



Zarządzanie danymi badawczymi w naukach o komunikacji społecznej i mediach

dr Leszek Szafrński
Uniwersytet Jagielloński
Biblioteka Jagiellońska

Zadanie realizowane przez Narodowe Centrum Nauki na podstawie
zlecenia Ministra Edukacji i Nauki dot. krajowej koordynacji partnerstwa
European Open Science Cloud w latach 2022-2023.



NARODOWE CENTRUM NAUKI



Ministerstwo
Edukacji i Nauki

Dlaczego zarządzanie danymi badawczymi jest ważne w naukach o komunikacji społecznej i mediach

1. Poprawa jakości badań
2. Oszczędność czasu i kosztów badań
3. Umożliwienie ponownego użycia danych
4. Archiwizacja pozwalająca na zachowanie danych w dłuższym okresie czasu
5. Umożliwienie weryfikacji i kontroli procesu badawczego

Charakterystyka danych w naukach o komunikacji społecznej i mediach

Zarządzanie informacją, bibliologia, informatologia, public relations, marketing polityczny, analiza zawartości mediów cyfrowych i tradycyjnych, kognitywistyka, genealogia dziennikarska, komunikacja: naukowa, społeczna, międzynarodowa, polityczna, wizualna.

1. Dane tekstowe (cyfrowe i tradycyjne)
2. Dane liczbowe, statystyczne
3. Kwestionariusze ankiet
4. Zdjęcia, wywiady, nagrania audio i video, transkrypcje nagrań
5. Oprogramowanie

Zasady FAIR

Findable: odnajdywalne

Dane powinny być możliwe do odnalezienia (opisane szczegółowymi metadanymi, mają przypisany unikalny i trwały identyfikator, zdeponowane w repozytorium).

Accessible: dostępne

Dane można odszukać za pomocą trwałej ścieżki dostępu, identyfikatora i innych informacji zawartych w metadanych np. słów kluczowych, opisu, powinny być możliwe do przeglądania i pobierania.

Interoperable: interoperacyjne

Metadane i pliki z danymi są zapisane w obowiązujących na świecie standardach umożliwiających wymianę informacji pomiędzy systemami, zawierają informacje uzupełniające, odniesienia i powiązania z innymi danymi, są odpowiednio cytowane.

Reusable: możliwe do ponownego wykorzystania

Dane udostępnione są na licencji, która pozwala na ich wykorzystanie, jasno opisane jest pochodzenie danych, są zapisane w trwałych i popularnych formatach, opisane są metadanymi pozwalającymi na wyszukiwanie i dokładne rozumienie danych.

Zasady otwierania danych badawczych

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

[jęz. ang.: as open as possible, as closed as necessary]

Plan Zarządzania Danymi

Ogólne Informacje

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o komunikacji społecznej i mediach?

- Opisuje zarządzanie danymi na każdym etapie cyklu badawczego,
- Ułatwia zarządzanie danymi badawczymi w projekcie,
- Pomaga monitorować postępy w realizacji badań,
- Wzmacnia kontrolę jakości danych na każdym etapie procesu badawczego,
- Wzmacnia kontrolę dostępu do danych,

Dlaczego Plan Zarządzania Danymi jest ważny w naukach o komunikacji społecznej i mediach?

- Wspiera zabezpieczenie danych przed utratą lub nieuprawnionym wykorzystaniem,
- Wspiera zabezpieczenie dostępu do wrażliwych danych,
- Ułatwia współpracę pomiędzy naukowcami pracującymi w danym projekcie,
- Pomaga zorganizować i skoordynować realizację badania,
- Wspiera zorganizowanie długoterminowej archiwizacji danych badawczych

Opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych

Opis danych

Sposób pozyskiwania lub wytwarzania nowych danych lub ponownego wykorzystywania danych istniejących.

Dane pierwotne (nowe):

- sposób pozyskiwania danych: wywiad, ankieta, kwerenda, obserwacja, eksperyment, dygitalizacja, analiza zawartości, analiza dokumentów zastanych.
- oprogramowanie do pozyskiwania danych: narzędzia do: OCR, transkrypcji, NLP, eye tracking, CAST (Systemu do analizy zawartości programów telewizyjnych).

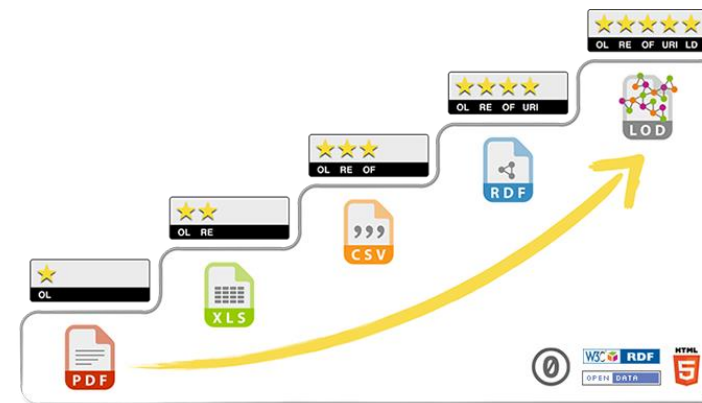
Dane wtórne (istniejące wcześniej):

- własność danych: zależna od umów,
- sposób udokumentowania pochodzenia danych: służą do tego metadane,
- kontrola jakości i spójności danych,
- digitalizacja danych analogowych.

Opis danych

Jakie dane będą pozyskiwane lub wytwarzane w projekcie?

- rodzaje danych: nagrania audio i video, teksty, notatki, oprogramowanie, dane liczbowe, zdjęcia, grafiki,
- format danych: (**formaty otwarte**: csv, odt, ods, odp, rtf, txt, html, xml (pliki tekstowe), png (pliki graficzne), flac (pliki dźwiękowe); **formaty zamknięte**: doc, docx (pliki tekstowe), pdf (pliki tekstowo-graficzne), tiff (pliki graficzne), xls (pliki tekstowe, bazy danych), mp3 (pliki dźwiękowe))
- format a wybór oprogramowania (oprogramowanie zamknięte a otwarte)
- szacunkowa objętość danych wytworzonych, pozyskanych, ponownie wykorzystanych.



Dokumentacja i jakość danych

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o komunikacji społecznej i mediach

- międzynarodowe standardy lub schematy służące do organizacji metadanych w dyscyplinie – np. Dublin Core,
- zakres metadanych:

Tytuł * ?	<input type="text" value="Wprowadź tytuł..."/>	
	<input type="button" value="Dodaj „Dane replikacji dla” do tytułu"/>	
Wersja językowa tytułu ?	<input type="text"/>	
Autor/Autorzy * ?	Nazwisko i imię lub nazwa organizacji * ? <input type="text" value="Szafrąński, Leszek"/>	Jednostka ? <input type="text" value="Jagiellonian University in Kraków"/>
		<input type="button" value="+"/> Schemat identyfikowania ? <input type="text" value="Wybierz..."/>
Identyfikator ?	<input type="text"/>	
Współtwórca ?	Nazwisko i imię lub nazwa Współtwórcy ? <input type="text" value="Nazwisko, imię lub organizacja"/>	Jednostka ? <input type="text"/>
		<input type="button" value="+"/> Schemat identyfikowania ? <input type="text" value="Wybierz..."/>
Identyfikator ?	<input type="text"/>	
Kontakt * ?	Osoba do kontaktu * ? <input type="text" value="Szafrąński, Leszek"/>	Jednostka * ? <input type="text" value="Jagiellonian University in Kraków"/>
		<input type="button" value="+"/> E-mail * ? <input type="text" value="l.szafranski@uj.edu.pl"/>

Opis * ?	To pole obsługuje tylko niektóre tagi HTML.	<input type="button" value="+"/> Tekst * ?
		<input type="text"/>
Abstrakt ?		<input type="button" value="+"/> Streszczenie projektu ?
		<input type="text"/>
Słowa kluczowe * ?		<input type="button" value="+"/> Termin * ?
		<input type="text"/>
Dziedzina nauki zgodnie z MEIN * ?		<input type="button" value="+"/> Dziedzina/Dyscyplina * ?
		<input type="text" value="Wybierz..."/>
Dziedzina nauki w OECD * ?		<input type="button" value="+"/> Dziedzina/Dyscyplina * ?
		<input type="text" value="Wybierz..."/>

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o komunikacji społecznej i mediach

- zakres metadanych:

Powiązane publikacje ?

Typ powiązania ?

Wybierz...

Autor/Autorzy ?

+

Tytuł powiązanej publikacji ?

Typ identyfikatora ?

Wybierz...

Identyfikator ?

URL ?

Wpisz kompletny URL, zaczynając od ht

Powiązany zbiór danych ?

Typ powiązania ?

Wybierz...

Autor/Autorzy ?

+

Tytuł ?

Typ identyfikatora ?

Wybierz...

Identyfikator ?

URL ?

Uwagi ?

Język *

Wybierz język *

Wybierz...

+

Producent ?

Nazwisko i imię lub nazwa organizacji ?

Nazwisko, imię lub organizacja

Jednostka ?

+

Skrót ?

URL ?

Wpisz kompletny URL, zaczynając od ht

Finansowanie ?

Jednostka finansująca ?

Wybierz...

Nazwa grantu ?

+

Numer grantu ?

ROR instytucji finansującej ?

Deponent ?

Szafrąński, Leszek

Data zdeponowania ?

2023-06-12

Okres gromadzenia danych ?

Początek ?

YYYY-MM-DD

Koniec ?

YYYY-MM-DD

+

Rodzaj danych w zbiorze * ?

Rodzaj danych w zbiorze * ?

Wybierz...

+

Dokumentacja i jakość

Metadane i dokumentacja dot. danych w naukach o komunikacji społecznej i mediach

- zakres metadanych:


Licencje 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

Pliki z ograniczeniami + warunki dostępu [▲](#)

Popros o dostęp [?](#) Włącz prośbę o dostęp

Warunki dostępu [?](#)

Miejsce dostępu do danych [?](#)

Oryginalne archiwum [?](#)

Stan dostępności [?](#)

Kontakt w sprawie dostępu [?](#)

Rozmiar kolekcji [?](#)

Zakończenie badania [?](#)

Dokumentacja i jakość

Środki kontroli jakości danych

Nauki o komunikacji społecznej i mediach

- Metody jakościowe i ilościowe,
- Weryfikacja danych na etapie kodowania danych (2 lub więcej koderów),
- Prowadzenie badań w terenie,
- Metody weryfikacji jakości danych (znormalizowane przechwytywanie danych, sprawdzanie poprawności wprowadzania danych, wzajemna ocena danych lub znormalizowane protokoły gromadzenia danych).

Przechowywanie i tworzenie kopii zapasowych podczas badań

Przechowywanie i kopie zapasowe

Przechowywanie danych i metadanych – gdzie będą przechowywane i zabezpieczane dane w trakcie procesu badawczego?

- **Zasada 3-2-1**
- Zaleca się przechowywanie danych w centralnych systemach gromadzenia informacji na macierzystej uczelni.
- Przechowywanie danych w repozytoriach np. RODBUK (Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich) <https://rodbuk.pl>
- Przechowywanie danych tylko w pamięciach masowych, komputerach przenośnych itp. może spowodować ich nieodwracalną utratę.

Przechowywanie i kopie zapasowe

W jaki sposób zostanie zapewnione bezpieczeństwo i ochrona danych wrażliwych w okresie trwania projektu?

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.

Przykład:

Uniwersytet Jagielloński wdrożył politykę ochrony danych wrażliwych oraz powołał Inspektora Ochrony Danych. Pomaga on także w opracowaniu planu zarządzania danymi pod kątem gromadzenia i przechowywania danych osobowych i wrażliwych. Zobacz: <https://iod.uj.edu.pl>.

Wymogi prawne, kodeksy postępowania

Wymogi prawne, kodeksy postępowania

Jeżeli będzie miało miejsce przetwarzanie danych osobowych, w jaki sposób zostanie zapewniona zgodność z przepisami dotyczącymi danych osobowych oraz ich ochrony?

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:
Adres:

Wymogi prawne, kodeksy postępowania

W jaki sposób zostanie zapewniona zgodność z innymi przepisami, takimi jak prawa własności intelektualnej i prawa własności? Jakie przepisy mogą znaleźć zastosowanie w przypadku omawianej dyscyplin/dziedziny?

- Do wszelkich pozyskiwanych i wytwarzanych danych należy wskazać właścicieli praw własności intelektualnej (więcej, przykład na UJ: Regulamin zarządzania własnością intelektualną oraz zasad komercjalizacji w Uniwersytecie Jagiellońskim - <https://bip.uj.edu.pl/dokumenty/regulamin-zarzadzania-wlasnoscia-intelektualna-oraz-zasad-komercjalizacji-w-uj> <https://www.uj.edu.pl/documents/1587933/6bc2fafc-cd80-4ec1-896d-858bd181ab27>). Trzeba określić czy istnieją jakiegokolwiek ograniczenia prawne dotyczące ponownego wykorzystania danych pochodzących od osób trzecich.
- Należy 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.

Wymogi prawne, kodeksy postępowania

- Prawo autorskie – teksty, publikacje, prezentacje, normy wykorzystywane jako dane badawcze (wyjątek: prawo cytatu: art. 29 - Ustawa z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (Dz.U.2022.2509))
- Prawo własności intelektualnej – przykład na UJ: Załącznik do uchwały nr 62/IV/2021 Senatu UJ z dnia 28 kwietnia 2021 roku: Regulamin zarządzania własnością intelektualną oraz zasad komercjalizacji w Uniwersytecie Jagiellońskim - <https://bip.uj.edu.pl/dokumenty/regulamin-zarzadzania-wlasnoscia-intelektualna-oraz-zasad-komercjalizacji-w-uj>
- Ustawa o ochronie baz danych - Open Data Commons
- Umowy o grant i konkursy finansowane przez NCN regulują zasady zarządzanie i udostępnianie danych badawczych w otwartym dostępie
- Licencje Creative Commons, Licencje do programów komputerowych
- Regulamin korzystania z RODBUK - <https://rodbuk.pl/terms-of-use>

Wymogi prawne, kodeksy postępowania

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](#)

Licencje do baz danych

Open Data Commons Open Database License (ODbL)

Public Domain Dedication and License (PDDL)

Open Data Commons Attribution License (ODC-BY)

Licencje do programów komputerowych

General Public License (GNU GPL)

Lesser General Public License (GNU LGPL)

Udostępnianie i długotrwałe przechowywanie danych

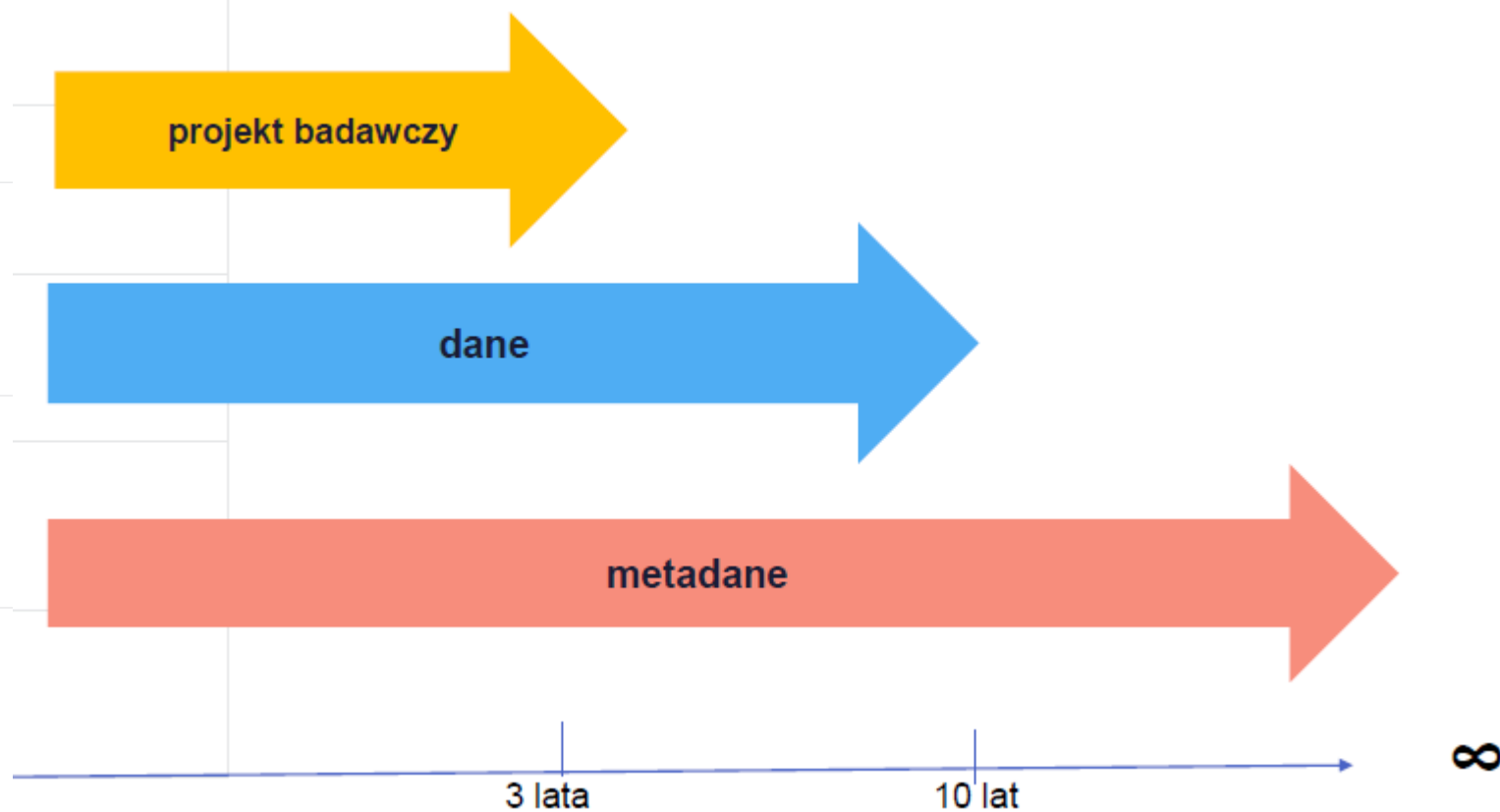
Udostępnianie i długotrwałe przechowywanie

Kiedy i w jaki sposób będą udostępniane dane z projektu? Czy istnieją ewentualne ograniczenia i zakazy dotyczące ich udostępniania?

- udostępnianie danych i metadanych – repozytorium danych badawczych,
- czas udostępnienia danych – w trakcie trwania, czy po zakończeniu badań, należy podać termin lub terminy udostępnienia oraz ograniczenia i przeszkody uniemożliwiające ich pełne lub częściowe udostępnienie,
- informacja na temat powiązania udostępnianych danych z publikacją – niektórzy wydawcy wdrażają obowiązek zdeponowania danych przed publikacją tekstu artykułu/monografii,
- zgoda uczestników badania – pisemna, jeżeli dane uczestników badania będą przechowywane lub udostępniane w otwartym dostępie.

Udostępnianie i długotrwałe przechowywanie

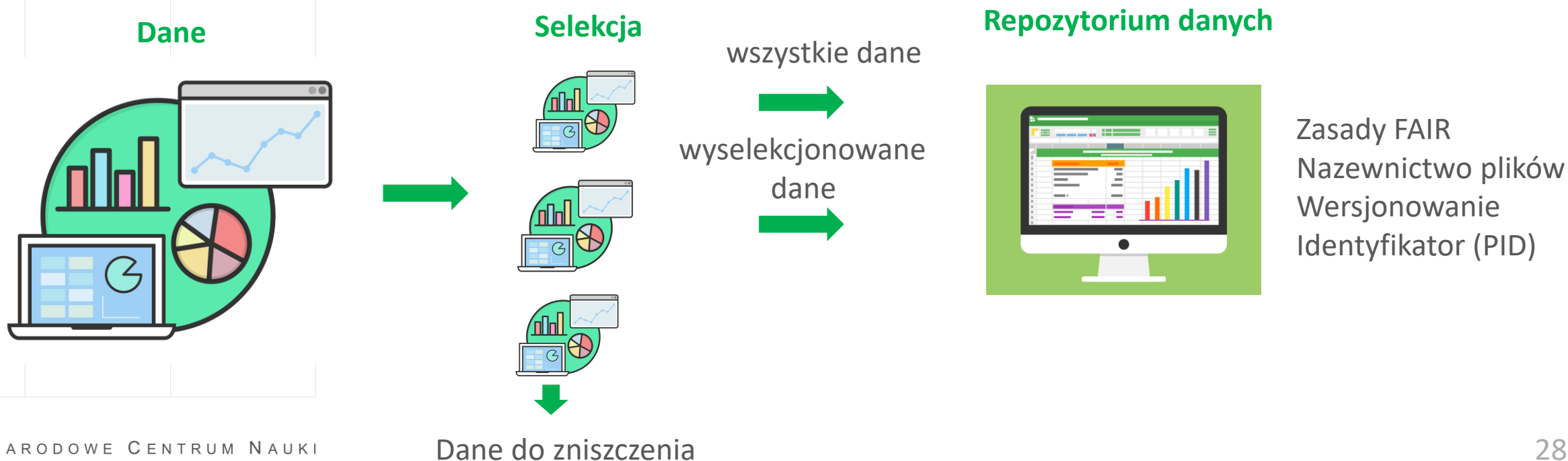
Minimalne okresy przechowywania danych i metadanych



Udostępnianie i długotrwałe przechowywanie

Jak będzie wyglądać selekcja danych przeznaczonych do utrwalenia i gdzie będą one długoterminowo przechowywane?

selekcja danych przeznaczonych do utrwalenia



Kryteria wyboru repozytorium

Typy repozytoriów

- repozytoria dziedzinowe vs ogólne
- repozytoria o wysokiej reputacji dla nauk o komunikacji społecznej i mediach?
- repozytoria posiadające certyfikat CoreTrust Seal

Wybór

- kwestie formalne wyboru (stosowanie FAIR, PID, licencja, czas przechowywania, embargo, rozmiar plików, inne)
- kwestie merytoryczne: deponowanie danych w repozytorium wpływa na widzialność, cytowalność, popularyzację dorobku naukowego, indeksowanie itp.

Rejestry repozytoriów

- Registry of Research Data Repositories [re3data.org]

Kryteria wyboru repozytorium

European Open Science Cloud (EOSC)

The screenshot shows the homepage of the European Open Science Cloud (EOSC). At the top, there is a navigation bar with links for "My Services", "Profile", "Contact Us", and "Login". Below this is the EOSC logo, which consists of a stylized globe with a cloud and the text "EUROPEAN OPEN SCIENCE CLOUD". To the right of the logo is a secondary navigation menu with links for "About", "Governance", "Services & Resources", "Policy", "EOSC in Practice", "Media", and "For providers".

The main content area features a large banner with a network-like background. On the left side of the banner, it says "EOSC relies on 6 action lines:" followed by a link to "Learn more in the EC Staff Working Document". On the right side of the banner, the text "ARCHITECTURE, DATA, SERVICES, ACCESS & INTERFACE, RULES, GOVERNANCE" is displayed in a dark, bold font.

Below the banner is a section titled "ACCESS EOSC SERVICES & RESOURCES" which contains four icons representing different services: "NETWORKING" (a network diagram), "COMPUTE" (a laptop with a gear), "STORAGE" (a server rack), and "SHARING & DISCOVERY" (a cloud with a magnifying glass).

Kryteria wyboru repozytorium

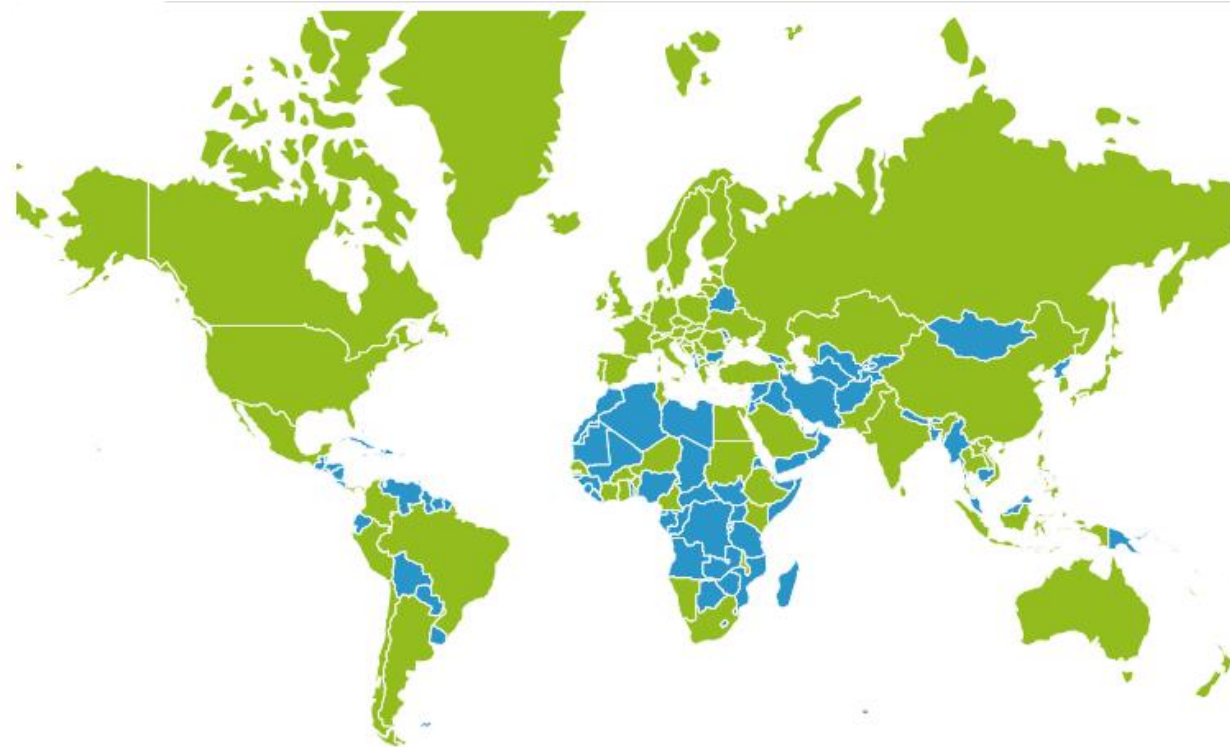


<https://www.re3data.org>

Kryteria wyboru repozytorium

REPOZYTORIA DANYCH BADAWCZYCH NA ŚWIECIE

4133 – zarejestrowane repozytoria



Kryteria wyboru repozytorium

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

RODBUK

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.

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

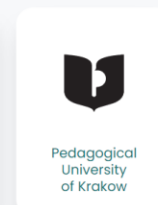
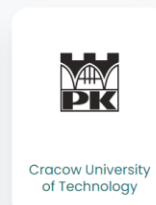


Search all dataverses...

Search

View all data >

Deposit data in your Institutional Collection:



Udostępnianie i długotrwałe przechowywanie

Jakie metody lub oprogramowanie umożliwiają dostęp do danych i korzystanie z danych?

Udostępnianie danych badawczych:

- otwarty dostęp do danych badawczych i metadanych w repozytorium,
- ograniczony dostęp do danych badawczych – metadane w repozytorium, dane badawcze udostępniane na żądanie.

Dane badawcze powinny być przekonwertowane do formatów otwartych przed udostępnieniem w repozytorium. Większość oprogramowania używanego do zapisu danych pozwala na konwersję do formatów edytowalnych.

Udostępnianie i długotrwałe przechowywanie

W jaki sposób zagwarantować stosowanie unikalnego i trwale przypisanego identyfikatora [ang. PID] dla każdego zbioru danych?

Przykład:

- cyfrowy identyfikator dokumentu elektronicznego DOI można uzyskać deponując dane badawcze w repozytorium danych np. RODBUK (Repozytorium Otwartych Danych Badawczych Uczelni Krakowskich)
<https://rodbuk.pl>

Zadania związane z zarządzaniem danymi oraz zasoby

Zadania związane z zarządzaniem danymi oraz zasoby

Kto będzie odpowiadał za zarządzanie danymi (tj. kto będzie ich opiekunem)?

Podział ról w projekcie badawczym – przykład UJ - obecnie:

- Wytwarzanie, gromadzenie i kontrola danych badawczych – naukowcy, pracownicy naukowo-techniczni,
- Kontrola metadanych i danych badawczych wprowadzanych przez naukowców do RODBUK - data stewardzi zatrudnieni w Bibliotece Jagiellońskiej w Oddziale Zbiorów Cyfrowych,
- Archiwizacja i długoterminowe zarządzanie danymi – data stewardzi zatrudnieni w Bibliotece Jagiellońskiej w Oddziale Zbiorów Cyfrowych we współpracy z informatykami z Cyfronet AGH,
- Plan Zarządzania Danymi (DMP) – przygotowanie, wersjonowanie i rozliczenie DMP - naukowcy we współpracy z data stewardami.

Zadania związane z zarządzaniem danymi oraz zasoby

Budżet na cele zarządzania danymi i zagwarantowanie przestrzegania zasad FAIR?

Przykład UJ:

- koszty utrzymania informatyków, pracowników zarządzających danymi w projektach naukowych i data stewardów,
- koszty utrzymania infrastruktury informatycznej (serwery, macierze dyskowe, oprogramowanie, RODBUK),
- koszty przechowywania i długoterminowej archiwizacji danych,
- obsługa prawna.

Wdrażanie i raportowanie Planów Zarządzania Danymi

Wdrażanie i raportowanie Planu Zarządzania Danymi

- DMP może zmieniać się w trakcie realizacji projektu (zalecane)
- Raport końcowy z wdrażania DMP – należy opisać stan faktyczny na koniec realizacji projektu: planowano vs. zrealizowano.
- W NCN, Raport końcowy z wdrażania DMP ocenia Zespół Ekspertów (aspekty merytoryczne) oraz Biuro NCN (aspekty formalne).
- Brak kompletnego raportu z realizacji DMP lub błędy formalne w raporcie lub udostępnionych zestawach danych spowodują wezwanie do złożenia wyjaśnień lub uzupełnień.
- Brak uzupełnień lub wyjaśnień dotyczących Raportu końcowego z realizacji DMP lub brak prawidłowo udostępnionych danych spowodują sankcje.



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



GRAMY DLA POLSKIEJ NAUKI





Narodowe Centrum Nauki Zespół ds. Otwartej Nauki otwarta.nauka@ncn.gov.pl



GRAMY DLA POLSKIEJ NAUKI

