**Informacje ogólne**

**Plan zarządzania danymi to:**

1. formalny dokument zawierający zarys tego, co będziemy robić z danymi w trakcie trwania danego projektu badawczego i po jego zakończeniu;
2. żywy dokument, który ma wspierać realizację projektu na każdym etapie.

**Co powinien zawierać plan zarządzania danymi (DMP)?**

* Jakie dane zostaną wytworzone lub zebrane? (rodzaje danych, formaty plików, szacunkowa objętość danych)
* W jaki sposób będą pozyskiwane lub wytwarzane dane? (standardy, metody, oprogramowania, narzędzia)
* Jak zostaną uporządkowane i opisane? (metadane, dokumentacja)
* Kwestie etyczne i prawne (kwestie związane z ochroną danych osobowych, danych niejawnych, etc.)
* W jaki sposób dane zostaną udostępnione? (jak, kiedy, komu)
* Które dane będą przechowywane długoterminowo? (gdzie, jak długo)

**Dobry plan zarządzania danymi powinien być:**

* prosty i konkretny
* realistyczny, oparty na dostępnych informacjach
* aktualny

**Uporządkowanie danych i ich publiczne udostępnienie ma istotne znaczenie ze względu na:**

1. Możliwość weryfikacji wyników –powtarzalność w nauce (reproducibility)

2. Możliwość ponownego wykorzystania –również komercyjnie (re-use)

**Zasady FAIR Data (**[**Inicjatywa FAIR Data**](https://www.go-fair.org/fair-principles/)**)**, służą jako wytyczne dla umożliwienia ponownego wykorzystania danych naukowych w wyraźnie opisanych warunkach zarówno przez ludzi, jak i przez maszyny. Dane, które ze względu na ochronę prywatności nie mogą zostać opublikowane w całości, również mogą spełniać wszystkie zasady FAIR:

*Findable*- możliwe do odnalezienia.

* Zbiór danych opatrzony jest metadanymi, które umożliwiają odnalezienie tego zbioru zarówno przez ludzi jak i programy komputerowe;
* Do zbioru przypisany jest unikalny identyfikator (np. DOI), który stanowi jednocześnie element metadanych opisujących ten zbiór;
* Metadane są indeksowane w ogólnodostępnych bazach danych umożliwiających ich przeszukiwanie.

*Accessible*- dostępne dla wszystkich. ”Dostępność” w FAIR nie oznacza otwartego dostępu bez ograniczeń; dostępność oznacza, że poprzez metadane określa się dokładne warunki, na jakich dane są udostępniane i możliwe do ponownego wykorzystania.

* Dostęp do zbioru danych, a przynajmniej do metadanych, możliwy jest bezpośrednio poprzez unikalny identyfikator i nie wymaga dodatkowych narzędzi ani oprogramowania;
* Metadane są zawsze dostępne, nawet jeśli sam zbiór danych został już usunięty lub przeniesiony

*Interoperable**– interoperacyjne*

* Dane oraz metadane dostarczone w formacie zapewniającym łatwy odczyt i przetwarzanie zarówno przez ludzi jak i komputery; zbiory danych i metadane je opisujące zawierają odnośniki do innych powiązanych z nimi zbiorów.

*Reusable*– możliwe do ponownego użycia;

* Metadane zawierają liczne atrybuty dokładnie opisujące zbiór danych i uławiające użytkownikom określenie ich przydatności dla ich własnych badań;
* Zbiór danych zawiera licencję określającą jednoznacznie warunki ponownego wykorzystania
i przetwarzania danych;
* Metadane wyraźnie określają autora oraz miejsce powstania danych;
* Metadane są skonstruowane według ogólnie przyjętych standardów specyficznych dla danej dyscypliny oraz rodzaju danych.

Wg wytycznych programu Horyzont 2020: **As open as possible, as closed as necessary** – trzeba wziąć pod uwagę ograniczenia prawne i etyczne oraz spójność z głównym celem prowadzonych badań. Nie wszystkie zbiory danych mogą posiadać otwarty charakter, w szczególności dotyczy to danych osobowych, komercjalizacji wyników badań oraz bezpieczeństwa narodowego.

Wg Polityki NCN dotyczącej otwartego dostępu do publikacji:

Dane powiązane (podstawowy zestaw danych) z opublikowanymi artykułami powinny być udostępniane w otwartym repozytorium, tam gdzie to możliwe, zgodnie z warunkami licencji CC0.

Wszystkie opublikowanie metadane muszą spełniać wytyczne podane przez OpenAIRE i zawierać adnotację o finansowaniu ze środków projektu (NCN, numer projektu).

**Wytyczne dla wnioskodawców do uzupełnienia PLANU ZARZĄDZANIA DANYMI w projekcie badawczym.**

1. **OPIS DANYCH ORAZ POZYSKIWANIE LUB PONOWNE WYKORZYSTANIE DOSTĘPNYCH DANYCH**

**Co to są dane badawcze?**

Zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

Dane zebrane, zaobserwowane lub wytworzone jako materiał do analizy, w celu uzyskania oryginalnych wyników naukowych.

Wszystko co zostało wyprodukowane lub wytworzone w ramach prowadzonych badań.

**Przykłady** danych badawczych:

* Dane liczbowe (arkusze kalkulacyjne, bazy danych)
* Dokumenty tekstowe (notatki z eksperymentów, dzienniki, protokoły laboratoryjne, opisy procedur, opisy metodologiczne, kwestionariusze, ankiety)
* próbki
* artefakty, obiekty
* nagrania audio lub wideo
* fotografie, obrazy
* oprogramowanie (skrypty, pliki wejściowe)
* wyniki symulacji komputerowych
* modele matematyczne i algorytmy

Są to zarówno surowe dane, czyli takie, które uzyskano bezpośrednio w wyniku zastosowania narzędzia badawczego, jak i takie, które już zostały poddane obróbce.

Sposób pozyskania/wytworzenia danych:

eksperyment, obserwacja, symulacja, ankiety, kwerendy, badania terenowe….

Format:

Wybierając format zapisu danych badawczych zalecane jest ograniczenie się do tzw. formatów otwartych, czyli takich, które można otworzyć przy użyciu darmowego oprogramowania.

Zalecane formaty plików do archiwizacji:

* Tabele: CSV, TSV, SPSS portable
* Teksty: HTML, RTF, PDF
* Media: MP4, Flacc
* Obraz: TIFF, JPGE2000, PNG
* Dane uporządkowane: XML, RDF
* Inne rekomendowane formaty można znaleźć na stronie: <https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx>

Oszacowanie rozmiaru wytwarzanych i udostępnianych danych konieczne jest do zaplanowania środków do ich krótko- i długoterminowego przechowywania. Jeżeli wielkość danych będzie znaczna, konieczne może okazać się zapewnienie lub stworzenie przeznaczonej specjalnie dla nich infrastruktury (np. macierzy dyskowych) oraz oprogramowania (umożliwiającego np. odfiltrowanie
i pobranie jedynie potrzebnej części danych albo zdalne prowadzenie analiz przez inne osoby zainteresowane ich wykorzystaniem). Nie chodzi o podanie bardzo konkretnej odpowiedzi,
a o oszacowanie skali.

1. **DOKUMENTACJA I JAKOŚĆ DANYCH**

Metadane: (np. nazwa zbioru danych, autorzy), stanowiące podstawowy opis umożliwiający odnalezienie, identyfikację i zrozumienie danych innym użytkownikom.

Odpowiednie przygotowanie, uporządkowanie i opisanie danych pozwoli na ich sprawne wyszukanie.
Dane powinny zostać opatrzone metadanymi w taki sposób, aby odbiorca wiedział jakiego rodzaju są to dane, w jaki sposób zostały wytworzone oraz na jakich zasadach może z nich korzystać.

Nie ma jednego powszechnie obowiązującego standardu opisu metadanych dla danych badawczych, dlatego warto zapoznać się ze standardami opisu metadanych stosowanymi w repozytorium,
w którym zamierzamy zdeponować dane.

Ogólne standardy metadanych to Dublin Core i Data Cite, czy też Data Documentation Initiative (DDI). Są one uniwersalne dziedzinowo i powszechnie stosowane.

Informacje dot. Repozytorium UP: oprogramowanie DSpace używa schematu metadanych opartego na QDC (Qualified Dublin Core). Instytucja może rozszerzyć ten schemat podstawowy lub dodać niestandardowe schematy podobne do QDC. DSpace może również importować lub eksportować metadane z innych głównych schematów metadanych, takich jak MARC lub MODS.

**Dokumentacja**powinna opisywać metodologię prowadzonych badań oraz ich kontekst i źródło. Informuje o sposobie organizacji danych w trakcie projektu np. przyjętej konwencji, wersji
i strukturze folderów. Często zawiera także dodatkowe pliki potrzebne do skorzystania z danych (np. skrypty) czy wykorzystane standardowe słowniki, definicje zmiennych czy jednostek pomiarowych. Może zostać opisana w  osobnym pliku ReadMe.txt.

W skrócie: co jest niezbędne dla właściwego zrozumienia i ponownego użycia danych?

Wskazówka:

W zależności od rodzaju danych powinny zostać one odpowiednio katalogowane:

- elektroniczne w usystematyzowanych folderach (np. nazwa projektu -> rodzaj eksperymentu -> miesiąc), odpowiednio nazwanych plikach (np. rodzaj eksperymentu\_nazwisko osoby odpowiedzialnej\_data) ,

- papierowe w opisanych teczkach/segregatorach/na półkach)

Odpowiednio nazwany plik/zbiór plików może znacząco ułatwić użytkownikowi korzystanie z danych. Wszystkie te elementy składają się na późniejsze efektywnie wykorzystanie danych we właściwym kontekście.

2.2 W jaki sposób spójność i jakość gromadzenia danych będzie kontrolowana i dokumentowana? Kontrola może obejmować procesy takie jak kalibracja, powtarzanie poboru próbek lub pomiary, znormalizowane przechwytywanie danych, walidacja wprowadzanych danych, sprawdzanie poprawności wprowadzanych danych (peer review) lub reprezentacja kontrolowanymi słownikami.

**Przykład:**

* Podczas całego okresu realizacji projektu jakość pozyskiwanych danych będzie monitorowana oraz oceniana na bieżąco przez kierownika projektu lub/oraz przez zespół/zaplanowanych
w projekcie ekspertów. Również na bieżąco będzie doskonalona metoda badawcza, aby osiągnąć jak najlepszej jakości rezultaty.
* Badania laboratoryjne będą przeprowadzane skrupulatnie/zgodnie z powszechnie uznanymi zasadami/standardami oraz w ilości pozwalającej zakładać, że ostateczny rezultat jest prawidłowy.
1. **PRZECHOWYWANIE I TWORZENIE KOPII ZAPASOWYCH PODCZAS BADAŃ.**

Wskazówki:

Bezpieczne przechowywanie danych zapewnia zastosowanie reguły 3-2-1:

- zawsze należy mieć trzy backupy,

- należy używać dwóch różnych technologii przechowywania danych (chmura, pendrive, zewnętrzny dysk twardy, taśma itp.)

- jeden backup należy przechowywać w innym miejscu niż dwa pozostałe np. poza uczelnią.

Zalecane jest tworzenie kopii danych badawczych w dwóch różnych lokalizacjach/miejscach.

**Przykład:**

* Dane badawcze będą przechowywane w:

- „chmurze” Microsoft One Drive, gdzie tworzenie kopii zapasowej jest automatyczne co kilka minut

- w urządzeniu pomiarowym lub komputerze oraz ich kopia na nośniku zewnętrznym (np. pendrive, dysk zewnętrzny ew. płyta)

* Kopie zapasowe danych będą tworzone po zakończeniu generowania danych w danym dniu; zostanie przygotowany harmonogram tworzenia kopii bezpieczeństwa oraz wyznaczone odpowiedzialne osoby;

3.2 Ochrona danych (dane wrażliwe)

Wskazówki:

Dane wrażliwe to dane ujawniające pochodzenie rasowe lub etniczne, poglądy polityczne, przekonania religijne lub światopoglądowe, przynależność do związków zawodowych oraz dane genetyczne, dane biometryczne jednoznacznie identyfikujące osoby fizyczne lub dane dotyczące zdrowia, seksualności lub orientacji seksualnej tej osoby.

Jeżeli badania zakładają gromadzenie i/lub przetwarzanie danych wrażliwych konieczne jest przemyślenie ich właściwej ochrony.

Należy rozważyć sposób zarządzania ryzykiem oraz odzyskania danych w przypadku niespodziewanego zdarzenia.

**Przykład:**

* Dane poufne będą przechowywane na komputerach/innych nośnikach bez dostępu do sieci.
1. **WYMOGI PRAWNE, KODEKSY POSTĘPOWANIA.**

4.1 Przetwarzanie danych osobowych

Uwagi ogólne:

Uniwersytet Pedagogiczny wprowadził politykę ochrony danych osobowych, w tym danych wrażliwych, zgodnie z przepisami prawa polskiego i europejskiego. Na uczelni został powołany Inspektor Ochrony Danych Osobowych, który monitoruje i kontroluje przestrzeganie przepisów prawa w zakresie ochrony danych, a także szkoli i doradza w sprawach związanych
z bezpieczeństwem danych osobowych.

Pracownicy administracyjni realizujący projekt posiadają odpowiednie uprawnienia do przetwarzania danych osobowych w związku z wykonywaniem swoich obowiązków oraz w odpowiednio ograniczonym zakresie. Pracownicy są przeszkoleni i zobowiązani do przestrzegania zasad bezpieczeństwa i poufności.

Na UP wprowadzone zostało zarządzenie nr R/Z.0201-011/2018 Rektora UP im. KEN w Krakowie
z dnia 22 maja 2018 roku w sprawie ochrony danych osobowych w Uniwersytecie Pedagogicznym, które reguluje kwestie związane z dostępem do danych osobowych, ich bezpieczeństwem oraz powiązanymi procedurami.

**Przykład:**

* W razie przetwarzania danych osobowych każdemu uczestnikowi badania zostanie przedłożona informacja o przetwarzaniu danych osobowych, przygotowana zgodnie
z właściwymi przepisami. W razie wątpliwości szczegóły informacji o przetwarzaniu danych osobowych zostaną uzgodnione z IOD
* Dane osobowe będą podlegać anonimizacji lub pseudonimizacji (klucz do szyfru znajduje się u kierownika projektu i nie podlega udostępnianiu).
* Dane osobowe będą zabezpieczone zgodnie z obowiązującymi na **Uniwersytecie Pedagogicznym** przepisami, ze szczególną dbałością o ich bezpieczeństwo.
* Pliki i komputery będą chronione silnym hasłem (zmienianym np. raz w miesiącu).
* Dane na nośnikach papierowych będą zamykane na klucz. Dane na nośnikach elektronicznych zostaną zabezpieczone silnym hasłem dostępu.
* Dostęp do danych osobowych będzie miał tylko kierownik/wyznaczona osoba.

4.2 Prawa własności intelektualnej i prawa własności

Na Uniwersytecie Pedagogicznym obowiązuje "Regulamin zarządzania prawami autorskimi, prawami pokrewnymi i prawami własności przemysłowej oraz zasad komercjalizacji w Uniwersytecie Pedagogicznym im. Komisji Edukacji Narodowej w Krakowie” , który określa ogólne zasady zarządzania prawami autorskimi i prawami własności przemysłowej odnoszącymi się do utworów
i projektów wynalazczych.

W razie konsorcjum należy przemyśleć kwestie związane z prawami kontroli dostępu do danych dla partnerów projektu, ustalenie współautorstwa i podziału praw do danych.

Należy wskazać właścicieli praw autorskich i praw własności intelektualnej do wszelkich pozyskiwanych i wytwarzanych danych. Trzeba określić czy istnieją jakiekolwiek ograniczenia prawne dotyczące ponownego wykorzystania danych pochodzących od osób trzecich.

Należy skonsultować z przepisami w jednostce finansującej.

**Przykład:**

* W miarę możliwości dane badawcze będą udostępniane na licencji Creative Commons za zgodą autorów.

Wskazówki:

Warto skorzystać z poniższych otwartych licencji:

* Licencje [**Creative Commons**](https://creativecommons.pl/poznaj-licencje-creative-commons/)(przykłady licencji: CC BY, CC BY-SA, CC BY-NC, CC BY-NC-ND, więcej: https://creativecommons.pl/poznaj -licencje-creative-commons)
* Licencje wolnego oprogramowania [**GNU**](https://www.gnu.org/licenses/licenses.html)
* Licencje baz danych [**Open Data Commons**](https://opendatacommons.org/)
1. **UDOSTĘPNIANIE I DŁUGOTRWAŁE PRZECHOWYWANIE DANYCH.**

**Przykładowe odpowiedzi dotyczące sposobu i terminu udostępniania danych:**

* Upowszechnienie wyników badań nastąpi w wyniku publikacji papierowej na zasadach określonych w umowie z wydawnictwem.
* W projekcie przewidziano środki finansowe na udostępnienie danych w formie Open Access.
* Wyniki badań zostaną udostępnione w repozytorium … po upływie odpowiedniego czasu przewidzianego w umowie z wydawnictwem/czasopismem. Dane surowe będą udostępnione przez kierownika projektu w formie brudnopisu/artykułu w nieostatecznej wersji przekazanej do druku/innej w otwartym repozytorium (sugerowane Repozytorium UP)
* Wyniki badań będą udostępnione w formie artykułu w czasopiśmie ukazującym się on-line (Gold Open Access)
* Dane zawierające roboczy materiał będą przechowywane na nośniku elektronicznym, notatki w formie papierowej zostaną zarchiwizowane w zbiorach autora.
* Dane będą dostępne po zakończeniu projektu (LUB dane będą udostępniane w trakcie realizacji projektu po wykonaniu poszczególnych zadań). Łatwa lokalizacja danych będzie możliwa dzięki DOI i adresom URL. Dane będą przechowywane w 2 egzemplarzach przez 10 lat.
* Nieprzetworzone dane (surowe) oraz opracowane (publikacja nt. uzyskanych wyników) będą udostępniane na prośbę skierowaną bezpośrednio do autora danych

Sugerowanym przez NCN okresem przechowywania danych jest minimum 10 lat.

Jeśli z jakichś powodów nie można udostępnić samych danych lub trzeba wycofać uprzednio udostępniony zbiór, należy zadbać o dostępność opisujących je metadanych. W pierwszym przypadku będą one sygnałem, iż dane w ogóle istnieją, co pozwoli uniknąć duplikacji badań, w drugim zaś – oznaką tego, że w określonym czasie były one udostępnione i mogły zostać zacytowane lub poddane wtórnym analizom.

5.2 Selekcja danych i długoterminowe przechowywanie

Należy umotywować jakie dane i dlaczego zostaną zdeponowane w repozytorium (często nie wszystkie dane wytworzone w czasie projektu trzeba, a nawet warto zamieszczać w repozytorium.)

Zalecanym repozytorium jest Repozytorium UP:

Uniwersytet Pedagogiczny zarządza Repozytorium UP ( http://rep.up.krakow.pl/), które zapewnia długotrwałe archiwizowanie zdeponowanych danych na serwerach uniwersyteckich. Bezpieczeństwo danych jest także zapewnione poprzez wykonywanie regularnych kopii zapasowych. Kopie bezpieczeństwa przechowywane są na serwerach Biblioteki Głównej.

Do zabezpieczenia danych wykorzystywana jest metoda snapshot (tzw. migawka - zrzut danych, rozumiany jako sposób zapewniania dostępu do starszych wersji danych przechowywanych
w pamięci masowej) robiony automatycznie codziennie, natomiast backup (kopia zapasowa) - 1x
w tygodniu manualnie przez informatyka odpowiedzialnego za infrastrukturę informatyczną.

5.3 Oprogramowanie i metody umożliwiające dostęp do danych

**Przykład:**

* Wszystkie dane zostaną zapisane w formatach …., do ich wykorzystania wystarczą ogólnodostępne programy freeware.
* W repozytorium tematycznym zostaną zdeponowane metadane oraz zostanie podany adres kontaktowy, na który będzie można przesłać zapytanie w sprawie dostępu do danych

5.4 Unikalny i trwale przypisany identyfikator dla każdego zbioru danych

**Przykład:**

* Trwały identyfikator, w postaci URI, zostanie nadany w momencie zdeponowania danych przez Autora w repozytorium UP.
* DOI zostanie pozyskane dla każdej publikacji projektowej
1. **ZADANIA ZWIĄZANE Z ZARZĄDZANIEM DANYMI ORAZ ZASOBY**

6.1 Osoba odpowiedzialna za zarządzanie danymi

* Na uczelni w pionie Rektora zostanie powołany pełnomocnik/ specjalista ds. zarządzania danymi.
* Osobą odpowiedzialną za zarządzanie danymi będzie kierownik projektu.

W przypadku konsorcjum należy opisać podział zadań pomiędzy partnerami.

6.2 Zasoby przeznaczone na cele zarządzania danymi i zagwarantowanie przestrzegania zasad FAIR

Wskazówki:

Deponowanie zbiorów danych w repozytorium UP jest bezpłatne.

Należy uwzględnić opłaty za Open Access

Pomocne linki:

<https://www.dcc.ac.uk/dmps> - na stronie można znaleźć gotowe plany zarządzania danymi, szablony, przewodniki, wytyczne, a także informacje na temat formatów zamieszczanych danych, metadanych, w zależności od dyscypliny oraz lista kontrolna (Checklist for a Data Management Plan) dotycząca zawartości DM

[**DMPTool**](https://dmptool.org/) – narzędzie online służące tworzeniu planów zarządzania danymi; **zawiera przykłady takich planów.**
[**DMPonline**](https://dmponline.dcc.ac.uk/) – kreator planów zarządzania danymi badawczym

<https://www.e-education.psu.edu/dmpt/> - tutorial nt. DMP University of Pennsylvania

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide>