



PLAN ZARZĄDZANIA DANYMI

DATA MANAGEMENT PLAN

dr Leszek Szafrński

16 listopada 2023 r.

Dane badawcze to wszelkiego rodzaju informacje (ilościowe i jakościowe) cyfrowe i analogowe, zaobserwowane, zbierane, przetwarzane lub wytworzone jako materiał do analizy, w celu uzyskania oryginalnych wyników naukowych.

Dane badawcze są własnością jednostki naukowej.

Rodzaje gromadzonych danych:

- 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

Dlaczego zarządzanie danymi jest ważne?

- Ochrona danych przed utratą
- Poprawa jakości danych i badań
- Niższy koszt badań
- Oszczędność czasu
- Ponowne użycie
- Kontrola (proces badawczy powinien być możliwy do zweryfikowania)

Koszty zarządzania danymi nie są tak duże jak koszty wytworzenia
nowych danych

FAIR DATA

W zarządzaniu danymi istotne jest przestrzeganie zasad FAIR Data, które oznaczają:

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

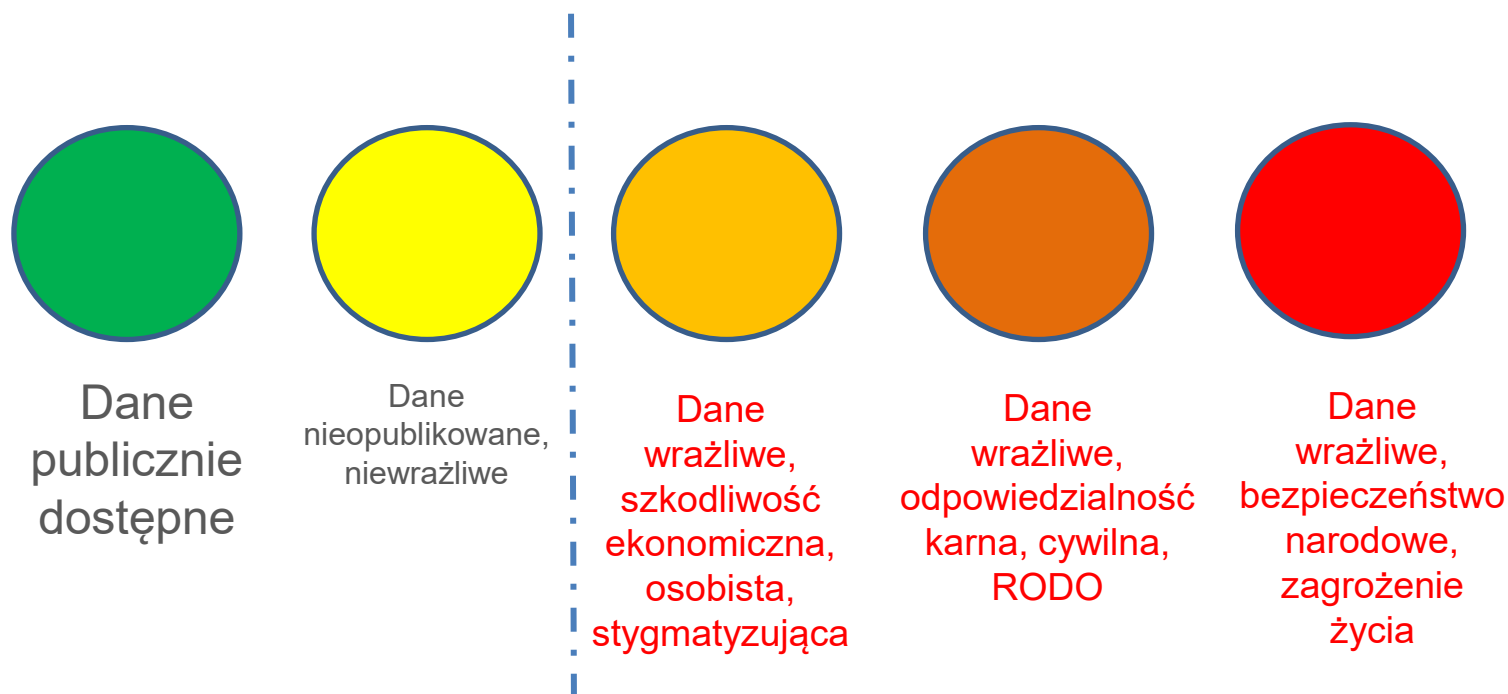
Więcej o FAIR Data można przeczytać tutaj: <https://www.go-fair.org/fair-principles>.

“As open as possible, as closed as necessary”

Jakie dane trzeba otwierać,
jakie można,
a jakich nie wolno?



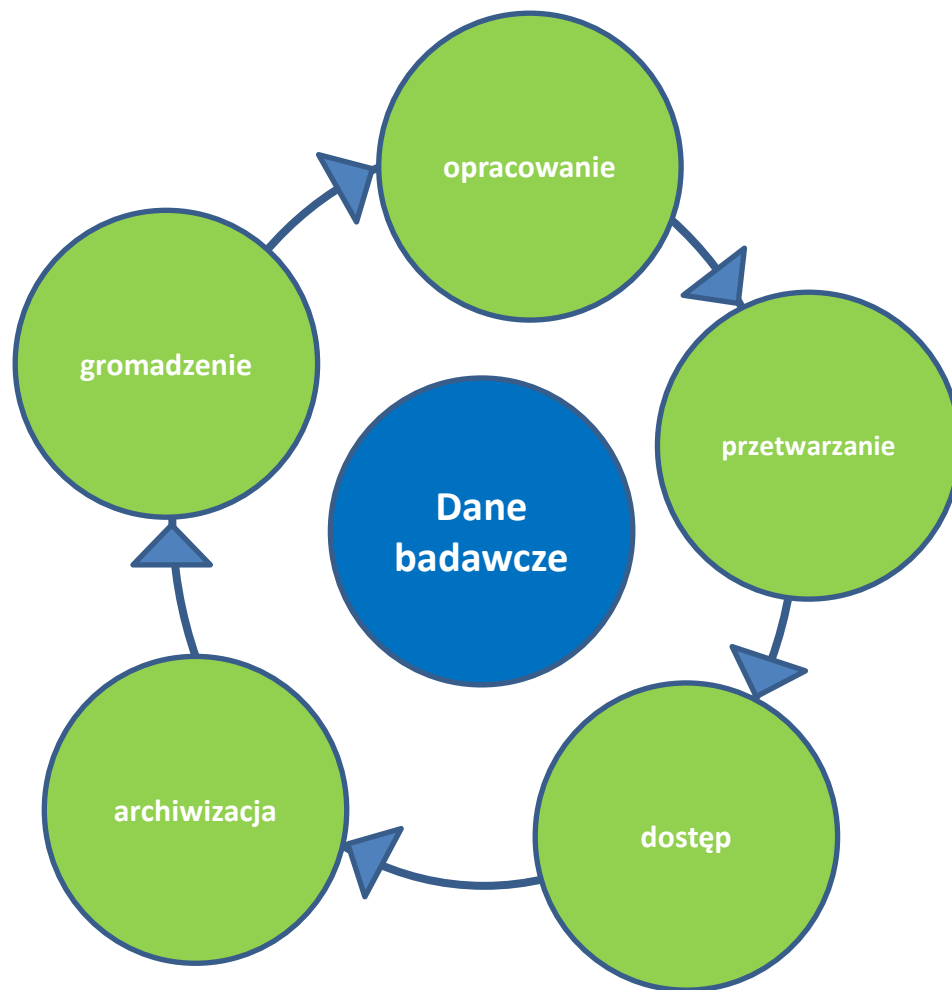
5 poziomów zabezpieczenia danych - Harvard University Information Security



Plan zarządzania danymi (ang.: Data Management Plan (DMP))

- Rodzaje zbieranych danych, metadane i dokumentacja
 - Przechowywanie i bezpieczeństwo danych
 - Ochrona danych (dane wrażliwe, dane chronione prawem autorskim, licencje)
 - Dostęp i ponowne użycie danych badawczych
 - Długoterminowa archiwizacja danych badawczych
 - Zarządzanie danymi (opiekun danych, zasady FAIR)
-

Plan zarządzania danymi (ang.: Data Management Plan (DMP))



Przykłady wypełnienia planu zarządzania danymi:

https://ruj.uj.edu.pl/xmlui/bitstream/handle/item/155290/Szafranski_Leszek_DMP_examples_2020.pdf?sequence=1&isAllowed=y

Plan zarządzania danymi (ang.: Data Management Plan (DMP))

Dlaczego jest ważny?

- Wspiera zabezpieczenie danych przed utratą
 - Ułatwia współpracę pomiędzy naukowcami w projekcie
 - Wspiera kontrolę jakości badań
 - Ułatwia zarządzanie danymi badawczymi w projekcie
 - Wspiera zabezpieczenie dostępu do wrażliwych danych
 - Wspiera zorganizowanie długoterminowej archiwizacji danych badawczych
-

Sposoby pozyskiwania i wytwarzania danych

Należy opisać standardy, metody i oprogramowanie, która posłużą do wytworzenia lub gromadzenia danych np. przyjętej konwencji, wersji, strukturze i organizacji plików/folderów, potrzebie digitalizacji danych analogowych, rodzaju, formacie i szacunkowej objętości plików.

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, strukturze i organizacji plików/folderów, potrzebie digitalizacji danych analogowych, rodzaju, formacie i szacunkowej objętości plików.

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, opis etc.). Dane badawcze muszą być udostępnione wraz z ich metadanymi.

Zakres metadanych:

Tytuł * Wprowadź tytuł...	Opis * To pole obsługuje tylko ni Tekst *	Powiązane publikacje *	Typ powiązania * Wybierz... Tytuł powiązanej publikacji * Wybierz... Identyfikator * Wybierz... Autor/Autorzy * Wybierz... Typ identyfikatora * Wybierz... URL * Wpisz kompletny URL, zaczynając od ht
Wersja językowa tytułu * Wybierz...	Nazwisko i imię lub nazwa organizacji * Szafrński, Leszek Jednostka * Jagiellonian University Schemat identyfikowai Wybierz... Abstrakt *	Streszczenie projektu * Powiązany zbiór danych *	Typ powiązania * Wybierz... Tytuł * Wybierz... Identyfikator * Wybierz... Autor/Autorzy * Wybierz... Typ identyfikatora * Wybierz... URL * Wpisz kompletny URL, zaczynając od ht
Współtwórcza * Nazwisko i imię lub nazwa Współtwórcy * Nazwisko, imię lub organizacja Jednostka * Wybierz... Schemat identyfikowai Wybierz... Słowa kluczowe *	Dziedzina nauki zgodnie z MEIN * Wybierz... Dziedzina/Dyscyplina * Wybierz... Termin * Wybierz... Uwagi *	Język * Wybierz...	Wybierz język * Wybierz...
Kontakt * Osoba do kontaktu * Szafrński, Leszek E-mail * l.szafranski@uj.edu.pl Jednostka * Jagiellonian University Dziedzina nauki w OECD * Wybierz... Dziedzina/Dyscyplina * Wybierz...			

Producent * Nazwisko i imię lub nazwa organizacji * Nazwisko, imię lub organizacja Jednostka * Wybierz... Skrót * Wybierz... URL * Wpisz kompletny URL, zaczynając od ht	Nazwisko i imię lub nazwa organizacji * Nazwisko, imię lub organizacja Jednostka * Wybierz... Skrót * Wybierz... URL * Wpisz kompletny URL, zaczynając od ht	Nazwa grantu * Wybierz... ROR instytucji finansującej * Wybierz...
Finansowanie * Jednostka finansująca * Wybierz... Nazwa grantu * Wybierz... ROR instytucji finansującej * Wybierz...		
Deponent * Szafrński, Leszek Data zdeponowania * 2023-06-12 Okres gromadzenia danych * Początek * YYYY-MM-DD Koniec * YYYY-MM-DD Rodzaj danych w zbiorze * Wybierz...		

Licencja i warunki

Warunki korzystania

Licencja

Zbiór danych udostępniany jest na podstawie niniejszej licencji. Pamiętaj, że licencję, na prawach której udostępniasz swoje dane badawcze możesz zmienić edytując wymagania dotyczące warunków.

CC BY-NC 4.0

CC BY-NC 4.0

Piki z ograniczeniami + warunki dostępu

Poproś o dostęp *

☒ Włącz prośbę o dostęp

Warunki dostępu *

Miejsce dostępu do danych *

Oryginalne archiwum *

Dokumentacja i jakość danych

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, żeby wyeliminować błędy pomiarowe i stronniczość.

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.

Przechowywanie i tworzenie kopii zapasowych podczas badań

Należy zastanowić się gdzie będą przechowywane i zabezpieczane dane w trakcie procesu badawczego (zalecane jest tworzenie kopii danych badawczych według **zasady 3-2-1**). 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.

Wymogi prawne, kodeksy postępowania

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.

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.

Uniwersytet Jagielloński wdrożył politykę ochrony danych wrażliwych 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>.

Wymogi prawne, kodeksy postępowania

Dane osobowe

Imię i nazwisko: Andrzej Nowak
Telefon: 999333444
Adres: ul. Mieszkalna 10/1,
32-447 Kraków

Pseudonimizacja

Imię i nazwisko: #12#\$@456!
Telefon: #\$65\$%&65
Adres: D435\$fg#\$^\$%%TY%^

Anonimizacja

Imię i nazwisko: mężczyzna
Telefon: [redacted]
Adres: [redacted]
[redacted]

Prawo autorskie i licencje

Do wszelkich pozyskiwanych i wytwarzanych danych należy wskazać właścicieli praw autorskich i praw własności intelektualnej. Więcej:

<https://www.uj.edu.pl/documents/1587933/6bc2fafc-cd80-4ec1-896d-858bd181ab27>).

Trzeba określić czy istnieją jakiekolwiek 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.

*Science Europe (2018). Science Europe Practical Guide to the International Alignment of Research Data Management. https://www.scienceurope.org/media/jezkhnno/se_rdm_practical_guide_final.pdf (dostęp: 25.10.2019).

Prawo autorskie i licencje

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.

„Creative Commons oferuje twórcom licencje, które pozwalają im zachować własne prawa i jednocześnie dzielić się swoją twórczością z innymi.”

Prawo autorskie i licencje

LICENCJE CREATIVE COMMONS



Uznanie autorstwa 4.0

Przystępne podsumowanie

Tekst licencji



Uznanie autorstwa-Na tych samych warunkach 4.0

Przystępne podsumowanie

Tekst licencji



Uznanie autorstwa-Użycie niekomercyjne 3.0 Polska

Przystępne podsumowanie

Tekst licencji



Uznanie autorstwa-Użycie niekomercyjne-Na tych samych warunkach 4.0

Przystępne podsumowanie

Tekst licencji



Uznanie autorstwa-Bez utworów zależnych 4.0

Przystępne podsumowanie

Tekst licencji



Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 4.0

Przystępne podsumowanie

Tekst licencji

Prawo autorskie i licencje



Open Data Commons Open Database License (ODbL)

Public Domain Dedication and License (PDDL)

Open Data Commons Attribution License (ODC-BY)

Prawo autorskie i licencje



General Public License (GNU GPL)

Lesser General Public License (GNU LGPL)

<https://www.gnu.org/licenses/gpl-3.0.html>

<https://www.gnu.org/licenses/lgpl-3.0.html>

Udostępnianie i długotrwałe przechowywanie danych

Dane powinny być tak otwarte, jak to możliwe i na tyle zamknięte, na ile to jest konieczne.

Zapewnienie dostępu do danych badawczych polega na ich udostępnieniu oraz opisanu. 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), czy udostępnienie danych wymaga zgody uczestników badań.

Ponowne użycie danych – zapewnienie trwałych identyfikatorów np. DOI, ORCID

Przykład: Ludzkie białko policystyn-1 ma unikatowy na całym świecie identyfikator podany przez bazę danych UniProt: <http://www.uniprot.org/uniprot/P98161>

*Science Europe (2018). Science Europe Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf (dostęp: 25.10.2019).

*FAIR Principles. <https://www.go-fair.org/fair-principles> (dostęp: 25.10.2019).

Udostępnianie i długotrwałe przechowywanie danych

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.

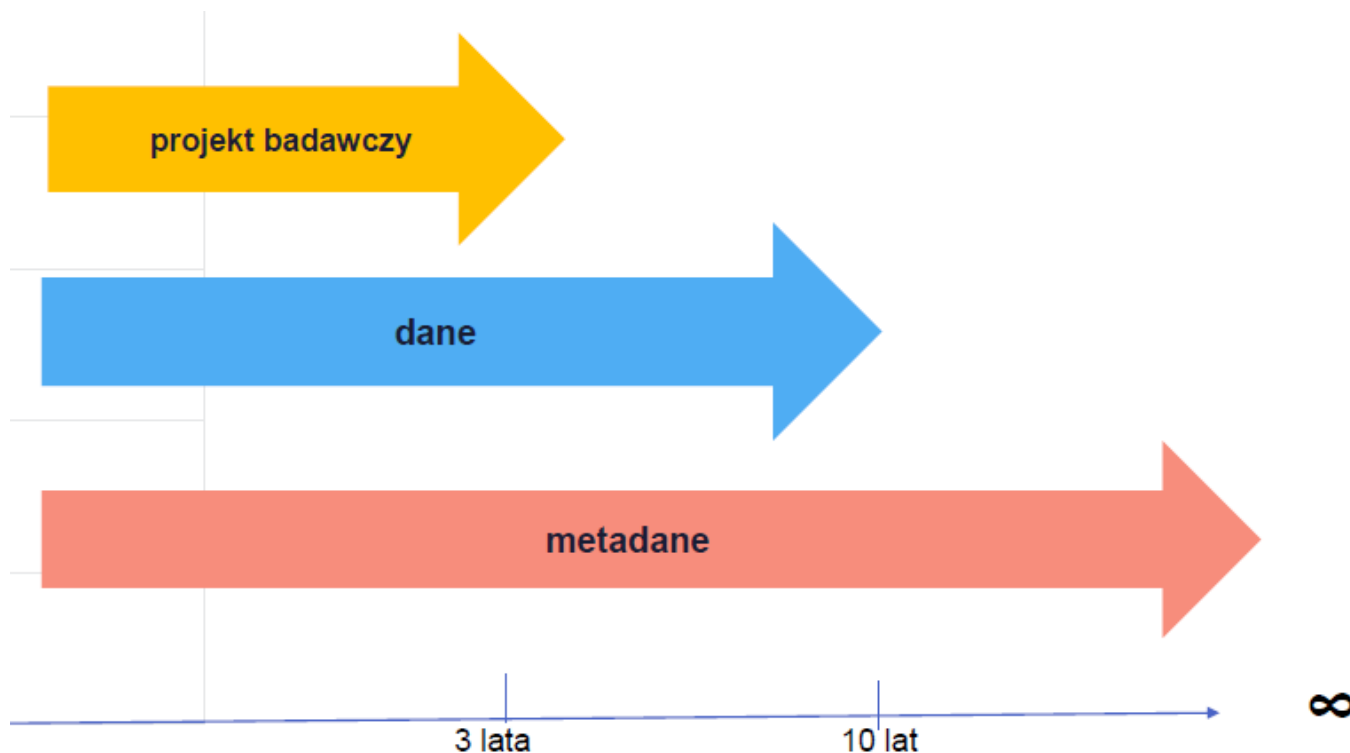
Uniwersytet Jagielloński zarządza RODBUK (<https://rodbuk.pl>), które zapewnia długotrwałe archiwizowanie zdeponowanych danych na serwerach zarządzanych przez CYFRONET AGH. Bezpieczeństwo danych jest także zapewnione poprzez wykonywanie regularnych kopii zapasowych.

*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 (dostęp: 25.10.2019).

*FAIR Principles. <https://www.go-fair.org/fair-principles> (dostęp: 25.10.2019).

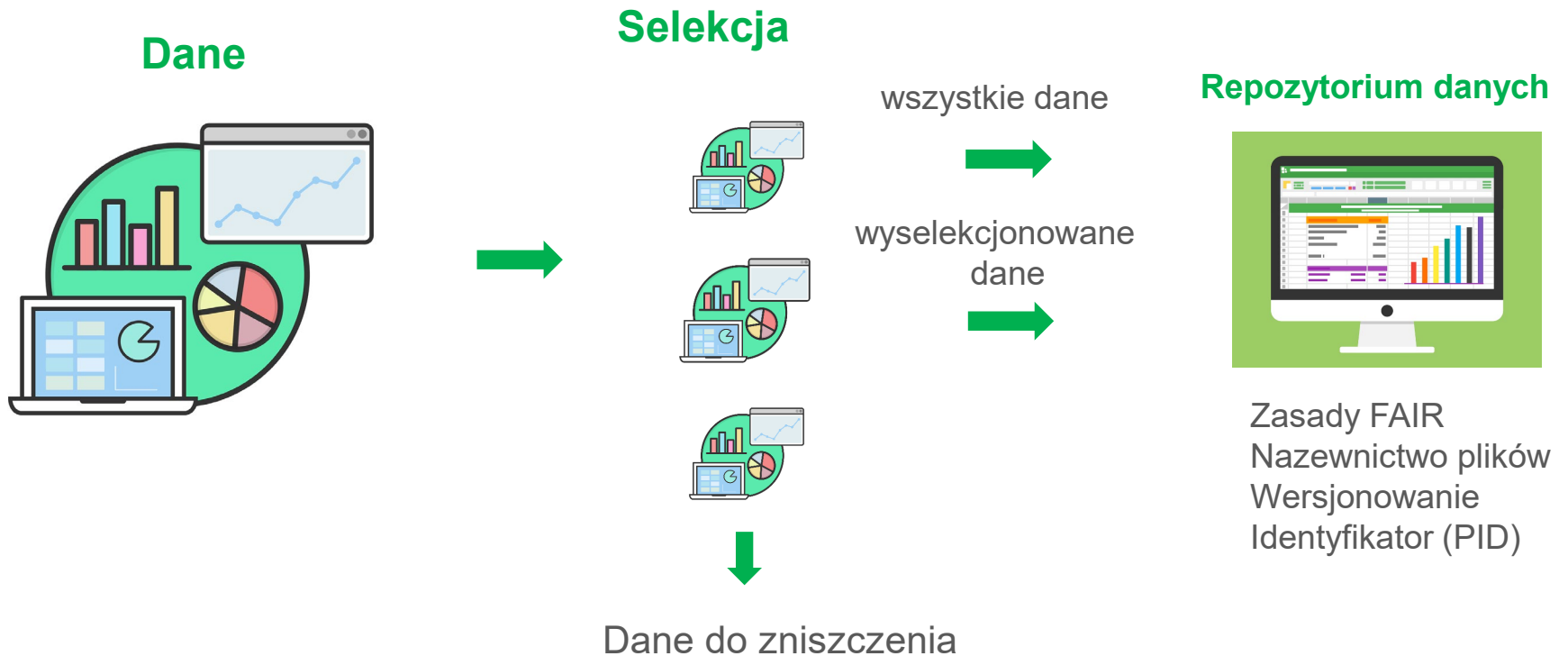
Udostępnianie i długotrwałe przechowywanie danych

Minimalne okresy przechowywania danych i metadanych



Udostępnianie i długotrwałe przechowywanie danych

Selekcja danych przeznaczonych do długoterminowej archiwizacji



Zadania związane z zarządzaniem danymi badawczymi

Opiekun danych - osoba lub instytucja, która docelowo zajmie się zarządzaniem danymi w dłuższym okresie czasu.

Przekazanie danych do repozytorium - jakość danych, przechowywanie i tworzenie kopii zapasowych, długoterminowa archiwizacja, udostępnianie danych oraz odpowiedzialność za zarządzanie danymi (np. instytucja, osoba).

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.

*Science Europe (2018). Science Europe Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf (dostęp: 25.10.2019).

*FAIR Principles. <https://www.go-fair.org/fair-principles> (dostęp: 25.10.2019).

Pismo Dyrektora NCN dotyczące planu zarządzania danymi

Zgodnie z wytycznymi Narodowego Centrum Nauki plan zarządzania danymi powinien być przygotowany w skróconej formie, uzupełniony na etapie składania raportu końcowego i po zakończeniu realizacji projektu zostanie oceniony przez instytucje finansującą.*

Rozliczenie DMP

- DMP może zmieniać się w trakcie projektu
- Na koniec projektu należy napisać raport końcowy
- DMP ocenia Zespół Ekspertów
- Brak kompletnego planu lub błędy spowodują wezwania do wyjaśnienia

*Narodowe Centrum Nauki (2019).

https://ncn.gov.pl/sites/default/files/pliki/2019_04_03_pismo_dyrektora_NCN_zarzadzanie_danymi_naukowymi.pdf

Polityka NCN dotycząca udostępniania danych

2 % kosztów pośrednich może być przeznaczona na opłaty związane z publikowaniem i archiwizacją danych badawczych w trakcie i po zakończeniu projektu.

Należy dążyć do upublicznienia danych w repozytoriach na licencjach Creative Commons, domyślnie CC0 (CC Public Domain) **lub** CC BY.



POMOCNE NARZĘDZIA W TWORZENIU DMP



Science Europe – przewodnik w wersji angielskiej z przykładowym szablonem planu zarządzania danymi

https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

*Science Europe (2018). Science Europe Practical Guide to the International Alignment of Research Data Management. https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf (dostęp: 25.10.2019).

Digital Curation Centre



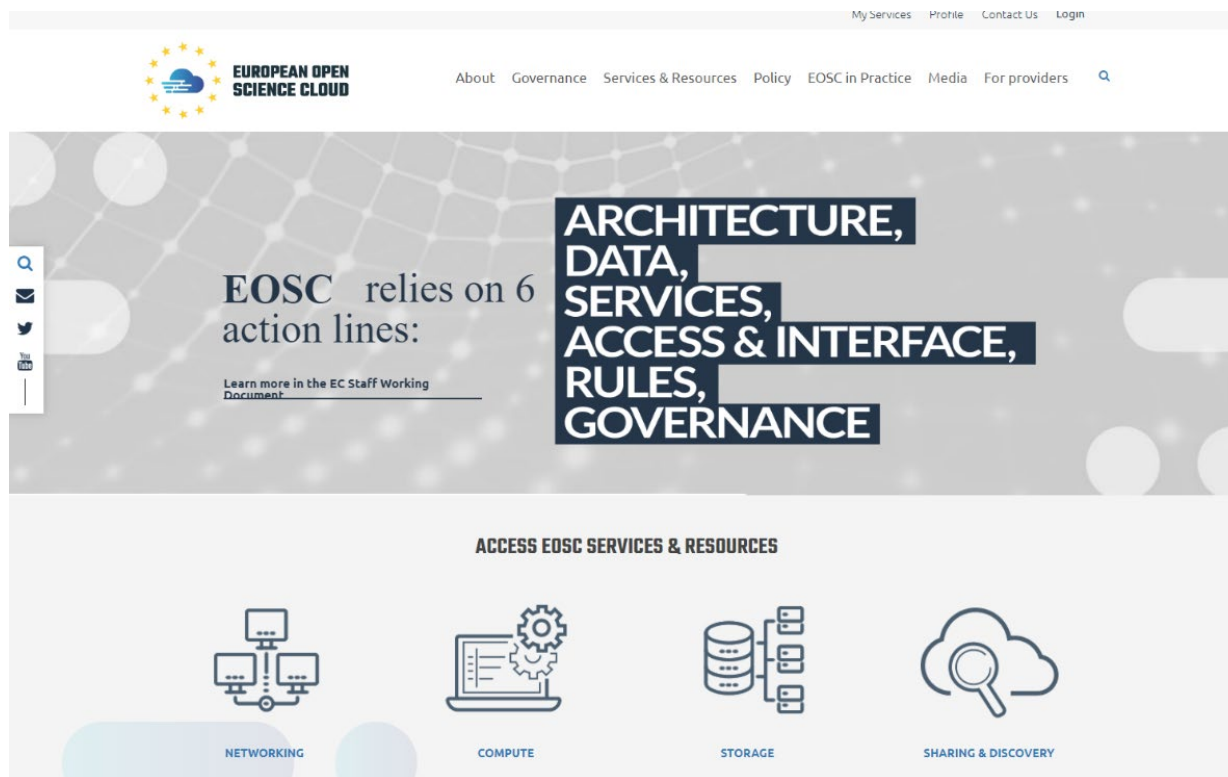
- Szablon planu zarządzania danymi z pomocniczym opisem i możliwością wyeksportowania uzupełnionego dokumentu w różnych formatach m. in. docx, pdf.
- Przykłady planów zarządzania danymi.

<https://dmponline.dcc.ac.uk>

AGREGATORY I REPOZYTORIA DANYCH BADAWCZYCH

European Open Science Cloud (EOSC)

W 2016 r. Komisja Europejska zaproponowała inicjatywę European Open Science Cloud (EOSC) w celu zbudowania konkurencyjnej gospodarki opartej na danych i wiedzy w Europie. Jest to katalog zasobów, narzędzi i usług dla otwartej nauki.





<https://www.re3data.org>

Agregatory i wyszukiwarki danych badawczych

- OpenAIRE Explore: <https://explore.openaire.eu>
- Google Dataset Search: <https://datasetsearch.research.google.com>
- Harvard Dataverse: <https://dataverse.harvard.edu>
- Mendeley Data: <https://data.mendeley.com>

RODBUK – Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich – <https://rodbuk.pl>

RODBUK

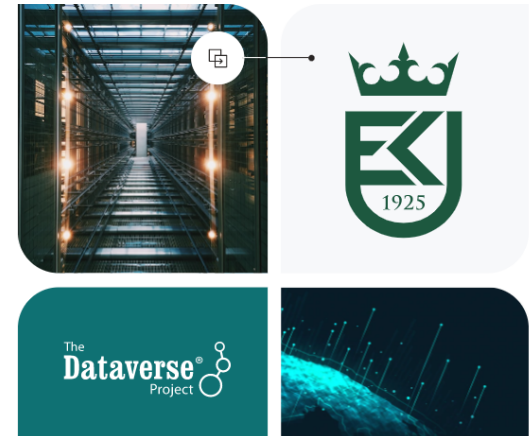
[User Guide](#) [Support](#) [English ▾](#)

RODBUK Cracow Open Research Data Repository

RODBUK was founded on the initiative of Board of Directors of the Cracow Library Group.

RODBUK is co-created by six Cracow universities:

- AGH University of Science and Technology,
- University of Physical Education in Krakow,
- Cracow University of Technology,
- Krakow University of Economics,
- Jagiellonian University in Kraków,
- Pedagogical University of Krakow.



Search

[View all data >](#)

Deposit data in your Institutional Collection:



AGH University
of Science
and Technology
in Krakow



Jagiellonian
University
in Kraków



Cracow University
of Technology



Krakow University
of Economics



Pedagogical
University
of Krakow



University
of Physical
Education
in Krakow

RODBUK umożliwia:

- Zapoznanie się z badaniami prowadzonymi w krakowskich ośrodkach naukowych,
- Przechowywanie i archiwizację różnego rodzaju zbiorów danych badawczych,
- Szybkie dotarcie do informacji poprzez przyjazny interfejs wyszukiwawczy,
- Najwyższy standard zapisu metadanych w formacie Dublin Core,
- Uzyskanie stałego identyfikatora DOI dla każdego zbioru danych,
- Znormalizowane cytowanie posiadanych danych,
- Udostępnianie danych na licencjach Creative Commons lub innych otwartych licencjach (licencje do baz danych, licencje do programów komputerowych) - <https://docs.cyfronet.pl/pages/viewpage.action?pageId=130288524>,
- Przechowywanie danych w ograniczonym dostępie,

DATA STEWARDS

DATA STEWARDS NA UNIWERSYTECIE JAGIELLOŃSKIM

- Szkolenia we współpracy z Centrum Wsparcia Nauki dotyczące przygotowania i rozliczania Planów Zarządzania Danymi
- Wsparcie naukowców w przygotowaniu i rozliczaniu Planów Zarządzania Danymi
- Rozwój i zarządzanie RODBUK
- Zarządzanie danymi badawczymi w RODBUK
- Przygotowywanie instrukcji i poradników

OTWARTA NAUKA NA UJ - <https://pod.uj.edu.pl>

Otwarta nauka na UJ
Uniwersytet Jagielloński

EN 



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

[STRONA GŁÓWNA](#) [AKTUALNOŚCI](#) [POLITYKA OTWARTOŚCI](#) [JAK PUBLIKOWAĆ](#) [SZKOLENIA](#) [FAQ](#) [KONTAKT](#)



DANE BADAWCZE

UNIWERSYTET JAGIELLOŃSKI > OTWARTA NAUKA NA UJ > STRONA GŁÓWNA

OTWARTA NAUKA NA UJ

<https://pod.uj.edu.pl/kursy-navoica>



[NAVOICA](#) to ogólnopolska platforma edukacyjna, należąca do Ministerstwa Edukacji i Nauki, którą rozwija Ośrodek Przetwarzania Informacji - Państwowy Instytut Badawczy (OPI PIB).

[Zarządzanie danymi badawczymi dla naukowców - kurs podstawowy](#)

[Zarządzanie danymi badawczymi dla naukowców - kurs średnio zaawansowany](#)

DZIĘKUJĘ ZA UWAGĘ

dr Leszek Szafrąński
l.szafranski@uj.edu.pl

16 listopada 2023 r.