

Wskazówki do [wytycznych NCN](#) dotyczących uzupełniania planów zarządzania danymi badawczymi

Zestaw wskazówek został przygotowany przez zespół Platformy Otwartej Nauki ICM UW z myślą o badaczach i badaczach planujących korzystanie z repozytorium [Dane Badawcze UW](#) – instytucjonalnego repozytorium danych badawczych Uniwersytetu Warszawskiego.

To roboczy dokument, który będzie uzupełniany o dodatkowe informacje.

Data ostatniej aktualizacji: maj 2026 r.

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

1.1 W jaki sposób będą pozyskiwane lub wytwarzane nowe dane lub ponownie wykorzystywane dane już istniejące?

W ramach jednego projektu badawczego można zarówno korzystać z już istniejących danych, jak i wytwarzać nowe dane.

Dane do ponownego wykorzystania można znaleźć w repozytoriach danych badawczych (np. [RepOD](#), [Zenodo](#), [RDS](#), [MX-RDR](#)), w bibliotekach i archiwach różnych instytucji (np. [Archiwum Akt Nowych](#)) czy w Internecie (np. [Europeana](#), witryny projektowe). Należy zapoznać się z warunkami korzystania z danych, np. z rodzajami licencji nadanymi poszczególnym plikom w repozytorium. Ponowne użycie danych może oznaczać powtórzenie przeprowadzonych na nich analiz lub wykonanie nowych, zestawienie danych z innymi materiałami, umieszczenie ich w nowym kontekście itd.

Nowe dane wytwarzane w ramach projektu mogą np. zostać zgromadzone w trakcie realizacji badań terenowych, wygenerowane w laboratorium przy pomocy konkretnego urządzenia i oprogramowania, spisane jako notatki, wypisy, tłumaczenia czy transkrypcje, nagrane jako materiał dźwiękowy lub wizualny, sfotografowane lub zeskanowane itd.

Jakie standardy, metody i oprogramowanie posłużą do pozyskiwania lub wytwarzania nowych danych?

W poszczególnych dziedzinach nauki istnieją mniej lub bardziej ustandaryzowane metody prowadzenia badań i pozyskiwania danych. Naukowcy mogą korzystać z obowiązujących, nazwanych standardów, metodologii i procedur, opracowywać własne metody badawcze, sięgać po istniejące oprogramowanie czy tworzyć własny kod.

Jeżeli stosowane są standardowe narzędzia i metody, należy je wymienić lub opisać, korzystając z obowiązującej w danej dziedzinie terminologii.

Jeżeli nie – należy stworzyć wyczerpujący opis elementów procesu wytwarzania danych.

Jakie dane już istniejące (własne lub będące w posiadaniu strony trzeciej) zostaną ponownie wykorzystane?

Nie w każdym projekcie będzie można wykorzystać istniejące dane. Jednak nie należy wytwarzać już istniejących danych ponownie (chyba że taki jest cel badania).

Zespół badawczy może posiadać własne dane i dokumentację wytworzone w przeszłości, a może sięgnąć po dane wytworzone lub zgromadzone przez inne osoby. Należy wskazać i nazwać te zasoby.

W jaki sposób zostanie udokumentowane pochodzenie danych?

Dokumentacja może towarzyszyć danym zdeponowanym w repozytorium lub zostać udostępniona przez twórców danych wraz z zezwoleniem na ich wykorzystanie. Badacz może też zawrzeć umowę z instytucją, archiwum, organizacją lub osobą prywatną, które są właścicielami danych. Zgromadzonym danym mogą towarzyszyć opisy bibliograficzne czy metodologiczne, raporty, elektroniczne dzienniki laboratoryjne, fotografie, współrzędne, treść podpisanych zgód itp.

Jak wyglądać będzie organizacja plików i zarządzanie ich różnymi wersjami?

W każdym projekcie należy opracować system organizacji plików i folderów, sposób segregowania i przechowywania materiałów, oznaczenia czytelne dla wszystkich osób zaangażowanych w projekt. Należy przemyśleć, jak będą oznaczane różne wersje plików, szczególnie jeśli pliki będą przesyłane między członkami zespołu. Zasady te powinny zostać spisane, by również osoby spoza projektu mogły się z nimi zapoznać.

Czy konieczna będzie digitalizacja danych analogowych lub wydanych w formie papierowej (takich jak mapy, fotografie, teksty)?

Dane najczęściej istnieją w formie elektronicznej, ale niekiedy konieczna jest digitalizacja materiałów takich jak mapy czy listy, fotografowanie obiektów, nagrywanie i filmowanie wydarzeń itp. Tego rodzaju opracowań należy dokonywać przy użyciu odpowiedniego (bezpiecznego i wysokiej jakości) sprzętu i oprogramowania, które zapewnią dostępność danych. Np. zapisanie skanów w formacie komercyjnym (zamkniętym) spowoduje, że nie będą one dostępne bez odpowiedniego oprogramowania, a jeśli ich konwersja do formatu otwartego okaże się stratna – jakość danych ucierpi.

1.2 Jakie dane (tj. rodzaje, formaty, objętości) będą pozyskiwane lub wytwarzane w projekcie?

Jaki rodzaj, format i objętość danych planują Państwo pozyskać, wytworzyć lub ponownie wykorzystać?

W planie zarządzania danymi prezentowane powinny być wstępne założenia dotyczące danych. Powinny one uwzględniać szacowaną objętość danych zgromadzonych i/lub przetworzonych w trakcie realizacji projektu (np. wytworzonych zostanie maksymalnie 5 GB danych surowych), rodzaje danych (np. dane tekstowe, wyniki pomiarów, skany) i formaty ich zapisu (np. format ODT, TIFF, CSV).

Np. nagrania 12 rozmów (pliki audio, format WAV, około 2 KB); transkrypcje 12 rozmów (pliki tekstowe, format ODT, około 100 MB); plan wywiadu (plik tekstowy, format PDF i ODT, około 100 MB); formularz zgody na udział w badaniu (plik tekstowy, format PDF, około 10 MB); notatki z wywiadów (plik tekstowy, format TXT, około 10 MB).

2. Dokumentacja i jakość danych

2.1 Jakie metadane i dokumentacja (np. metodologia oraz sposoby pozyskiwania i organizacji danych) będą towarzyszyć danym w projekcie?

Metadane to informacje o danych. Tworzenie opisu metadanych danych odbywa się najczęściej za pośrednictwem repozytorium i polega na wypełnianiu pól metadanych, np. pola *autor*, pola *afiliacja autora*, pola *opis danych* itd. W ten sposób można przekazać podstawowe informacje o danych.

Dokumentacja może zawierać dodatkowe informacje o danych i o projekcie lub badaniu. Może mieć postać plików dołączonych do danych w repozytorium. Mogą to być pliki tekstowe z opisem procesu powstawania danych, raporty z kontroli jakości danych, dzienniki laboratoryjne, słowniki, legendy czy notatki.

Główny plik z dokumentacją to najczęściej tzw. plik README. Plik README przygotowuje zwykle autor danych i umieszcza w nim informacje takie jak opis metodologii badania, opis procedur, zawartości plików, informacje o zastosowanym oprogramowaniu, objaśnienia użytych skrótów, symboli i terminologii, objaśnienia dotyczące brakujących elementów danych czy nieudanych pomiarów. Plik powinien być zapisany w ogólnodostępnym formacie, najlepiej TXT.

Szablon pliku README.txt jest dostępny na tej stronie:

<https://data.research.cornell.edu/data-management/sharing/readme/>. Można go pobrać i przystosować do swoich potrzeb.

Jakie informacje należy uzupełnić, aby potencjalni użytkownicy (komputery lub ludzie) byli w stanie w przyszłości odczytać i zinterpretować zebrane dane? Czy możliwe jest ich komputerowe odczytanie?

Potencjalny użytkownik może potrzebować objaśnień dotyczących informacji zawartych w plikach, opisu ustawień urządzeń gromadzących dane, dokładnego opisu źródeł danych, np. numerów katalogowych czy identyfikatorów, objaśnień zmiennych, zastosowanych skrótów i symboli, objaśnień dotyczących metodologii badania.

Komputery, za pomocą których dane i metadane będą odczytywane, mogą wymagać podobnych informacji, jednak podanych w inny sposób (np. afiliacja autora danych powinna być jednoznacznie określona identyfikatorem ROR instytucji, a metadane zmiennych dostępne w postaci pliku XML). Nieprzydatne będą natomiast oznaczenia kolorystyczne zawarte np. w plikach XLSX. Oprogramowanie, którego użyto do zapisania danych, powinno być w stanie przetrwać próbę czasu, to znaczy jego dokumentacja powinna być dostępna (komercyjne oprogramowanie może nie być dostępne w dłuższej perspektywie czasowej, więc dane nie będą mogły zostać odczytane).

W jaki sposób powstanie odpowiednia dokumentacja?

Dokumentacja powinna być tworzona w trakcie trwania projektu, dlatego należy ustalić, kto będzie odpowiadał za realizację tego zadania lub jakiego rodzaju materiały – np. generowane automatycznie przez urządzenia – będą stanowić dokumentację badania.

Jakie standardy środowiskowe (jeżeli w ogóle) posłużą do objaśnienia (meta)danych? Jakie międzynarodowe standardy lub schematy (tj. Dublin Core, DDI) posłużą do organizacji metadanych?

Metadane umieszczane w repozytorium będą zgodne ze schematem przyjętym w tym repozytorium. Repozytorium Dane Badawcze UW umożliwia wprowadzenie metadanych, to znaczy wypełnienie pól formularza metadanych, a następnie wyeksportowanie ich do jednego z kilku formatów, m.in. DublinCore, DDI, OpenAIRE, Datacite.

2.2 Jakie planują Państwo zastosować środki kontroli jakości?

Użytkownik, który korzysta z danych, powinien mieć pewność, że w procesie badawczym dbano o kontrolę jakości danych. Środkami kontroli jakości mogą być np. powtarzanie pomiarów, stosowanie dwóch urządzeń nagrywających rozmowę, wykonywanie dodatkowych testów, sprawdzanie poprawności wprowadzanych informacji przez innych członków zespołu itd.

W jaki sposób metody pozyskiwania, analizy i przetwarzania danych mogą wpływać na ich jakość?

Niekiedy metody pracy z danymi czy np. konwersja danych powodują trwałe zmiany w tych danych – należy to opisać. Istotne jest przechowywanie i udostępnianie danych surowych wraz z opisem przeprowadzonych działań badawczych, a także wersji plików reprezentujących poszczególne etapy procesu.

W jaki sposób można wyeliminować błędy pomiarowe i problem stroniczości? W jaki sposób zminimalizować ryzyko dotyczące poprawności danych?

Weryfikacja pomiarów wykonywanych przez innych członków grupy badawczej, konsultacje z ekspertami, odpowiedni dobór grupy badanych (zgodnie ze standardami obowiązującymi w danej dziedzinie), tworzenie dokumentacji fotograficznej – to przykłady działań pozwalających zminimalizować ryzyko błędów.

Stosowanie kontroli wersji plików, powtarzanie kolejnych etapów badania, rejestrowanie wykonywanych działań, stosowanie kontroli dostępu do plików, praca w trybie śledzenia zmian – również tego rodzaju działania minimalizują ryzyko pomyłek.

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

3.1 W jaki sposób w trakcie projektu będą przechowywane dane i metadane? W jaki sposób będą tworzone ich kopie zapasowe?

Jaką dysponują Państwo pojemnością pamięci i gdzie będą przechowywane dane? Jak wyglądają procedury tworzenia kopii zapasowych?

Dane i metadane należy przechowywać zgodnie z zasadą 3-2-1: 3 kopie danych zapisane na 2 nośnikach, w tym 1 znajdującym się w innej lokalizacji.

Kopie zapasowe mogą być tworzone w różny sposób na poszczególnych nośnikach, dlatego procedury ich tworzenia również będą się różnić. Pojemność pamięci i miejsca przechowywania danych także zależą od stosowanych rozwiązań (zob. [opis miejsc przechowywania danych](#)).

Na Uniwersytecie Warszawskim dostępne są poniższe rozwiązania (szczegóły: <https://it.uw.edu.pl/pl/start/>).

W trakcie projektu dane i towarzyszące im metadane mogą być przechowywane:

- na dysku Google zarządzanym przez UW (Usługi Google UW). Limit objętości przechowywanych danych dla pracowników i doktorantów wynosi 200 GB. Pracownicy UW mogą wnioskować o utworzenie dysku współdzielonego z osobnym limitem dostępnej przestrzeni (200 GB). Kopie zapasowe tworzone są automatycznie. Więcej informacji: <https://it.uw.edu.pl/pl/uslugi/UslugiHelpdeskDysk/>
- na komputerze służbowym – odpowiednio zabezpieczonym i z możliwością kontroli dostępu do plików. Za tworzenie kopii zapasowych odpowiada pracownik.
- na urządzeniach przenośnych, jako dodatkowa kopia zapasowa przechowywana w innej lokalizacji niż komputer służbowy. Urządzenia tego rodzaju należy regularnie kontrolować i aktualizować, a także stosować bezwzględną ochronę hasłem. Za tworzenie kopii zapasowych odpowiada pracownik.
- z wykorzystaniem usługi UW Cloud – to rozwiązanie chmurowe, umożliwiające korzystanie z zasobów informatycznych centrum przetwarzania danych Uniwersytetu Warszawskiego. W ramach UW Cloud można uruchamiać: maszyny wirtualne (dwie), przestrzeń dyskową (200 GB + udział NFS 1 TB), obiektowy system plików, usługę backupu. Można zgłosić większe zapotrzebowanie w systemie DocSense: generator.ds.uw.edu.pl
- z wykorzystaniem zasobów chmurowych Amazon Web Services (AWS). Dostępne są m.in. usługi przetwarzania danych, przechowywania plików, baz danych i analizy danych. Zakres i opis dostępnych usług znajduje się na stronie: aws.amazon.com. Rozliczenie kosztów odbywa się zgodnie z realnym zużyciem. Zapotrzebowanie na dostęp do chmury AWS można zgłosić w systemie DocSense: generator.ds.uw.edu.pl.

Procedury tworzenia kopii zapasowych powinny zostać opracowane (i spisane) w zespole, a ich realizacja zaplanowana w regularnych odstępach czasowych. Należy wyznaczyć w zespole osobę lub osoby, które będą odpowiedzialne za tworzenie kopii zapasowych, ich kontrolę, tworzenie raportów z kontroli i zgłaszanie ewentualnych incydentów.

Czy potrzebne będą specjalne metody przenoszenia danych z urządzeń mobilnych, stanowisk terenowych lub sprzętu domowego na główny serwer w miejscu pracy?

Niektóre dane mogą wymagać specjalnych metod przenoszenia między miejscem ich pozyskania a miejscem przechowywania i archiwizacji. Jeżeli standardowe urządzenia przenośne okażą się niewystarczające, należy zaplanować zastosowanie innych rozwiązań,

które mogą się wiązać z kosztami (np. zakup dysku przenośnego, zakup dodatkowej przestrzeni dyskowej, zakup dostępu do odpowiedniego narzędzia).

3.2 W jaki sposób zostanie zapewnione bezpieczeństwo i ochrona danych wrażliwych w okresie trwania projektu?

W jaki sposób można będzie odzyskać dane utracone w wyniku incydentu? Kto uzyska dostęp do danych w czasie trwania projektu i jak wyglądać będzie kontrola dostępu do danych, zwłaszcza w przypadku współpracy między kilkoma partnerami?

RODO opisuje zasady przetwarzania szczególnych kategorii danych osobowych „ujawniających pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz przetwarzania danych genetycznych, danych biometrycznych”. Strategię przechowywania i przesyłania danych wrażliwych należy odpowiednio zaplanować, zaczynając od odpowiedzi na pytanie, czy przechowywanie i przesyłanie tego rodzaju danych jest absolutnie konieczne.

Następnie należy określić miejsca przechowywania danych (opis kilku rozwiązań jest [dostępny tutaj](#)) oraz ewentualne rozwiązania dotyczące kopiowania, modyfikowania i wersjonowania danych oraz ich trwałego usuwania. Można ustalić różne poziomy dostępu do danych i odpowiednie poziomy ich ochrony w zależności od możliwych ryzyk związanych z ujawnieniem, uszkodzeniem czy utratą danych.

4. Wymogi prawne, kodeksy postępowania

Zachęcamy do zapoznania się z [Poradnikiem](#) oraz [broszurą](#), a także skrótowym [FAQ prawnym](#).

4.1 Jeżeli będzie miało miejsce przetwarzanie danych osobowych, w jaki sposób zostanie zapewniona zgodność z przepisami dotyczącymi danych osobowych oraz ich ochrony?

Czy w procesie pozyskiwania danych niezbędna będzie ich anonimizacja?

Czy należy usunąć informacje umożliwiające identyfikację osób lub zataić tożsamość uczestników (np. za pomocą pseudonimizacji) przed udostępnieniem danych?

Zapewnienie zgodności z przepisami RODO jest konieczne, jeżeli zbiór danych zawiera dane osobowe. Najlepiej wraz ze zgodą na udział w badaniu uzyskać odpowiednie zgody osób, których dane dotyczą – zgody na zbieranie i przetwarzanie danych osobowych, a także na udostępnienie danych w repozytorium w sposób otwarty.

W wypadku niewyrażenia lub nieuzyskania takiej zgody, najprostszą drogą do udostępnienia danych jest anonimizacja, rozumiana jako usunięcie ze zbioru określonych informacji powodujących, że dane są danymi osobowymi.

Przetwarzanie danych osobowych szczególnych kategorii w celu jednoznacznego zidentyfikowania osoby fizycznej lub danych dotyczących zdrowia, seksualności lub orientacji seksualnej tej osoby jest generalnie zakazane. Możliwe jest tylko w szczególnych wypadkach, np. gdy dane te zostały w sposób oczywisty upublicznione przez osobę, której dotyczą. Prowadzenie badań naukowych samo w sobie nie zwalnia z konieczności stosowania RODO i dochowania zasad ochrony tych danych.

Anonimizacja to proces usuwania z danych informacji pozwalających na ustalenie tożsamości poszczególnych osób, a także odniesień do innych danych, które umożliwiają ustalenie ich tożsamości.

Pseudonimizacja oznacza przetwarzanie danych osobowych w taki sposób, by bez użycia dodatkowych informacji dane nie mogły zostać powiązane z osobami, których dotyczą. Dodatkowe informacje muszą być przechowywane osobno i podlegać technicznym oraz organizacyjnym środkom ochrony, tak by zapewnić, że identyfikacja osób pozostanie niemożliwa. W odróżnieniu od anonimizacji pseudonimizacja jest procesem odwracalnym, co znaczy, że osoby, których dane dotyczą, będzie można zidentyfikować, o ile umożliwiony zostanie dostęp do dodatkowych informacji.

4.2 W jaki sposób planują Państwo zapewnić zgodność z innymi przepisami, takimi jak prawa własności intelektualnej i prawa własności? Jakie przepisy znajdują w tym przypadku zastosowanie?

W zbiorach danych mamy najczęściej do czynienia z ochroną wynikającą z **praw autorskich** i/lub z **praw do baz danych**, a czasem także z **innymi prawami**, na przykład związanymi z osobami, których dobra lub interesy pozostają w określonej relacji do zbioru danych.

Jeżeli sądzimy, że jakieś elementy naszego zbioru danych badawczych są objęte prawami autorskimi, prawami do baz danych lub innymi prawami, to powinniśmy ustalić, komu te prawa przysługują. W tym celu musimy uwzględnić przepisy ustawowe (np. w prawie autorskim jest zasada, że autorskie prawa majątkowe przysługują twórcy, ale istnieją też wyjątki od tej zasady) oraz ustalenia poczynione w dokumentach poszczególnych jednostek naukowych, przede wszystkim w regulaminach zarządzania własnością intelektualną itp. (zob. [Regulamin zarządzania własnością intelektualną na Uniwersytecie Warszawskim](#)).

Kto będzie właścicielem danych?

Jakie będą ograniczenia ponownego wykorzystania danych pochodzących od osób trzecich? Czy przed udostępnieniem danych konieczne będzie uzyskanie odpowiedniego zezwolenia w zakresie praw autorskich?

Prawa autorskie dzielą się na osobiste i majątkowe. Prawa osobiste przysługują zawsze twórcy (lub współtwórcom), natomiast prawa majątkowe mogą być przenoszone na inne osoby, a w szczególności – w wypadku wyników badań naukowych – mogą czasami przysługiwać instytucji naukowej, w której badania są prowadzone.

Należy zapoznać się z [Regulaminem zarządzania własnością intelektualną na Uniwersytecie Warszawskim](#).

Zgodnie z Ustawą o prawie autorskim pracodawca ma prawo pierwszeństwa publikacji pracowniczych utworów naukowych, zaś autorskie prawa majątkowe pozostają w całości przy twórcy.

Natomiast na podstawie obowiązującego obyczaju oraz częstych postanowień w regulaminach jednostek naukowych można przyjąć, że w większości wypadków instytucje naukowe w Polsce nie korzystają z pierwszeństwa publikacji i zezwalają autorom na samodzielne decydowanie o publikowaniu wyników badań (odnosi się to zarówno do artykułów naukowych, jak i do tych elementów zbiorów danych, które uznajemy za utwory). W tej sytuacji pamiętać należy jednak o tym, by udostępnienie danych uzgodnić ze wszystkimi współautorami (wystarczy ustnie) oraz by respektować inne ograniczenia prawne i etyczne w udostępnianiu danych.

Programy komputerowe oraz twórcze bazy danych spełniające przesłanki utworu zbiorowego tworzone przez pracowników naukowych to przykłady wyników naukowych, do których majątkowe prawa autorskie najprawdopodobniej posiada pracodawca. Jeżeli więc chcemy udostępnić dane o takim właśnie charakterze, to powinniśmy skontaktować się z ośrodkiem transferu technologii lub dyрекcją jednostki by upewnić się, że wolno nam dany program lub bazę udostępnić. W praktyce, jeżeli baza danych czy program komputerowy nie posiadają żadnego potencjału komercyjnego (np. skrypt komputerowy służący jedynie do przetwarzania naszych danych), będzie to zapewne jedynie formalność.

Prawa do baz danych przysługują na mocy ustawy podmiotowi, który poczynił w tę bazę znaczące nakłady inwestycyjne, a zatem zazwyczaj będzie to pracodawca. Aby bazę udostępnić, musimy więc mieć jego zgodę – powinniśmy skontaktować się z ośrodkiem transferu technologii lub dyрекcją jednostki. Uzyskanie zgody w wypadku bazy danych, która nie ma żadnego potencjału komercyjnego, zwykle jest formalnością.

Prawa osób trzecich. Osobom, które w jakiś sposób zostały zaangażowane w zbieranie danych lub osobom, których dotyczą zebrane dane, mogą oczywiście przysługiwać różne prawa. Jeżeli tak jest, to i z nimi musimy uzgodnić, czy możemy udostępnić dane. Jeżeli badanie jest prowadzone z udziałem osób, istotna jest treść podpisanej przez uczestników zgody na udział w badaniu.

Jakie zostaną zastosowane licencje?

Licencje nadawane są całemu zbiorowi danych lub poszczególnym plikom z danymi w obrębie jednego zbioru poprzez wybór odpowiedniej opcji w repozytorium (zob. [instrukcja](#)). Wybór licencji należy do autora danych. Warto jednak zwrócić uwagę na kilka kwestii.

Po pierwsze, instytucje finansujące badania naukowe zwykle formułują wyraźne zalecenia dotyczące licencji. Np. NCN i Komisja Europejska rekomendują stosowanie licencji CC0 lub CC BY.

Po drugie, zalecane są licencje wolne, wśród których najpopularniejsze są licencje Creative Commons (CC). Licencje CC są tak skonstruowane, że osoba udzielająca licencji nie zrzeka się żadnych praw w imieniu osób trzecich, a jedynie tych praw, którymi sama dysponuje. Ponadto są to licencje międzynarodowe, a zatem ich konstrukcja zapewnia, że we wszystkich jurysdykcjach ich działanie jest tak zbliżone, jak to tylko możliwe.

Po trzecie, metadane są zwykle udostępniane na licencji CC0 na mocy regulaminów repozytoriów (zob. [regulamin repozytorium Dane Badawcze UW](#)). Przekazanie metadanych

do domeny publicznej ułatwia wyszukiwanie udostępnionych zbiorów, bo metadane mogą być swobodnie przekazywane do innych serwisów. Szczegółowe informacje na ten temat można znaleźć w [poradniku](#).

Warto zapoznać się z [poradnikiem prawnym w formie FAQ](#).

5. Udostępnianie i długotrwałe przechowywanie danych

5.1 Kiedy i w jaki sposób będą udostępniane dane z projektu?

Zwykle – m.in. w projektach finansowanych przez NCN – dane należy opublikować możliwe szybko, najpóźniej w momencie publikacji wyników badań (czyli zazwyczaj opublikowania pracy w czasopiśmie naukowym).

Właściwą metodą udostępnienia danych jest umieszczenie ich w odpowiednim repozytorium danych badawczych. Polityka UW zaleca deponowanie danych w repozytoriach dziedzinowych lub w repozytorium instytucjonalnym [Dane Badawcze UW](#).

Czy istnieją ewentualne ograniczenia i zakazy dotyczące ich udostępniania? Czy istnieją jakieś ograniczenia i przeszkody uniemożliwiające ich pełne lub częściowe udostępnienie?

Ograniczenia dotyczące części lub całości danych mogą wynikać m.in. z praw, którym podlegają dane wykorzystywane ponownie (np. materiały w posiadaniu instytucji, która zabrania ich dalszego udostępniania); z niemożności przeprowadzenia skutecznej anonimizacji danych; ze zgód, które wyrazili lub których nie wyrazili uczestnicy badania (np. wyrazili zgodę na udział w badaniu, ale nie na udostępnianie i ponowne wykorzystywanie swoich nieanonimizowanych danych).

Jednak w większości wypadków odpowiednio przygotowane dane czy zastosowanie rozwiązań takich jak odpowiednio przemyślane formularze zgód na udział w badaniu sprawiają, że nie wystąpią ograniczenia dotyczące udostępniania danych.

Inne ograniczenia mogą np. dotyczyć planowanej komercjalizacji wyników badań i związanego z tym opóźnienia pełnego udostępnienia danych.

W jaki sposób o Państwa danych dowiedzą się potencjalni użytkownicy?

Umieszczenie danych w repozytorium sprawia, że stają się one możliwe do wyszukania w sieci, w tym w systemach agregujących informacje czy specjalistycznych wyszukiwarkach. Prawdopodobieństwo odnalezienia danych zdeponowanych w repozytorium jest znacznie wyższe niż tych umieszczonych np. na stronach internetowych. Informacja o danych (wraz z identyfikatorem konkretnego zbioru) powinna również towarzyszyć publikacji naukowej, która powstała na podstawie tych danych.

Jak długo będą one przechowywane?

Informacji o okresie przechowywania danych należy szukać na stronach repozytorium danych, z którego planuje się skorzystać.

Repozytorium Dane Badawcze UW zakłada bezterminowe przechowywanie zdeponowanych danych. Informacja ta znajduje się w [regulaminie repozytorium](#). Jest to zgodne m.in. z zaleceniami NCN (wg. NCN okres przechowywania danych powinien wynosić minimum dziesięć lat).

Czy wydawcy czasopism będą wymagać składowania danych na poparcie ustaleń publikacji?

Wielu wydawców wymaga dostępu do danych badawczych na etapie recenzji pracy. Często wymagane jest też uzupełnienie sekcji publikacji dotyczącej dostępu do danych (sekcje te mogą mieć różne nazwy, np. *data availability statement*). Ta część planu zarządzania danymi służy zaplanowaniu odpowiednich działań na wczesnym etapie, tak by zapewnienie dostępu do wymaganych informacji nie było kłopotliwe dla badacza na późniejszym etapie projektu.

Warto zwrócić uwagę na narzędzia oferowane przez repozytorium: [embargo](#) i [prywatny link](#). Umożliwiają one udzielenie recenzentom publikacji wczesnego dostępu do niedostępnych publicznie danych.

Czy udostępnianie danych wymaga zgody uczestników badania?

Zgodnie z ogólnymi regulacjami Ustawy o ochronie danych osobowych, jeżeli dane są zanonimizowane (czyli istotnie nie można odtworzyć tożsamości osób), to nie są chronione jako dane osobowe i można je swobodnie udostępniać (także w repozytorium).

Warto (a niekiedy trzeba, np. w wypadku danych medycznych) przyjąć podejście bezpieczne – uprzedzić osoby badane, nawet jeśli planowana jest anonimizacja danych, o planowanym udostępnieniu danych w repozytorium. Należy zatem uzyskać zgodę badanych na udział w badaniu **oraz** uzyskać zgodę na udostępnienie (zanonimizowanych lub nie) danych z tego projektu, podając informację gdzie, kiedy, w jakim zakresie i celu, na jakiej licencji oraz komu dane zostaną udostępnione. Przykładowy formularz zgody dla badanych jest dostępny [tutaj](#).

5.2 Jak będzie wyglądać selekcja danych przeznaczonych do utrwalenia i gdzie będą one długoterminowo przechowywane (np. w repozytorium danych, archiwum)?

Które dane trzeba zachować, a które zniszczyć z przyczyn wynikających z umów, przepisów prawnych lub regulacji?

Należy zaplanować procedurę selekcji danych przeznaczonych do udostępnienia, ponieważ nie wszystkie dane można, warto i trzeba archiwizować i/lub upubliczniać.

Część danych może wymagać trwałego usunięcia, np. w celu przeprowadzenia skutecznej anonimizacji. Niekiedy umowy zawarte z firmami zewnętrznymi lub dodatkowe regulacje mogą zobowiązywać badacza do zachowania lub usunięcia części danych.

W jaki sposób zostanie podjęta decyzja o tym, które dane zachować? Jaka będzie procedura selekcji przeznaczonych do utrwalenia danych?

Decyzję podejmuje zwykle zespół badaczy zaangażowanych w projekt, samodzielnie kierownik projektu lub wyznaczony członek zespołu. Proces decyzyjny może obejmować konsultacje czy spotkania zespołu. Procedura selekcji danych do długoterminowego przechowywania i otwartego udostępniania powinna zostać opracowana odpowiednio wcześniej.

W proces decyzyjny niekiedy może być zaangażowany inspektor ochrony danych lub inna osoba wewnątrz lub na zewnątrz instytucji/jednostki.

Z jakiego repozytorium będą Państwo korzystać? Czy przestrzega ono zasad FAIR Data?

Dane powinny zostać umieszczone w repozytorium dziedzinowym lub tematycznym, jeśli jest taka możliwość. Serwisy tego rodzaju to np. Repozytorium Danych Społecznych (RDS), repozytorium danych krystalograficznych (MX-RDR), Protein Data Bank. Te i inne zaufane repozytoria z rozmaitych dziedzin można znaleźć w katalogu re3data.org.

Jeżeli nie istnieje repozytorium dziedzinowe ani specjalistyczne lub z jakiegoś powodu autor danych nie chce z takiego korzystać, pracownicy UW powinni deponować swoje dane w repozytorium [Dane Badawcze UW](#). Repozytorium spełnia wszystkie wymagania nakładane na tego typu serwisy, m.in. jest indeksowane przez systemy agregujące i wyszukiwarki oraz nadaje trwałe identyfikatory (DOI) wszystkim zbiorom danych. Jest również w trakcie procesu certyfikacji.

Czy Państwa instytucja tworzy regularnie kopie zapasowe?

Repozytorium Dane Badawcze UW jest prowadzone przez Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego UW, które dba o tworzenie kopii zapasowych przechowywanych danych badawczych.

Za zachowanie spójności i poprawności plików z danymi odpowiada oprogramowanie macierzy S3, na której są one składowane. Istnieje również możliwość przywrócenia utraconych zasobów z dodatkowej kopii zapasowej składowanej w osobnej lokalizacji. Repozytorium okresowo sprawdza czy zasoby, które zgodnie z informacjami znajdującymi się w bazie danych repozytorium powinny znajdować się na macierzy S3, rzeczywiście się na niej znajdują. W wypadku stwierdzenia rozbieżności odzyskiwana jest kopia zapasowa plików, znajdująca się na osobnej macierzy w innej lokalizacji. Kopie zapasowe danych plikowych oraz bazy danych metadanych tworzone są raz na dobę. W odrębnej lokalizacji przechowywane są kopie bazy danych z ostatnich 30 dni oraz kopie danych z ostatnich 2 dni.

W wypadku innych repozytoriów należy zapoznać się z regulaminem i dokumentacją i odszukać w nich informacje na temat zasad tworzenia kopii zapasowych.

Kwestie kopii zapasowych danych przechowywanych w czasie realizacji projektu są omówione w sekcji 3.

5.3 Jakie metody lub oprogramowanie umożliwiają dostęp do danych i korzystanie z danych?

Czy aby dane mogły być przechowywane przez dłuższy okres i zachowały długi okres ważności należy je przekształcić do formatu standardowego lub otwartego?

Sposób, w jaki zapisane są dane, może wpłynąć na ich żywotność. Jeżeli dane są zapisane w formatach zamkniętych, komercyjnych (np. takich, które wymagają uzyskania płatnej licencji), może dojść do sytuacji, w której firma zarządzająca tym oprogramowaniem przestanie wspierać stworzony przez siebie format danych. W takiej sytuacji dane staną się niedostępne.

Najlepszym sposobem zapisywania i archiwizowania danych jest stosowanie formatów otwartych, to znaczy korzystających z otwartego, ogólnodostępnego oprogramowania i funkcjonujących niezależnie od konkretnych firm. Tego typu formaty mają dużo większą szansę, by przetrwać próbę czasu i pozostać dostępne dla użytkowników.

Dane należy więc albo tworzyć i zapisywać od razu w formatach otwartych, albo tworzyć ich dodatkowe kopie w formatach otwartych.

Lista zalecanych formatów plików [jest dostępna tutaj](#).

Czy do skanowania lub konwersji niezbędny będzie dodatkowy sprzęt lub oprogramowanie?

Zapis w formatach otwartych lub konwersja do formatów otwartych mogą wymagać użycia dodatkowego oprogramowania. Należy sprawdzić, jakie możliwości daje narzędzie, którego używamy do pracy, np. czy pozwala wyeksportować dane w formacie oryginalnym (zwykle zamkniętym) oraz otwartym. Jeżeli konieczne będzie zastosowanie innego rozwiązania, by dane zapisać w odpowiedni sposób, należy je objaśnić.

Jaki mechanizm posłuży do udostępniania danych (np. odpowiedzi na żądanie, repozytorium)?

To pytanie dotyczy sposobu, w jaki użytkownik będzie mógł uzyskać dostęp do zdeponowanych danych.

Dane powinny być udostępniane za pośrednictwem repozytorium w sposób otwarty. W wyjątkowych wypadkach, na życzenie autora danych i o ile repozytorium to umożliwia, mogą zostać zastosowane ograniczenia dostępu, takie jak ograniczenie możliwości pobierania plików. Użytkownik musi w tej sytuacji wysłać prośbę o dostęp (najczęściej robi to bezpośrednio z poziomu repozytorium) lub w inny sposób zasygnalizować chęć pobrania danych. Należy tę procedurę opisać.

W wypadku repozytorium Dane Badawcze UW wszystkie zdeponowane dane są dostępne w sposób otwarty, a ewentualne ograniczenia mogą polegać na nałożeniu embarga (ograniczenia dostępu do plików na z góry wyznaczony czas, maksymalnie 36 miesięcy).

5.4 W jaki sposób zagwarantują Państwo stosowanie unikalnego i trwale przypisanego identyfikatora (np. takiego jak cyfrowy identyfikator dokumentu elektronicznego (DOI)) dla każdego zbioru danych?

Czy trwale przypisany identyfikator zostanie uzyskany?

Identyfikator jest najczęściej automatycznie nadawany przez repozytoria danych badawczych. Zwykle jest to identyfikator DOI (digital object identifier).

Repozytorium Dane Badawcze UW przypisuje identyfikator DOI automatycznie i bezpłatnie każdemu zbiorowi danych. Można go zobaczyć już na etapie utworzenia wersji roboczej zbioru w postaci linku, który jest elementem cytowania zbioru. Raz przypisany identyfikator nie ulega zmianie (ani po opublikowaniu zbioru, ani po wprowadzeniu modyfikacji do zbioru).

Jakie istniejące trwale identyfikatory pozostaną w użyciu (np. cyfrowe identyfikatory dokumentu elektronicznego, numery dostępne)?

Niektóre już istniejące dane czy materiały mogą mieć własne identyfikatory (np. pochodzące z innego repozytorium, archiwum itp.) Należy je uwzględnić w dokumentacji.

6. Zadania związane z zarządzaniem danymi oraz zasoby

6.1 Kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem)?

W projektach badawczych najczęściej to kierownik projektu odpowiada za zarządzanie danymi badawczymi.

W niektórych, zwłaszcza bardzo dużych projektach, практикуje się zatrudnianie osoby odpowiedzialnej za zarządzanie danymi badawczymi. Wówczas to tę osobę można wskazać jako (współ)opiekuna danych.

Czy Państwa instytucja posiada personel odpowiedzialny za zarządzanie danymi w trakcie trwania projektu, tj. od ich utworzenia, początkowego przechowywania, aż do archiwizacji i długoterminowego przechowywania danych?

Uniwersytet Warszawski nie zatrudnia personelu odpowiedzialnego za zarządzanie danymi w trakcie trwania projektu (to znaczy aktywnie dbającego o zarządzanie danymi w trakcie konkretnego badania). Zatrudnia natomiast osoby, które mogą udzielić badaczom [konsultacji w zakresie deponowania danych w repozytorium](#).

Jeśli NIE to kto jest odpowiedzialny za archiwizację i długoterminowe zarządzanie danymi w Państwa Jednostce?

Za archiwizację danych i zarządzanie danymi odpowiedzialny jest kierownik projektu lub osoba przez niego wyznaczona.

Jeśli TAK to jakie zajmuje stanowisko w obrębie Jednostki?

Jeżeli konkretne jednostki/instytuty/działy lub uczelnie współuczestniczące w projekcie zatrudniają taką osobę, można ją wskazać.

Jaką rolę pełni Kierownik projektu w zarządzaniu danymi?

W praktyce zwykle to kierownik projektu odpowiada za cały proces zarządzania danymi – od tworzenia planu zarządzania danymi, poprzez aktywne dbanie o poprawność działań wykonywanych na danych, po decyzje i czynności związane z długoterminową archiwizacją danych i ich otwartym udostępnianiem.

6.2 Jakie zasoby zostaną przeznaczone na cele zarządzania danymi i zagwarantowanie przestrzegania zasad FAIR?

Czy potrzebne są dodatkowe zasoby do zarządzania danymi, takie jak osoby, czas, sprzęt lub oprogramowanie?

Dodatkowe zasoby, np. finansowe, mogą zostać przeznaczone na zatrudnienie w projekcie osoby odpowiedzialnej za zarządzanie danymi badawczymi zgodnie z zasadami FAIR; na zakup sprzętu potrzebnego do przechowywania lub przesyłania danych, takiego jak dysk zewnętrzny lub dodatkowa przestrzeń chmurowa; na pokrycie opłat za korzystanie z repozytorium (jeśli takie są – repozytorium Dane Badawcze UW jest bezpłatne).

Koszty należy planować z pomocą Biura Obsługi Badań UW.

Jakie koszty związane będą z zapewnieniem standardów FAIR w projekcie? W jaki sposób zostaną opłacone?

Korzystanie z repozytorium Dane Badawcze UW jest bezpłatne. Nie ma potrzeby planowania żadnych kosztów w tym zakresie.

Jeśli konieczny jest zakup sprzętu albo zatrudnienie dodatkowej osoby w projekcie, koszty należy zaplanować wspólnie z Biurem Obsługi Badań UW.

Treść wskazówek objęta jest licencją Creative Commons Uznanie Autorstwa 4.0 (CC BY 4.0): creativecommons.org/licenses/by/4.0.