



**PORTAL
ZARZĄDZANIA
WIEDZĄ UJ CM**

Projekt „Portal Zarządzania Wiedzą i Potencjałem Naukowym Uniwersytetu Jagiellońskiego – Collegium Medicum
jako moduł Polskiej Platformy Medycznej

Dane badawcze

DANE BADAWCZE

Dane badawcze (Research Data) to zarejestrowane materiały o charakterze faktograficznym (w postaci liczbowej, tekstowej, graficznej czy dźwiękowej), powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.

Dane badawcze dzielić można na:

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

Repozytorium i pozycjonowanie wydawnictw KUL. Poradnik - Plan zarządzania danymi badawczymi, s. 2

OTWARTE DANE BADAWCZE

- otwarte dane to dane, do których każdy ma dostęp; można je ponownie wykorzystywać, modyfikować i redystrybuować;
- dane pozyskiwane w procesach badawczych;
- specyfikacja dziedzinowa – niektóre obszary nauki nie mają dostępnych danych w postaci cyfrowych ect.;
- dane surowe - dane zebrane, ale jeszcze nieprzeanalizowane;
- era „wielkich danych” (ang. „big data”);
- dostęp do treści naukowych (np. w ramach licencji Creative Commons) pozwala na lepsze „zrozumienie” dołączonych do treści danych (kontekst, metodologia pozyskania);
- w programach grantowych, np. Horyzont 2020 wymagane jest stworzenie planu zarządzania danymi - **Data Management Plan**;
- firma Google uruchomiła nową wyszukiwarkę, która pozwala na wyszukiwanie zbiorów danych badawczych udostępnianych w tysiącach repozytoriów na świecie - **Dataset Search**.

Biblioteka Politechniki Gdańskiej

CECHY OTWARTOŚCI DANYCH PUBLICZNYCH

standardy otwartości danych

- **dostępne** - udostępnione bez żadnych ograniczeń szerokiemu gronu użytkowników do dowolnych celów,
- **upublicznione w wersji źródłowej** - dostępne w oryginalnej i niezmienionej formie,
- **kompletne**,
- **aktualne** - zachowana wartość danych,
- **odczytywalne maszynowo** - udostępnione w formatach do odczytu maszynowego, czyli CSV, XML, arkusz kalkulacyjny,
- **udostępnione niedyskryminująco** - dostępne dla każdego bez konieczności rejestracji, weryfikacji tożsamości – hasła i loginy,
- **dostępne bez ograniczeń prawnych** - dane niebędące przedmiotem praw autorskich, patentów, znaków towarowych lub tajemnicy handlowej, wykorzystywane w dowolnych celach bez konieczności ubiegania się o jakąkolwiek zgodę na ich używanie,
- **niezastrzeżone** - dostępne w formacie szeroko stosowanym, niekontrolowanym przez żaden podmiot.

Ministerstwo Cyfryzacji. Otwieranie danych. Podręcznik dobrych praktyk, 2018 s.10-11

SYSTEM 5 POZIOMÓW OTWARTOŚCI

- * udostępnianie danych w sieci Web (w dowolnym formacie) na warunkach otwartej licencji
- ** udostępnianie danych w formie ustrukturyzowanej (np. arkusz kalkulacyjny zamiast zeskanowanego obrazu tabeli)
- *** używanie formatów otwartych (np. CSV zamiast arkusza kalkulacyjnego)
- **** używanie URI do oznaczania zasobów, aby możliwe było ich wyszukiwanie
- ***** łączenie danych, aby zapewnić kontekst

Zapisywanie danych w jednolitym standardzie umożliwi porównywanie, analizę lub łączenie danych ze sobą, a także tworzenie na ich podstawie wizualizacji, produktów i usług.

Ministerstwo Cyfryzacji. Otwieranie danych. Podręcznik dobrych praktyk, 2018 s.13-14

OGRANICZENIA KORZYSTANIA Z DANYCH BADAWCZYCH (w internecie)

- finansowe - opłata za korzystanie,
- techniczne - formaty zamknięte lub wymagające specjalistycznego oprogramowania,
- prawne.

Strzelczyk E. Otwarte dane badawcze – kolejny krok do otwierania nauki, 2017, s. 4

KORZYŚCI OTWIERANIA DANYCH BADAWCZYCH

- obniżenie kosztów pracy naukowej (niepowielanie tych samych wysiłków),
- wzbogacenie danych tworzeniem nowego materiału badawczego,
- weryfikacja danych (wychwytywanie nierzetelności),
- współpraca interdyscyplinarna (nawiązywanie i rozwój współpracy między badaczami, instytucjami),
- zwiększenie liczby kanałów dostępu dla użytkownika (zwiększenie wyszukiwalności),
- szerszy krąg odbiorców wyników pracy badawczej,
- wzrost wartości marki,
- realizacja misji społecznej (dorobek ze środków publicznych dostępny dla publiczności),
- wykorzystanie danych przez podmioty gospodarcze,
- szybszy przepływ informacji (medycyna),
- bezpieczna archiwizacja.

Strzelczyk E. Otwarte dane badawcze – kolejny krok do otwierania nauki, 2017, s. 16-17

KONIECZNE ZMIANY

- uregulowania prawne na poziomie krajowym,
- wypracowanie odpowiedniego modelu prawnego,
- aprobowanie wiedzy jako dobra publicznego,
- jasne wytyczne dotyczące przechowywania i udostępniania danych (licencje, standardy metadanych, trwałe identyfikatory DOI),
- włączenie publikacji danych badawczych do systemu oceny dorobku naukowego,
- pokonywanie barier technicznych (automatyczne metody wyszukiwania i opracowania danych),
- wspieranie i zachęcanie naukowców do znajdowania oraz wielokrotnego użycia danych do ich dalszych badań.

Strzelczyk E. Otwarte dane badawcze – kolejny krok do otwierania nauki, 2017, s. 17-18

REPOZYTORIA DANYCH BADAWCZYCH INSTYTUCJI NAUKOWYCH

Clarín-PL – repozytorium lingwistyczne, Politechnika Wrocławska

RepOD – repozytorium danych badawczych prowadzone przez ICM UW.

e-Science – Politechnika Wrocławska

Azon – uczelnie wrocławskie

Narodowy Korpus Języka Polskiego – Uniwersytet Łódzki

Archiwum Danych Społecznych – UW, PAN Warszawa

Serwis „Uwolnij Naukę”

PLAN ZARZĄDZANIA DANYMI BADAWCZYMI

- Plan zarządzania danymi (Data Management Plan–DMP) określa, w jaki sposób dane badawcze mają być zarządzane zarówno w czasie trwania projektu badawczego, jak i po jego zakończeniu. Plan zarządzania danymi uznawany jest za część planu badań.
- Instytucje i programy finansujące badania naukowe, czasopisma coraz częściej wymagają od naukowców przedstawienia DMP na etapie składania i oceny wniosków grantowych czy przed publikacją artykułu (np. Horyzont 2020, NCN).

Repozytorium i pozycjonowanie wydawnictw KUL. Poradnik. Plan zarządzania danymi badawczymi, s. 2

PLAN ZARZĄDZANIA DANYMI

Plan zarządzania danymi powinien zawierać informacje dotyczące

▪ danych,

- jakie dane zostaną wytworzone lub zebrane (format i typ plików, liczba danych),
- jak zostaną uporządkowane i opisane (metodologia, standardy, metadane),
- kwestie etyczne i prawne (własność intelektualna, prawa autorskie, dane niejawne),
- sposób udostępniania danych (jak, kiedy, komu),
- które dane będą przechowywane długoterminowo (kwestia sposobu przechowywania i ochrony danych)

▪ licencji (aspekty prawne)

- Creative Commons – CC0, CC-BY, CC-BY-SA,
- Open Data Commons – licencje wyłącznie do baz danych – posiadają skomplikowane rozwiązania prawne; obejmują całą bazę danych, ale nie obejmują jej części składowych.

Repozytorium i pozycjonowanie wydawnictw KUL. Poradnik. Plan zarządzania danymi badawczymi, s. 3



**PORTAL
ZARZĄDZANIA
WIEDZĄ UJ CM**

otwarty.cm-uj.krakow.pl

Projekt „Portal Zarządzania Wiedzą i Potencjałem Naukowym Uniwersytetu Jagiellońskiego – Collegium Medicum
jako moduł Polskiej Platformy Medycznej

