



Zarządzanie danymi badawczymi

Tworzenie planu zarządzania danymi (DMP)

na potrzeby Narodowego Centrum Nauki

Agnieszka Milewska

Biblioteka Główna GUMed

dane badawcze

to dane zaobserwowane, zebrane lub wytworzone w ramach prowadzenia badań:

- dokumenty tekstowe, notatki
- dane liczbowe (w formie nieprzetworzonej lub możliwej do odczytania przez komputer)
- kwestionariusze, ankiety, wyniki badań naukowych
- nagrania audio i video, zdjęcia
- modele matematyczne, algorytmy
- oprogramowanie (skrypty, pliki wejściowe)
- wyniki symulacji komputerowych
- protokoły laboratoryjne, opisy metodologiczne
- próbki, artefakty, obiekty
- metadane (opisy)

Dlaczego warto/należy udostępniać dane badawcze

- Możliwość weryfikacji wyników
 - Ułatwienie prowadzenia badań
 - Możliwość ponownego wykorzystania
 - Argumenty etyczne
 - Bezpieczna i trwała archiwizacja
- Dobre dane świadczą o jakości badań
 - Dane mogą zostać zacytowane
 - Dostępność danych zwiększa szanse na zacytowanie publikacji powstałej na ich bazie

Kiedy udostępnienie danych badawczych bywa konieczne?

1. Wymagania nakładane przez wydawców naukowych: konieczność udostępniania danych
2. Wymagania nakładane w umowach grantowych: konieczność tworzenia Planów Zarządzania Danymi i/lub udostępniania danych

Otwarte dane badawcze:

- dostępne w postaci cyfrowej w domenie publicznej,
- upowszechniane bez żadnych ograniczeń, czyli mogą być dowolnie **używane, rozpowszechniane i przetwarzane** przez kogokolwiek, gdziekolwiek w dowolnym celu
- otwartość danych w odróżnieniu od otwartości publikacji związana jest nie tylko z ich dostępem, ale głównie **z możliwością ich ponownego wykorzystania**

As open as possible, as closed as necessary

Licencje Creative Commons

otwarty dostęp **gratis** i otwarty dostęp **libre**

Cztery podstawowe warunki licencji CC to:



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.



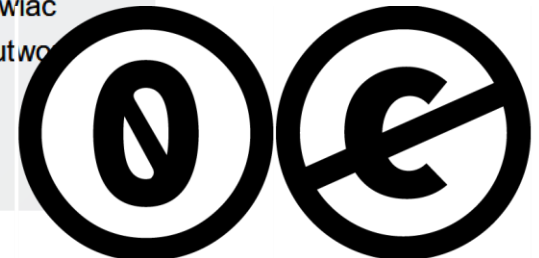
Użycie niekomercyjne. Wolno kopiować, rozprowadzać, przedstawiać i wykonywać objęty prawem autorskim utwór oraz opracowane na jego podstawie utwory zależne jedynie do celów niekomercyjnych.



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.



FAIR data

- **Findable** - żeby było łatwo je znaleźć
- **Accessible** - żeby były dostępne dla wszystkich
- **Interoperable** - żeby można było je połączyć z innymi danymi
- **Reusable** - żeby dało się je ponownie wykorzystać

Najważniejsze kwestie związane z udostępnianiem danych badawczych:

- Kwestie prawne
- Przygotowanie danych do udostępnienia – dobre praktyki
 - Formaty plików
 - Nazywanie plików i katalogów
 - Selekcja danych
 - Metadane i dokumentacja
- Miejsce i sposób archiwizacji oraz udostępnienia

Kwestie prawne

- Komu przysługują prawa do konkretnego zbioru danych? Kto ma prawo podjąć decyzję o jego udostępnieniu?
- Co wolno użytkownikowi danych? Kto i w jaki sposób o tym decyduje?
- Prawa autorskie – jeżeli zbiór danych ma cechy utworu
- Prawa ochrony baz danych – jeżeli zbiór danych ma cechy bazy danych
- Prawa osób trzecich – np. jeżeli zbiór zawiera dane osobowe pacjentów, mają oni prawo do ochrony tych danych.

Przygotowanie danych do archiwizacji

- Formaty plików
- Nazywanie plików i katalogów
- Selekcja danych
- Metadane i dokumentacja (danych i projektu)

Formaty plików

5 star open data



dobre praktyki

- Nazywanie plików i katalogów
- Selekcja danych
- Metadane
- Dokumentacja (projektu i danych)

Readme.txt

dobre praktyki

- Nazywanie plików i katalogów
- Selekcja danych
 - Do czego jesteśmy zobowiązani? (przez grantodawcę, pracodawcę, wydawcę)?
 - Jaka jest ich wartość naukowa?
 - Na ile dane są unikalne?
 - Czy łatwo można je pozyskać ponownie?
 - Czy dane są dobrej jakości i czy łatwo będzie innym z nich skorzystać?
 - Jakie koszty wiążą się z zarządzaniem i ich przechowywaniem?
 - Czy dysponujemy kompletną i poprawną dokumentacją danych?

Readme.txt

dobre praktyki

- Nazywanie plików i katalogów
- Selekcja danych
- Metadane
 - międzynarodowe standardy lub schematy (tj. Dublin Core, DDI, Datacite)
 - metadane powinny być takie, którymi posługują się największe huby wyszukiwawcze
 - tytuł powinien w sposób adekwatny opisywać zbiór
 - marketing

Readme.txt

dobre praktyki

- Nazywanie plików i katalogów
- Selekcja danych
- Metadane
- Dokumentacja (projektu i danych)
 - Opis celu i kontekstu badań
 - Opis metodologiczny sposobu pozyskania danych
 - Struktura plików z danymi, relacje pomiędzy plikami
 - Linki do publikacji powiązanych z danymi
 - Formularze zgody na udział w badaniu
 - Zastosowane środki kontroli danych
 - Informacja o wersjach zbiorów danych
 - Informacja o zbiorach danych, których nie można udostępnić
 - Nazwy i opisy zmiennych/rekordów
 - Opisy schematów klasyfikacyjnych
 - Informacje o urządzeniach pomiarowych, ich ustawieniach i metodach kalibracji
 - Oznaczenia brakujących wartości oraz wyjaśnienie, dlaczego ich brakuje
 - W przypadku badań jakościowych – lista przypadków lub obiektów objętych badaniem
 - Wyniki testów jakościowych

Readme.txt

Gdzie udostępniać zbiory danych?

- **Repozytoria danych – specjalistyczne i ogólne**
- Polska Platforma Medyczna
- Most Danych
- Zenodo
- Re3data.org

Data management plan

- Pozyskiwanie danych
- Dobór formatów i ich nazewnictwo
- Krótko i długoterminowe przechowywanie
- Zasady dostępu do danych, możliwość ich ponownego wykorzystania
- Prawne i etyczne aspekty zarządzania zbiorem danych
- Zasoby potrzebne do zarządzania danymi (infrastruktura, kompetencje, środki finansowe)

Data management plan - NCN

- Informacje podstawowe
- Wnioskodawca
- Podmioty realizujące
- Pomoc publiczna
- Współpraca międzynarodowa
- Streszczenie
- Streszczenie popularnonaukowe
- Harmonogram
- Zespół badawczy
- Kwestie etyczne
- Plan zarządzania danymi**
- Zbliżone zadania badawcze
- Opis
- Wynagrodzenia i stypendia
- Aparatura
- Inne koszty
- Koszty - zestawienia
- Oświadczenia administracyjne
- Ochrona danych osobowych
- Redaktor wniosku
- Redaktorzy pomocniczy
- Czytelnicy
- Helpdesk

Plan zarządzania danymi

Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych:	Brak odpowiedzi na wszystkie pytania.
Dokumentacja i jakość danych:	Brak odpowiedzi na wszystkie pytania.
Przechowywanie i tworzenie kopii zapasowych podczas badań:	Brak odpowiedzi na wszystkie pytania.
Wymogi prawne, kodeks postępowania:	Brak odpowiedzi na wszystkie pytania.
Udostępnianie i długotrwałe przechowywanie danych:	Brak odpowiedzi na wszystkie pytania.
Zadania związane z zarządzaniem danymi oraz zasoby:	Brak odpowiedzi na wszystkie pytania.

Wymogi formalne

Sekcja wypełniana w języku angielskim.

NIE DOTYCZY na wszystkie

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

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

[Edytuj](#) [Nie dotyczy](#)

Słów: 0, Znaków: 0

Pozyskiwane lub opracowywane dane (np. rodzaj, format, ilość)

[Edytuj](#) [Nie dotyczy](#)

Data management plan - NCN

1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

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

- Jakie standardy, metody i oprogramowanie posłużą do pozyskiwania lub wytwarzania nowych danych?
- Jakie zostaną wdrożone procesy kontroli jakości?
- Jakie dane już istniejące (własne lub będące w posiadaniu strony trzeciej) zostaną ponownie wykorzystane?
- W jaki sposób zostanie udokumentowane pochodzenie danych?

Pozyskiwane lub opracowywane dane (np. rodzaj, format, ilość)

- Jakie dane zostaną wygenerowane (teksty, obrazy, zdjęcia, próbki fizyczne, kody??)
- W jakich formatach będą pliki (np. .xls)?
- Jakiej objętości będą te pliki?

Data management plan - NCN

2. Dokumentacja i jakość danych

Metadane i dokumenty (np. metodologia lub pozyskiwanie danych oraz sposób porządkowania danych) towarzyszące danym

- Jakie informacje należy uzupełnić, aby potencjalni użytkownicy (komputery lub ludzie) mogli odczytać i zinterpretować zebrane dane?
- W jaki sposób powstanie odpowiednia dokumentacja?
- Jakie standardy środowiskowe posłużą do objaśnienia (meta)danych?
- Jakie międzynarodowe standardy lub schematy (tj. Dublin Core, DDI) posłużą do organizacji metadanych?

Stosowane środki kontroli jakości danych

- W jaki sposób metody pozyskiwania, analizy i przetwarzania danych mogą wpływać na ich jakość?
- W jaki sposób można wyeliminować błędy pomiarowe i problem stroniczości?
- W jaki sposób zminimalizować ryzyko dotyczące poprawności danych

Data management plan - NCN

3. Przechowywanie i tworzenie kopii zapasowych podczas badań

Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

- Jaką dysponują Państwo pojemnością pamięci i gdzie będą przechowywane dane?
- Jak wyglądają procedury tworzenia kopii zapasowych?
- Czy potrzebne będą specjalne metody przenoszenia danych z urządzeń mobilnych, stanowisk terenowych lub sprzętu domowego na główny serwer w miejscu pracy?
- Jakie wymogi związane z przechowywaniem danych muszą być wypełnione?

Sposób zapewnienia bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

- 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?

Data management plan - NCN

4. Wymogi prawne, kodeks postępowania

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

- Czy 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?

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

- Kto będzie właścicielem danych?
- Jakie zostaną zastosowane licencje?
- Jakie będą ograniczenia ponownego wykorzystania danych pochodzących od osób trzecich?
- Czy przed udostępnieniem danych konieczne będzie uzyskanie odpowiedniego zezwolenia w zakresie praw autorskich?

Data management plan - NCN

5. Udostępnianie i długotrwałe przechowywanie danych

Sposób i termin udostępnienia danych. Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga

- Jak długo będą one przechowywane?
- Czy istnieją jakieś ograniczenia i przeszkody uniemożliwiające ich pełne lub częściowe udostępnienie?
- Czy wydawcy czasopism będą wymagać składowania danych na poparcie ustaleń publikacji?
- Czy udostępnianie danych wymaga zgody uczestników badania?

Sposób wyboru danych przeznaczonych do przechowania oraz miejsce długotrwałego przechowywania danych (np. repozytorium lub archiwum danych)

- Które dane trzeba zachować, a które zniszczyć z przyczyn wynikających z umów, przepisów prawnych lub regulacji?
- Jaka będzie procedura selekcji przeznaczonych do utrwalenia danych?
- z jakiego repozytorium będą Państwo korzystać? Czy przestrzega ono zasad FAIR Data3? Czy Państwa instytucja tworzy regularnie kopie zapasowe?

Metody lub narzędzia programowe umożliwiające 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?
- Czy do skanowania lub konwersji niezbędny będzie dodatkowy sprzęt lub oprogramowanie?
- Jaki mechanizm posłuży do udostępniania danych (np. odpowiedzi na żądanie, repozytorium)?

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

- Czy trwale przypisany identyfikator zostanie uzyskany?
- Jakie istniejące trwałe identyfikatory pozostaną w użyciu (np. cyfrowe identyfikatory dokumentu elektronicznego, numery dostępne)?

Data management plan - NCN

6. Zadania związane z zarządzaniem danymi oraz zasoby

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

- Jaką rolę pełni w Państwa instytucji opiekun danych? Data steward
- Kto będzie odpowiedzialny za wdrożenie planu zarządzania danymi?

Środki (np. finansowe i czasowe) przeznaczone do zarządzania danymi i zapewnienia możliwości odnalezienia, dostępu, interoperacyjności i ponownego wykorzystania danych

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

Usługi archiwizacji danych

Centrum Informatyczne TASK

- CI TASK oferuje usługi IT realizowane przez infrastrukturę TASK oraz usługi doradcze dotyczące wykorzystania tej infrastruktury
- w ramach projektu PL-Grid oraz Krajowego Magazynu Danych istnieje możliwość archiwizacji danych

Kontakt: noc@task.gda.pl

lub telefonicznie w godzinach 8:00-16:00 pod numerem +48 58 347 10 77



Dziękuję za uwagę.

*Agnieszka Milewska
Biblioteka Główna
amilewska@gumed.edu.pl*