

Dane badawcze (Research Data) – są to zebrane, w wyniku zastosowania różnych technik badawczych, materiały o charakterze faktograficznym (w postaci liczbowej, tekstowej, graficznej czy dźwiękowej), uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

Powszechnie wyróżnia się dane badawcze surowe, nieprzeanalizowane, czyli takie, które uzyskano bezpośrednio w wyniku zastosowania narzędzia badawczego, w różnych przedsięwzięciach naukowych lub też zgromadzone na potrzeby konkretnych projektów oraz dane, które poddane zostały obróbce.

## **1. Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych**

Należy wyjaśnić, jakie metodologie lub oprogramowanie zostaną wykorzystane, do gromadzenia danych badawczych. Podać wszelkie ograniczenia, jeżeli takie istnieją, dotyczące ponownego wykorzystania istniejących danych.

Rodzaje gromadzonych danych są bardzo różnorodne. Zależne od dziedziny nauki oraz przyjętej metodologii badań. Są to m.in.:

- Dokumenty tekstowe, notatki
- Dane liczbowe
- Kwestionariusze, ankiety, wyniki badań ankietowych
- Nagrania audio i video, zdjęcia
- Zawartość baz danych (video, audio, teksty, obrazy)
- Modele matematyczne, algorytmy
- Oprogramowanie (skrypty, pliki wejściowe...)
- Wyniki symulacji komputerowych
- Protokoły laboratoryjne, opisy metodologiczne
- próbki, artefakty, obiekty\*

Kwestią wartą przemyślenia są typy danych, sposób ich gromadzenia i/lub przetwarzania, ilość i częstotliwość występowania.

Formaty plików mogą być dowolne, jednak dbając o powszechny dostęp i otwartość, dobrze korzystać z formatów, które nie wymagają komercyjnego oprogramowania do odczytu danych. W jednym opisie można dodać wiele plików. Jeżeli plików jest dużo, dobrym rozwiązaniem jest ich pogrupowanie i spakowanie, np. do postaci .zip. Należy także dobrze przemyśleć nazewnictwo plików. 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. Dokumentacja i jakość danych**

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. Można np. utworzyć osobny plik ReadMe.txt, który będzie zawierał dokumentację pozyskiwania danych badawczych, licencje, prawa autorskie itp. Jeżeli istnieje już publikacja naukowa, w której opisana jest dokumentacja badawcza, należy podać odnośnik do niej w polu adres URL.

Metadane umożliwiają nam scharakteryzowanie danych badawczych, tak by potencjalny użytkownik wiedział jakiego rodzaju są to dane. Metadane charakteryzują opis całego zbioru danych (autor, tytuł, data powstania, licencja, dyscyplina naukowa etc.). Dane badawcze muszą być udostępnione wraz z ich metadanymi.

W metadanych opisujących dane badawcze powinien znajdować się:

Autor lub autorzy danych,

Ujednolicony tytuł,

Data udostępnienia danych,

Opis, w którym osoba wprowadzająca dane badawcze powinna krótko scharakteryzować ich zawartość, pochodzenie, stosowane metody badawcze, kontekst badań i inne,

Zakres czasowy: należy podać datę początkową i datę końcową określającą czas prowadzonych badań, który często jest tożsamy z okresem trwania grantu,

Dostawca danych: do jakiej instytucji lub osoby należą dane,

Obszar badań: należy wskazać Dzielnicę nauki/sztuki z zawężeniem do Dyscypliny naukowej/artystycznej. Wyboru dokonujemy spośród klasyfikacji dziedzin nauki i dyscyplin naukowych oraz dyscyplin artystycznych wskazanych w Rozporządzeniu Ministra Nauki i Szkolnictwa Wyższego z dnia 20 września 2018 r.,

Unikalny identyfikator danych badawczych: należy zastanowić się czy repozytorium, w którym deponowane są dane nadaje im unikalne identyfikatory np. DOI (np. w Repozytorium UJ jest to możliwe).

W gromadzeniu danych istotna jest kontrola jakości na każdym etapie prowadzonych badań. Należy wyjaśnić w jaki sposób spójność i jakość gromadzonych danych będzie kontrolowana i udokumentowana. Opis może zawierać np. procesy takie jak kalibracja, powtarzanie próbek lub pomiarów, znormalizowane przechwytywanie danych, sprawdzanie poprawności wprowadzania danych, wzajemna ocena danych lub reprezentacja z kontrolowanymi słownikami.

### **3. Przechowywanie i tworzenie kopii zapasowych podczas badań**

Kwestię bezpieczeństwa oraz przechowywania danych należy gruntownie przemyśleć dla całego procesu gromadzenia i ewentualnego przetwarzania danych badawczych. Należy przeanalizować kwestię dostępu do danych (szczególnie jeżeli zawierają dane wrażliwe), by zapobiec niewłaściwemu dostępowi do poufnych danych. Konieczne jest też opracowanie planu tworzenia kopii zapasowych, by zapobiec utracie danych w wyniku np. awarii sprzętu. Należy zastanowić się gdzie będą przechowywane i zabezpieczane dane w trakcie procesu badawczego (zalecane jest tworzenie kopii danych badawczych w dwóch różnych lokalizacjach/miejscach). Zaleca się przechowywanie danych w centralnych systemach gromadzenia informacji na macierzystej uczelni. Przechowywanie danych w pamięciach masowych, komputerach przenośnych itp. może

spowodować ich utratę. Należy opisać w jaki sposób dane zostaną odzyskane w przypadku awarii sprzętu.

#### **4. Wymogi prawne, kodeksy postępowania**

Ochrona danych (dane wrażliwe, dane chronione prawem autorskim)

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 danej osoby.

Jeżeli badania zakładają gromadzenie i/lub przetwarzanie danych wrażliwych konieczne jest przemyślenie ich właściwej ochrony, ustalenie, kto będzie miał dostęp do danych, kto jest właścicielem danych itp. Jak będą chronione dane wrażliwe w przypadku badań prowadzonych w wielu instytucjach naukowych jednocześnie.

Uniwersytet Jagielloński wdrożył politykę ochrony danych wrażliwych i oraz powołał Inspektora Ochrony Danych do jej przestrzegania. Pomoże on także w opracowaniu planu zarządzania danymi, który rozwiąże wszelkie potencjalne problemy związane z prywatnością lub prawem. Zobacz: <https://iod.uj.edu.pl>.

Prawa autorskie i licencje

Należy wskazać właścicieli praw autorskich i praw własności intelektualnej (więcej: <https://www.uj.edu.pl/documents/1587933/6bc2fafc-cd80-4ec1-896d-858bd181ab27>) do wszelkich pozyskiwanych i wytwarzanych danych. Trzeba określić czy istnieją jakiegokolwiek ograniczenia prawne dotyczące ponownego wykorzystania danych pochodzących od osób trzecich.

Należy także wskazać licencje dla udostępnianych danych badawczych. Rekomenduje się korzystanie z otwartych licencji Creative Commons, możliwe jest też udostępnienie danych na zasadach domeny publicznej. Autor deponujący dane badawcze w repozytorium odpowiada za uzyskanie wszelkich zgód na udostępnienie danych, jak również odpowiada za anonimizację/pseudonimizację danych osobowych i wrażliwych. Należy pamiętać, że w przypadku prawa o ochronie danych osobowych (tj. RODO) konieczne będzie uzyskanie świadomej zgody uczestników na utrwalanie i udostępnianie ich danych osobowych.

#### **5. Udostępnianie i długotrwałe przechowywanie danych**

Dostęp i ponowne użycie danych badawczych

Zapewnienie dostępu do danych badawczych polega na ich udostępnieniu oraz opisanie. Należy określić kiedy dane zostaną udostępnione (czy w trakcie trwania, czy po zakończeniu badań, należy podać termin lub terminy udostępnienia) oraz czy dostęp będzie pełny czy ograniczony (w tym przypadku należy wskazać ograniczenia i przeszkody uniemożliwiające ich pełne/częściowe udostępnienie).

Ponowne użycie danych badawczych w innym kontekście powinno być zapewnione poprzez zastosowanie unikalnego i trwale przypisanego identyfikatora np. DOI.

Długoterminowa archiwizacja danych badawczych

Długoterminowa archiwizacja to przechowywanie danych badawczych w dłuższym okresie czasu. W planie zarządzania danymi należy uwzględnić gdzie będą przechowywane dane. W przypadku wyboru instytucji zewnętrznej, która udostępnia repozytorium danych badawczych, istotne jest uwzględnienie m. in.: czy posiada plan przechowywania danych w dłuższym okresie czasu, czy pliki, w których są zapisane dane, można opisać metadanymi, kto jest odpowiedzialny za dostęp do danych np. za 10 lub 15 lat oraz kto finansuje repozytorium i jakie są warunki przechowywania.

Uniwersytet Jagielloński zarządza Repozytorium UJ (<https://ruj.uj.edu.pl>), które zapewnia długotrwałe archiwizowanie zdeponowanych danych na serwerach uniwersyteckich zarządzanych przez Centrum Rozwoju Systemów Zintegrowanych Uniwersytetu Jagiellońskiego. Bezpieczeństwo danych jest także zapewnione poprzez wykonywanie regularnych kopii zapasowych.

## 6. Zadania związane z zarządzaniem danymi oraz zasoby

Należy przemyśleć kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem) w trakcie i po skończeniu badań? Opiekun rozumiany jest tutaj jako osoba lub instytucja, która docelowo zajmie się zarządzaniem danymi w dłuższym okresie czasu.

Jeżeli dane zostaną przekazane do wybranego repozytorium (instytucjonalnego, centralnego, dziedzinowego itp.), to należy opisać jak będzie w nim zapewniona: jakość danych, przechowywanie i tworzenie kopii zapasowych, długoterminowa archiwizacja, udostępnianie oraz kto będzie odpowiadał za zarządzanie danymi (np. instytucja, osoba).

Jeżeli dane nie zostaną przekazane do żadnego repozytorium, to podobnie jak wyżej należy zastanowić się jak będą przechowywane i kto będzie odpowiadał za ich zarządzanie w trakcie i po zakończeniu badań.

Należy rozważyć jakie zasoby (np. ludzkie, finansowe, czasowe) będą potrzebne do zarządzania danymi zbieranymi w trakcie własnych badań. Jeżeli badania prowadzone są przez kilka instytucji, to kto w danym projekcie będzie odpowiedzialny za zarządzanie danymi i kto poniesie koszty finansowe. Mogą to być koszty: przechowywania i archiwizacji, zakupu sprzętu, opłacenia personelu, przygotowania danych, związane z opłatami depozytowymi, utrzymania repozytorium. Należy oszacować potrzebne koszty i ustalić w jaki sposób zostaną opłacone.

W trakcie wyboru sposobu zarządzania danymi należy ustalić czy gwarantuje on przestrzeganie zasad FAIR lub jakie koszty trzeba ponieść, żeby spełnić te zasady.

Zasady FAIR Data w rozwinięciu oznaczają:

- Findable - łatwo znajdowane i wyszukiwane.
- Accessible - dostępne dla wszystkich.
- Interoperable - interoperacyjne, tak aby można było je połączyć z innymi danymi.
- Reusable - wielokrotnego użytku.

Zasady FAIR Data służą jako wytyczne dla umożliwienia ponownego wykorzystania danych naukowych w wyraźnie opisanych warunkach, zarówno przez ludzi, jak i przez maszyny. Więcej o FAIR Data można przeczytać tutaj: <https://www.go-fair.org/fair-principles>.

Narzędzie do szybkiej oceny czy dane spełniają zasady FAIR znajduje się pod tym linkiem: <https://www.ands-nectar-rds.org.au/fair-tool>.

Science Europe przygotowała przewodnik w wersji angielskiej, w którym jest przykładowy szablon planu zarządzania danymi: [https://www.scienceurope.org/media/jezkhnoo/se\\_rdm\\_practical\\_guide\\_final.pdf](https://www.scienceurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf)

Zapraszamy na stronę <https://ruj.uj.edu.pl/xmlui/help/0067>, gdzie można zapoznać się z informacjami dotyczącymi przechowywania danych badawczych w Repozytorium Uniwersytetu Jagiellońskiego, które umożliwia m. in. deponowanie danych badawczych, publikacji naukowych i dydaktycznych.

#### **Bibliografia**

1. Dane badawcze. Repozytorium Uniwersytetu Jagiellońskiego <https://ruj.uj.edu.pl/xmlui/help/0067> (dostęp: 25.10.2019).
2. FAIR Principles. <https://www.go-fair.org/fair-principles> (dostęp: 25.10.2019).
3. Open research data. Definitions. <https://www.fosteropenscience.eu/content/open-research-data-definitions> (dostęp: 25.10.2019).
4. Narodowe Centrum Nauki. Wytyczne dla wnioskodawców do uzupełnienia Planu Zarządzania Danymi w projekcie badawczym. [https://www.ncn.gov.pl/sites/default/files/pliki/regulaminy/wytyczne\\_zarządzanie\\_danymi.pdf](https://www.ncn.gov.pl/sites/default/files/pliki/regulaminy/wytyczne_zarządzanie_danymi.pdf) (dostęp: 25.10.2019).
5. Science Europe (2018). Science Europe Practical Guide to the International Alignment of Research Data Management. [https://www.scienceurope.org/media/jezkhnoo/se\\_rdm\\_practical\\_guide\\_final.pdf](https://www.scienceurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf) (dostęp: 25.10.2019).