

Wskazówki dotyczące zarządzania danymi badawczymi

Opracowanie: Marta Błaszczynska, Mateusz Franczak, Maciej Maryl

Niniejszy dokument zawiera wytyczne dotyczące sposobów planowania, gromadzenia, organizowania, przetwarzania i publikowania danych badawczych w projekcie naukowym. Radzimy, co należy wziąć pod uwagę przy wypełnianiu planu zarządzania danymi w odniesieniu do danych badawczych w naukach humanistycznych. Należy pamiętać, że każdy plan zarządzania danymi będzie wyglądał inaczej w zależności od projektu. Plan zarządzania danymi może być również modyfikowany w trakcie trwania projektu według potrzeb badaczy. Wytyczne są zgodne z wymogami określonymi przez Narodowe Centrum Nauki i powinny być stosowane z uwzględnieniem obowiązujących w Instytucie przepisów kancelaryjno-archiwalnych.

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

1.1 Sposób pozyskiwania i opracowywania nowych danych i/lub ponownego wykorzystania dostępnych danych

Krótko opisz, w jaki sposób planujesz zbierać dane (dobór próby lub selekcja, organizacja pracy, procedury zapewniające jakość danych - zob. pkt 2.2).

Omów wszelkie specjalne programy lub narzędzia, które mogą być stosowane do gromadzenia, organizowania lub wykorzystywania danych. Przykładem w badaniach literackich może być wykorzystanie systemu baz danych lub menedżera bibliografii (np. Zotero) do organizacji bibliografii.

Jeśli planujesz wykorzystać istniejące już zbiory danych, podaj kilka podstawowych informacji o tym, kto je zebrał (Ty, współpracownik lub osoba trzecia) oraz gdzie zostały zdeponowane. Jeśli zgłosiłeś/aś się już z prośbą o pozwolenie na korzystanie z nich (i jeśli będzie to konieczne), należy o tym poinformować.

Jeśli gromadzisz dane samodzielnie, możesz rozważyć wymienienie możliwych źródeł danych, np. w przypadku przeglądu literatury: platformy, repozytoria, bazy danych i bibliografie, które chcesz wykorzystać.

Przykłady danych badawczych w humanistyce i sposoby zarządzania nimi w projekcie opisane są w raporcie ALLEA [Sustainable and FAIR Data Sharing in the Humanities](#).

1.2 Dane pozyskiwane lub opracowywane

Opisz dane, które zamierzasz zbierać: ich rodzaj, format, przewidywany rozmiar.

Dane mogą zawierać jeden z poniższych elementów:

- Teksty (zestawy tekstów, skany, notatki, adnotacje, raporty);
- Tabele (np. zestawienia liczbowe, listy osób, miejsc czy pojęć)
- Metadane;
- Mapy;
- Dane liczbowe;
- Obrazy (zdjęcia, rysunki, obrazy);
- Transkrypcje wywiadów, nagrania;
- ...i wiele innych.

Więcej szczegółów można znaleźć w artykule [Research Data Management Explained](#) przygotowanym przez Leeds University Library.

Poniższe kategorie mogą pomóc w dokładnym opisie danych:

- Cyfrowe lub niecyfrowe;
- Surowe lub przetworzone.

Należy wspomnieć, że dane będą sformatowane w sposób, który pozwoli na ich wymianę i wykorzystanie przez innych. Dlatego też format ten powinien być szeroko stosowany i standardowy.

Kilka przykładów formatów danych:

- Tekstowe (docx, txt itp.);
- Tabelaiczne (csv, xlsx itp.);
- Pliki wideo (avi, mp4 itp.);
- Pliki audio (mp3, ogg itp.).

Zaleca się wyjaśnienie, dlaczego wybrany został konkretny format danych i oszacowanie wielkości danych, które będziesz zbierać.

Jeśli planujesz udostępniać swoje dane w przyszłości, możesz zapoznać się z artykułem Komisji Europejskiej [Choosing the right format for open data](#).

2. Dokumentacja i jakość danych

2.1 Metadane i dokumentacja (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych)

Opisz sposób zarządzania metadanymi i wewnętrzną dokumentację dotyczącą danych, aby nowy członek zespołu lub osoba korzystająca z danych w przyszłości mogła je zrozumieć.

Dokumentacja projektu powinna być zorganizowana w taki sposób, aby główny badacz i członkowie zespołu mogli ją z łatwością odnaleźć. Dokumenty powinny być uporządkowane w sposób uzgodniony na początku projektu (na przykład według chronologii lub nazwy). Można określić konwencję nazewnictwa, która będzie używana.

Możesz skorzystać z artykułu [Best practices for file naming](#) przygotowanego przez Stanford Libraries.

2.2 Stosowane środki kontroli jakości danych

Na początku projektu można dostosować proces zapewniający wysoką jakość danych i standardy spójności w całym zespole. Może on obejmować podział obciążenia pracą, wzajemną kontrolę, uzgodnienie spójnych standardów gromadzenia, organizacji i prezentacji danych.

Jakość danych powinna być regularnie sprawdzana przez cały czas trwania projektu. Porady na temat możliwych procesów zapewniania jakości można znaleźć w [przewodniku](#) Open University.

Jeśli jest to przydatne, można rozważyć przeprowadzenie badań pilotażowych, podczas których można sprawdzić, czy wszyscy członkowie zespołu rozumieją gromadzenie i analizę danych w ten sam sposób.

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

3.1 Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

Opisz swoją strategię tworzenia kopii zapasowych danych.

Jako źródło procedur Instytutu możesz wskazać dokument *Polityka bezpieczeństwa przetwarzania danych osobowych w Instytucie Badań Literackich Polskiej Akademii Nauk* (patrz 3.2).

Instytut posiada dostęp do G Suite for Education, który umożliwia użytkownikom przechowywanie danych na Google Drive w formie indywidualnej i zespołowej.

Utracone dane będą mogły zostać odtworzone poprzez zapisanie (co najmniej) dwóch kopii w różnych miejscach, np. jednej na dysku Google i jednej na dysku twardym. Dotyczy to w szczególności dokumentacji projektu. Należy pamiętać, aby w każdej lokalizacji zapisać najbardziej aktualną wersję pliku. W Google Drive można aktywować funkcję, która automatycznie tworzy kopie zapasowe wybranych folderów na dysku twardym.

Jeśli planujesz wygenerować szczególnie duży zbiór danych, można rozważyć włączenie do budżetu projektu urządzenia do przechowywania danych, np. dysku twardego z dużą ilością pamięci i zakupić go na potrzeby projektu. Jeśli planujesz taki zakup, należy skontaktować się z działem odpowiedzialnym za nabycie odpowiedniego sprzętu.

Przydatny może być również przewodnik po dobrych praktykach ochrony danych osobowych Europejskiego Instytutu Uniwersyteckiego - [Guide on good data protection practice in research](#).

3.2 Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

Instytut posiada dokument wewnętrzny *Polityka bezpieczeństwa przetwarzania danych osobowych w Instytucie Badań Literackich Polskiej Akademii Nauk*, który jest głównym źródłem procedur bezpieczeństwa danych.

W przypadku przypadkowej utraty jakichkolwiek danych, projekt będzie realizowany zgodnie z protokołem opisanym w tym dokumencie.

Polityka bezpieczeństwa przetwarzania danych osobowych opisuje również procesy związane z kontrolą dostępu do danych w Instytucie Badań Literackich Polskiej Akademii Nauk.

4. Wymogi prawne, kodeks postępowania

4.1 Sposób zapewnienia zgodności z przepisami dotyczącymi danych osobowych i bezpieczeństwa danych w przypadku przetwarzania danych osobowych

Jeśli w projekcie są gromadzone i przetwarzane [dane osobowe](#), należy upewnić się, że cały proces jest zgodny z Rozporządzeniem o ochronie danych osobowych (RODO) konsultując się z Inspektorem Ochrony Danych Instytutu. Warto tu wspomnieć że, że pojęcie “dane osobowe” zgodnie z definicją RODO dotyczy wyłącznie osób żyjących.

Proces gromadzenia, wykorzystywania, przechowywania i zarządzania danymi osobowymi w ramach projektu będzie zgodny z RODO. Każdy nowy proces gromadzenia danych osobowych będzie konsultowany z Inspektorem Ochrony Danych Osobowych w IBL PAN, w celu zapewnienia zgodności z RODO oraz przestrzegania procedur przyjętych w Instytucie (tj. polityki przetwarzania danych osobowych IBL PAN).

Przetwarzanie danych musi być zgodne z prawem - zgodnie z art. 6, ust. 1 lit. a RODO. Osoby, których dane dotyczą (np. rozmówcy, uczestnicy badania itp.), muszą wyrazić świadomą zgodę na przetwarzanie swoich danych osobowych w jednym lub kilku konkretnych celach. Badacz musi przekazać odpowiedni formularz zgody RODO (wzór musi zostać zatwierdzony przez Inspektora Danych Osobowych IBL PAN) każdej osobie, której dane dotyczą. Podpisując formularz, osoba potwierdza, że została poinformowana m.in. o tym, jakie dane są od niej zbierane, jaka jest podstawa prawna ich przetwarzania oraz w jaki sposób jej dane będą zarządzane i przechowywane. Formularze są gromadzone i archiwizowane w ramach dokumentacji projektu.

Wszystkie cyfrowe pamięci masowe (dyski, dokumenty) zawierające dane osobowe lub w inny sposób poufne będą szyfrowane/chronione hasłem. Przechowywanie drukowanych kopii dokumentacji i wymiennych nośników zawierających dane osobowe będą chronione w zabezpieczonej/zamkniętej szafie, do której dostęp będzie miał wyłącznie upoważniony personel projektowy.

Zgodnie z zasadą minimalizacji danych, gromadzona będzie minimalna niezbędna ilość danych osobowych.

Jeśli dane osobowe zostaną [opatrzone pseudonimem i zanonimizowane](#), należy o tym poinformować, opisać proces i jego przyczyny. [Ten przewodnik](#) może być do tego przydatny.

W przypadku, gdy Twój projekt w dużym stopniu opiera się na danych osobowych, prosimy o zapoznanie się z opracowaniem Parlamentu Europejskiego [How the General Data Protection Regulation changes the rules for scientific research](#) oraz skontaktowanie się z Inspektorem Ochrony Danych Osobowych IBL PAN przed podjęciem decyzji o wykorzystywaniu, gromadzeniu, zarządzaniu i przechowywaniu danych.

4.2 Sposób zarządzania innymi kwestiami prawnymi, np. prawami własności intelektualnej lub własnością. Obowiązujące przepisy

Zachęcamy do korzystania z licencji Creative Commons (przeczytaj o różnych licencjach [tutaj](#)), ponieważ mogą one sprawić, że Twoja praca będzie widoczna i przydatna dla innych badaczy.

Należy wskazać, które licencje zostaną wykorzystane i jak inni będą mogli ponownie wykorzystać Twoje dane.

Można wspomnieć, że żadne z działań podjętych w czasie trwania projektu nie będzie sprzeczne z zaadaptowanym przez Instytut dokumentem wewnętrznym *Regulamin zarządzania prawami autorskimi i prawami pokrewnymi, zarządzania prawami własności przemysłowej oraz zasad komercjalizacji wyników badań naukowych i prac rozwojowych w Instytucie Badań Literackich Polskiej Akademii Nauk*.

5. Udostępnianie i długotrwałe przechowywanie danych

5.1 Sposób i termin udostępnienia danych.

Jeżeli zamierzasz upublicznić dane, opisz gdzie je umieścisz. Rekomenduje się korzystanie z Repozytorium RCIN, oraz innych certyfikowanych repozytorium danych jak Repozytorium Otwartych Danych CeON czy Zenodo.

Należy opisać ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga (późniejszego udostępnienia), zob. 5.2.

5.2 Sposób wyboru danych przeznaczonych do przechowania oraz miejsce długotrwałego przechowywania danych

W przypadku projektów generujących mniejsze zbiory danych, dane będą mogły być zapisywane zarówno na instytucjonalnym dysku Google, jak i na dysku twardym Instytutu. W przypadku większych projektów, dane będą mogły być zapisywane zarówno na dysku Google, jak i na dysku twardym Instytutu (patrz 3.1).

W razie potrzeby możesz zdecydować się na tworzenie kopii zapasowych danych w ramach projektu przez cały czas jego trwania z wykorzystaniem ww. usług.

Po zakończeniu projektu dane badawcze wraz z pozostałą dokumentacją projektu powinny zostać przekazane do Archiwum zgodnie z zasadami określonymi w Instrukcji kancelaryjnej IBL PAN .

Należy przestrzegać zasady, aby dane były tak otwarte jak to możliwe i na tyle zamknięte, na ile to konieczne. Oznacza to, że nie musisz otwierać swoich danych dla wszystkich we wszystkich przypadkach, ale powinieneś starać się utrzymywać je otwarte dla innych, jeśli jest to możliwe.

Jeżeli zdecydujesz się nie udostępniać swoich danych szerszemu gronu odbiorców, wyjaśnij dlaczego.

5.3 Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie z danych

Wyjaśnij, czy do uzyskania dostępu do danych potrzebne będą specjalistyczne lub nietypowe programy czy narzędzia. Jeśli tak, należy wyjaśnić przyczynę korzystania z tego konkretnego rozwiązania.

Możesz rozważyć dołączenie w metadanych informacji o najlepszych programach do obsługi danych.

5.4 Sposób zapewniający stosowanie unikalnego i trwałego identyfikatora (np. cyfrowego identyfikatora obiektu (DOI)) dla każdego zestawu danych

Jeżeli zbiór danych będzie przechowywany w certyfikowanym repozytorium (np. Zenodo), stały identyfikator zostanie dostarczony do zbioru danych bez żadnych opłat.

Dodatkowo, IBL PAN pokrywa coroczną subskrypcję [Crossref](#), która umożliwia pracownikom przydzielanie cyfrowych identyfikatorów dokumentów elektronicznych (DOI – digital object identifier) obiektom cyfrowym, takim jak artykuły, książki, zbiory danych itp., przechowywanym w repozytorium IBL PAN (RCIN).

6. Zadania związane z zarządzaniem danymi oraz zasoby

6.1 Osoba (np. funkcja, stanowisko i instytucja) odpowiedzialna za zarządzanie danymi

Lider(ka) projektu wybiera członka zespołu odpowiedzialnego za dane w projekcie. Osoba ta może mieć tytuł: opiekun(ka) danych (data steward) i może pełnić w projekcie równocześnie inne role. Lider(ka) projektu może również pełnić rolę opiekuna lub opiekunki danych.

6.2 Środki przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych

Zaleca się, by dane badawcze były zbierane, przechowywane i wykorzystywane zgodnie z międzynarodowym standardem [FAIR Data](#). FAIR jest akronimem czterech wytycznych wobec danych:

“Findable” - można je z łatwością wyszukać;

“Accessible” - są dostępne dla wszystkich;

“Interoperable” - można je połączyć z innymi danymi;

“Reusable” - istnieje możliwość ich ponownego wykorzystania.

Możesz skorzystać z narzędzia [FAIR self-assessment tool](#), które pomaga w weryfikacji czy dane są zgodne ze standardami FAIR.

Prosimy również zapoznać się z przykładem [procesu dostosowywania danych do standardu FAIR, przyjętym przez GO FAIR](#).

Opisz procesy, które pomogą Ci zapewnić, że Twoje dane będą zgodne ze standardami FAIR, np. czas, który zostanie poświęcony na zarządzanie danymi, tworzenie metadanych itp. Zadania te mogą zostać włączone do obowiązków członków zespołu i dodane do ich tygodniowego/miesięcznego harmonogramu.

Instytut posiada G Suite for Education, który pozwala pracownikom na przechowywanie dużych zbiorów danych bez dodatkowych kosztów (patrz 3.1) oraz Crossref do rejestracji DOI (patrz 5.4).

Dodatkowe źródła

[Dobre praktyki publikowania danych badawczych](#) (artykuł) – Anna Małgorzata Kamińska. Biuletyn EBIB, n. 177, p. 1-13, Dec. 2017.

[Otwarte dane badawcze w warsztacie pracy naukowca](#) (artykuł) – Małgorzata Rożniakowska-Kłosińska. Biuletyn EBIB, n. 183, p. 1-6, Dec. 2018.

[Otwarte udostępnianie danych badawczych](#) (prezentacja) – Marta Hoffman-Sommer, CC BY.

[PORADNIK – Plan zarządzania danymi badawczymi](#) (poradnik) – Repozytorium i Pozycjonowanie Wydawnictw KUL

[Warsztaty z zarządzania danymi badawczymi](#) (prezentacja) – Natalia Gruenpeter, CC BY.