

Procedura tworzenia **PLANU ZARZĄDZANIA DANYMI** dla dyscypliny Technologia żywności i żywienia

Plan zarządzania danymi uznawany jest za część planu badań. Określa, w jaki sposób dane badawcze mają być zarządzane zarówno w czasie trwania projektu badawczego, jak i po jego zakończeniu.

W planie zarządzania danymi wprowadza się następujące definicje:

Dane badawcze (research data) - zarejestrowane materiały o charakterze faktograficznym (w postaci liczbowej, tekstowej czy graficznej), powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

Dane surowe (raw data), dane, które uzyskano bezpośrednio w wyniku zastosowania narzędzia badawczego, w różnych przedsięwzięciach naukowych lub też zgromadzone na potrzeby konkretnych projektów, nieprzeanalizowane.

Wyniki badań – dane, które poddane zostały obróbce.

Otwarte dane badawcze (Open Research Data) – dane, do których każdy ma dostęp. Można je ponownie wykorzystywać, modyfikować, redystrybuować i udostępniać z poszanowaniem prawa.

Definicje te obejmują wszystkie możliwe formaty, zarówno cyfrowe jak i nie cyfrowe (np. próbki, wypełnione kwestionariusze itd.)

Dane badawcze powinny być łatwe do znalezienia. W planie zarządzania danymi powinny znaleźć się informacje na temat sposobów pozyskiwania danych, ich walidacji, pochodzenia, możliwości ponownego wykorzystania oraz ich rodzaj.

Plan zarządzania danymi powinien zawierać informacje nt.:

1. Identyfikacja projektu badań

Uzyskane wyniki badań, objęte polityką otwartości danych powinny być opatrzone informacją zapewniającą szybką identyfikowalność w tym:

- tytuł projektu i źródło finansowania
- kierownika projektu
- streszczenie projektu
- okres realizacji projektu oraz datę utworzenia/modyfikacji planu
- osobę zarządzającą danymi

2. Opis danych w projekcie badawczym

Opis danych powinien dotyczyć zarówno danych surowych oraz przetworzonych wyników badań i powinien dostarczać niezbędnych informacji dotyczących:

- sposobu pozyskiwania i wytwarzania danych (sprzęt, przyjęte metodyki)
- określenia czy będą tworzone nowe dane czy wykorzystywane dane już istniejące
- rodzaj wytworzonych lub zebranych danych (np. format, tekst, zdjęcia itp.)
- sposobu kontrolowania jakości danych (próbkiowanie, kalibracja, walidacja, recenzje, ekspertyzy)
- sposobu zapisywania i utrwalania danych oraz organizacji plików z danymi

3. Dokumentacja dotycząca danych

Wymagane jest określenie rodzaju metadanych mających ułatwić użytkownikom odnalezienie i identyfikację poszukiwanych informacji. Należy określić następujące informacje:

- źródła pochodzenia danych w tym numeru identyfikacyjnego autora (np. ORCID)
- formatu metadanych, dzięki któremu będzie możliwa identyfikacja i odczytanie poszukiwanych informacji
- określenie definicji zmiennych oraz jednostek pomiarowych
- określenie struktury folderów
- należy wskazać przyjęte mechanizmy mające na celu zabezpieczenie danych przed nieuprawnioną modyfikacją
- jakie międzynarodowe standardy lub schematy posłużą do organizacji metadanych ¹

4. Przechowywanie danych

Wymagane jest określenie sposobu przechowywania danych i metadanych oraz sposobu tworzenia kopii zapasowych w szczególności:

- osoby zarządzającej danymi
- miejsca przechowywania i składowania danych²
- wymaganej pojemności pamięci
- procedury tworzenia kopii zapasowych
- częstotliwości aktualizacji danych
- określenie możliwości digitalizacji danych analogowych w tym czy występuje konieczność wytworzenia odpowiednich stanowisk (skaner)
- dokumentacji danych wytworzonych w terenie lub na sprzęcie domowym oraz metod ich przenoszenia z urządzeń mobilnych
- określenie polityki ochrony danych zgodnie z wdrożoną polityką instytucji

5. Wymogi prawne

Należy zapewnić ochronę danych osobowych zgodnie z przepisami dotyczącymi danych osobowych oraz i ich ochrony. W szczególności należy określić:

- kto jest właścicielem danych
- jak chronione są dane osobowe (RODO)
- czy konieczne jest uzyskanie zgody na przetwarzanie i udostępnianie danych osobowych
- czy konieczna jest anonimizacja danych
- jakie są zastosowane licencje
- jak będą zabezpieczone prawa autorskie i prawa własności intelektualnej do pozyskiwanych i wytworzonych danych

-czy będą istniały ograniczenia prawne dotyczące ponownego wykorzystania danych przez osoby trzecie

6. Udostępnianie i długotrwałe przechowywanie danych

Dane należy udostępnić możliwie szybko – najpóźniej w momencie publikacji wyników badań. Należy określić kiedy i w jaki sposób będą udostępniane dane z projektu i czy istnieją ewentualne ograniczenia i zakazy dotyczące udostępniania. W szczególności należy określić:

- jakie dane chcemy przechowywać po zakończeniu projektu
- jak długo będziemy je przechowywać³
- gdzie zostaną zdeponowane dane do przechowywania długoterminowego
- w jaki sposób będzie można uzyskać informację o danych
- czy istnieją ograniczenia w pełnym udostępnianiu danych, jeżeli tak to jakie?
- w jakim repozytorium dane zostaną zarchiwizowane
- czy udostępnianie danych wymaga zgody uczestników badań

7. Zadania związane z zarządzaniem danymi oraz zasoby

Wymagane jest określenie właściciela danych oraz ich opiekuna. Niezbędne jest opisanie ról i obowiązków w zakresie zarządzania danymi i wskazanie kto odpowiada za wdrożenie planu zarządzania, a także jego weryfikację i korektę. W przypadku projektu opartego na współpracy, należy wyjaśnić sposób koordynacji w zakresie zarządzania danymi pomiędzy poszczególnymi partnerami. Należy wskazać:

- osobę odpowiedzialną za zarządzanie danymi
- koszty związane z archiwizacją danych

¹ Digital Curation Centre (DCC) prowadzi listę szeroko stosowanych standardów metadanych. Można ją znaleźć pod adresem <http://www.dcc.ac.uk/resources/metadata-standards>.

² Cyfrowe repozytoria danych: specjalistyczne, instytucjonalne, szeroko zakrojone tematycznie, ogólne.

Wyszukiwarka re3data.org (Registry of Research Data Repositories) to rejestr repozytoriów danych badawczych ze wszystkich dyscyplin akademickich. Jest to darmowe narzędzie, które oferuje naukowcom, organizacjom finansującym badania, bibliotekom i wydawcom przegląd repozytoriów danych badawczych. Umożliwia wyszukiwanie repozytoriów według dziedziny wiedzy, kraju oraz typu danych badawczych.

Istnieje wiele międzynarodowych systemów certyfikacji określających poziom wiarygodności poszczególnych repozytoriów danych. Podstawowy zestaw kryteriów to Data Seal of Approval. Zaleca się korzystanie w pierwszej kolejności z repozytoriów certyfikowanych lub szeroko uznanych w danej dyscyplinie, a jeżeli takie repozytoria nie istnieją, kierować się w swoim wyborze kryteriami zawartymi w dokumencie Practical Guide to the international alignment of research data management (https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf).

Polecane repozytoria

Zenodo

Repozytorium zostało opracowane dzięki inicjatywie OpenAIRE i CERN. Przeznaczone dla tzw. małych danych. Umożliwia proste archiwizowanie i dzielenie się opracowanymi przez siebie danymi przez naukowców prowadzących badania we wszystkich dziedzinach wiedzy.

RepOD

Repozytorium otwartych danych badawczych opracowane przez ICM UW. Przeznaczone dla tzw. małych danych. Umożliwia archiwizowanie i udostępnianie różnych rodzajów danych wytworzonych przez naukowców z różnych dziedzin wiedzy.

³ Dane surowe i przetworzone powinny być przechowywane przez okres odpowiedni dla danej dyscypliny i zastosowanej metodologii. W rozumieniu NCN uzasadniony okres przechowywania danych to minimum 10 lat.

Narzędzia pomocne podczas tworzenia Planu zarządzania danymi – data management plan DMP

DMPtool (US) – narzędzie przygotowujące szablony DMP dostosowane do wymagań amerykańskich grantodawców

DMPonline (UK) – narzędzie bardzo podobne do DMPtool zawierające jednak bazę instytucji finansującej naukę z Wielkiej Brytanii

DCC Data Management Plan Content Checklist – lista kontrolna zawartości planu zarządzania danymi, szablon projektu

http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf – pozwala szybko określić, jakich informacji może brakować w przygotowywanym DMP

Serwis brytyjskiej instytucji specjalizującej się w zarządzaniu danymi badawczymi – The Data Curation Center – znajdują się tam gotowe plany zarządzania danymi, szablony, przewodniki, wytyczne, informacje na temat formatów zamieszczanych danych badawczych, metadanych w zależności od dyscypliny oraz lista kontrolna (Checklist for a Data Management Plan) dotycząca zawartości DMP, która pozwala szybko określić, jakich informacji może brakować w przygotowywanym DMP.

Przykładowy plan zarządzania danymi

Pole	Opis	Elementy
Informacje ogólne	Zawiera informacje ogólne dotyczące projektu i gromadzenia danych	1. Wnioskodawca 2. Afiliacja 3. ORCID 4. Projekt i opis 5. Finansowanie
Opis zbioru danych lub wykorzystanie istniejących danych	A/ Zawiera opis jak będą wyglądać nowe dane zebrane lub wyprodukowane i czy/lub zostaną ponownie wykorzystane	1. Jakie metody lub oprogramowanie zostanie użyte do zebrania lub wytworzenia danych? 2. Czy przewidywane są ograniczenia dotyczące ponownego wykorzystania danych. 3. Dokumentacja dotycząca pochodzenia danych.
	B/ Zawiera opis, jakie dane (np. rodzaj, format) zostaną zebrane lub wyprodukowane	1. Szczegóły dotyczące rodzaju danych. 2. Szczegóły dotyczące formatu danych. 3. Uzasadnienie dla użycia konkretnych formatów. 4. Szczegóły dotyczące użycia otwartych formatów. 5. Szczegóły dotyczące wielkości woluminów.
Dokumentacja i jakość danych	A/ Zawiera opis, jakie metadane i dokumentacja potrzebne są przy wybranych danych	1. Jakie metadane zostaną udostępnione? 2. Określić standardy metadanych (najlepiej użyć tych najbardziej powszechnych). 3. Określić sposób, w jaki dane zostaną zorganizowane i uporządkowane podczas projektu. 4. Wskazać opcjonalnie inną dokumentację potrzebną do

		ponownego użycia.
	B\ Zawiera opis, jakie środki kontroli jakości danych zostaną zastosowane	1. Wyjaśnienie, jak spójność i jakość zebranych danych będzie kontrolowana i udokumentowana.
	C\ Zawiera opis, w jaki sposób, gdzie dane będą udostępniane i w jakim stopniu otwarte	1. Restrykcje dotyczące ponownego wykorzystania danych, ich udostępniania i przekształcania.
Przechowywanie i kopia zapasowa podczas badań	A\ Zawiera opis, gdzie dane będą archiwizowane i na jak długo	1. Metody zastosowane do archiwizowania i zapewnienia długotrwałego dostępu. 2. Ochrona danych i ich trwałość.
	B\ Zawiera opis, jakie działania należy wykonać, żeby zapewnić bezpieczeństwo danych	1. Procedury odzyskiwania danych. 2. Dostęp do danych, zwłaszcza kontrolowany danych wrażliwych. 3. Ochrona danych instytucjonalnych.
Wymogi prawne i etyczne, kodeks postępowania	A\ Zawiera opis, jakie dane osobowe są przetwarzane (zgodność działań z ustawą o ochronie danych osobowych)	1. Dane osobowe, zgody na ich przetwarzanie, zachowanie i udostępnianie. 2. Zastosowanie anonimizacji, pseudoanonimizacji.
	B\ Zawiera opis, jak jest rozwiązany problem z własnością intelektualną	