

Spis treści

Czym są dane badawcze?	3
Dlaczego udostępniamy nasze dane badawcze?	6
Rozwój nauki i lepszy świat	6
Finansowanie badań i ewaluacja nauki.....	7
Większa widoczność i rzetelność badań	8
Zasady FAIR w nauce	10
F jak FINDABLE, czyli znajdź dane bez problemu	11
A jak ACCESSIBLE, czyli miej łatwy dostęp do danych	12
I jak INTEROPERABLE, czyli łącz z innymi.....	13
R jak REUSABLE, czyli użyj ponownie.....	13
Warto być FAIR	14
Przygotowanie danych do udostępniania	16
Analiza i selekcja danych badawczych	16
Zgoda na gromadzenie, przetwarzanie i udostępnianie danych	16
Anonimizacja i pseudonimizacja danych	17
Wybór formatów plików	17
Nadawanie odpowiednich nazw plikom i folderom.....	18
Wersjonowanie	18
Dane danych, czyli metadane	19
Stan prawny danych i licencjonowanie.....	20
Bezpieczeństwo danych	20
Miejsce deponowania danych – repozytoria.....	20
Koszty związane z zarządzaniem danymi	21

Plan Zarządzania Danymi	23
Dlaczego warto nauczyć się pisać Plan Zarządzania Danymi?.....	23
Co zawiera Plan Zarządzania Danymi?.....	24
Narzędzia do tworzenia i edycji PZD, wskazówki, instrukcje.....	26
Przykładowe Plany Zarządzania Danymi.....	27
Skonsultuj z nami Plan Zarządzania Danymi!	27
Gdzie deponować i publikować dane badawcze?	29
Repozytoria instytucjonalne.....	29
Repozytoria danych badawczych.....	30
REPOZYTORIA OGÓLNE.....	30
REPOZYTORIA DZIEDZINOWE.....	32
Czasopisma publikujące dane badawcze (ang. data journals)	35
A po co naukowcom serwisy społecznościowe?.....	35
Cytujmy też dane badawcze!	36
Aspekty prawne zarządzania danymi badawczymi.....	37
Dane badawcze nie rządzą się tylko własnymi prawami	37
Jak wybrać licencję?	38
Szkolenia i kursy	42

Czym są dane badawcze?

Czy dla biologa dane badawcze to analizowane próbki? A może ich opis?

Czy matematyk za dane badawcze uzna wyłącznie dane liczbowe?

Czy dane badawcze archeologa to badane przedmioty czy ich opis? A co, jeśli bada budynki?

Czy historyk bada fakty, a dane badawcze to dla niego dokumenty?

Czy badacz dokonujący analizy twórczości literackiej ma do czynienia z danymi badawczymi?

Czy dane badawcze będące utworami są przedmiotem prawa autorskiego?

Czy badacz ma prawo do danych badawczych opartych na przypadkach pacjentów?

Na początku drogi naukowej, a często i później, pytań jest bardzo wiele i wszystkie są uzasadnione. Zadają je sobie badacze na całym świecie, w Polsce również, a tym intensywniej, im częściej frazy „dane badawcze” i „plan zarządzania danymi” pojawiają się we wnioskach o dofinansowanie. Co więcej, coraz częściej towarzyszy im pojęcie „otwartości”, z którym dopiero osławiamy się na gruncie publikacji wyników badań. Naszym zadaniem jest przejść tę drogę razem z pracownikami, doktorantami i studentami Uniwersytetu Łódzkiego. Zebranie w jednym miejscu informacji na temat danych badawczych i sukcesywne ich uzupełnianie pozwala nam być na bieżąco z tym, co w nauce aktualnie się dzieje, i z problemami, jakie napotykają nasi badacze, niezależnie od ich wieku, doświadczenia czy uprawianej dyscypliny.

Istnieje co najmniej kilka popularnych definicji danych badawczych („Research Data”). Spotykamy je na stronach zagranicznych jednostek naukowych, w dokumentach rządowych i pozarządowych organizacji. W [Dyrektywie Parlamentu Europejskiego](#) z 20 czerwca 2019 roku czytamy:

„Ilość generowanych danych badawczych rośnie w gwałtownym tempie, a możliwości ich ponownego wykorzystywania wykraczają poza środowisko naukowe. Aby umożliwić sprostanie rosnącym wyzwaniom społecznym w wydajny i całościowy sposób, niezwykle ważną i pilną kwestią jest zapewnienie dostępu do danych pochodzących z różnych źródeł, jak również sektorów i dyscyplin, oraz możliwości syntezy i ponownego wykorzystywania tych danych. Dane badawcze obejmują dane statystyczne, wyniki eksperymentów, pomiarów, obserwacji prowadzonych w terenie, wyniki ankiet, nagrania wywiadów i zdjęcia. Obejmują one także metadane, specyfikacje i inne obiekty cyfrowe.”

Już w przytoczonym zakresie definicji danych badawczych widać, jak szerokie jest to pojęcie. Obejmuje każdą dyscyplinę naukową, niezależnie od tego, jakim rodzajem danych się ona posługuje, a sformułowanie „inne obiekty cyfrowe” sprawia, iż nie mamy wątpliwości co do otwartości definicji. Parlament Europejski [precyzuje](#):

„(...) *dane badawcze* oznaczają dokumenty w formie cyfrowej, inne niż publikacje naukowe, które są gromadzone lub produkowane w ramach działalności badawczo-naukowej i są wykorzystywane jako dowody w procesie badawczym bądź też są powszechnie akceptowane w środowisku naukowym jako konieczne do weryfikacji poprawności ustaleń i wyników badań.”

Wiele mówiąca jest także definicja, a raczej jej brak w przewodniku opublikowanym przez Australian National Data Service. Autorzy [piszą bowiem](#), że „podanie miarodajnej definicji danych badawczych jest trudne, gdyż będzie ona zależała od kontekstu, w jakim o nią pytamy”. Oczywiście, możemy uznać, że taka definicja ucina wszelkie dyskusje i świadczy tylko o mętności zagadnienia, jednak istnieją uznawane przez środowisko naukowe definicje danych badawczych, które nieco rozjaśniają ten obraz. Na przykład [na stronie](#) The Queensland University of Technology czytamy:

„Dane badawcze to dane w postaci faktów, obserwacji, obrazów, rezultatów uzyskiwanych przez programy komputerowych, nagrań, pomiarów lub doświadczeń, na których opiera się argument, teoria, badanie, hipoteza lub inny wynik. Odnosi się to do danych generowanych, gromadzonych lub wykorzystywanych podczas projektów badawczych, a w niektórych przypadkach także do samych wyników badań. Dane mogą być liczbowe, opisowe, wizualne lub namacalne. Mogą być surowe, oczyszczone lub przetworzone i przechowywane w dowolnym formacie czy na dowolnym nośniku. Dane badawcze w wielu dyscyplinach siłą rzeczy mogą obejmować oprogramowanie, algorytm, model i/lub parametry użyte do uzyskania wyniku badania – oprócz samych surowych danych, do których stosuje się to oprogramowanie, algorytm lub model.” (tłum. własne)

Definicje danych badawczych odnoszą się zarówno do zbierania (uzyskiwania danych), jak i do obserwacji czy wytwarzania danych w procesie badawczym. Ich zadaniem jest zapewnienie otrzymania oryginalnych wyników, które można zweryfikować.

W interesującej dla nas części [formularza](#) Narodowego Centrum Nauki z czerwca 2020 roku, wymaganego w procedurze wnioskowania o finansowanie projektu, znajdujemy takie stwierdzenie:

„(...) pod pojęciem „dane” NCN rozumie zarówno dane zebrane i dotąd nieprzetworzone, jak i dane wytworzone i poddane analizie, inne niż publikacje naukowe. Definicja ta obejmuje wszystkie możliwe formaty, zarówno cyfrowe, jak i niecyfrowe (np. próbki, wypełnione kwestionariusze, nagrania dźwiękowe, itd.).”

Ta szeroka definicja nie zawiera konkretów, które są niezbędne do stworzenia własnego opisu danych badawczych. Należy zatem dodać, że dane to fakty, obserwacje, doświadczenia, na których opiera się nasza argumentacja – dane w każdym stadium obróbki i różnego typu. Mogą to być raporty laboratoryjne albo terenowe, kwestionariusze, nagrania audio i wideo, modele, fotografie oraz symulacje; nasz zbiór danych może zawierać slajdy, okazy, filmy, artefakty czy próbki; może być zbiorem metadanych i zawierać informacje o pochodzeniu danych badawczych: jak, kiedy i gdzie je pozyskaliśmy, a nawet w jaki sposób i za pomocą jakich przyrządów; możemy uwzględnić rodzaj oprogramowania, z którego korzystamy (jego kod, opis). W przypadku dyscyplin artystycznych dane badawcze obejmują wszelkie artefakty będące efektem ludzkiej pracy – teksty, dzieła sztuk wizualnych, performatywnych, muzyki. Co istotne, dane badawcze to zbiór „żywy”,

powinniśmy postrzegać go jako stale ewoluujący i wielopłaszczyznowy i obserwować każdy etap jego powstawania, a wyniki i sam proces obserwacji traktować jako element zbioru. Powszechnie wyróżnia się surowe dane badawcze – czyste, uzyskane bezpośrednio w czasie badań – oraz dane przetworzone, czyli poddane obróbce. Każda dyscyplina wytwarza specyficzne dane badawcze.

Rosnące wymagania dotyczące ewaluacji nauki nakładają na badacza obowiązek przyglądania się własnej pracy z szerszej perspektywy. Mamy nadzieję, że zgromadzone tutaj informacje pomogą usystematyzować wiedzę na temat danych badawczych i pozwolą przejść swobodnie przez proces ich gromadzenia, opisu i udostępniania.

Dowiedz się więcej...

- [*Practical Guide To The International Alignment of Research Data Management*](#)
- [*Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities*](#)
- [*What is this about Research Data?* – Helen Cooper \(University of Kent\)](#)
- [*Dyrektywa Parlamentu Europejskiego I Rady \(UE\) 2019/1024...*](#)
- [*The EU's open science policy*](#)
- [*Open access takes flight* – Jeffrey Brainard](#)
- [*A Big Science Publisher Is Going Open Access. But at What Cost?* – Grigori Guitchounts](#)
- [*Otwarty dostęp* – Elsevier](#)
- [*Rekomendowane bazy danych* – Curtin University](#)
- [*Narzędzia dla badaczy od Digital Science & Research*](#)
- [*What Is Humanities Research Data?* – Lafayette](#)

Dlaczego udostępniamy nasze dane badawcze?

Do niedawna wydawało nam się, że dużym skokiem w otwartość badań naukowych jest publikowanie ich wyników w przestrzeni wirtualnej, która byłaby dostępna dla każdego bez konieczności udzielania opłaty czy podawania informacji na swój temat i prośby o zgodę. Tak rzeczywiście było. Teraz zmierzamy jednak o krok dalej i pozwalamy nie tylko czytać wyniki naszych badań podane w formie publikacji, ale także weryfikować je w oparciu o nasze dane badawcze.

Rozwój nauki i lepszy świat

Badania nad wirusem SARS-CoV-2 [pokazały](#), jak ważne i pożyteczne może być dzielenie się danymi badawczymi w przestrzeni wirtualnej. Mądrzejsi o doświadczenia związane z rozprzestrzenianiem się wirusa Ebola, naukowcy postanowili współpracować dla dobra ludzkości, by móc szybciej i skuteczniej zahamować rozwój nowej epidemii. Brytyjska organizacja Wellcome Trust nawoływała do udostępniania wyników badań w celu rozpowszechnienia wiedzy i ulepszenia terapii – odzew otrzymała od około stu innych organizacji, w tym uniwersytetów czy wydawców. [„The Guardian” podaje](#), iż obecnie można oszacować globalny odsetek danych w otwartym dostępie na 37%, co wydaje się dobrym wynikiem jak na nie tak długo trwającą kampanię otwierania nauki.

Otwarte Dane Badawcze (ang. Open Research Data) to takie, do których każdy ma dostęp, bezpłatny i niewymagający zawierania żadnej umowy. Mogą być ponownie wykorzystywane, modyfikowane i udostępniane, choć oczywiście z poszanowaniem prawa. Jedną z najważniejszych praktyk Otwartej Nauki jest właśnie otwieranie danych, które nie tylko przyspiesza uzyskiwanie kolejnych wyników, ale i wpływa korzystnie na rozwój gospodarczy. Na ideę Otwartej Nauki składają się ponadto: otwarty dostęp do publikacji (ang. Open Access) oraz otwarta komunikacja naukowa (ang. Open Scholarly Communication). Nazywane [„nową walutą w dobie otwartej nauki”](#), dane badawcze stanowią podstawę weryfikacji jakości i przejrzystości procesu badawczego. Ich udostępnienie pozwala stwierdzić, czy nasze badania są wartościowe – nie tylko innym, ale i nam samym.

Podstawowa korzyść z udostępniania danych to dobro społeczne i rozwój nauki, jest jednak wiele innych, mniej ważnych powszechnie, ale istotnych dla samych badaczy powodów, o których warto pamiętać.

Finansowanie badań i ewaluacja nauki

O ile do tej pory otwartość danych badawczych była mile widzianym dodatkiem, tak dziś zaczyna stanowić jeden z obligatoryjnych punktów oceny każdego projektu naukowego. Na posiedzeniu Rady Narodowego Centrum Nauki z kwietnia 2019 roku wyraźnie zasygnalizowano kierunek, w jakim zmierzać będzie finansowanie badań naukowych na gruncie polskim. W [raporcie](#) z posiedzenia czytamy bowiem:

„Rada NCN w porozumieniu z dyrekcją Centrum, w związku z planowanym wprowadzeniem przez NCN polityki otwartego dostępu do danych wytworzonych lub wykorzystywanych w ramach realizacji projektów finansowanych przez NCN, zdecydowała się na wprowadzenie w konkursach MAESTRO i SONATA BIS po raz pierwszy obowiązku załączania do wniosków informacji dotyczących planu zarządzania danymi powstałymi lub wykorzystanymi w trakcie realizacji projektu badawczego.”

Decyzje te nie zapadały rzecz jasna w próżni, lecz wynikają z ogólnoświatowych tendencji do wprowadzania polityki otwartości w jednostkach naukowych. Warto wiedzieć, co na temat udostępniania danych badawczych [mówi Parlament Europejski](#):

„Dane badawcze różnią się od artykułów naukowych relacjonujących i komentujących ustalenia wynikające z badań naukowych będących ich źródłem. Przez wiele lat otwarta dostępność i możliwość ponownego wykorzystywania wyników danych z badań naukowych finansowanych ze środków publicznych były przedmiotem konkretnych inicjatyw politycznych. Otwarty dostęp rozumiany jest jako praktyka zapewniania użytkownikom końcowym przez internet bezpłatnego dostępu do wyników badań, bez ograniczeń w zakresie korzystania i ponownego wykorzystywania, wykraczającego poza ewentualny wymóg podania autora. Polityka otwartego dostępu ma w szczególności na celu zapewnienie naukowcom i ogółowi społeczeństwa dostępu do danych badawczych na jak najwcześniejszym etapie procesu rozpowszechniania oraz ułatwienie korzystania z nich i ich ponownego wykorzystywania. Otwarty dostęp przyczynia się do zwiększenia jakości, ograniczenia konieczności zbędnego powielania badań, do przyspieszenia postępu naukowego, do zwalczania oszustw w dziedzinie nauki, a także może ogólnie sprzyjać wzrostowi gospodarczemu i innowacyjności.”

Publikacja naukowa jest w pełni i naprawdę otwarta dopiero wtedy, gdy towarzyszą jej dane badawcze, na podstawie których wysnuwano wnioski – stąd dążenie do wykształcenia w środowisku naukowym przekonania o słuszności czy wręcz konieczności dzielenia się danymi. Informacja o istnieniu takich danych pozwala uniknąć powielania się badań, zaoszczędzić czas kolejnym badaczom, dowieść, że dane są prawdziwe i posłużyły do opisanych w publikacji celów.

Różnego rodzaju agencje finansujące badania coraz częściej wymagają od aplikujących badaczy Planu Zarządzania Danymi (ang. Data Management Plan) i to już na wczesnym etapie składania wniosku. Również redakcje czasopism niekiedy oczekują od autorów udostępnienia danych, z których korzystali, przygotowując pracę. [Polityka otwartości Unii Europejskiej](#) jest pod tym względem bardzo zaawansowana, a wszystkie podmioty, które chcą korzystać ze środków finansowych UE, muszą brać to pod uwagę.

Większa widoczność i rzetelność badań

Otwarte Dane Badawcze to nie tylko dawanie czegoś innym, ale również korzyści dla badacza, który swoje dane udostępnił. Popularyzacja badań naukowych jest niezwykle ważna, a im bardziej widoczny naukowiec w sieci, tym lepiej dla niego i jego jednostki. Dlaczego warto dzielić się swoimi danymi badawczymi?

1. **Weryfikacja** – udostępnianie danych badawczych służy przede wszystkim ocenie jakości nauki. Jeżeli prowadzimy rzetelne badania, nie bójmy się dzielić nie tylko wynikami, ale także danymi, na podstawie których powstaje nasza praca.
2. **Interakcja** – umieszczenie danych badawczych w dostępnej dla każdego przestrzeni wirtualnej może sprawić, że otrzymamy o wiele więcej informacji zwrotnych, niż gdybyśmy zachowali je dla siebie. Interakcja ze środowiskiem naukowym zawsze niesie obopólne korzyści.
3. **Cytowalność** – im łatwiej dotrzeć do naszych danych badawczych, tym więcej innych autorów z nich skorzysta, a każde ponowne wykorzystanie danych zostawia ślad. Udostępniamy swoją pracę na określonej licencji, a uznanie autorstwa jest podstawą dalszego życia naszych danych.
4. **Widoczność** – dzieląc się danymi badawczymi, zezwalając na ich ponowne wykorzystanie, stajemy się bardziej rozpoznawalni w środowisku naukowym i postrzegani jako badacze rzetelni, którzy nie obawiają się weryfikacji swoich badań, a z ewentualnych interakcji chcą czerpać wiedzę. To dobra marka.
5. **Wymiana** – dzielenie się danymi badawczymi sprawia, że zaczynamy odkrywać nowe rejony nauki, z których potem sami korzystamy. Repozytoria danych badawczych czy *data journals* umożliwiają deponowanie danych, ale i pobieranie danych od innych badaczy. Wymiana informacji to podstawa dobrej komunikacji naukowej.
6. **Oszczędność** – możliwość wykorzystywania już istniejących zasobów pozwala nam zarówno zaoszczędzić czas, jak i obniżyć koszty własnych badań.
7. **Rozwój** – korzystanie z udostępnionych danych badawczych może zaowocować całkiem nowymi badaniami na ich podstawie. Dane, przekazane w sposób zunifikowany, można łączyć z innymi danymi, osiągając ciekawe rezultaty.
8. **Bezpieczeństwo** – deponowanie danych w repozytoriach danych badawczych pozwala na ich bezpieczne przechowywanie. W razie utraty własnego sprzętu, nośnika czy problemów z serwerem uczelni, dane mają dodatkową, bezpieczną lokalizację.
9. **Dostęp** – nierzadko zebrane przez badacza dane są unikalne i nie można zebrać ich ponownie. Dzielenie się nimi zwiększa szanse na ich lepszą eksploatację i dłuższe życie.
10. **Satysfakcja** – udostępniając dane badawcze w takim miejscu jak repozytorium, mamy w ręku narzędzia umożliwiające monitorowanie ich recepcji. Możemy sprawdzić, ile razy zostały pobrane i śledzić cytowania.

Zasady, na jakich udostępniamy, otwieramy dane badawcze, zostały dogłębnie przemyślane, a służące do tego narzędzie zaprojektowane tak, by badacz dzielący się własną pracą niczego nie tracił. Zachęcamy do zapoznania się z dalszymi informacjami i odkrywania korzyści, jakie daje Otwarta Nauka.

Dowiedz się więcej...

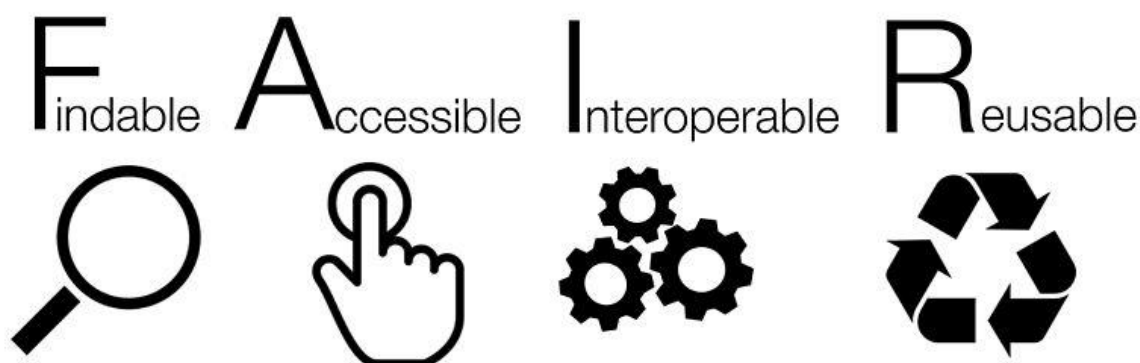
- [What is this about Research Data? – Helen Cooper \(University of Kent\)](#)
- [Dyrektywa Parlamentu Europejskiego I Rady \(UE\) 2019/1024...](#)
- [The EU's open science policy](#)
- [Otwarta Nauka w walce z pandemią koronawirusa SARS-CoV-2](#)
- [Data rules – OK? – Paul Ayris \(University College London\)](#)
- [Otwieranie danych. Podręcznik dobrych praktyk – Ministerstwo Cyfryzacji](#)
- [Sorbonne declaration on research data rights](#)

Zasady FAIR w nauce

Dane powinny być otwarte tak, jak to możliwe, a zamknięte na tyle, na ile to konieczne.

Powyższe zdanie stanowi wartą zapamiętania esencję tego, co powinniśmy wiedzieć o otwartych danych badawczych. Zachowanie równowagi to coś, na czym zależy wszystkim podmiotom nauki, nie tylko badaczom. Dzielenie się danymi badawczymi powinno przebiegać według określonych zasad, które pozwolą udostępnić je w najlepszy sposób i możliwie najszerzej, przy zachowaniu wszelkich standardów bezpieczeństwa. Prawo i rozsądek nakazują chronić zarówno interesy badaczy i agencji finansujących badania, które liczą na wymierne korzyści, jak i osób postronnych uczestniczących w procesie wytwarzania danych, np. w roli ankietowanych, pacjentów, rozmówców, właścicieli badanych utworów czy artefaktów.

Należy wiedzieć nie tylko, dlaczego warto i trzeba udostępniać dane badawcze, ale i jak to robić. Badaczom służą pomocą międzynarodowe zasady właściwej pracy z danymi, tzw. zasady FAIR. Nie są to sztywne standardy, których musimy się trzymać, ale niewątpliwie stanowią przydatne narzędzie do oceny, czy dzielimy się danymi w skuteczny i dobry dla wszystkich sposób. Dane, które z różnych przyczyn muszą pozostać zamknięte, wciąż mogą spełniać kryteria bycia **FAIR**.



Autorstwo grafiki: [SangyaPundir](#)

Creative Commons [Attribution-Share Alike 4.0 International](#)

F jak FINDABLE, czyli znajdź dane bez problemu

Dane powinny być „wyszukiwalne”, tzn. wykrywalne dzięki metadansom, identyfikowalne i możliwe do zlokalizowania za pomocą standardowych mechanizmów.

1. **Trwały identyfikator** (ang. Persistent Identifier, PID) to najlepsza gwarancja, że nasze dane badawcze będą zawsze rozpoznawalne. Jego głównym zadaniem jest dostarczenie informacji niezbędnych do identyfikacji danych – obiektu, publikacji czy osoby. Może być powiązany z zestawem metadanych danego obiektu, niekoniecznie z samym obiektem. To bardzo ważny element udostępniania danych, gdyż jednoznacznie je identyfikuje i ułatwia cytowanie. Warto upewnić się, że miejsce, w którym udostępniamy dane, zapewnia taki trwały identyfikator. Najczęściej spotykane w nauce unikalne identyfikatory należą do dwóch typów – jedne odnoszą się do obiektów, czyli publikacji, danych, oprogramowania itd. (np. DOI, URN, Handle, ARK), drugie natomiast do osób (np. ORCID, ISNI). Ogromną zaletą tego standardowego mechanizmu identyfikowania danych jest możliwość dotarcia do nich nawet wtedy, gdy zmieni się adres internetowy zbioru.
2. **Metadane** (ang. Metadata) stanowią drugi najważniejszy element prawidłowo opisanych i udostępnionych danych badawczych. Bez nich znalezienie naszego zbioru przez ludzi czy maszyny nie będzie możliwe. Należy pamiętać o trzymaniu się standardów tworzenia metadanych właściwych dla wybranej dziedziny – gdy nie zostały one określone, możemy stosować się do ogólnych schematów. Metadane indeksowane są w ogólnodostępnych bazach, które przeszukujemy sami, ale warto brać pod uwagę również możliwość automatycznej analizy, stąd istotna jest dbałość o zachowanie określonych standardów (zob. m.in. [Dublin Core](#)).
3. **Opis i aktualizacja danych** to uzupełnienie procesu wyszukiwania i identyfikacji danych. Sporządzając opis danych, należy pamiętać o konwencjach przyjętych w danej dziedzinie i stosowaniu słowników dziedzinowych. Istotnym elementem są słowa kluczowe, które pozwolą w prosty sposób wyszukać zbiór danych w wyszukiwarkach typu [Google Dataset Search](#). Ważne jest, by nie usuwać poprzednich wersji naszych danych – mogły zostać już zacytowane i mają przypisany trwały identyfikator.

Znalezienie danych badawczych umieszczonych w sieci jest warunkiem sensowności całego procesu ich udostępniania. Sporządzenie prawidłowego opisu danych, przygotowanie ich do dalszych etapów, czyli zdeponowania w repozytorium lub innym przeznaczonym do tego miejscu, to podstawa pracy naukowej, jeśli myślimy o otwartości. Więcej wskazówek zamieszczamy w rozdziale dotyczącym przygotowywania danych do udostępniania.

A jak ACCESSIBLE, czyli miej łatwy dostęp do danych

W kwestii dostępności danych badawczych umieszczanych w repozytoriach czy specjalizujących się w tym czasopismach zawsze pojawia się wiele wątpliwości. Dotyczą przede wszystkim postulatów otwartego dostępu do danych – łatwego, bezpłatnego i umożliwiającego dalszą obróbkę. Tymczasem otwartość nie musi oznaczać całkowitego uwolnienia naszej własności intelektualnej czy wręcz materialnej – istnieją przesłanki, by ograniczyć lub zamknąć dostęp dla osób postronnych.

1. **Ograniczenie dostępu** jest prawem badacza. Możemy nie udostępniać całości lub części danych, jeśli przeszkodę stanowią przepisy prawa (choćby RODO), zobowiązania będące wynikiem zawartych umów albo sam proces wykorzystywania przez nas danych w trakcie trwania projektu i później. Warto pamiętać o istotnym zdaniu: dane powinny być otwarte tak, jak to możliwe, a **zamknięte na tyle, na ile to konieczne**. Nikt nie odmawia nam prawa do zamknięcia dostępu w pewnych okolicznościach. Jeśli dostęp musi zostać ograniczony, należy jednak zaznaczyć to w opisie zbioru i pamiętać o otwarciu przynajmniej metadanych – będą wizytówką naszego zbioru danych, dostępną nawet po przeniesieniu danych lub ich usunięciu. Dostęp do metadanych to plan, który możemy określić jako minimum otwartości.
2. **Przechowywanie danych** to kluczowa kwestia i dotyczy zarówno danych otwartych w pełni, jak i tych z ograniczonym dostępem. Miejsce zdeponowania danych jest niezwykle ważne, ponieważ dane (lub przynajmniej metadane) muszą być zawsze dostępne i osiągalne, np. gdy znamy numer DOI, nic nie powinno już stać pomiędzy nami a zbiorem danych, który nas interesuje. Wybór odpowiedniego **repozytorium otwartych danych** nie jest łatwe, gdy mamy ich do dyspozycji coraz więcej, zarówno dziedzinowych (DARIAH, Azon, Copernicus), jak i ogólnych (Zenodo, Mendeley Data czy RepOD). Więcej informacji na temat deponowania danych zamieściliśmy w innym rozdziale, ale na początku warto zwrócić uwagę przede wszystkim na wymagania prawne i techniczne konkretnych repozytoriów (dostępne formaty plików, wielkość danych). Należy sprawdzić ponadto, czy właściciel repozytorium dba o bezpieczeństwo danych, właściwie je opisuje (metadane) oraz dodaje informację na temat licencji. Istnieją spisy repozytoriów, z których możemy skorzystać, takie jak witryna [Registry of Research Data Repositories](#).
3. **Pełny dostęp do danych badawczych**, gdy tylko jest możliwy, stanowi idealne rozwiązanie – w tym planie maksimum dostęp do zbioru danych odbywa się bezpośrednio dzięki metadansom i trwałemu identyfikatorowi, bez potrzeby instalowania dodatkowego oprogramowania czy narzędzi. Jeżeli zbiór wymaga wykorzystania jakiegoś oprogramowania, jego dokumentacja także powinna być dostępna. Najlepiej, by potrzebne oprogramowanie było otwarte, a protokoły używane do pobierania danych standardowe.

Warto dowiedzieć się więcej na temat repozytoriów danych badawczych i czasopism, w których możemy dane publikować, a także przeanalizować prawne aspekty udostępniania naszej pracy. Szerzej piszemy o nich w dalszych rozdziałach.

I jak INTEROPERABLE, czyli łączyć z innymi

Interoperacyjność danych badawczych oznacza, że można je wymieniać, łączyć z innymi danymi i aplikacjami. Dane powinny być przygotowane i opisane w taki sposób, by kolejni badacze mogli z nich skorzystać bez konieczności wprowadzania wielu zmian czy kupowania zastrzeżonego oprogramowania. Ponowne wykorzystanie danych może nastąpić w dowolnej części świata i warto mieć to na uwadze.

1. Dane i metadane powinny być udostępnione **w formacie, który zapewni ich łatwy odczyt i przetwarzanie**. Najlepiej, żeby nie wymagały konwersji ani mapowania, a oprogramowanie służące do ich przetwarzania nie było zastrzeżone.
2. Przygotowując dane, należy **trzymać się standardów** danej dziedziny oraz stosować prawidłową metodologię. Ważne, by były one czytelne dla badacza (lub urzędnika) posługującego się typowymi dla naszej dziedziny słownikami. Dotyczy to również metadanych.
3. Interoperacyjność danych polega także na **powiązaniu ich z pozostałymi zbiorami danych**, jeśli dotyczą jednego projektu – należy pamiętać o odpowiednich odnośnikach i informacjach.

Pewna uniwersalizacja danych badawczych jest konieczna, by ułatwić łączenie ich z innymi zbiorami i ocenę potencjalnej przydatności. Dbanie o zachowanie standardów danej dziedziny czy naukowości jako takiej, jeśli nasz dziedzina nie wypracowała odpowiednich słowników lub metod, będzie korzystne nie tylko dla badaczy z innych instytucji czy krajów, ale również dla nas samych, gdy zechcemy do tych danych powrócić i wykorzystać je w kolejnym projekcie.

R jak REUSABLE, czyli użyj ponownie

Ponowne wykorzystanie udostępnionych danych to ogromny potencjał naukowy, który pozwala zaoszczędzić czas i środki finansowe. Należy więc zadbać o to, by zasady „przetwarzalności” naszych danych były jasne i wyraźnie zawarte w ich opisie.

1. **Dokładny opis danych i rozbudowane metadane** to podstawa, gdyż ułatwiają użytkownikom określenie przydatności zbioru dla dalszych badań. Powinniśmy udostępnić pełną dokumentację danych, jeśli tylko jest to możliwe. Każdemu badaczowi zależy na sygnowaniu własnej pracy, dlatego warto pamiętać o wyraźnym wskazaniu pochodzenia naszych danych – to ważna informacja dla kogoś, kto zechce ich użyć.
2. Kluczowa jest też **informacja o licencji**, na jakiej udostępniamy zbiór danych – to ona określa, w jaki sposób kolejni użytkownicy mogą używać naszych danych. Najbardziej pożądane są oczywiście licencje mniej restrykcyjne, czyli umożliwiające jak najszersze ponowne wykorzystanie danych (np. w programie „Horyzont 2020” zaleca się CC-0 lub CC-BY). Na temat licencji piszemy więcej w rozdziale poświęconym prawnym aspektom zarządzania danymi.

3. Warto do dokumentacji **dodać plik README**, aby upewnić się, że kolejni użytkownicy bez problemu znajdą najistotniejsze informacje i poprawnie je zinterpretują. Może to być najzwyczajniejszy plik tekstowy lub dokument PDF z informacjami, takimi jak: nazwy plików z naszego zbioru wraz z krótkimi opisami (opisy tabel, rysunków itd., jeśli takie są), stosowane kody danych i jednostki miary, opis etapów przetwarzania danych (mogą mieć wpływ na interpretację danych), wskazanie innych miejsc zdeponowania tych konkretnych danych (jeżeli są), kontakt do kogoś, kto odpowie na ewentualne pytania.

Każdemu badaczowi zależy na tym, by jego dane badawcze były udostępniane w sposób bezpieczny i prawidłowy, gwarantujący mu możliwość zastrzeżenia sobie praw do pierwszeństwa czy własności. Dbłość o wskazanie odpowiedniej licencji oraz opatrzenie zbioru pełną dokumentacją to elementy ważne zarówno dla właściciela, jak i dla kolejnych użytkowników. Pomocą służą nam poradniki i kreatory licencji, takie jak [EDUAT](#) czy strona [The Digital Curation Centre](#) (DCC), a często wytyczne dołączane do konkretnych projektów i programów finansowania.

Warto być FAIR

Zasady FAIR nie są standardem naukowym, którego musimy przestrzegać bez żadnych odstępstw – to zbiór informacji pomocny w zrozumieniu całego procesu udostępniania danych badawczych i prawidłowym przygotowaniu ich do deponowania. Trzymanie się zasad FAIR pozwala o wiele łatwiej stworzyć Plan Zarządzania Danymi (ang. Data Management Plan) – dokument wymagany obecnie przez wiele podmiotów finansujących badania naukowe.

Przestrzeganie zasad FAIR **nie oznacza**, że nasze dane muszą być udostępnione w sposób całkowicie otwarty dla wszystkich. Będą prawidłowo opisane i zdeponowane, ale otwarte na tyle, na ile tego chcemy. Podobnie jak całkowite otwarcie danych nie oznacza, że będą one zgodne z zasadami FAIR – jeżeli nie opiszemy ich w sposób właściwy lub nie dodamy odpowiedniej dokumentacji i informacji o licencji, będą one otwarte tylko pozornie, a tak naprawdę nie w pełni użyteczne. Warto więc zadbać o to, by nasz zbiór danych można było znaleźć, uzyskać do niego dostęp, połączyć z innymi danymi i ponownie wykorzystać.

Zasady FAIR wciąż są rozwijane i precyzowane przy udziale międzynarodowej społeczności, która bierze pod uwagę indywidualne doświadczenia badaczy z różnych krajów i reprezentujących różne dziedziny nauki. Nie ma więc obaw, że postępując zgodnie z wytycznymi, stale aktualizowanymi przez inicjatywę [GO FAIR](#), przeoczymy istotne zmiany technologiczne czy prawne. Warto zaglądać na stronę GO FAIR Initiative również ze względu na oferowane tam webinaria i szkolenia.

Dowiedz się więcej...

- [Making FAIR data a reality... and the challenges of interoperability and reusability – Simon Hodson \(CODATA\)](#)
- [The FAIR Guiding Principles for scientific data management and stewardship – Mark. D. Wilkinson etc.](#)
- [Horizon 2020 Commission expert group on Turning FAIR data into reality](#)
- [Guidelines on FAIR Data Management in Horizon 2020](#)
- [FAIR Principles](#)
- [The PID Forum](#)
- [How to make your data FAIR?](#)
- [Dublin Core Metadata Initiative](#)
- [List of Metadata Standards](#)
- [Registry of Research Data Repositories](#)
- [EDUAT. Store and Publish Your Research Data](#)
- [How to License Research Data?](#)

Przygotowanie danych do udostępniania

Decyzja o udostępnieniu danych powinna zawsze zostać poprzedzona staranną analizą. Składają się na nią konkretne kroki, które opisujemy poniżej, a które pomogą nie tylko wybrać przyszłe miejsce deponowania zbiorów danych, ale także przygotować Plan Zarządzania Danymi. Poznajmy nasze dane badawcze – to najlepszy sposób, by zacząć nimi zarządzać.

Analiza i selekcja danych badawczych

Czy musimy udostępniać wszystkie nasze dane? Oczywiście nie, co jest zgodne z zasadami FAIR i podstawową formułą: otwarte tak, jak to możliwe, a zamknięte tak, jak to konieczne. **Wstępna selekcja danych** pozwoli nie tylko uniknąć problemów, wynikających z umów, ochrony danych osobowych itd., ale także oszczędzić pracy i kosztów. Odpowiedzmy sobie więc na kilka pytań:

- Jakie są wymagania agencji finansującej nasze badania?
- Czy wszystkie nasze dane mają wartość naukową?
- Czy są one wyjątkowe, unikalne?
- Czy może takie same dane już zostały udostępnione i możemy z nich skorzystać?
- Czy dany zbiór jest kompletny i umożliwia replikację wyników badań?
- Jakie koszty będzie generowało zarządzanie danymi i ich przechowywanie?

Dopiero, gdy wybierzemy te dane, które z całkowitą pewnością chcemy i możemy udostępnić, przekonamy się, ile ich jest, jakiego są rodzaju i jak dalej nimi zarządzać. To oczywiste, że na tym etapie dane badawcze nie będą idealne – pisząc Plan Zarządzania Danymi zwykle nawet jeszcze tych danych nie mamy! – więc musimy liczyć z odstępstwami od naszych założeń i lukami. Zawsze warto zaznaczyć, że takie braki są lub mogą się zdarzyć i wskazać ich przyczyny.

Zgoda na gromadzenie, przetwarzanie i udostępnianie danych

Ten punkt wiąże się z naszymi uprawnieniami, wymaganiami podmiotu finansującego badania i prawami osób trzecich, takich jak np. ankietowani czy pacjenci biorący udział w projekcie. Należy pamiętać o uzyskaniu zgody od każdej takiej osoby we wstępnej fazie współpracy. Nawet jeśli okaże się, że nikt od nas takiej zgody nie wymaga, lepiej być przygotowanym.

Anonimizacja i pseudonimizacja danych

Oprócz tego, że musimy uzyskać zgody na przetwarzanie danych, naszym obowiązkiem jest usunięcie danych wrażliwych, czyli tych umożliwiających identyfikację badanych osób. Proces ten może być nieodwracalny, jak anonimizacja, lub odwracalny – pseudonimizacja.

- **Anonimizacja** – to proces polegający na takim przekształceniu danych osobowych, by uniemożliwić przypisanie poszczególnych informacji do określonej osoby ani nawet możliwej do zidentyfikowania.
- **Pseudonimizacja** – polega na przetworzeniu danych tak, by nie udało się ich przypisać do osoby, której dotyczą, bez posiadania dodatkowych informacji.

To badacz musi zdecydować, jakich środków ostrożności wymagają wygenerowane w procesie badawczym dane.

Wybór formatów plików

Formaty plików mogą co prawda być dowolne, ale otwarte dane badawcze to dbałość o łatwy i powszechny dostęp – dlatego najlepiej skoncentrować się na formatach ogólnodostępnych i unikać takich, które będą wymagały użycia komercyjnych narzędzi do odczytu. Wybrany przez nas format powinien wykorzystywać standardowe kodowanie (jak ASCII, UTF-8). Jeżeli reprezentowana przez nas dyscyplina preferuje jakieś formaty plików, warto wziąć to pod uwagę i trzymać się ustalonych, nawet niepisanych zasad. Kolejni badacze sięgający po nasze zbiory danych nie będą wówczas musieli poddawać plików konwersji, co zawsze ma wpływ na jakość danych. **Najczęściej wybierane format plików** dla różnego typu danych to:

- teksty / dokumentacja / skrypty – XML, PDF/A, HTML, Plain Text;
- nieruchome obrazy / zdjęcia – TIFF, JPEG 2000, PNG, JPEG/JFIF, DNG, BMP, GIF;
- dane geoprzestrzenne – Shapefile (SHP, DBF, SHX), GeoTIFF, NetCDF;
- pliki graficzne rastrowe / obrazy rastrowe – TIFF, JPEG2000, PNG, JPEG/JFIF, DNG, BMP, GIF;
- pliki graficzne wektorowe / obrazy wektorowe – Scalable vector graphics, AutoCAD Drawing Interchange Format, Encapsulated Postscripts, Shape files;
- pliki kartograficzne – Most complete data, GeoTIFF, GeoPDF, GeoJPEG2000, Shapefile;
- pliki dźwiękowe / audio – WAVE, AIFF, MP3, MXF, FLAC;
- filmy / video – MOV, MPEG-4, AVI, MXF;
- bazy danych / zbiory danych – XML, CSV, TAB.

Jeśli mamy możliwość, warto dążyć do jak najmniejszych wielkości plików – takie są łatwiejsze w deponowaniu i pobieraniu, ale co istotne, są preferowane przez wyszukiwarki przy wyświetlaniu wyników. W przypadku pliku PDF najlepiej mieścić się w 5 MB.

Istnieją narzędzia pomagające oczyszczać dane i tym samym dopasowywać je do wymagań określonego formatu – jeśli nie zadbamy bowiem o prawidłowy opis danych, ich odczytanie będzie utrudnione. Dbłość o pisownię, gramatykę, skrupulatne umieszczanie identyfikatorów obiektów czy kontekstów jest bardzo istotna. W niektórych przypadkach uporać się z bałaganem w danych pomoże np. [OpenRefine](#), bezpłatne narzędzie do „czyszczenia” danych.

Nadawanie odpowiednich nazw plikom i folderom

Każdy, kto kiedykolwiek szukał wśród swoich plików jednego konkretnego, a nazywa swoje artykuły naukowe „Artykuł” albo „Tekst”, wie, że czas zaoszczędzony na wpisaniu pełnego tytułu i daty upomni się o siebie w przyszłości i wtedy stracimy go z nawiązką. Nadawanie odpowiednich nazw folderom i plikom jest naprawdę istotne. Nazwa powinna mówić coś o zawartości, najlepsza praktyka to stosowanie nazw opisowych. Wiemy, że pliki z danymi nierzadko mają kilka wersji, nadpisujemy w nich nowe informacje, więc warto by odnotowywać zmiany również w nazwie.

Możemy **opracować własny system nazewnictwa**: każdy **plik** może mieć w nazwie skrót od nazwy projektu, opis zawartości, numer wersji i inicjały autora, jeśli nie pracujemy sami. Stworzymy wówczas niepowtarzalny na zewnątrz system, za to powtarzalny i użyteczny w obrębie projektu czy dla danego zespołu badawczego. Podobne zasady dotyczą **folderów**, przy czym należy dodatkowo zadbać o odpowiednią strukturę katalogów, z zachowaną konwencją ich nazywania. Tych samych struktur powinniśmy używać do tworzenia **kopii zapasowych**. Niezalecane jest stosowanie znaków diakrytycznych czy interpunkcyjnych, najlepiej ograniczyć się do liter, cyfr i podkreślników, a daty zapisywać w standardowym formacie. Warto też przewidzieć, jak dużo plików może mieścić się w jednym folderze i numerując je, zastosować sekwencję z zerami poprzedzającymi liczbę, np. gdy zakładamy, że będziemy mieć 98 plików, zacząć numerowanie od 001 aż do 098 – nie będzie problemu, gdybyśmy jednak przekroczyli setkę, bo zachowamy sekwencję.

Jeśli nie uda nam się samodzielnie od początku zadbać o pliki z danymi, np. korzystamy z gotowych zestawów danych lub otrzymaliśmy je od innego członka zespołu, możemy posłużyć się narzędziami do masowej zmiany nazw plików, takich jak: [Renamer 5](#) (Mac – program płatny), [PSRenamer](#) (Linux, Mac, Windows – program darmowy), [den4b ReNamer](#) (Linux, Mac, Windows – wersja darmowa lub płatna).

Wersjonowanie

Zespoły badawcze liczą zwykle minimum kilku członków, rzadkością jest praca w pojedynkę, dlatego warto zadbać o **kontrolowanie wersji naszych plików**. Jeśli nad danym plikiem pracuje więcej niż jedna osoba, trzeba dbać o wersjonowanie – to najlepszy sposób na śledzenie zmian w całym projekcie lub zbiorze danych. Można korzystać ponadto ze wsparcia programów do śledzenia zmian i kontroli wersji, jak [Subversion](#) czy [Git](#).

Dane danych, czyli metadane

Metadane są najważniejszą częścią całego procesu przygotowywania danych do udostępnienia, ponieważ to one często decydują o tym, czy nasze dane w ogóle zostaną odnalezione i staną się częścią otwartej nauki. Opatrzanie zbiorów danych odpowiednim opisem w postaci właśnie metadanych to warunek ich zindeksowania, wyszukania i ponownego użycia, ponadto pomagają zrozumieć same dane i prawidłowo z nich korzystać.

Możemy wyróżnić **trzy podstawowe typy metadanych**:

- **metadane opisowe** – zawierają takie elementy, jak: tytuł, streszczenie, autor i słowa kluczowe; służą dostarczaniu informacji niezbędnych do odnalezienia i zidentyfikowania zbioru danych;
- **metadane strukturalne** – informują o relacjach i zależnościach pomiędzy zbiorami danych i/lub poszczególnymi elementami zbiorów; ułatwiają nawigację;
- **metadane administracyjne** – służą zarządzaniu danym zasobem; zawierają takie informacje, jak: datę i sposób utworzenia zasobu, typ pliku, dostęp.

Do metadanych administracyjnych włącza się te, które dotyczą praw własności intelektualnej albo zawierają informacje dotyczące archiwizacji zasobu.

Nasze metadane powinny zawierać informacje o strukturze danych, o istniejących ograniczeniach czy o zalecanym sposobie cytowania. Wprowadzając prawidłowe i przemyślane metadane, realnie wpływamy na odbiór naszych danych badawczych i ich dalsze życie.

Podstawowe pola metadanych są nam wszystkim znane z deponowania publikacji naukowych, które również mają metadane. Przeważnie są nietłumaczone i funkcjonują w języku angielskim: Dataset name (nazwa zbioru danych), Version (wersja), Licence (licencja), Keywords (słowa kluczowe), DOI (niepowtarzalny identyfikator nadawany przez niektóre wydawnictwa lub repozytoria), Discipline (dyscyplina)... Wszystkie te pola stanowią metryczkę naszego zbioru danych.

Dążenie do standaryzacji w tym zakresie ma na celu usystematyzowanie sposobu opisu danych tak, by był on zrozumiały zarówno dla ludzi, jak i dla programów komputerowych. Można wyróżnić standardy dziedzinowe i ogólne, takie jak Dublin Core i Data Cite albo Data Documentation Initiative (DDI). Do dziedzinowych można zaliczyć m.in. EML (ekologia), TEI i CDWA (nauki humanistyczne) czy DC (nauki o życiu). Istnieją inicjatywy, które dążą do sformalizowania specyfikacji metadanych, takie jak: Research Data Alliance ([RDA](#)), [OpenAire](#) i [Metadata 2020](#).

Do metadanych zaliczymy także opis zmiennych czy słowniki danych, takie jak: Unified Medical Dictionary (UMD), Central Europe Glossary (CEG), UNESCO thesaurus.

Przydatnym narzędziem, nie tylko do edycji metadanych, może okazać się [PDF Candy](#) – bezpłatny edytor PDF online. Można za jego pomocą skompresować PDF, dokonać konwersji najpopularniejszych formatów do PDF czy pociąć plik PDF.

Stan prawny danych i licencjonowanie

Zagadnienia natury prawnej zawsze są mocno zindywidualizowane i należy je przemyśleć w związku z każdym konkretnym projektem osobno. Warto, by takie dyskusje odbyły się na wstępnym etapie, przed opublikowaniem danych badawczych. Oprócz wspomnianego dopilnowania kwestii umów i zgód na przetwarzanie danych ważna jest kwestia odpowiedniej licencji i uzgodnienia dotyczące praw własności intelektualnej. Więcej informacji na ten temat zamieściliśmy w rozdziale „Aspekty prawne zarządzania danymi badawczymi”.

Bezpieczeństwo danych

Popularną i dobrą radą w kwestii bezpieczeństwa danych badawczych jest stosowanie tzw. **reguły 3-2-1**. Polega ona na tworzeniu trzech kopii zapasowych (3) na dwóch różnych nośnikach (2) w tym jednej kopii fizycznej (1), np. w chmurze lub w innym budynku. Dbłość o bezpieczeństwo danych stanowiących podstawę naszej pracy nad projektem to priorytet.

Warto na wstępie opracować **organizację systemu przechowywania danych**, wspólną dla wszystkich uczestników projektu. Będzie do niej należeć, oprócz reguły 3-2-1, wspomniane już wersjonowanie i ukrycie danych wrażliwych, przyznanie poszczególnym badaczom odpowiednich uprawnień, szyfrowanie, ustalona droga udostępniania plików innym osobom, rozwiązanie kwestii etycznych (jeśli dotyczy).

Na ilu nośnikach przechowujemy aktualnie pisany artykuł naukowy? Albo doktorat? Na jednym laptopie? Na dysku zewnętrznym? W chmurze? Może warto we wszystkich tych miejscach naraz. Dobre praktyki archiwizacji danych służą przede wszystkim samym badaczom.

Miejsce deponowania danych – repozytoria

Wybór odpowiedniego miejsca przechowywania naszych danych badawczych wiąże się ze wszystkimi powyższymi punktami. O repozytoriach danych badawczych szerzej piszemy w poświęconym im rozdziale, ale warto pamiętać o kilku istotnych kwestiach. **Otwarte repozytoria danych badawczych** coraz częściej pełnią funkcje archiwów i jest ich też coraz więcej, a nie wszystkie będą pasowały do specyfiki naszych danych. Jak wybrać odpowiednie repozytorium, jeśli instytucja finansująca badania nie narzuca nam konkretnego miejsca? Na co zwrócić uwagę?

Nasza znajomość z wybranym repozytorium będzie długa, jeśli zdeponowany zbiór danych otrzyma stały identyfikator i pod nim właśnie będzie potem „znajdywalna”. Dlatego:

- choć pewnie rzadko to robimy, tym razem **naprawdę sumiennie czytamy regulamin** korzystania z serwisu – jeśli repozytorium nie spełnia naszych wymagań, powinniśmy dowiedzieć się o tym już na tym etapie;
- szukamy informacji o **zabezpieczeniach i zasadach przechowywania danych**, a także o czasie ich przechowywania;
- sprawdzamy, czy repozytorium rzeczywiście oferuje naszemu zbiorowi **trwały identyfikator** (DOI) i czy można w nim **powiązać plik z autorem** (możliwość wprowadzenia swojego numeru ORCID lub innego rodzaju identyfikacji osobowej);
- warto wpisać w wyszukiwanie kilka znanych nam nazwisk badaczy z bliskich dyscyplin i sprawdzić, czy ktoś nam znany deponował w tym repozytorium albo czy w ogóle **tematyka z naszej dyscypliny jest w nim popularna**;
- oceniamy także **standard metadanych**;
- upewniamy się, czy deponowanie i późniejsze przechowywanie plików **na pewno jest bezpłatne** (niektóre repozytoria wprowadzają opłaty – tzw. Data Processing Charge);
- na koniec możemy jeszcze sprawdzić, czy zasoby z wybranego repozytorium są **indeksowane w największych bazach**, takich jak Mendeley Data, Google Dataset Search czy Data Citation Index.

Pomocne podczas wyszukiwania odpowiedniego repozytorium może okazać się skorzystanie z wyszukiwarki typu [RE3DATA](#). Więcej na temat [Repozytorium Uniwersytetu Łódzkiego](#) oraz innych repozytoriów – w dalszej części poradnika.

Koszty związane z zarządzaniem danymi

Koszty naszych badań to zagadnienie, które niewątpliwie bardzo interesuje agencje finansujące badania, zarówno przed przyznaniem środków, jak i w trakcie trwania projektu. Zarządzanie danymi badawczymi również generuje pewne koszty, o których należy pamiętać. Podstawowe rodzaje kosztów, o jakich powinniśmy pomyśleć, to:

- **koszty sprzętu i infrastruktury**, czyli dostęp do jakiegokolwiek specjalistycznej aparatury, np. do dokonywania wysokowydajnych obliczeń czy badań, ale także koszty przechowywania danych w chmurze, nośniki danych, koszty składowania danych na serwerach instytucjonalnych (warto zawsze porozmawiać ze specjalistą IT z danej jednostki);
- **koszty oprogramowania** – jeśli planujemy zakup licencji na oprogramowanie wspierające zarządzanie danymi, musimy to przewidzieć;
- **koszty publikacji** – dotyczą zarówno publikacji wyników badań, jak i danych badawczych, a o ile repozytorium danych badawczych może być bezpłatne, to już publikacja wyników w dobrym czasopiśmie czy bazie niekoniecznie; warto zapoznać się z ofertą naszej jednostki badawczej i umowami, jakie podpisała z zewnętrznymi wydawcami (w przypadku Uniwersytetu Łódzkiego informacje te można znaleźć na stronie Biblioteki UŁ).

Wstępna selekcja i przygotowanie danych badawczych do udostępniania to praca, którą warto wykonać i która nie będzie stratą czasu – wszystkie zgromadzone w ten sposób informacje badacz może wykorzystać podczas sporządzania Planu Zarządzania Danymi. Więcej na ten temat w kolejnym rozdziale.

Dowiedz się więcej...

- [Co to jest anonimizacja i pseudonimizacja?](#)
- [Elsevier na temat danych badawczych](#)
- [OpenRefine – narzędzie do czyszczenia danych](#)
- [Backup skuteczny, czyli zasada 3-2-1 – Jakub Stachecki](#)
- [Registry of Research Data Repositories](#)
- [Selekcja i przygotowywanie danych badawczych do udostępniania – broszura informacyjna DRODB](#)
- [Dane badawcze w pigułce. Poradnik – Polska Platforma Medyczna](#)
- [Otwarte dane badawcze w warsztacie pracy naukowca – Małgorzata Rożniakowska-Kłosińska](#)

Plan Zarządzania Danymi

Otwieranie publikacji i danych badawczych powoli zaczyna zmieniać swój status. Jeszcze niedawno było postrzegane jako dziwne novum, z czasem awansowało na mile widziane udogodnienie, a dziś, gdy [w życie wchodzi Plan „S”](#) i coraz więcej instytucji finansujących badania oczekuje otwartości, zaczyna być normą, a przynajmniej dobrą praktyką. Znakiem tej zmiany jest pojawienie się w wymaganiach grantowych formularza zwanego **Planem Zarządzania Danymi** (ang. Data Management Plan). Krótko mówiąc, to dokument opisujący, w jaki sposób dane badawcze będą w czasie trwania projektu generowane lub/i wykorzystywane. W planie można (a czasem trzeba) zawrzeć informacje na temat gromadzenia, zarządzania, przechowywania i udostępniania danych po zakończeniu działań projektowych.

Należy zdawać sobie sprawę z tego, że Plan Zarządzania Danymi to „żywy” dokument, „work in progress” – może zmieniać się w trakcie trwania projektu i agencje finansujące badania zdają sobie z tego sprawę. Mimo to dobrze przygotowany PZD pozwala uniknąć wielu niechcianych sytuacji, skłania do zastanowienia się nad wieloma kwestiami wcześniej. **Jest formalnym dokumentem, który zawiera zarys tego, co będziemy robić z danymi badawczymi w trakcie i po zakończeniu projektu.**

Dlaczego warto nauczyć się pisać Plan Zarządzania Danymi?

Możemy dużo pisać o korzyściach wynikających z przygotowywania takiego planu, ale warto uświadomić sobie jedno: może się zdarzyć, że tego planu po prostu będzie się od nas wymagać. Nie wszystkie instytucje oczekują od swoich pracowników naukowych przedstawienia PZD, ale wiele z nich już tak, a należy raczej spodziewać się tendencji wzrostowej w tym przypadku. Zewnętrzne instytucje finansujące badania już jasno określiły swoje preferencje – Narodowe Centrum Nauki opublikowało [Wytyczne dla wnioskodawców do uzupełnienia PLANU ZARZĄDZANIA DANymi w projekcie badawczym](#), a Unia Europejska przedstawiła wymóg PZD w programie grantowym Horyzont 2020, wymaga go także Economic and Social Research Council czy Natural Environmental Research Council. Dlatego **warto wiedzieć, czego tak naprawdę oczekuje się od badacza i nie zrażać się samym wyglądem formularza** we wniosku. Plan Zarządzania Danymi może okazać się bardzo użyteczny i pomóc wydajnie, efektywnie i bezpiecznie zarządzać danymi badawczymi.

Co zawiera Plan Zarządzania Danymi?

Dążenie do standaryzacji przejawia się również w PZD i wszystkie będą do siebie bardzo podobne – różnice wynikają przede wszystkim ze specyfiki dziedzin, jakie reprezentują naukowcy. Jest oczywiste, że inaczej będzie wyglądał plan biologa, inaczej socjologa, a jeszcze inaczej religioznawcy, a w obrębie dyscyplin także spotkamy różne warianty. Dlaczego? Wszystko zależy od tego, jakie dane badawcze zamierzamy gromadzić i udostępniać. Wszystkie plany mają jednak **pewne elementy wspólne**, a są to:

- informacje o sposobie wytwarzania/pozyskiwania danych,
- informacje o plikach (formaty, nazewnictwo, przechowywanie),
- opis dostępu do danych,
- aspekty prawne (niekiedy także etyczne),
- kwestie finansowe, infrastrukturalne, kompetencyjne.

Widać więc, że praktycznie wszystkie te zagadnienia poruszyliśmy na etapie przygotowania danych do udostępniania. Najłatwiej przedstawić taki **PZD według formularza NCN-u**, a potem odesłać do przykładowych, wypełnionych już planów z różnych dyscyplin, które poniżej wskażemy. Spójrzmy więc na pola wymagane przez NCN i omówmy je krótko. Czego oczekuje się od badacza w każdym z nich?

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

- Opis metodologii lub oprogramowania, jakie będzie użyte do zebrania nowych danych lub wykorzystania już istniejących.
- Opis rodzaju danych, np. teksty, zdjęcia, grafiki wektorowe, dane liczbowe, wywiady, kody, pomiary, skany.
- Formaty plików – muszą być oczywiście dopasowane do typów plików i najlepiej, by były standardowe, łatwe do odczytu, zapewniające powszechny dostęp.
- Planowana objętość danych.
- Sposób organizacji danych (struktura katalogów, wersjonowanie, nazewnictwo).

2. Dokumentacja i jakość danych

- Metodologia prowadzonych badań, kontekst i źródło.
- Organizacja danych – struktura folderów i wersji, dokumentacja, np. plik ReadMe.txt.
- Metadane – ich organizacja, standardy opisu, np. Dublin Core, DDI.
- Kontrola jakości danych – w jaki sposób będzie sprawdzana poprawność danych.

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

- Miejsce przechowywania danych w czasie projektu.
- Sposoby przechowywania danych, np. reguła 3-2-1, dane analogowe czy zdigitalizowane itd.
- Dane poufne – dbałość o bezpieczeństwo (warto zwrócić uwagę na [wskazówki dot. ochrony danych](#) od Inspektora Ochrony Danych na UŁ).

4. Wymogi prawne, kodeks postępowania

- Wskazanie właściciela praw autorskich i praw własności intelektualnej do wszelkich danych pozyskiwanych i przetwarzanych w projekcie.
- Wszystkie ograniczenia w dalszym użytkowaniu danych.
- Anonimizacja/pseudonimizacja danych osobowych i danych wrażliwych.
- RODO – opis zasad przetwarzania.
- Dla Uniwersytetu Łódzkiego zostały stworzone odpowiednie uchwały i regulaminy dotyczące m.in. praw własności intelektualnej i korzystania z infrastruktury: [Regulamin zarządzania prawami autorskimi](#), [Regulamin korzystania z infrastruktury badawczej](#).

5. Udostępnianie i długotrwałe przechowywanie danych

- Opisać, kiedy dane będą udostępniane – przed czy po ukazaniu się publikacji wyników badań? Czy jest przewidziane embargo?
- Informacja o ograniczeniu dostępu do danych i powodach ograniczenia – jeśli jest.
- Wskazanie okresu przechowywania danych (NCN rekomenduje 10 lat). Dane nie muszą być przechowywane wiecznie – to decyzja zespołu.
- Opis selekcji danych.
- Wskazanie miejsca przechowywania danych po zakończeniu projektu – repozytorium/archiwum – i określić, czy przestrzega zasad FAIR.
- Wskazać wszelkie dodatkowe narzędzie potrzebne do odczytu danych, np. dodatkowe oprogramowanie.
- Informacja o cyfrowym identyfikatorze obiektu (DOI).

6. Zadania związane z zarządzaniem danymi oraz zasoby

- Wskazanie opiekuna danych (ang. Data steward) – może to być osoba lub instytucja odpowiedzialna za proces zarządzania danymi w trakcie i po zakończeniu projektu. Uniwersytet Łódzki nie wyznaczył takiej osoby na stałe – należy ją wskazać spośród członków zespołu, może to być kierownik zespołu lub inna osoba, w zależności od miejsca przechowywania danych.
- Wskazanie osoby odpowiedzialnej za weryfikację i ewentualne poprawienie całego procesu zarządzania danymi – osobą odpowiedzialną za realizację zarządzania danymi. Należy wpisać imię i nazwisko, tytuł, stanowisko, miejsc pracy itd.
- Środki finansowe – określenie, jakie zasoby dodatkowe będą potrzebne do realizacji planu: sprzęt, wiedza specjalistyczna, co stanie się z danymi za 10 lat – czy ich przechowywanie obciąża projekt kosztami?

Odpowiedzi na niemal wszystkie te pytania przygotowujemy już na wstępie – tak jak pisaliśmy w części o przygotowaniu danych do udostępniania, zrobienie tego wcześniej pozwala uniknąć przykrych niespodzianek. Gromadzenie i opis danych zbieranych podczas trwania projektu (rodzaj zbieranych danych, jakość danych, metadane/formaty plików), metodologia i dokumentacja, regulacje prawne, zagadnienia etyczne, przechowywanie danych i tworzenie kopii zapasowych, selekcja, długotrwałe przechowywanie danych i udostępnianie, które/jakie dane zostaną

udostępnione i gdzie po zakończeniu projektu, cyfrowy identyfikator obiektu, opiekun danych, środki finansowe – **wszystkie te informacje to podstawa naszej pracy z danymi i gdy mamy tę wiedzę, umieszczenie jej w Planie Zarządzania Danymi to tylko kwestia użycia odpowiednich sformułowań**. Pomóc mogą nam specjalne narzędzia do tworzenia PZD oraz przykłady, które możemy znaleźć na stronach jednostek naukowych.

Narzędzia do tworzenia i edycji PZD, wskazówki, instrukcje

Istnieje wiele przydanych narzędzi, które ułatwiają tworzenie Planu Zarządzania Danymi. Zwykle są to formuły dostosowane do wymagań grantodawców zagranicznych, ale wspomniane przez nas wielokrotnie dążenie do standaryzacji sprawia, że większość elementów jest powtarzalna. Wiele instytucji tworzy **poradniki, filmy instruktażowe czy artykuły** na ten temat, do których można zajrzeć. Poniżej lista kilkunastu wybranych:

- [DMPonline \(UK\)](#) – przygotowuje szablony PZD dopasowane do wymagań instytucji finansujących badania z Wielkiej Brytanii, University of Edinburgh (ang.)
- [DMPtool \(US\)](#) – przygotowuje szablony PZD zgodnie z oczekiwaniami grantodawców amerykańskich, University of California Digital Centre. To najpopularniejsze narzędzie online do tworzenia PZD. Strona zawiera także przykłady czy opisy dobrych praktyk (ang.)
- [DCC Data Management Plan Content Checklist](#) – narzędzie służące do szybkiego sprawdzenia, jakich informacji może brakować w naszym PZD – lista kontrolna (ang.)
- [The Data Curation Center](#) – brytyjski serwis naukowy – specjalizuje się w zarządzaniu danymi i udostępnia m.in. przykładowe PZD, przewodniki, informacje i wytyczne (ang.)
- [Plan zarządzania danymi badawczymi NCN – najczęściej zadawane pytania](#) (pl.)
- [Poradnik Katolickiego Uniwersytetu Lubelskiego](#) – (pl.)
- [Wskazówki dot. PZD dla historyków](#) – artykuł Susan L. Collins (ang.).
- [Filmik szkoleniowy o Planach Zarządzania Danymi](#) przygotowany przez Research Data Netherlands (ang.)
- [Poradnik](#) opracowany przez Consortium of European Social Science Data Archives (ang.)
- [Szkolenie](#) pomagające przygotować PZD dla lingwistów, University of Pittsburgh (ang.)
- [Porady i wskazówki od CLARIN-u](#) – kreator PZD (ang.)
- [Kreator PZD](#) oferuje również ARGOS (ang.)
- [Praktyczny przewodnik](#) od Science Europe (ang.)

Istnieją ponadto **kursy online** do nauki tworzenia PZD:

- **MANTRA** – bezpłatny kurs online adresowany do studentów, pracowników i bibliotekarzy, przygotowujący do zarządzania danymi i tworzenia PZD (University of Edinburgh; ang.)
- **Research Data Management and Sharing** – bezpłatny pięcioletni kurs online na platformie Coursera – na zaliczenie przygotowuje się pełny PZD oceniany przez innych uczestników; kurs jest wznawiany cyklicznie (University of North Carolina at Chapel Hill, University of Edinburgh; ang.)

Przykładowe Plany Zarządzania Danymi

Wielu naukowców dzieli się swoimi Planami Zarządzania Danymi, co pokazuje, jak wiele dobrego może dać nam pozytywne nastawienie do Otwartej Nauki. Wyszukaliśmy kilka przykładów różnych PZD, żeby łatwo można było dostrzec ich specyfikę i znaleźć przydatne rozwiązania. **Zachęcamy również pracowników Uniwersytetu Łódzkiego do dzielenia się swoimi pomysłami i wzorami planów!** Chętnie je zamieścimy na stronie. Kontakt: openaccess@lib.uni.lodz.pl.

- [Przykłady zagranicznych planów z różnych dyscyplin](#) (The Digital Curation Centre)
- [A technical plan about developing the Scots syntactic atlas](#) (University of Glasgow)
- [Doctoral Dissertation Research: An Agent-Based Model of Population Changes in a Vulnerable Coastal Environment](#) (Louisiana State University)
- [Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013](#) (Mauna Loa Observatory)
- [PZD – nauki o kulturze i religii – konkurs OPUS-19](#) (dr Marek Wojtaszek, Uniwersytet Łódzki)
- [PZD – filozofia – konkurs OPUS-19](#) (dr hab. Robert Podkoński, Uniwersytet Łódzki)
- [Opis PZD dla literaturoznawców](#) (Uniwersytet Rzeszowski)
- [Przykładowe odpowiedzi do PZD dla NCN-u](#) (Uniwersytet Warszawski)
- [Przykładowy PZD dla ekonomistów](#) (Uniwersytet Mikołaja Kopernika w Toruniu)
- [Przykłady autorstwa dra Leszka Szafrąńskiego](#) (Uniwersytet Jagielloński)
- [Przykładowy PZD dla biologów](#) (Uniwersytet Mikołaja Kopernika w Toruniu)

Skonsultuj z nami Plan Zarządzania Danymi!

Zachęcamy pracowników, doktorantów i studentów Uniwersytetu Łódzkiego do korzystania z **bezpłatnych konsultacji** – w miarę możliwości pomożemy w przygotowaniu Planu Zarządzania Danymi. Prosimy o kontakt na co najmniej dwa tygodnie przed terminem składania wniosków do danego grantu czy projektu, żebyśmy mieli realne szanse wspólnie przemyśleć plan.

Kontakt e-mail: openaccess@lib.uni.lodz.pl

Polecamy także szkolenia i spotkania Centrum Nauki UŁ -> [strona](#).

Dowiedz się więcej...

- [Repozytorium Uniwersytetu Łódzkiego](#)
- [Elsevier na temat danych badawczych](#)
- [*Data Management Plans for historians. How to Document and Protect Your Research*, Susan L. Collins, Carnegie Mellon University, 2017, Perspectives on History, the newsmagazine of the American Historical Association](#)
- [*Open access takes flight* – Jeffrey Brainard](#)
- [Extension Of The Open Research Data Pilot In Horizon 2020](#)
- [Narzędzia do zarządzania danymi od CESSDA ERIC](#)

Gdzie deponować i publikować dane badawcze?

Mając już przygotowane i opisane dane badawcze, powinniśmy zastanowić się nad miejscem odpowiednim do ich zdeponowania. Wybór takich miejsc jest naprawdę duży – istnieją repozytoria ogólne, repozytoria instytucjonalne, repozytoria danych badawczych, różne serwisy społecznościowe, a nawet czasopisma przeznaczone do publikowania wyłącznie danych badawczych. Wszystkie te miejsca mają określoną specyfikę, którą warto poznać przed podjęciem wyboru. Badaczom przychodzi do głowy wiele pytań... Czy mogę publikować tylko w repozytorium własnej jednostki naukowej? Czy muszę w ogóle to robić? Czy jeśli mam profil na Twitterze albo dodaję swoje publikacje na Academia.edu albo ResearchGate, to wystarczy, a moje teksty i badania są widoczne? Czy grantodawcy preferują jakieś konkretne repozytoria? Jak sprawdzić, czy repozytorium jest dobre? A jeśli chciał(a)bym połączyć publikację danych i wyników badań, to co mam zrobić?

Najlepiej zacząć od zapoznania się z wszystkimi możliwościami. Można skorzystać z istniejących wyszukiwarek repozytoriów, spośród których najpopularniejszą jest [RE3DATA](#) – Registry of Research Data Repositories, czyli miejsce, w którym znajdziemy liczące się repozytoria (możemy wyszukiwać je [według dziedziny](#), kraju czy typów danych). Innym takim katalogiem repozytoriów jest [OpenDOAR](#). Kolejne narzędzie, powiązane z RE3DATA, to [Repository Finder](#), która to wyszukiwarka wiąże się z projektem [FAIRsFAIR](#). Korzystając z tych stron, możemy znaleźć repozytorium odpowiednie do deponowania naszych danych. Spośród jakich rozwiązań wybieramy?

Repozytoria instytucjonalne

Repozytoria instytucjonalne, jak sama nazwa wskazuje, są związane z konkretnymi jednostkami naukowymi i koncentrują się raczej na deponowaniu treści, których autorami są pracownicy danej jednostki lub takich, które związane są z samą jednostką, np. z wydawcą lub projektem realizowanym w ramach działania tej jednostki. Warto sprawdzić, czy nasza jednostka posiada takie repozytorium, ponieważ zwykle jest to miejsce spełniające wymagania agencji finansujących badania, programów, takich jak granty NCN-u czy Horyzont 2020, a także wydawców, a przy tym możemy liczyć na bezpośredni kontakt z osobami odpowiedzialnymi za zatwierdzanie danych w bazie i szybką pomoc. **Repozytoria instytucjonalne zapewniają szerokie indeksowanie, archiwizację, także długotrwałą, a co najważniejsze – trwały identyfikator dla naszych danych.** Ponadto są nastawione na model niekomercyjny, więc mamy gwarancję bezpłatnej usługi.

Pracownicy, doktoranci i studenci Uniwersytetu Łódzkiego mają możliwość deponowania treści, w tym również danych badawczych, w [Repozytorium Uniwersytetu Łódzkiego](#) (RUŁ).

- RUŁ umożliwia deponowanie danych **zgodnie z zasadami FAIR**, o których pisaliśmy.
- Metadane są dostępne za pośrednictwem protokołu OAI-PMH, a ich opis zachowuje standard Dublin Core.
- RUŁ pełni rolę tzw. repozytorium małych danych – ze względów technicznych musimy brać pod uwagę wielkości plików.
- Zapewniamy **trwały cyfrowy identyfikator zbioru** – *handle* – który jest bezpłatny dla autorów (koszty ponosi Biblioteka Uniwersytetu Łódzkiego).
- Możliwa jest **długotrwała archiwizacja danych**.
- RUŁ dba o **bezpieczeństwo danych** – kopie zapasowe wykonywane są regularnie.
- Repozytorium jest **zintegrowane z Sherpa Romeo** (polityka wydawnicza), autor ma możliwość **wyboru licencji CC i przypisania numeru ORCID**.
- **Dane indeksowane** są w takich miejscach, jak: [OpenAIRE](#), [OpenDOAR](#), [CORE](#), [DART-Europe E-theses Portal](#), [Google Scholar](#).

Zachęcamy do kontaktu e-mailowego: repozytorium@lib.uni.lodz.pl

Repozytoria danych badawczych

Oferta repozytoriów niezwiązanych z jedną konkretną instytucją i otwartych dla wszystkich badaczy okazuje się naprawdę szeroka i problemem nie jest znalezienie miejsca dla swoich danych, lecz wybranie tego najdogodniejszego. Poniżej proponujemy zestawienie największych repozytoriów ogólnych i dziedzinowych.

REPOZYTORIA OGÓLNE

- [RepOD](#) (pl.) – Repozytorium Otwartych Danych – to pierwsze polskie repozytorium tego typu, istnieje od 2015 roku, a prowadzone jest przez zespół Platformy Otwartej Nauki Interdyscyplinarnego Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego. Jest w pełni bezpłatne i zapewnia zbiorom danych numery DOI. Wielkość deponowanych plików nie może przekraczać 5 GB dla każdego pliku.
- [Zenodo](#) (ang.) – międzynarodowe repozytorium danych badawczych, rekomendowane dla tzw. małych danych. Oferuje zasobom numery DOI, umożliwia wersjonowanie, zamykanie

i otwieranie dostępu oraz statystyki użytkownika. Obsługiwane jest przez CERN European Organization for Nuclear Research.

- **Mendeley Data** (ang.) – repozytorium istniejące od 2015 roku pod patronatem firmy Elsevier, otwarte dla społeczności międzynarodowej. Umożliwia śledzenie wykorzystania deponowanych zasobów, oferuje identyfikator DOI, gwarantuje bezpieczeństwo danych.
- **Portal Otwartych Danych EU** (pl.) – oferuje dostęp do otwartych danych publikowanych przez instytucje i organy Unii Europejskiej; jest w pełni otwarte, a znalezione w nim dane można swobodnie wykorzystywać w celach komercyjnych i niekomercyjnych. Stroną zarządza Urząd Publikacji UE.
- **Figshare** (ang.) – repozytorium działające od 2012 roku, umożliwia deponowanie dowolnych formatów plików do 5 GB, oferuje numer DOI, a nawet jego rezerwację przez opublikowaniem danych. Zarządza nim firma Digital Science.
- **E-SCIENCE.PL** (pl./ang.) – to platforma stworzona przez Politechnikę Wrocławską oferująca miejsce do zarządzania danymi badawczymi. Dysponuje takimi narzędziami, jak repozytorium cyfrowe, menadżer bibliografii czy repozytorium kodu.
- **Joint Research Centre Data Catalogue** (ang.) – działa pod banderą Komisji Europejskiej i ma na celu zapewnienie wsparcia technicznego naukowcom, zwłaszcza w związku z programem Horyzont 2020.
- **The Dataverse Project** (ang.) – oferuje oprogramowanie *open source* do gromadzenia, udostępniania, cytowania danych badawczych. Zapewnia trwały identyfikator danych. Projekt związany z Harvard University.
- **DRYAD** (ang.) – repozytorium gromadzące dane różnego typu i w różnych formatach, które można swobodnie popierać i przetwarzać na licencji CC-0. Zapewnia identyfikator DOI, długotrwałe przechowywanie danych, statystyki, oprogramowanie *open source*. To repozytorium ogólnego przeznaczenia, choć z przewagą danych z zakresu nauk medycznych i przyrodniczych. Projekt ten był prowadzony przez British Library i University of Oxford, aktualnie to większe konsorcjum związane z Zenodo.
- **F1000Research** (ang.) – brytyjska otwarta platforma dla naukowców umożliwiająca szybkie publikowanie i deponowanie danych badawczych, oferująca identyfikator DOI. Zasoby indeksowane są w Google Scholar. Platforma należy do Taylor & Francis Group.

REPOZYTORIA DZIEDZINOWE

Nauki ścisłe, przyrodnicze i rolnicze

- **Copernicus** (ang.) – program zarządzany przez Komisję Europejską zapewniający pełny i bezpłatny dostęp do danych badawczych głównie z zakresu nauk o Ziemi: zmiany klimatu, zarządzanie kryzysowe, bezpieczeństwo, badanie atmosfery, lądów i wód itd.
- **AZON** (pl.) – Atlas Zasobów Otwartej Nauki – platforma prowadzona przez Politechnikę Wrocławską przy współpracy z innymi podmiotami, finansowana przez UE. Repozytorium gromadzi dane z zakresu nauk medycznych, farmaceutycznych, inżynieryjnych czy technicznych.
- **IMPACT** (ang.) – Information Marketplace for Policy and Analysis of Cyber-risk & Trust – gromadzi dane z zakresu cyberbezpieczeństwa i cybernetyki. Platforma powstała z inicjatywy US Department of Homeland Security (DHS).
- **PCoE** (ang.) – Prognostic Centre of Excellence – repozytorium danych prognostycznych pod banderą NASA, gromadzi dane pozyskane od różnych firm, instytucji, agencji, jednostek naukowych.
- **International Directory Network** (ang.) – to międzynarodowa inicjatywa służąca zbieraniu danych głównie z zakresu nauk o Ziemi, umożliwiającą dodawania i aktualizowanie zbiorów danych. Zarządzana przez CEOS – Committee on Earth Observation Satellites.
- **„Nature” poleca repozytoria** (ang.) gromadzące dane z zakresu biologii, medycyny, chemii, nauk o Ziemi czy fizyki.
- **dataINRAE** (fr.) – francuskie repozytorium udostępniające głównie dane badawcze dotyczące środowiska, żywienia, żywności i rolnictwa. Nie gromadzi danych zewnętrznych, ale udziela otwartego dostępu do metadanych i wielu danych badawczych.

Nauki społeczne

- **DARIAH** (ang.) – platforma powstała w 2014 roku jako European Research Infrastructure Consortium, integruje badania z zakresu nauk humanistycznych i sztuki cyfrowej z całego świata. Repozytorium DARIAH-DE obejmuje nie tylko prace związane z projektem, ale także dane niezrzeszonych badaczy. Oferuje trwały identyfikator i długotrwałe archiwizowanie.
- **Repozytorium Danych Społecznych** (pl.) – gromadzi dane z zakresu nauk społecznych, zarówno dodawane przez nowych użytkowników, jak i zamieszczone wcześniej w dwóch miejscach: [Archiwum Danych Jakościowych](#) Instytutu Filozofii i Socjologii PAN (ADJ) oraz [Archiwum Danych Społecznych](#) (ADS). Stworzyły one wspólną inicjatywę mającą na celu ujednolicenie sposobu archiwizacji danych społecznych i ułatwienie procesu zarządzania danymi. RDS służy deponowaniu zbiorów danych oraz dokumentacji techniczno-metodologicznej.
- **Archeology Data Service** (ang.) – to akredytowane brytyjskie repozytorium cyfrowe deklarujące długoterminową ochronę powierzonych danych, skoncentrowane na szeroko rozumianych treściach z zakresu archeologii i historii, historii kultury.

- **ReShare** (ang.) – repozytorium należące do UK Data Service, gromadzi dane z zakresu szeroko rozumianych nauk społecznych.
- **DC Data Catalogue** (ang.) – norweska platforma umożliwiająca wyszukiwanie i przetwarzanie danych z zakresu nauk humanistycznych i społecznych; należy do europejskiego konsorcjum CESSDA ERIC.

Nauki humanistyczne

- **DARIAH** (ang.) – platforma powstała w 2014 roku jako European Research Infrastructure Consortium, integruje badania z zakresu nauk humanistycznych i sztuki cyfrowej z całego świata. Repozytorium DARIAH-DE obejmuje nie tylko prace związane z projektem, ale także dane niezrzeszonych badaczy. Oferuje trwały identyfikator i długotrwałe archiwizowanie.
- **CLARIN-PL** (pl./ang.) – polskie konsorcjum naukowe złożone z sześciu jednostek prezentuje repozytorium danych oraz szereg innych narzędzi adresowanych przede wszystkim do lingwistów. Deponuje dane z zakresu nauk humanistycznych i społecznych, od deponenta wymaga podpisania umowy licencyjnej.
- **Prometheus** (ang./niem.) – to cyfrowe archiwum obrazów z dziedziny historii, sztuki, archeologii, łączy bazy z różnych instytucji kultury, muzeów, placówek badawczych. Można w nim przechowywać i publikować własne obrazy, a także wygodnie przeszukiwać bazy i zapisywać obrazy. Należy do Universität zu Köln.
- **Narodowy Korpus Języka Polskiego** (pl.) – zawiera korpus referencyjny polszczyzny, który można przeszukiwać za pomocą wyszukiwarki PELCRA, obejmuje typowe użycia słów i konstrukcji zdaniowych oraz informacje o ich znaczeniu i funkcji. Stanowi wsparcie w badaniach językoznawczych. To wspólna inicjatywa badaczy z Uniwersytetu Łódzkiego, Polskiej Akademii Nauk oraz Wydawnictwa Naukowego PWN.
- **Archeology Data Service** (ang.) – to akredytowane brytyjskie repozytorium cyfrowe deklarujące długoterminową ochronę powierzonych danych, skoncentrowane na szeroko rozumianych treściach z zakresu archeologii i historii, historii kultury.
- **DaSCH** (niem./ang.) – Data and Service Center for the Humanities – jest prowadzone przez Schweizerische Akademie der Geistes- und Sozialwissenschaften od 2017 roku, to platforma danych gromadząca badania z zakresu nauk humanistycznych.
- **WALS** (ang.) – The World Atlas of Language Structures – ta ogromna baza gromadzi dane strukturalne dotyczące języków, działa pod patronatem Max Planck Institute for Evolutionary Anthropology.
- **arthistoricum.net** (niem.) – portal adresowany głównie do historyków sztuki, gromadzący książki, artykuły, czasopisma tematyczne, ale również dane badawcze różnego typu. Deponowane pliki otrzymują identyfikatory DOI. Sygnowany przez dwie niemieckie biblioteki naukowe: Sächsische Landesbibliothek oraz Universitätsbibliothek Heidelberg.
- **DRH** (ang.) – The Database of Religious History – repozytorium danych powstałe z inicjatywy Cultural Evolution of Religion Research Consortium oraz University of British Columbia in Vancouver. Zbiera dane z zakresu nauk teologicznych.

- **DC Data Catalogue** (ang.) – norweska platforma umożliwiająca wyszukiwanie i przetwarzanie danych z zakresu nauk humanistycznych i społecznych; należy do europejskiego konsorcjum CESSDA ERIC.
- **E-LIS** (ang.) – międzynarodowe repozytorium zawierające treści z zakresu bibliotekoznawstwa i informacji naukowej zarządzane przez Università di Napoli Federico II.

Nauki medyczne

- **Vivli** (ang.) – Center for Global Clinical Research Data – platforma koncentrująca się na gromadzeniu i udostępnianiu wyników badań klinicznych. Dysponuje niezależnym repozytorium danych badawczych. Organizacja non-profit, która zarządza platformą, powstała z inicjatywy The Multi-Regional Clinical Trials Center of Brigham and Women's Hospital oraz MRCT Center (Harvard).
- **GHO** (ang.) – Global Health Observatory – repozytorium danych należące do WHO, koncentrujące się na najważniejszych tematach dotyczących zdrowia publicznego.
- **AZON** (pl.) – Atlas Zasobów Otwartej Nauki – platforma prowadzona przez Politechnikę Wrocławską przy współpracy z innymi podmiotami, finansowana przez UE. Repozytorium gromadzi dane z zakresu nauk medycznych, farmaceutycznych, inżynieryjnych czy technicznych.
- **BRENDA** (ang.) – The Comprehensive Enzyme Information System – baza danych skoncentrowana na tych danych badawczych, które dotyczą enzymów, prowadzona przez Technische Universität Braunschweig.

Nauki inżynieryjne i techniczne

- **PCoE** (ang.) – Prognostic Centre of Excellence – repozytorium danych prognostycznych pod banderą NASA, gromadzi dane pozyskane od różnych firm, instytucji, agencji, jednostek naukowych.
- **AZON** (pl.) – Atlas Zasobów Otwartej Nauki – platforma prowadzona przez Politechnikę Wrocławską przy współpracy z innymi podmiotami, finansowana przez UE. Repozytorium gromadzi dane z zakresu nauk medycznych, farmaceutycznych, inżynieryjnych czy technicznych.
- **IMPACT** (ang.) – Information Marketplace for Policy and Analysis of Cyber-risk & Trust – gromadzi dane z zakresu cyberbezpieczeństwa i cybernetyki. Platforma powstała z inicjatywy US Department of Homeland Security (DHS).
- **International Directory Network** (ang.) – to międzynarodowa inicjatywa służąca zbieraniu danych głównie w zakresie nauk o Ziemi, umożliwiająca dodawania i aktualizowanie zbiorów danych. Zarządzana przez CEOS – Committee on Earth Observation Satellites.

Czasopisma publikujące dane badawcze (ang. data journals)

Wartym uwagi miejscem przeznaczonym do udostępniania danych badawczych są zyskujące na popularności czasopisma, tzw. **data journals**. Jest to sposób na połączenie publikacji i danych badawczych – **publikacje w takim recenzowanym czasopiśmie opisują zbiory danych i zawierają bezpośrednie odnośniki do nich**. Artykuły w takich czasopismach, tzw. data descriptors, oprócz odnośników czy załączników z danymi mogą zawierać informacje jakościowe i ilościowe o danych badawczych, metodologii, licencjach czy źródłach i sposobach pozyskania danych. Może być to zatem świetne uzupełnienie zdeponowania danych w repozytorium i sposób na popularyzację danego zbioru i dorobku badacza.

Wśród *data journals* możemy wyróżnić **czasopisma interdyscyplinarne**, jak „[Scientific Data](#)” czy „[Data in Brief](#)” (teraz połączone z „[Genomics Data](#)”), oraz **czasopisma skoncentrowane na określonych dyscyplinach**: „[Journal of Open Humanities Data](#)” (historia, historia sztuki, literaturoznawstwo, językoznawstwo, filozofia i in.), „[Open Health Data](#)” (medycyna), „[The Journal of Data Mining & Digital Humanities](#)” (psychologia, historia, językoznawstwo i in.), „[Biodiversity Data Journal](#)” (biologia i pokrewne), [Research Data Management in Economic Journals](#) (ekonomia), „[The ZBW Journal Data Archive](#)” (ekonomia i zarządzanie), „[Journal of Open Research Software](#)” (nauki fizyczne, informatyka, matematyka i in.), „[Earth System Science Data](#)” (nauki o Ziemi). Skorzystanie z dziedzinowych czasopism może posłużyć jako bilet do grona badaczy zainteresowanych tymi samymi zagadnieniami. *Data journals* niekiedy udostępniają listy repozytoriów polecanych dla danej dyscypliny, zwłaszcza gdy specjalizują się w jednym obszarze badawczym.

A po co naukowcom serwisy społecznościowe?

Zapewne wszystkim badaczom znane są **naukowe portale społecznościowe**, takie jak [ResearchGate](#), [Impactstory](#) czy [ACADEMIA](#), w których można zamieszczać swoje prace lub ich fragmenty. Zapewniają one możliwość stworzenia przydatnych sieci kontaktów, poznania nowości z interesujących nas dyscyplin, śledzenia konkretnych badaczy... Stanowią niekwestionowane **wsparcie w promowaniu naukowego dorobku i popularyzacji nauki**, podobnie jak media społecznościowe – wielu badaczy dzieli się wynikami swoich badań na [Twitterze](#), niekiedy jest to pierwsze miejsce, w którym ogłasza się ważne odkrycie. Każdy, kto lubi być na bieżąco ze światem nauki i ceni komunikację społeczną, świetnie odnajdzie się w świecie serwisów społecznościowych – tych mniej i bardziej naukowych.

Takie miejsca **nie zastąpią jednak repozytoriów czy czasopism naukowych** w kwestii publikacji ani zdeponowania danych badawczych. Prócz tego, że nie posiadają niektórych ważnych funkcji (trwały identyfikator, bezpieczeństwo danych itd.), to przede wszystkim nie spełniają założeń rzeczywistego otwartego publikowania danych i nie wpisują się w oczekiwania instytucji finansujących badania. Poza tym są nastawione na model komercyjny. Mogą jednak stanowić dopełnienie całego procesu udostępnienia danych – poinformować o nim obserwujące nas osoby i **wpłynąć na zwiększenie cytowalności**. Oprócz tradycyjnych wskaźników bibliometrycznych

spotykamy także **alternatywne wskaźniki cytowalności (ang. altmetrics)**, dotyczące zarówno publikacji, jak i danych badawczych. Kluczowa jest dla nich częstotliwość wspominania o danym zbiorze w mediach i serwisach społecznościowych dla naukowców. Sprawdźmy, [kto mówi o naszych badaniach...](#)

Cytujmy też dane badawcze!

Deponujemy nasze dane badawcze nie tylko po to, by ktoś inny mógł z nich skorzystać i zweryfikować wyniki, ale również z myślą o rozpowszechnianiu własnych badań. Sięgając po udostępnione przez kogoś dane badawcze, musimy pamiętać o ich odpowiednim zacytowaniu – tak jak w przypadku publikacji. W końcu sami chcielibyśmy na pewno, by nasze dane były poprawnie cytowane.

Wygląd opisu bibliograficznego danych badawczych będzie zależał od przyjętego w publikacji **stylu cytowania**. Istnieje wiele takich modeli cytowania i o tym, czy skorzystamy ze stylu APA, Chicago czy IEEE, w dużej mierze decyduje wydawca. Niezależnie jednak od przyjętych norm, opisując dane badawcze, powinniśmy zamieścić najważniejsze informacje o nich, czyli: autora, rok, tytuł, wersję, nazwę repozytorium (lub innego miejsca zdeponowania danych) oraz trwały identyfikator. Jeżeli nasz zbiór danych ma już przydzielony numer DOI, możemy skorzystać z **generатора cytowań**, który pozwala wybrać spośród setek możliwości określenia stylu i stworzyć opis bibliograficzny – [DOI Citation Formatter](#). Jeżeli dzielimy się danymi badawczymi także w mediach społecznościowych czy portalach dla naukowców, warto pamiętać o wspomnianych już [alternatywnych wskaźnikach cytowalności](#).

Każde prawidłowe zacytowanie danych przyczynia się do zwiększenia widoczności samych danych, ale i badacza, pozwala łatwo zlokalizować zbiór danych, poprawia wizerunek badacza i zwiększa jego wiarygodność, a poza tym – co prozaiczne, ale i ważne – po prostu zwiększa liczbę cytowań publikacji, z którymi powiązane są nasze dane.

Dowiedz się więcej...

- [Standard techniczny. Standardy otwartości danych – Ministerstwo Cyfryzacji](#)
- [Otwieranie małych danych badawczych – Marta Hoffman-Sommer](#)
- [Jak korzystać z zasobów w repozytoriach danych](#)
- [Numeryczny styl cytowania w tekście – poradnik Politechniki Gdańskiej](#)
- [Dziedziny Repozytoria Otwartych Danych Badawczych](#)
- [Sharing research data – Elsevier](#)
- [Standardy metadanych – Dublin Core – Marek Zieliński](#)
- [General Repository Comparison – FAIRsharing.org](#)
- [Repozytorium Danych Społecznych - film](#)

Aspekty prawne zarządzania danymi badawczymi

Każdy z etapów prowadzonych przez nas badań podlega lub może podlegać przepisom prawa – różnego typu. Zawsze musimy działać w granicach prawa powszechnie obowiązującego, w którym mieści się m.in. ochrona danych osobowych, brać pod uwagę kwestie etyczne i politykę naszej jednostki naukowej, wszelkiego rodzaju dotyczące nas regulaminy, np. grantowe, prawa osób trzecich i innych członków zespołu badawczego, prawo własności intelektualnej... W ramach projektu „Otwarte Repozytoria Danych Badawczych” realizowanego przez Uniwersytet Warszawski powstał [podręcznik](#), który w klarowny sposób zbiera informacje na temat praw, z jakimi zetknie się badacz na każdym etapie realizacji projektu. Z większości z nich intuicyjnie zdajemy sobie sprawę, jednak należy pamiętać, że specyfika danego projektu badawczego może sprawiać, iż pojawią się kolejne aspekty prawne, choćby kwestia tajemnicy lekarskiej czy prawa patentowego.

Dane badawcze nie rządzą się tylko własnymi prawami

Już na etapie przygotowywania danych badawczych do udostępniania powinniśmy zdawać sobie sprawę z tego, jaki regulacjom one podlegają. Oprócz absolutnej podstawy, czyli prawa powszechnie obowiązującego, możemy spotkać się w procesie badawczym z wieloma prawami szczegółowymi, zawartymi w różnych typach dokumentów, takich jak: ustawy, rozporządzenia, akty prawa UE, umowy (grantowe, zespołów badawczych...), regulaminy (konkursów, projektów, repozytoriów...), kodeksy (np. etyczny), wewnętrzne regulacje jednostek naukowych i finansujących badania, licencje (np. na oprogramowanie). Nikodem Rycko [wymienia](#) m.in. takie **prawa regulujące udostępnianie danych badawczych**:

- prawo własności intelektualnej,
- ochrona danych osobowych,
- ochrona dóbr osobistych, wizerunku, korespondencji,
- przepisy o komercjalizacji wyników badań,
- obowiązek zachowania poufności, tajemnicy,
- zobowiązania wynikające z umów,
- przepisy o badaniach klinicznych,
- przepisy o ochronie zwierząt...

Jedną z najczęściej pojawiających się kwestii są jednak **prawa autorskie**. Z zapisu [Ustawy o prawie autorskim i prawach pokrewnych](#) wynika, że:

„Przedmiotem prawa autorskiego jest każdy przejaw działalności twórczej o indywidualnym charakterze, ustalony w jakiejkolwiek postaci, niezależnie od wartości, przeznaczenia i sposobu wyrażenia (utwór).”

Ochronie podlegają więc takie dane, jak pojedyncza fotografia, nagranie czy utwór literacki, a także zbiory danych, np. dane liczbowe i ich zestawienia. W Ustawie wymienia się ponadto dane tworzone na podstawie danych, utwory pracownicze, utwory naukowe i programy komputerowe; ustawodawca porusza również kwestię utworów zbiorowych oraz współautorstwa.

Specyficznej ochronie podlegają **bazy danych – oprócz prawa autorskiego obejmuje je prawo sui generis**. Pierwsze z nich chroni strukturę bazy danych, natomiast drugie, na podstawie dowodu dokonania znacznych inwestycji przez właściciela bazy, chroni zawartość bazy przed jej nieuprawnionym pobieraniem lub wtórnym wykorzystaniem. Obszerniejsze wyjaśnienia, stale aktualizowane, można znaleźć na stronie [European IP Helpdesk](#).

Jednym z praw najbardziej przestrzeganych w ostatnich latach, zwłaszcza w Unii Europejskiej, jest **ochrona danych osobowych**. Warto o niej pamiętać na każdym etapie zarządzania danymi badawczymi, ponieważ dotyczy nas wszystkich i z danymi osobowymi mamy do czynienia wszędzie. Niektóre projekty badawcze operują ponadto danymi wrażliwymi, czyli takimi, które podlegają szczególnej ochronie, ponieważ dotyczą sfery prywatności lub intymności konkretnych osób. Najczęściej mają z nimi do czynienia nauki społeczne i medyczne (nieanonimowe ankiety, wywiady pogłębione, karty chorób...). [Ogólne rozporządzenie o ochronie danych](#) z 2016 roku, znane wszystkim jako RODO, dodatkowo podkreśla potrzebę ochrony takich danych, a jego wejście w życie sprawiło, że staliśmy się na ich punkcie, nomen omen, *wrażliwi*. O konieczności dbania o poufność danych wspominaliśmy przy okazji przygotowywania danych do udostępniania, pisząc o anonimizacji i pseudonimizacji danych osobowych. Warto pamiętać, że istnieją takie pomocne rozwiązania jak [Amnesia](#), czyli narzędzie do anonimizacji danych (proces uniemożliwiający identyfikację osób według danych), stworzone przez [OpenAIRE](#).

Od badacza piszącego Plan Zarządzania Danymi oczekuje się nie tylko wiedzy o wszystkich tych przepisach prawa, którym dane podlegają, ale również **określenia zasad, na jakich będzie można korzystać z udostępnionych danych badawczych**. Choć można takie zasady opracować samodzielnie, rekomendujemy skorzystanie z gotowych licencji.

Jak wybrać licencję?

Najczęściej wykorzystywanymi w zastrzeganiu (lub uwalnianiu) praw do danych badawczych udostępnianych w ramach otwierania nauki są **licencje Creative Commons (CC)**. Ich podstawowym narzędziem są licencje prawne, które zastrzegają prawa częściowo zamiast całościowo, jednocześnie nie zwalniając z obowiązku poszanowania prawa autorskiego. [Na stronie Creative Commons Polska](#) czytamy:

„Twórca korzystając z licencji zawsze zachowuje prawa autorskie, jednocześnie umożliwia innym kopiowanie i rozpowszechnianie, dodatkowo może określić czy ich wykorzystywanie może odbywać się wyłącznie w warunkach niekomercyjnych lub ograniczyć możliwości tworzenia utworów zależnych.”

Badacz, stając przed dylematem, na jakich zasadach udostępnić swoje dane, musi przede wszystkim zdecydować, czy zgadza się na komercyjne użycie tych danych w ich dalszym życiu i czy zgadza się na tworzenie na ich podstawie utworów zależnych (a jeśli tak, to czy na takiej samej licencji jak oryginał). Nie musimy się obawiać, że wraz z nadaniem licencji przestaje obowiązywać prawo cytatu i dane badawcze powędrują w świat bez naszego nazwiska – **licencje CC nie naruszają wolności wynikającej z prawa autorskiego**. Każda kopia naszych danych musi zostać opatrzona poprawnymi informacjami o autorze i licencji.



Źródło grafik: [Creative Commons Polska](#) CC 4.0

Grafiki oznaczające [podstawowe warunki licencji CC](#) są z pewnością wszystkim znane:

Uznanie autorstwa  „Wolno kopiować, rozprowadzać, przedstawiać i wykonywać objęty prawem autorskim utwór oraz opracowane na jego podstawie utwory zależne pod warunkiem, że zostanie przywołane nazwisko autora pierwowzoru”.

Użycie niekomercyjne  „Wolno kopiować, rozprowadzać, przedstawiać i wykonywać objęty prawem autorskim utwór oraz opracowane na jego podstawie utwory zależne jedynie do celów niekomercyjnych”.

Na tych samych warunkach  „Wolno rozprowadzać utwory zależne jedynie na licencji identycznej do tej, na jakiej udostępniono utwór oryginalny”.

Bez utworów zależnych  „Wolno kopiować, rozprowadzać, przedstawiać i wykonywać utwór jedynie w jego oryginalnej postaci – tworzenie utworów zależnych nie jest dozwolone”.

Wszystkie licencje Creative Commons są darmowe i wszystkie opierają się na uznaniu autorstwa – z tych czterech powyższych podstawowych warunków stworzono **sześć kombinacji licencyjnych, o różnym stopniu otwartości**. Uznanie autorstwa przestaje obowiązywać dopiero wtedy, gdy

chcemy umieścić nasz utwór **w domenie publicznej**, co nie jest zalecane dla danych badawczych – zależy nam widoczności i cytowalności, a więc na śledzeniu dalszych losów naszych badań, nierzadko wiążą nas też umowy niepozwalające na zrzeczenie się praw do danych.



Poniższa grafika jest symbolem Znaku Domeny Publicznej 1.0.

Źródło: [Creative Commons Polska CC 4.0](#)

W polskim prawodawstwie nie spotykamy zrzeczenia się autorskich praw osobistych, jednak w innych krajach niekiedy istnieje taka możliwość. Jej odpowiednikiem wśród licencji CC jest licencja **Creative Commons Zero (CC0)**, której wskazanie oznacza zrzeczenie się przez autora autorskich praw majątkowych i przekazanie utworu do domeny publicznej. Wówczas, korzystając z takiego utworu, nie musimy informować o źródle.

Na stronie Creative Commons Polska znajdziemy [szczegółowe opisy licencji](#), zarówno w postaci pełnego tekstu licencji, jak i w wersji uproszczonej. Warto zapoznać się przynajmniej z opisami wszystkich licencji CC, gdyż spotykamy się z nimi często – wykorzystują je chociażby repozytoria. Prócz najpopularniejszych licencji CC możemy spotkać także inne, jak choćby [GNU General Public License](#) (stosowane jako licencje wolnego oprogramowania) czy licencje [MIT](#).

Warto wiedzieć, że istnieją **licencje zaprojektowane specjalnie dla zarządzania danymi** – powstały w ramach projektu [Open Data Commons](#), prowadzonego przez Open Knowledge Foundation. Są one przeznaczone do licencjonowania baz danych:

- **Open Data Commons Open Database License (ODbL)** – licencja zezwalająca na dalsze przetwarzanie oraz rozpowszechnianie danych pod warunkiem uznania autorstwa i rozpowszechnienia na tych samych warunkach.
- **Open Data Commons Attribution License (ODC-By)** – licencja zezwalająca na kopiowanie i modyfikowanie danych pod warunkiem uznania autorstwa.
- **Open Data Commons Public Domain Dedication and License (PDDL)** – zakłada nieograniczoną możliwość kopiowania, modyfikowania, dalszego udostępniania baz danych; domena publiczna.

Należy pamiętać, że korzystając z udostępnionych zbiorów danych, musimy zamieścić odpowiednią informację (chyba że mamy do czynienia z domeną publiczną i jest to wyraźnie oznaczone), czyli imię i nazwisko autora danych, tytuł zbioru, źródło, typ licencji i link do licencji, jeśli nim dysponujemy. **Naruszenie zapisów licencji wskazanej przez autora danych może skutkować roszczeniami** zgodnie z zapisem Ustawy o prawie autorskim i prawach pokrewnych.

Wybór licencji powinno poprzedzić rozeznanie w naszej sytuacji jako badacza – czy tylko my mamy prawa do naszych danych badawczych? Czy dotyczy to osób trzecich? Jakim regulacjom podlegamy jako pracownicy danej jednostki naukowej albo uczestnicy konkursu grantowego? Warto zapoznać się z polecanymi materiałami, a w razie wątpliwości skontaktować się z działem prawnym jednostki.

Dowiedz się więcej...

- [Czy „punktoza” wpływa na nieetyczne zachowania naukowców? Sprawdzili to psychologowie z UŚ – Ludwika Tomala](#)
- [Prawne aspekty otwierania danych badawczych – podręcznik – Krzysztof Siewicz, Nikodem Rycko](#)
- [Udostępnianie danych badawczych – zagadnienia prawne – prezentacja, Nikodem Rycko](#)
- [Ochrona baz danych](#)
- [Ochrona własności intelektualnej](#)
- [Bazy danych – aspekty prawne](#)
- [Ochrona baz danych – European IP Heldesk](#)
- [Urząd Ochrony Danych Osobowych](#)
- [Poznaj licencje Creative Commons](#)
- [Licencje Open Data Commons](#)
- [Regulamin ochrony własności intelektualnej na UŁ](#)
- [Commercialisation of research results – cooperation between science and business – Dariusz M. Trzmielak](#)
- [Przewodnik po prawie autorskim – Legalna Kultura](#)
- [Domena publiczna – Koalicja Otwartej Edukacji](#)
- [Centrum Cyfrowe](#)

Szkolenia i kursy

- Szkolenia oferowane przez Platformę Otwartej Nauki (pl.) -> [strona](#).
- Szkolenia oferowane w ramach programu Działalności Repozytoria Otwartych Danych Badawczych (pl.) -> [strona](#).
- Szkolenia oferowane przez EMBL-EBI (biologia) (ang.) -> [strona](#).
- Szkolenie „Zarządzanie danymi badawczymi” – Marta Hoffman-Sommer (pl.) -> [materiały](#).
- Szkolenie „Zarządzanie danymi badawczymi” – Bożena Bednarek-Michalska (pl.) -> [prezentacja](#).
- Warsztaty z zarządzania danymi badawczymi – Natalia Gruenpeter (pl.) -> [prezentacja](#).
- Szkolenie „Otwarte dane badawcze w humanistyce” – Natalia Gruenpeter, Michał Starczewski (pl.) -> [prezentacja](#).
- Szkolenie „Udostępnianie danych badawczych – zagadnienia prawne” (pl.) – Nikodem Rycko -> [prezentacja](#).

W razie jakichkolwiek pytań skontaktuj się z nami!

openaccess@lib.uni.lodz.pl