

# Jak napisać Plan Zarządzania Danymi Badawczymi

Opracowanie (na podstawie wytycznych NCN)  
Agnieszka Wasilewska



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

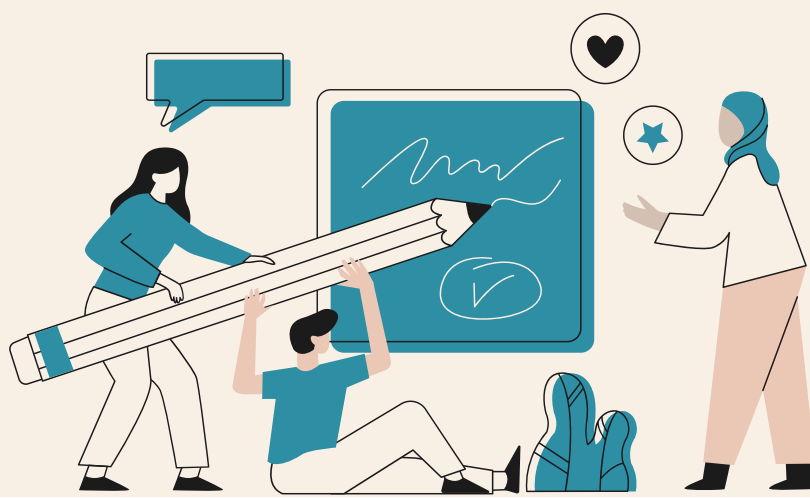
Opisz dane które zbierzesz lub utworzysz, z uwzględnieniem treści, zasięgu i typu danych. Należy również pokrótce opisać ich rodzaj, np. teksty, obrazy, zdjęcia, pomiary, statystyki, próbki fizyczne czy kody.

**Nowe dane** - ogólny opis, jak te dane będą pozyskiwane, jak będą pobierane, opis metodologii, użytego sprzętu, zastosowanych procesów.

- Zastanów się, w jaki sposób Twoje dane mogą się uzupełniać i integrować z istniejącymi danymi, lub czy są jakieś istniejące dane lub metody, których możesz użyć ponownie.
- Wskaż, które dane mają wartość długoterminową i powinny być udostępniane i / lub chronione.

**Ponowne wykorzystanie istniejących danych** – należy wskazać źródło danych, warunki i ewentualne ograniczenia ich ponownego wykorzystania.

- Należy wykazać, że dane pochodzą z wiarygodnych źródeł.
- Czy właściciel danych przekaże prawa do ponownego użycia?
- Dane z archiwów, bibliotek – czy nie ma ograniczeń ich ponownego wykorzystania, na jakich zasadach można wykorzystać, czy trzeba mieć zgodę na zbieranie i ponowne wykorzystanie danych?



# Pozyskiwane lub opracowane dane ( np. rodzaj, format, ilość)

## Objętość:

- Szacunkowa wielkość danych w MB / GB / TB.
- Wskaż proporcje danych surowych, danych przetworzonych i innych wyników wtórnych (np. raportów).
- Rozważ konsekwencje ilości danych w zakresie przechowywania, dostępu i ochrony. Czy musisz uwzględnić dodatkowe koszty?

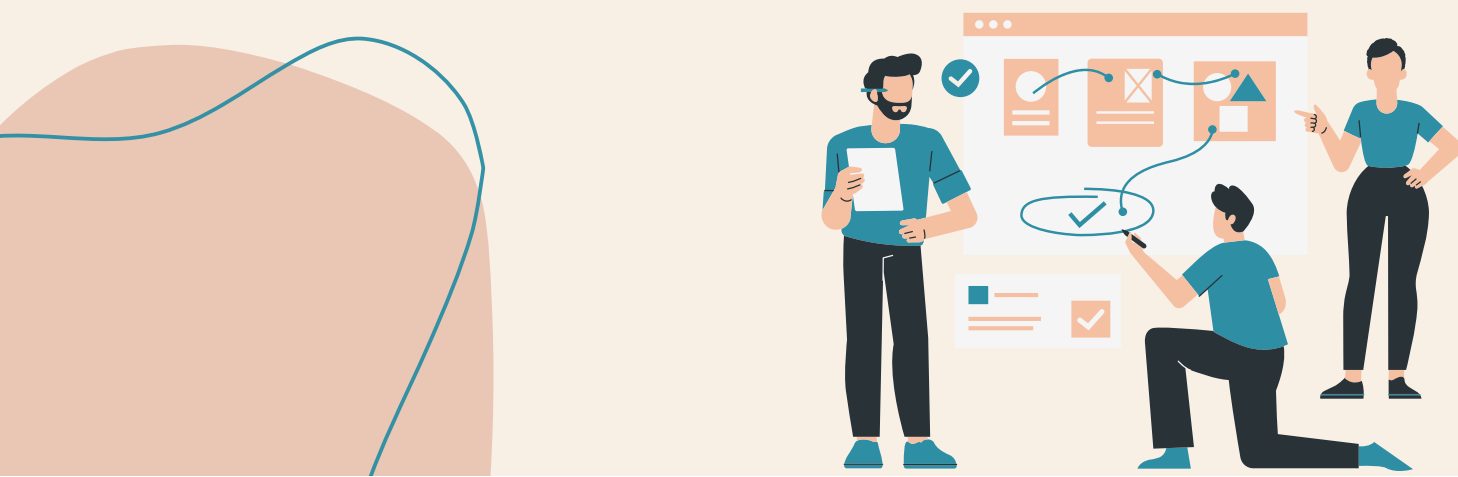
## Format:

- W jakim formacie będą znajdować się dane - należy stosować formaty otwarte i standardowe.
- Tabele: zaleca się użycie plików o formatach: CSV, TSV, SPSS portable. Nie zaleca się plików: XLS, XLSX.
- Teksty: zaleca się użycie plików w formatach: HTML, RTF, PDF. Nie zaleca się plików: DOC, DOCX.
- Media: zaleca się użycie plików w formatach: MP4, Flacc. Nie zaleca się plików: QTFF.
- Obraz: zaleca się użycie plików o formatach: TIFF, JPGE2000, PNG. Nie zaleca się plików: GIF, JPG.
- Dane uporządkowane: zaleca się użycie plików o formatach: XML, RDF. Nie zaleca się plików: RDBMS.



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

- **Standardy metadanych** – dzięki temu dane będą możliwe do odnalezienia nie tylko przez ludzi ale i przez algorytmy wyszukiwawcze. Repozytoria danych badawczych posiadają odpowiednie formatki do tworzenia metadanych. Dane umieszczone w uznanym, dziedzinowym repozytorium danych badawczych będą posiadały opis metadanych zgodny ze standardem.
- **Dokumentacja** - opisz typy dokumentacji, które będą towarzyszyć danym, aby pomóc użytkownikom w ich zrozumieniu i ponownym użyciu. Dokumentacja powinna opisywać ogólny kontekst zebranych danych - informacje dotyczące źródła pochodzenia danych, informacje o metodologii, o procedurach, definicje zmiennych czy jednostek pomiarowych.
- **Organizacja folderów i plików** – stosowanie spójnej konsekwentnej struktury, nazewnictwo plików odzwierciedlające zawartość, standardowy format daty – tak aby łatwo można było odnaleźć i posortować zawartość.
- **Konwencja nazewnictwa** - stosowanie słowników kontrolowanych używanych w danej dyscyplinie.



# Stosowane środki kontroli jakości danych

Jakość danych – subiektywna i obiektywna przydatność, dokładność, kompletność, spójność, aktualność, wyjątkowość

## Dane wytworzone:

- Stosowane procedury – rodzaj środków kontroli wytwarzania/zbierania danych i ich częstotliwość.
- Należy opisać, w jaki sposób planuje się kontrolować i dokumentować jakość i spójność danych np. procesy kalibracji sprzętu, wielokrotne pomiary, znormalizowane metody wraz z formularzami, szablony kwestionariuszy, pytania kontrolne.

## Ponownie wykorzystane dane:

- Należy opisać źródło.
- Dane powinny pochodzić z wiarygodnego źródła.



# Przechowywanie i tworzenie kopii zapasowych danych i metadanych podczas badań

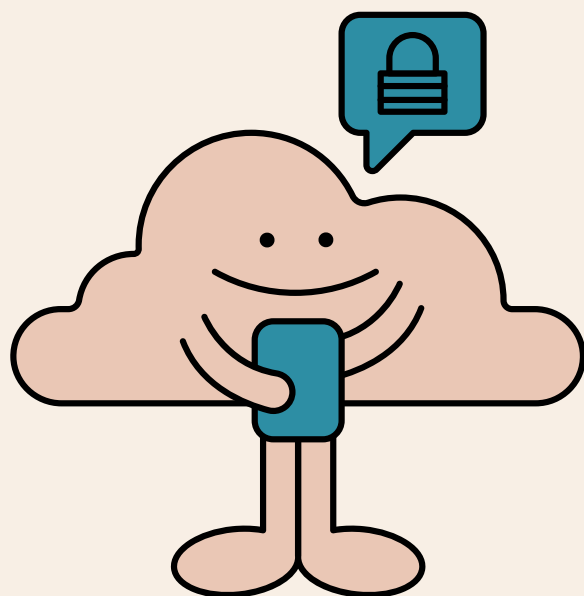
Gdzie będą przechowywane dane i kopie zapasowe podczas badań?

- Należy starannie wybrać sprzęt na którym dane będą zbierane i przechowywane. Dotyczy to zarówno serwerów, jak i przestrzeni dyskowej (nie zaleca się przechowywania danych na prywatnych komputerach, laptopach, dyskach przenośnych, urządzeniach USB).
- Przechowywanie w chmurze – z zastrzeżeniem, że serwery będą na terenie Europy. Nie zaleca się używania prywatnych chmur komercyjnych. Najlepiej żeby usługa była regulowana umową dwustronną z uczelnią (tj. OneDrive).
- Centrum Informatyczne TASK – oferuje przechowywanie danych dla projektów NCN realizowanych przez pracowników UG – 100 GB + 200 GB back-up.
- Kopie zapasowe – zaleca się stosowanie zasady 3-2-1 – co najmniej 3 kopie, na dwóch różnych nośnikach, jeden z nośników powinien się znajdować w zupełnie innej lokalizacji.
- Jak często będzie robiony backup, czy będzie to proces automatyczny/ręcznie kontrolowany, gdzie i przez kogo, co ile hasło będzie zmieniane.



# Sposób zapewniania bezpieczeństwa danych oraz ochrony danych wrażliwych podczas badań

- Kto będzie miał dostęp do danych podczas trwania projektu, jak dostęp będzie kontrolowany?
- Przewidywanie możliwych zagrożeń i ryzyka oraz sposobów postępowania.
- Jeśli dane są poufne/wrażliwe należy przedstawić wszelkie odpowiednie środki bezpieczeństwa i odnotować wszelkie formalne standardy, które będziesz przestrzegać np. szyfrowanie, anonimizacja, ściśle kontrolowany dostęp.
- **Zabezpieczenie fizyczne** – przed nieuprawnionym dostępem, zniszczeniem, uszkodzeniem nośników, kradzieżą.
- **Zabezpieczenie cyfrowe** – back-up, silne hasło, antywirus, firewall.



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

Jeżeli nie były pozyskiwane dane osobowe, zaznaczamy nie dotyczy (no personal data will be processed).

Powołujemy się na Politykę ochrony danych w instytucji:

- Inspektor Ochrony Danych UG.
- Dane będą gromadzone i przetwarzane zgodnie z rozporządzeniem o ochronie danych osobowych: Zarządzenie Rektora UG nr 163/R/21 w sprawie Polityki ochrony danych osobowych w Uniwersytecie Gdańskim.
- Polityka ochrony danych osobowych w Uniwersytecie Gdańskim.
- Czy uzyskano świadomą zgodę uczestników badania na udział w badaniu i przetwarzanie danych?
- Zabezpieczenie - anonimizacja, pseudonimizacja, szyfrowanie.





## Sposób zarządzania innymi kwestiami prawnymi, np. prawa własności intelektualnej

- Wskazanie właścicieli praw do wszelkich pozyskiwanych i wytwarzanych danych.
- Współpraca z inną jednostką lub korzystanie ze sprzętu innej jednostki - należy ustalić z jednostką współpracującą podział praw własności intelektualnej.
- Na jakich zasadach będą udostępnione dane - zaleca się stosowanie standardowych licencji Creative Commons.
- Czy istnieją ograniczenia prawne ponownego wykorzystania istniejących danych - uwzględnienie praw autorskich. Jeśli korzysta się z danych z bibliotek czy archiwów - stosowanie się do zasad i regulaminów.
- Aspekt etyczny - czy istnieją przeciwwskazania natury etycznej, czy potrzebna jest ocena komisji etycznej, zgoda Komisji Bioetycznej.
- Zgody i pozwolenia na prowadzenie badań – czy badania wymagają zezwoleń specjalistycznych organów.



# Sposób i termin udostępniania danych.

## Ewentualne ograniczenia w udostępnianiu danych lub przyczyny embarga

**Gdzie** dane będą udostępniane - w jaki sposób o danych dowiedzą się potencjalni użytkownicy?

- Miejsce udostępnienia – zalecane repozytorium danych badawczych.
- Repozytorium powinno zapewnić DOI lub inny trwały, unikalny identyfikator i spełniać warunki zasady FAIR.

**Kiedy** udostępnisz dane? – czy jest embargo (jak długo).

- Dane należy udostępnić możliwie szybko – w momencie publikacji wyników badań.
- Embargo - należy uzasadnić przyczynę.

Jeśli jakieś zbiory nie mogą zostać udostępnione lub dostęp do nich musi zostać ograniczony, w DMP należy wyjaśnić, dlaczego jest to konieczne (np. ograniczenia o charakterze prawnym).

Dane powinny być "as open as possible, as closed as necessary" tak otwarte jak to możliwe i tak zamknięte jak to konieczne.

- Ograniczenia w udostępnianiu wynikające z przyczyn prawnych, potencjału komercjalizacji, dane poufne, wrażliwe, tajemnica przedsiębiorstwa itp.



# Sposób wyboru danych przeznaczonych do przechowywania oraz miejsce długotrwałego przechowywania (repozytoria, archiwum)

## Selekcja:

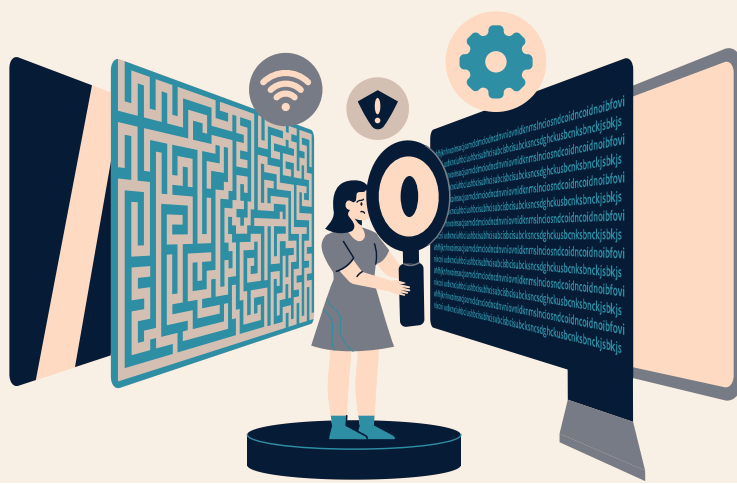
Jakie dane należy zachować a jakie zniszczyć? Na jakiej podstawie zdecydujesz jakie dane zachować np. jeśli dane mogą zostać ponownie wykorzystane w celu potwierdzenia wyników badań, przeprowadzenia nowych badań. Czy chcesz udostępniać dane z eksperymentu który się nie powiódł, żeby inni nie powielali błędów? Zdecyduj, które dane zachować i na jak długo.

## Które dane będą udostępnione?

- Nie wszystkie dane muszą być udostępnione w repozytorium – tylko te wyselekcjonowane, wyczyszczone, najlepszej jakości.
- Wymóg udostępnienia dotyczy danych stanowiących podstawę publikacji, służące do weryfikacji wyników badań.
- Czy będą udostępniane również dane surowe? – jeśli nie to oryginalne surowe dane powinny być zarchiwizowane.

## Archiwizacja:

- Jak długo będą przechowywane - w rozumieniu NCN uzasadniony okres przechowywania danych to minimum 10 lat.



# Metody lub narzędzia programowe umożliwiające dostęp do danych i korzystanie

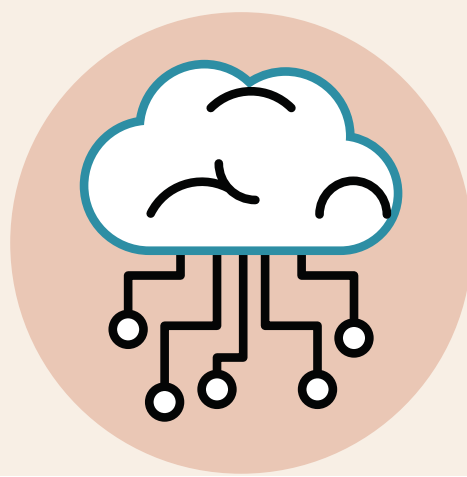
- Czy potencjalni użytkownicy danych będą potrzebowali określonych narzędzi aby uzyskać dostęp do danych i ponownie je wykorzystać?
- Czy do skanowania lub konwersji niezbędny będzie dodatkowy sprzęt lub oprogramowanie?
- Należy opisać oprogramowanie używane w trakcie zbierania danych, narzędzia (np. MatLab).
- Zasada FAIR (interoperacyjność) - czy jest możliwe wykorzystanie przez inne osoby i podmioty (w tym maszyny). Jeśli wykorzystanie zbioru wymaga opracowania translatora, sposobu mapowania czy konwersji formatów, świadczy to o jego ograniczonej interoperacyjności. O pełnej interoperacyjności możemy więc mówić wtedy gdy zdeponowany przez nas zbiór dostępny jest za pośrednictwem API, a jego metadane dodatkowo za pośrednictwem protokołu OAI-PMH.



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

Identyfikator nadawany jest w repozytorium – przy wyborze repozytorium danych badawczych należy zwrócić uwagę czy dane repozytorium nadaje DOI lub inny unikalny, trwały identyfikator.

- Trwały identyfikator - umożliwia dotarcie do zbioru nawet kiedy zmieni się adres url, pozwala na trafną i wydajną lokalizację danych, służy jako odniesienie oraz pozwala na łatwe śledzenie cytowań i wielokrotne użycie.



# Zadania związane z zarządzaniem danymi badawczymi - osoba odpowiedzialna za zarządzanie danymi

- Kto będzie odpowiedzialny za każde działanie związane z zarządzaniem danymi?
- Jak zostaną rozdzielone obowiązki między strony partnerskie w ramach wspólnych projektów badawczych?
- Przedstaw role i obowiązki dotyczące poszczególnych działań, np. przechwytywanie danych, tworzenie metadanych, jakość danych, przechowywanie i tworzenie kopii zapasowych, archiwizacja i udostępnianie danych. Zastanów się, kto będzie odpowiedzialny za zapewnienie przestrzegania odpowiednich zasad.



## Środki (finansowe i czasowe) przeznaczone do zarządzania danymi

Środki potrzebne do realizacji planu. Mogą one obejmować koszty przechowywania, sprzęt, oprogramowanie, czas pracy personelu, koszty przygotowania danych do depozytu i opłaty repozytorium (przeważnie repozytoria są bezpłatne). Tam, gdzie potrzebne są dedykowane zasoby, należy je nakreślić i uzasadnić konieczność ich użycia.

- Jakimi zasobami dysponujemy, a jakie dopiero musimy pozyskać?
- Jakie będą koszty związane będą z zapewnieniem standardów FAIR w projekcie? (koszty związane z długotrwałym przechowywaniem i udostępnianiem danych).



# Sekcja Zarządzania Danymi Badawczymi i Otwartej Nauki

## O nas:

- Udzielamy konsultacji w zakresie zarządzania danymi badawczymi.
- Oferujemy pomoc przy tworzeniu DMP.
- Doradzamy w zakresie bezpieczeństwa danych.
- Organizujemy szkolenia indywidualne i grupowe.
- Zarządzamy platformą Uniwersyteckie Czasopisma Naukowe.

## Zespół:

Agnieszka Wasilewska

Zbigniew Ruszczyk

Agnieszka Kranich-Lamczyk

Katarzyna Cherbetko-Zygmuntowicz

## Skontaktuj się z Data Stewardem

Biblioteka Główna, pok. 0.44

tel. 58 523-32-67

e-mail: [dane.badawcze@ug.edu.pl](mailto:dane.badawcze@ug.edu.pl)

