danych badawczych.

Licencje Creative Commons – popularne wzory licencji opracowane przez organizację Creative Commons.

Metadane – ustrukturyzowane informacje opisujące zasoby informacji, np. zbiory danych badawczych. Metadane zawierają informacje o formie i treści zasobów, dzięki czemu pozwalają na ich wyszukiwanie i identyfikację oraz zarządzanie nimi.

Ograniczony dostęp – model dostępu, w którym dane udostępniane są jedynie określonym osobom (np. tym, które uzyskały zgodę dysponenta danych) lub kategoriom osób (np. prowadzącym badania naukowe lub zatrudnionym w konkretnej instytucji).

Otwarte dane badawcze – dostępne za pośrednictwem internetu dane badawcze, które można wykorzystywać bez ponoszenia opłat oraz bez istotnych ograniczeń technicznych i prawnych.

Plan zarządzania danymi (*data management plan, DMP*) – dokument opisujący to, co będzie działo się z danymi w trakcie projektu badawczego i po jego zakończeniu. Ma on charakter „żywego dokumentu”, który może i powinien zmieniać się wraz ze zmianami pojawiającymi się w innych obszarach projektu badawczego.

Ponowne wykorzystanie – ogólny termin odnoszący się do technicznych, prawnych i metodologicznych uwarunkowań użycia danych przez dowolne osoby i/lub instytucje, w szczególności te, które nie były zaangażowane w ich wytworzenie.

Repozytorium danych – serwis internetowy służący do deponowania (umieszczania), przechowywania i udostępniania za pośrednictwem internetu danych badawczych w formie cyfrowej.

Wstęp

Niniejsza broszura jest przewodnikiem dla osób chcących korzystać z danych badawczych udostępnianych przez naukowców. Zawarto w niej informacje o tym, gdzie i w jaki sposób szukać takich danych oraz jak korzystać z repozytoriów danych, biorąc pod uwagę warunki dostępu oraz wykorzystania zasobów. Broszura uwzględnia zarówno podstawowe kwestie techniczne, takie jak wyszukiwanie, pobieranie i cytowanie danych, jak i zarys prawnych aspektów korzystania z danych badawczych. Przybliżyła ponadto konkretne rozwiązania, takie jak czasopisma naukowe publikujące artykuły opisujące zbiory danych, wyszukiwarki zbiorów danych, a także narzędzia prawne umożliwiające szeroki zakres wykorzystania danych (wolne licencje).

Część 1. Sposoby wyszukiwania danych badawczych

Data journals

Jednym z rozwiązań mających na celu zachęcanie badaczy do udostępniania danych są *data journals*, czyli recenzowane czasopisma z artykułami na temat zbiorów danych badawczych. Publikowane w nich artykuły (*data descriptors*) mogą zawierać załączniki z omawianymi danymi lub wskazywać miejsca przechowywania tych danych – zwykle repozytoria. W zależności od dziedziny nauki lub przyjętych przez redakcję wytycznych publikacje mogą ponadto zawierać takie elementy, jak: ilościowe i jakościowe informacje o zbiorach danych, opis ich źródła lub metodologii ich pozyskiwania, opis ich znaczenia, a także informacje na temat warunków korzystania z danych (rodzaju licencji). Zwykle *data journals* udostępniają też listy rekomendowanych repozytoriów, z których korzystać mogą naukowcy publikujący dane badawcze. Listy te mogą być pomocne również w wyszukiwaniu danych, zwłaszcza kiedy czasopismo specjalizuje się w określonym obszarze badawczym.

Przykładowe *data journals*

Tytuł	Wydawca	Zakres tematyczny	Czy publikuje listę polecanych repozytoriów danych?
„Data in Brief” (https://www.sciencedirect.com/journal/data-in-brief)	Elsevier	Wszystkie obszary wiedzy	Tak (https://www.elsevier.com/authors/author-resources/research-data/data-base-linking#repositories)

Tytuł	Wydawca	Zakres tematyczny	Czy publikuje listę polecanych repozytoriów danych?
„Scientific Data” (https://www.nature.com/sdata)	Springer Nature	Nauki przyrodnicze, ale także nauki społeczne, w szczególności jeżeli mają potencjał w badaniach interdyscyplinarnych	Tak (https://www.nature.com/sdata/policies/repositories)
„Data” (https://www.mdpi.com/journal/data)	MDPI	Dane badawcze	Nie
„Journal of Open Archeology Data” (https://openarchaeologydata.metajnl.com/)	Ubiquity Press	Archeologia	Tak (https://openarchaeologydata.metajnl.com/about/) prowadzi też własne repozytorium (https://dataverse.harvard.edu/dataverse/JOAD)
„Research Data Journal for the Humanities and Social Sciences” (https://brill.com/view/journals/rdj/rdj-overview.xml)	Brill	Nauki społeczne i humanistyczne	Tak (https://brill.com/fileasset/downloads_products/Author_Instructions/RDJ.pdf)
„Geoscience Data Journal” (https://rmets.onlinelibrary.wiley.com/journal/20496060)	Wiley dla Royal Meteorological Society	Nauki o Ziemi	Tak (https://rmets.onlinelibrary.wiley.com/hub/journal/20496060/features/data-centres-and-repositories)

Repozytoria danych badawczych

Repozytoria danych badawczych, podobnie jak repozytoria publikacji, mogą mieć charakter dziedzinowy lub instytucjonalny, Mogą też łączyć funkcje obu typów, jeżeli prowadząca je instytucja zajmuje się badaniami w określonym obszarze lub prowadzi repozytorium z myślą o konkretnym projekcie badawczym.

Repozytoria dziedzinowe gromadzą dane z konkretnych obszarów wiedzy, co dla poszukujących danych użytkowników jest istotne. Przeszukiwanie właściwego repozytorium zwiększa bowiem prawdopodobieństwo znalezienia szybko interesujących danych.

Przykładowe repozytoria dziedzinowe

Dryad

<https://datadryad.org>

Dryad,

nauki medyczne, nauki ścisłe i przyrodnicze

Archeology Data Services

<https://archaeologydataservice.ac.uk>

University of York,

archeologia

EarthChem

<http://www.earthchem.org/>

Lamont-Doherty Earth Observatory, Columbia University,
geochemia, petrologia, geochronologia

Repozytoria instytucjonalne gromadzą dane badawcze wytworzone i opracowane przez pracowników konkretnej instytucji lub grantobiorców realizujących finansowane przez nią badania. Dla szukających danych naukowców mogą być przydatne zwłaszcza wtedy, gdy prowadząca repozytorium instytucja specjalizuje się w określonym obszarze.

Przykładowe repozytoria instytucjonalne

DataCat: The Research Data Catalogue

<http://datacat.liverpool.ac.uk>

University of Liverpool

Open Data LMU

<https://data.ub.uni-muenchen.de>

Ludwig-Maximilians-Universität München

CaltechDATA

<https://data.caltech.edu/>

California Technical Institute of Technology

Repozytoria ogólnego przeznaczenia

Dane badawcze deponowane są również w repozytoriach ogólnego przeznaczenia, takich jak Zenodo (<https://zenodo.org>) czy Figshare (<https://figshare.com>).

Rejestry repozytoriów

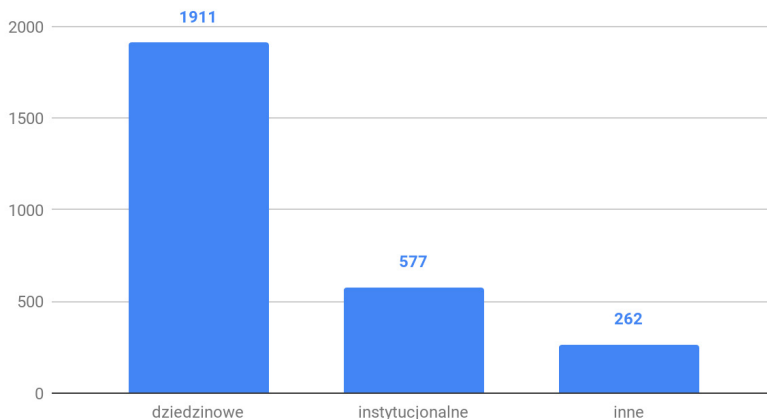
Poszukując właściwego repozytorium, zarówno do udostępnienia, jak i wyszukiwania danych badawczych, warto skorzystać z rejestru repozytoriów.

re3data.org

Najpełniejszym źródłem wiedzy na temat repozytoriów danych jest obecnie międzynarodowy rejestr repozytoriów danych badawczych re3data.org (Registry of Research Data Repositories, <https://www.re3data.org>), w którym znajduje się ich ponad 2300 (dane z lipca 2019 roku). Zasoby tego rejestru można przeszukiwać według wpisanych słów bądź przeglądać, korzystając z trzech podstawowych kryteriów: kraju, obszaru wiedzy oraz typu danych. Wyniki wyszukiwania lub przeglądania można

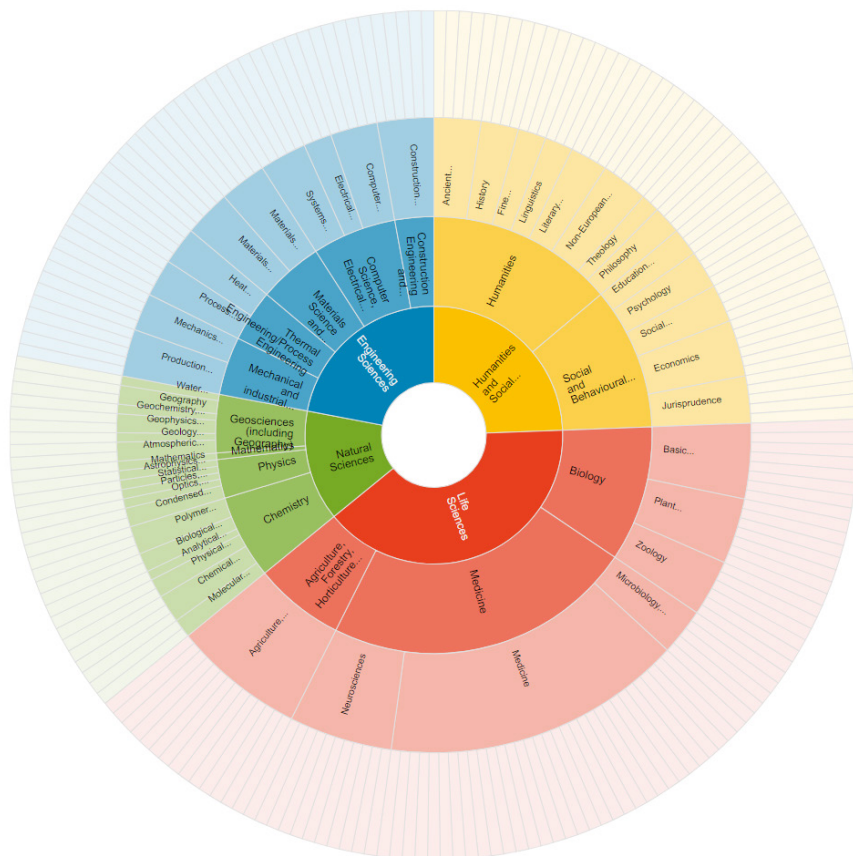
następnie zawęzić za pomocą blisko 30 filtrów, uwzględniających m.in. kwestie prawne (np. warunki dostępu i korzystania z repozytorium oraz z danych), instytucjonalne (m.in. rodzaj instytucji prowadzącej repozytorium), techniczne (jak np. oprogramowanie repozytoryjne, rodzaj API) czy zakres przedmiotowy (przedmiot, słowa kluczowe).

Repozytoria w rejestrze re3data.org



Większość zarejestrowanych w bazie re3data.org repozytoriów ma charakter dziedzinowy. Aby znaleźć repozytorium właściwe dla konkretnego obszaru wiedzy, można skorzystać z odpowiednich kryteriów wyszukiwania, takich jak obszar wiedzy (*subject*) i słowa kluczowe (*keywords*) bądź z interaktywnego koła, uwzględniającego różne poziomy specjalizacji badań.

Obszary wiedzy w rejestrze re3data.org



Źródło: <https://www.re3data.org/browse/by-subject/>
<http://re3data.org> – Registry of Research Data Repositories,
<https://doi.org/10.17616/R3D>

Drugą pod względem wielkości kategorią są repozytoria instytucjonalne, które również – ze względu na profil prowadzących je instytucji – mogą mieć charakter zbliżony do repozytoriów dziedzinowych. W kategorii „innych repozytoriów” znajduje się zaś ponad 260 serwisów, wśród których są np. repozytoria prowadzone przez sieci i projekty badawcze, a także serwisy z otwartymi danymi publicznymi, udostępnianymi przez państwowe urzędy i instytucje. Zostały one

ujęte w rejestrze, ponieważ informacje sektora publicznego również mogą być wykorzystywane w badaniach naukowych. W Polsce dane publiczne udostępniane są za pomocą serwisu <https://dane.gov.pl>, czyli Centralnego Repozytorium Danych Publicznych.

Wyszukiwarki danych badawczych

Stosunkowo nowym narzędziem do wyszukiwania danych badawczych są specjalne wyszukiwarki, które – podobnie jak inne wyszukiwarki internetowe – nie przechowują danych, tylko indeksują zasoby innych serwisów.

Data Cite Search

search.datacite.org

Data Cite Search to serwis służący do wyszukiwania danych badawczych, prowadzony przez organizację DataCite, która dostarcza trwałe identyfikatory DOI (Digital Object Identifier) dla zbiorów danych oraz innych obiektów powstających w toku badań. Serwis gromadzi metadane obiektów poprzez identyfikatory DOI, które odsyłają do zbiorów danych z ponad 1700 centrów badawczych, instytucji i serwisów. Z jego pomocą można znaleźć różnego rodzaju rezultaty badań, w tym artykuły naukowe, ale użytkownicy poszukujący danych badawczych mogą zaznaczyć opcję „zbiór danych” (*dataset*), zawężając w ten sposób wyniki wyszukiwania. Aktualnie baza zawiera metadane ponad 6 mln zbiorów (dane z lipca 2019 roku), do których odsyłają DOI zapisane jako hiperlinki. Nie wszystkie zasoby są otwarte, a wyszukiwarka nie umożliwia, niestety, sortowania wyników pod tym kątem. Serwis ułatwia natomiast cytowanie danych, ponieważ generuje rekordy bibliograficznego w ośmiu formatach: APA, Harvard, MLA, Vancouver, Chicago, IEEE, BibTex, RIS.

Elsevier DataSearch

datasearch.elsevier.com

Działający w wersji beta serwis Elsewiera to prototyp wyszukiwarki otwartych danych badawczych. Serwis indeksuje różnego rodzaju

zasoby, takie jak zbiory danych deponowane w otwartych repozytoriach, wybranych m.in. na podstawie liczby użytkowników oraz łatwości indeksowania danych, zbieranych zgodnie ze standardowymi rozwiązaniami, tzn. poprzez API, a czasami również poprzez import ich baz. Oprócz zasobów udostępnianych w otwartych repozytoriach danych wyszukiwarka indeksuje metadane z wybranych serwisów, a także wykresy, tabele i innego rodzaju załączniki do artykułów naukowych (zbierane są one z dwóch źródeł: arXiv oraz prowadzonego przez Elseviera serwisu ScienceDirect).

Google Dataset Search

toolbox.google.com/datasetsearch

Kolejną wyszukiwarką danych działającą na razie w wersji beta jest serwis uruchomiony przez firmę Google jesienią 2018 roku. Umożliwia wyszukiwanie informacji o zbiorach powszechnie dostępnych danych gromadzonych w repozytoriach. Informacje zbierane są na podstawie metadanych udostępnianych przez repozytoria w formacie schema.org/Dataset, wzbogacanych następnie o inne indeksowane przez Google dane, zaczerpnięte m.in. z Google Scholar.

Część 2. Najpopularniejsze funkcjonalności repozytoriów

Wyszukiwanie i pobieranie

Repozytoria danych badawczych umożliwiają użytkownikom zarówno przeszukiwanie (*search*), jak i przeglądanie (*browse*) zasobów. Wyniki obydwu działań można dodatkowo zawęzić, korzystając z oferowanych przez repozytoria filtrów. Najczęstsze to opcja przeszukiwania lub przeglądania według:

- a) możliwości uzyskania dostępu do danych (*terms of access* lub *access rights*);
- b) warunków korzystania z danych, ujętych zwykle w formule licencji (*terms of use*);
- c) obszarów wiedzy oraz dziedzin i dyscyplin naukowych (w repozytoriach ogólnego przeznaczenia) bądź słów kluczowych;
- d) rodzaju danych;
- e) formatu danych;
- f) źródła danych (informacje o autorach, zespołach czy instytucjach finansujących i prowadzących badania).

Dostęp do danych badawczych realizowany jest poprzez zapewnienie użytkownikom możliwości pobrania plików na dysk komputera. Dodatkową opcją usprawniającą proces wyszukiwania danych jest funkcja widoku czy podglądu (*preview, explore*), umożliwiająca zapoznanie się z zawartością pliku w przeglądarce, bez konieczności jego pobierania. Dobrą praktyką stosowaną w większości repozytoriów jest interfejs API – umożliwiający dostęp maszynowy – oraz możliwość pobierania metadanych przez OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting), co pozwala na automatyczne ich pobieranie i gromadzenie w ramach innych serwisów.

Udostępnianie i cytowanie

Użytkownikom, którzy znaleźli interesujące ich dane i chcą z nich skorzystać, repozytoria oferują narzędzia ułatwiające ich cytowanie, dzielenie się nimi w mediach społecznościowych czy zamieszczanie ich na stronach internetowych. Sugerowana formuła cytowania zwykle dostępna jest na stronie z opisem danych. Wystarczy ją skopiować i wkleić do dowolnego dokumentu, a w razie potrzeby – dostosować do przyjętego sposobu formatowania przypisów bądź bibliografii. Badacze korzystający z narzędzi do zarządzania bazami bibliograficznymi często mogą także skorzystać z opcji wygenerowania rekordu bibliograficznego w jednym z popularnych formatów, takich jak EndNote XML, RIS czy BibTex. Repozytoria dają ponadto możliwość szybkiego i bezpośredniego udostępnienia plików w mediach społecznościowych bądź bezpośredniego osadzenia ich na stronie internetowej poprzez wykorzystanie fragmentu kodu źródłowego (opcja *embed* – osadź).

Kontrola wersji

Funkcjonalnością ważną ze względów merytorycznych jest kontrola zamieszczanych w repozytorium wersji zbiorów danych badawczych, która polega na odnotowaniu wszelkich zmian wraz z datą ich wprowadzenia. To istotna funkcja, ponieważ dane lub opisujące je metadane mogą w toku badań ulegać zmianom (np. narastać). Naukowcy mogą też udostępniać kolejne, coraz lepiej opracowane wersje zbioru, które mają zastąpić poprzednie wersje. Nie znaczy to jednak, że poprzednie wersje powinny tym samym zniknąć, jako że w międzyczasie mogły one już zostać zacytowane lub wykorzystane przez innych badaczy do realizacji ich własnych celów badawczych. Korzystając z danych zamieszczonych w repozytorium, warto zapoznać się z polem „wersje” (*versions*), a cytując dane, należy wskazać odpowiednią.

Część 3. Dostęp do danych

Dostęp do repozytorium a dostęp do danych badawczych

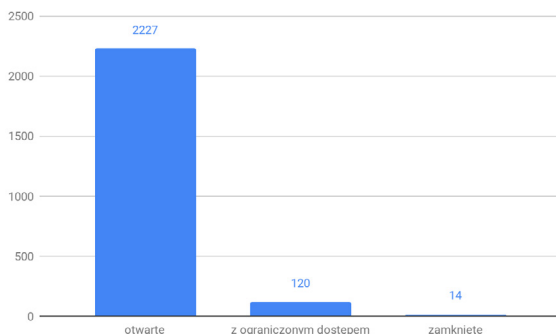
Rejestr repozytoriów danych badawczych re3data.org rozróżnia dostęp do repozytorium, czyli do bazy (*database access*) oraz dostęp do przechowywanych w repozytorium danych, czyli do zawartości bazy (*data access*). Pierwsze kryterium pozwala wyróżnić trzy rodzaje repozytoriów:

1. otwarte – taki charakter ma zdecydowana większość z nich: dostęp do nich nie jest w żaden sposób ograniczony, a umieszczone tam zasoby przeglądać może każdy,

2. zamknięte – dostęp do nich ma ograniczona grupa osób, zwykle pracownicy danej instytucji, osoby należące do danego zespołu, stowarzyszenia czy sieci badawczej,

3. z ograniczonym dostępem – dostęp może wiązać się z koniecznością rejestracji w portalu, wniesienia opłaty bądź z ograniczeniami innego typu, na przykład z koniecznością wypełnienia wniosku o udzielenie dostępu, przejścia procesu autoryzacji czy szkolenia (może to dotyczyć w szczególności repozytoriów przechowujących dane zebrane w toku badań klinicznych).

Rodzaje repozytoriów zarejestrowanych w re3data.org



Repozytoria oznaczone w rejestrze re3data.org jako otwarte umożliwiają użytkownikom przeglądanie zasobów, jednak nie oznacza to, że wszystkie zdeponowane w nich dane są dostępne w sposób otwarty. Większość repozytoriów umożliwia ustanowienie embarga, czyli okresu, w którym dane po ich zdeponowaniu w repozytorium pozostają zamknięte. Niezależnie od embarga, przeglądający zasoby użytkownicy mogą zapoznać się z opisem danych (metadanymi) oraz zidentyfikować badaczy, którzy zamieścili je w repozytorium (lub instytucję, która to zrobiła). W takiej sytuacji można skontaktować się z osobami odpowiedzialnymi za dane, aby dowiedzieć się, czy istnieje możliwość zapoznania się z nimi przed upływem okresu embarga.

Inne możliwe ograniczenia w dostępie do danych to m.in:

- konieczność wniesienia opłaty za dostęp,
- wymóg rejestracji w serwisie, czasami połączony z autoryzacją,
- uwierzytelnienie, np. konieczność potwierdzenia tożsamości i afiliacji użytkownika chcącego zapoznać się z danymi.

W takich sytuacjach na stronie z opisem zbioru danych znajduje się zwykle informacja o możliwości uzyskania dostępu, np. odnośnik do formularza kontaktowego lub kontakt do osób odpowiedzialnych za udostępnianie danych (*data access committee*).

Część 4. Korzystanie z danych

Prawne aspekty korzystania z danych badawczych

Uzyskanie dostępu do danych nie jest jednoznaczne z możliwością wykorzystania ich w dowolny sposób. Korzystanie z nich musi odbywać się zgodnie z przypisanymi im ograniczeniami prawnymi. Na przykład z danych udostępnionych bez wskazania konkretnej licencji bądź udostępnionych z wyraźnym zastrzeżeniem praw można korzystać jedynie w granicach swobód określonych prawem.

Chodzi tu przede wszystkim o przepisy prawa autorskiego oraz ustawy o ochronie baz danych – dane badawcze mogą podlegać prawu autorskiemu, prawu *sui generis* do baz danych bądź obu tym reżimom jednocześnie. Konsekwencją takiej ochrony jest wąski zakres dopuszczalnego ponownego wykorzystania (dozwolony użytek oraz jego okrojony odpowiednik w ustawie o ochronie baz danych, obejmujący użytek osobisty, użytek w celach dydaktycznych lub badawczych oraz cele państwowe: bezpieczeństwo wewnętrzne oraz postępowania sądowe lub administracyjne).

Wolne licencje

Z danych udostępnionych na określonej licencji można korzystać zgodnie z zakresem swobód określonych w tej licencji. Istotne są w tym kontekście zwłaszcza wolne licencje. Są one nieodpłatne, pozwalają na bardzo szeroki zakres wykorzystania i nie nakładają na licencjobiorców zobowiązań idących dalej niż obowiązki przekazywania odbiorcom określonych informacji (np. o autorze, źródle, licencji – tzw. klauzule uznania autorstwa) lub obowiązki stosowania takiej samej licencji w przypadku rozpowszechniania modyfikacji utworu (tzw. klauzule *copyleft* lub *sharealike*).

Wśród najbardziej popularnych licencji Creative Commons, wolne licencje to:

1. Creative Commons Uznanie autorstwa (CC-BY) – licencja ta pozwala na praktycznie dowolne wykorzystywane objętego nią materiału, wymaga od użytkowników jedynie zachowania oznaczeń autorstwa, poszanowania innych praw osobistych (np. zakaz przypisywania sobie poparcia autora dla określonego wykorzystania utworu) oraz przekazania dalej informacji o licencji.

2. Creative Commons Uznanie autorstwa – Na tych samych warunkach (CC-BY-SA) – W porównaniu do licencji CC-BY licencja ta zawiera jedno dodatkowe wymaganie, a mianowicie zobowiązuje użytkownika do opublikowania na takiej samej licencji jego własnych modyfikacji oryginalnego materiału, jeżeli zdecyduje się on je rozpowszechniać.

3. CC0 – wzorzec oświadczenia, na podstawie którego uprawniony zrzeka się w zasadzie wszelkich uprawnień, jakie mogłyby mu przysługiwać w odniesieniu do udostępnianego materiału. Na wypadek, gdyby takie zrzeczenie było nieskuteczne, we wzorcu przewidziano udzielenie szerokiej, nieodpłatnej licencji, bez żadnych wymagań dla użytkownika. Udostępniający zobowiązuje się nie egzekwować przysługujących mu uprawnień.

Warto podkreślić, że istnieją także licencje CC zawierające ograniczenia korzystania komercyjnego (klausule NC) lub zakaz modyfikacji (klausule ND). Licencje te nie są wolnymi licencjami. Choć nie doprowadzą one do otwarcia danych zgodnie z przyjętym rozumieniem tej otwartości (wymagającym braku takich ograniczeń), dają jednak użytkownikowi większą swobodę niż ta, która wynika z samych przepisów ustawy.

Poza ustaleniem tego, na jakich zasadach udostępniony jest określony zbiór danych, z praktycznego punktu widzenia istotne jest jeszcze sprawdzenie dwóch kwestii. Po pierwsze, czy dane zostały udostępnione na określonej licencji przez osobę uprawnioną. Po drugie, jakiego elementu zbioru dokładnie dotyczy określony zestaw wymagań (ustalenie zakresu licencji).

Prawa do danych powstają samoistnie i nie ma publicznych rejestrów, w których można by było sprawdzić, komu one przysługują. Z kolei umowy przenoszące te prawa są zazwyczaj dostępne tylko stronom. W związku z tym przeciętny użytkownik ma bardzo niewielkie możliwości sprawdzenia, czy udostępniający jest do tego uprawnionym. Opiera się to przede wszystkim na zaufaniu. Praktyczna realizacja kwestii pierwszej sprowadza się zatem do analizy wiarygodności samego repozytorium, ale też konkretnego udostępnionego w nim zestawu danych. Zbyt ogólnikowe metadane (w tym zbyt ogólne wskazanie licencji) to jedna z rzeczy, które powinny wzbudzić czujność prawną użytkownika i skłonić go do dokładniejszego upewnienia się, czy dane zostały udostępnione legalnie, a udostępniająca je osoba świadomie wybrała określoną licencję.

Natomiast ustalenie zakresu uzyskanej licencji wymaga przede wszystkim uświadomienia sobie, że prawna ochrona danych może dotyczyć zarówno ich zestawu (zbioru) jako całości, jak i jego poszczególnych elementów. Na tych różnych poziomach prawa mogą przysługiwać innym osobom (prawa *sui generis* do bazy jako całości przysługiwać będą producentowi, czyli np. jednostce naukowej, podczas gdy np. prawa do poszczególnych zawartych w bazie fotografii, wywiadów itd. mogą przysługiwać poszczególnym naukowcom lub nawet osobom trzecim, jak np. prawo do wizerunku fotografowanej osoby). Objęcie niektórych z tych praw (np. dóbr osobistych lub danych osobowych osób badanych, o których informacje znajdują się w zestawie danych) wolną licencją może w ogóle nie być możliwe. Z kolei w przypadku innych elementów lub praw wymagać to po prostu będzie dokładniejszego opisanie w metadanych przez samego udostępniającego (powinien on doprecyzować, czy udzielana licencja dotyczy tylko zbioru, czy także poszczególnych jego elementów).

Podsumowanie

Korzystając z udostępnionych w repozytoriach danych badawczych należy uwzględnić zarówno aspekty techniczne, np. formaty czy wielkość plików, jak i prawne, związane z warunkami dostępu oraz korzystania z danych. W obydwu tych kwestiach możliwość ponownego wykorzystania danych badawczych zależy od tego, w jaki sposób zbiory przygotowane zostały przez udostępniających. Stosowanie standardowych rozwiązań, w szczególności zgodnych z zasadami FAIR, nie tylko ułatwia wyszukanie danych, lecz także świadczy o ich rzetelnym przygotowaniu, umożliwia jednoznaczne ustalenie źródła danych oraz odpowiednie ich zacytowanie. To z kolei jest ważne z punktu widzenia badacza chcącego skorzystać z danych badawczych w sposób uczciwy i rzetelny.