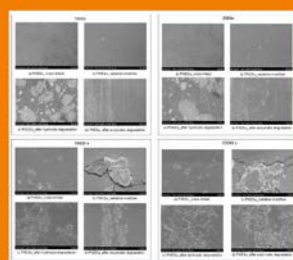
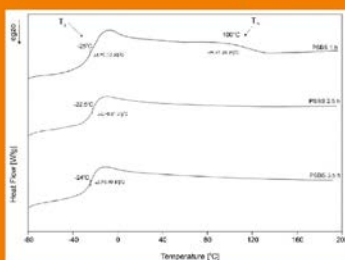
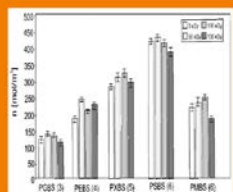
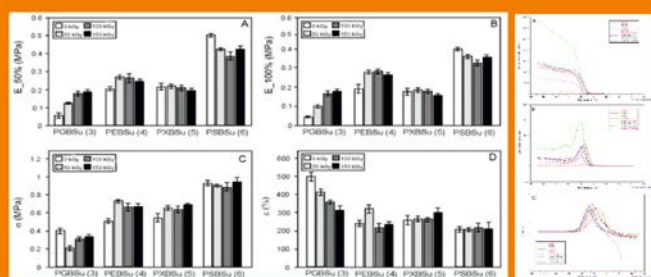


OTWARTE DANE BADAWCZE (OPEN RESEARCH DATA)

Repozytorium otwartych danych badawczych



OTWARTE

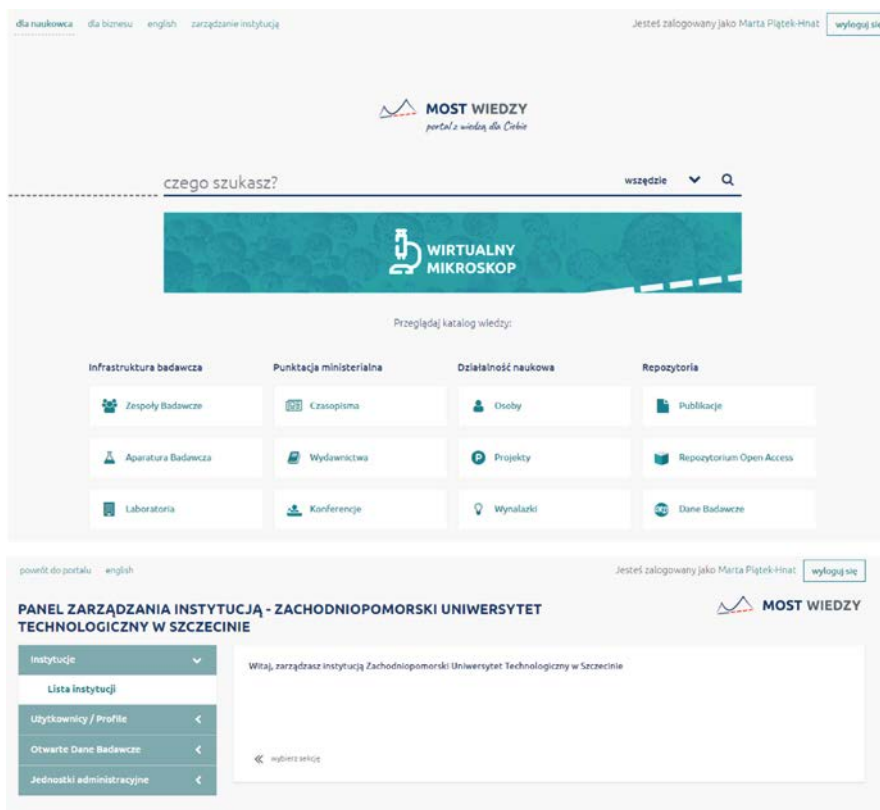


DANE
BADAWCZE

W dniach 24-30 października 2022 obchodzony był Międzynarodowy Tydzień Otwartego Dostępu. Poprzez wiele inicjatyw „ruch” Open Access propaguje publikowanie artykułów w otwartym dostępie oraz udostępnianie danych badawczych wytworzonych w ramach działalności naukowców na całym świecie.

Ważnym narzędziem informatycznym służącym do deponowania i udostępniania w Internecie zarówno artykułów, jak i danych badawczych są repozytoria. Bieżący dorobek instytucji naukowych zamieszczany jest w repozytoriach instytucjonalnych. I w tym miejscu należy przypomnieć, że Zachodniopomorski Uniwersytet Technologiczny w Szczecinie posiada własne repozytorium instytucjonalne <https://oa.zut.edu.pl/>, które pozwala na deponowanie artykułów naukowych, prac dydaktycznych, materiałów konferencyjnych, w tym posterów, prac doktorskich i innych. Zbiory Repozytorium ZUT znajdują się m. in. w Agregatorze Centrum Otwartej Nauki Ceon <http://agregator.ceon.pl/>, oraz jest ono indeksowane w Directory of Open Access Repositories OpenDOAR <https://v2.sherpa.ac.uk/id/repository/10558> jako jedno z 5600 otwartych repozytoriów. Niewątpliwą zaletą deponowania swoich prac naukowych w repozytorium otwartym jest to, że stają się one bardziej widoczne. Poza tym istnieje większa szansa na wzrost cytowań i promocji swoich badań na świecie. Oprócz repozytoriów, w których deponowane są artykuły naukowe, bardzo duże znaczenie mają repozytoria otwartych danych badawczych. Wiąże się to z wymogami deponowania wyników badań naukowych, jakie stawiają grantodawcy (Narodowe Centrum Nauki, Komisja Europejska) czy coraz częściej wydawnictwa. Związane to jest również z publicznym i nieograniczonym dostępem do wyników badań naukowych.

Wychodząc naprzeciw tym wymaganiom, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie przystąpił do projektu wdrożeniowego Repozytorium Danych Badawczych „Most Danych”, którego twórcą jest Politechnika Gdańska, na naszej Uczelni. Już w najbliższym czasie po dostosowaniu platformy i po migracji danych będą Państwo poinformowani o możliwościach tego repozytorium. Warto w tym miejscu nadmienić, że dane przechowywane w przedmiotowym repozytorium są indeksowane w Data Citation Index, który jest częścią międzynarodowej bazy Web of Science. W ten sposób dorobek naukowy autora danych jest rozpowszechniany, a on sam staje się bardziej rozpoznawalny w środowisku naukowym rozpatrywanym w skali globalnej.



Pisząc o danych badawczych można zadać sobie pytanie: **Czym tak naprawdę one są ?** **Dane badawcze** to wszelkie dane, które zostały zebrane, zaobserwowane bądź stworzone podczas procesu badawczego, mającego na celu otrzymanie oryginalnych wyników naukowych. W zależności od tego w jaki sposób bądź w jakich celach powstały, możemy wyróżnić m.in. dane obserwacyjne, eksperymentalne, symulacje, kompilacyjne, czy też dane referencyjne. Każda dyscyplina nauki tworzy własne, specyficzne dla niej dane badawcze. W biologii mogą to być filmy dokumentujące zachowania zwierząt, w inżynierii lądowej i ochronie środowiska dane geograficzne lub przestrzenne, a w przypadku historii dokumenty archiwalne. Do danych badawczych zaliczymy również wszelkie opisy procedur, dzienniki laboratoryjne czy notatki z eksperymentów. Aby dane badawcze mogły zostać określone jako otwarte, należy je udostępnić w taki sposób, aby każdy miał do nich nieograniczony dostęp oraz mógł je dowolnie wykorzystywać, modyfikować i upowszechniać.

Korzyści z szerokiego udostępniania danych badawczych są następujące:

- Lepsza komunikacja i wymiana informacji pomiędzy specjalistami, reprezentującymi różne dyscypliny nauki na całym świecie

- Możliwość przeprowadzania analiz opartych o unikalne dane, których z różnych względów nie można ponownie zebrać.
- Wzrost liczby cytowań zarówno samych danych, jak i publikacji, które są na nich oparte.
- Możliwość oceny rzetelności prowadzonych badań.
- Otwarty dostęp pozwala wykorzystać istniejące zasoby, a tym samym obniżyć koszty badań.

Ogólna zasada dotycząca udostępniania danych badawczych brzmi: *Dane powinny być tak otwarte, jak to możliwe i na tyle zamknięte, na ile to jest konieczne.*

Aby pomóc naukowcom w odpowiednim przygotowaniu oraz udostępnianiu danych badawczych sformułowano zasady FAIR. FAIR to akronim czterech angielskich przymiotników opisujących to, jakie powinny być dane badawcze: **F**indable, **A**ccessible, **I**nteroperable, **R**eusable. Nie są to sztywne zasady, a raczej wytyczne do właściwej pracy z danymi. Zasady FAIR zostały stworzone i są w dalszym ciągu rozwijane przede wszystkim po to, aby uczynić dane dostępnymi zarówno dla użytkowników, jak i oprogramowania komputerowego przeszukującego bazy danych bez udziału człowieka.



Zasady FAIR udostępniania danych badawczych są następujące:

1. **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 przeszukiwanie.

2. **ACCESSIBLE** – DOSTĘPNE

- Dostęp do zbioru danych, a przynajmniej do metadanych, możliwy jest bezpośrednio poprzez unikalny identyfikator i nie wymaga dodatkowego oprogramowania

- Metadane są zawsze dostępne, nawet jeśli sam zbiór danych został już usunięty lub przeniesiony.

3. INTEROPERABLE – INTEROPERACYJNE

- Dane oraz metadane są 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.

4. REUSABLE – MOŻLIWE DO PONOWNEGO UŻYCIA

- Metadane dokładnie opisują zbiór danych i ułatwiają użytkownikom określenie ich przydatności dla ich własnych badań.
- Zbiór danych zawiera licencję, która pozwala określić w sposób jednoznaczny warunki ponownego wykorzystania i przetwarzania danych.
- Metadane wskazują jednoznacznie autora oraz miejsce powstania danych.
- Metadane są skonstruowane według ogólnie przyjętych standardów, specyficznych dla danej dyscypliny oraz rodzaju danych.

Aby ułatwić sobie pracę z danymi badawczymi sporządza się **Plan Zarządzania Danymi Badawczymi (DMP) czyli Data Management Plan**. Jest to dokument opisujący czynności wykonywane na każdym etapie pracy z danymi badawczymi. Taki plan powinien powstać już na wstępnym etapie badań naukowych. Aktualnie jest on wymagany przez instytucje i agencje przyznające środki finansowe na badania naukowe (Narodowe Centrum Nauki, Komisja Europejska, Economic and Social Research Council, Natural Environmental Research Council). DMP ułatwia planowanie procedur związanych z pozyskiwaniem, przetwarzaniem i dzieleniem się danymi badawczymi oraz ich bezpieczeństwem.

Plan Zarządzania Danymi Badawczymi DMP powinien zawierać następujące informacje:

- ocenę danych już dostępnych, opis braków i potrzeb,
- opis sposobu gromadzenia danych (np. poprzez ankietę, instrument badawczy) i ich rodzaju (np. dane eksperymentalne lub obserwacyjne),
- dokumentację i standardy opisu danych (metadane),
- informacje o tym, kto będzie właścicielem praw autorskich i praw własności intelektualnej danych, a także kto będzie odpowiedzialny za zarządzanie nimi,
- wymogi i procedury związane z aspektami etycznymi pozyskiwanych danych,

- opis procedur zapewniających kontrolę nad jakością danych (opis powinien zawierać podział obowiązków i czynności związanych z nadzorowaniem i kontrolą poprawności danych),
- plan dotyczący dostępu do danych oraz dzielenia się nimi (określenie na jakiej licencji dane będą dostępne),
- krótko- oraz długoterminową strategię przechowywania i ochrony danych,
- określenie, jakie środki finansowe będą potrzebne do tego żeby DMP zrealizować

Dobrze sporządzony **Plan Zarządzania Danymi Badawczymi DMP** wiąże się z następującymi korzyściami:

- dzięki dobremu planowaniu i zarządzaniu pozyskiwanymi danymi wzrasta jakość oraz wiarygodność samych badań,
- dane mogą zostać łatwo zlokalizowane, plan ułatwia również jednoczesną pracę wielu osób przy projekcie,
- efektywne zarządzanie danymi, zapobieganie niepotrzebnemu ich powielaniu bądź nadpisywaniu,
- poprawa bezpieczeństwa danych,
- ułatwione przygotowanie danych do późniejszego udostępnienia.

Przygotowanie danych badawczych do udostępnienia nie jest procesem łatwym. Cykl życia danych obejmuje ich wytwarzanie/ pozyskiwanie, opisywanie i uporządkowanie, następnie ich przechowywanie i archiwizację oraz na końcu ich udostępnienie.

Kiedy pozyskaliśmy już dane zgodnie z zaplanowanym procesem badawczym oraz przyjętą metodologią, możemy zastanowić się, czy chcemy, oraz, czy możemy je udostępnić. Zanim to jednak zrobimy, warto wykonać poniższe kroki:

1. **SELEKCJA** – nie wszystkie dane muszą zostać udostępnione. Dobierając zbiory danych do archiwizacji i udostępniania warto kierować się takimi czynnikami jak:

- wymagania instytucji sponsorujących badania naukowe,
- wartość naukowa danych badawczych,
- wyjątkowość – warto sprawdzić, czy dane nie duplikują się z innymi istniejącymi już zbiorami danych,
- możliwość replikacji wyników badań – czy dane zawierają wszystkie parametry

umożliwiający powtórzenie eksperymentu,

- kwestie ekonomiczne – jakie koszty wiążą się z zarządzaniem i przechowywaniem danych i czy są one uzasadnione.

2. USUNIĘCIE DANYCH WRAŻLIWYCH, umożliwiających identyfikację badanych osób:

- anonimizacja polega na przekształceniu danych osobowych w sposób uniemożliwiający przyporządkowanie poszczególnych informacji do określonej lub możliwej do zidentyfikowania osoby,
- pseudonimizacja – to przetworzenie danych tak, by nie można ich było przypisać osobie, której te dane dotyczą, bez użycia dodatkowych informacji.

3. WYBÓR FORMATÓW PLIKÓW

Dane powinny zostać opublikowane w ogólnodostępnym formacie, który nie wymaga komercyjnego oprogramowania i wykorzystuje standardowe kodowanie (ASCII, UTF-8). Warto również zastanowić się, jakie formaty plików funkcjonują w naszej dyscyplinie, aby użytkownicy po pobraniu zbioru danych nie byli zmuszeni do ich dodatkowej konwersji, która może prowadzić do utraty jakości danych.

4. NADANIE ODPOWIEDNICH NAZW FOLDEROM ORAZ PLIKOM

Przy nadawaniu nazw warto zadać sobie pytanie: Jakie nazwy plików i jaka ich struktura byłyby dla mnie najbardziej użyteczne, gdybym sam chciał ponownie je wykorzystać? Co powinny zawierać nazwy, aby bezproblemowo można było odnaleźć konkretny zestaw danych? Podstawową zasadą organizacji plików jest konsekwencja, należy trzymać się raz przyjętej struktury i nazewnictwa.

5. OPATRZENIE ZBIORÓW DANYCH ODPOWIEDNIM OPISEM W POSTACI METADANYCH – tu należy pamiętać, że mamy metadane opisowe, strukturalne i administracyjne.

Metadane opisowe – dostarczają informacji niezbędnych do odszukania czy też identyfikacji zbioru danych. Mogą zawierać elementy takie jak: tytuł, streszczenie, autor i słowa kluczowe.

Metadane strukturalne – opisują relacje i zależności pomiędzy poszczególnymi zbiorami oraz elementami tych zbiorów w celu np. ułatwienia nawigacji.

Metadane administracyjne – zawierają informacje pomocne w zarządzaniu danym zasobem. Zawierają informacje takie jak: sposób i data jego utworzenia, typ pliku i informacje dotyczące dostępu. Istnieje kilka podzbiorów danych administracyjnych.

Występuje wiele standardów metadanych. Wyszczególnić można standardy ogólne, dziedzinowe i instytucjonalne. Ogólne standardy metadanych to Dublin Core i Data Cite, czy też Data Documentation Initiative (DDI). Są one uniwersalne dziedzinowo i powszechnie stosowane. Repozytorium Most Danych wykorzystuje standard metadanych Data Cite.

Długoterminowa strategia przechowywania, archiwizacji i ochrony danych powinna być opisana w **Planie Zarządzania Danymi DMP**. Archiwizacja ma na celu ograniczenie kosztów związanych z przechowywaniem danych oraz utrzymanie dostępności danych. Do archiwizacji danych badawczych są wykorzystywane Repozytoria.

Aby wyszukać odpowiednie repozytorium dla naszych danych, można skorzystać z wyszukiwarki repozytoriów np. RE3DATA (www.re3data.org). Przy wyborze powinniśmy kierować się kilkoma wskazówkami:

- należy dokładnie zapoznać się z warunkami korzystania z repozytorium i sprawdzić czy spełnia on nasze wymagania.
- warto dowiedzieć się, gdzie i na jakich zasadach będą przechowywane nasze dane oraz w jaki sposób będą zabezpieczone.
- aby zbiór danych spełniał wymogi zasad FAIR, musi być możliwy do odnalezienia, dlatego należy upewnić się, że dane repozytorium zapewnia przypisanie naszym zbiorom identyfikatora DOI. Ważne jest również odpowiednie powiązanie zbioru z jego autorami, dlatego istotna jest opcja wprowadzenia do opisu identyfikatora naukowca (numer ORCID).
- warto też sprawdzić, czy inni naukowcy z naszej dyscypliny korzystają z danego repozytorium.
- Należy dowiedzieć się, czy repozytorium wspiera używany w naszej dyscyplinie standard metadanych.
- niektóre repozytoria mogą pobierać opłatę za archiwizację danych – tzw. Data Processing Charge, dlatego należy to sprawdzić przed planowaniem kosztów w przypadku prac realizowanych z projektu badawczego.
- jako, że ważnym aspektem aktywności naukowej jest popularyzacja dorobku naukowego, dlatego warto upewnić się, że zasoby wybranego przez nas repozytorium znajdują się w bazach indeksujących repozytoria danych badawczych (Data Citation Index, Mendeley Data, Google Dataset Search)

Publikacja artykułu naukowego oraz udostępnienie danych badawczych np. poprzez repozytorium, to nie wszystko, co należy zrobić aby upowszechnić wyniki swoich badań.

Obecne technologie umożliwiają szereg działań popularyzujących dorobek naukowy. Możliwe jest publikowanie artykułów w tzw **Data Journals** (np Scientific Data, Data in Brief) , czyli w czasopismach recenzowanych, gdzie prace naukowe opisują poszczególne zbiory danych. Publikacje te zawierają bezpośredni link do zasobu. W takich artykułach opisywane są szczegółowo założenia teoretyczne, warunki, metody oraz parametry towarzyszące powstaniu danych.

Wśród możliwości popularyzacji swojego dorobku naukowego, w tym swoich danych badawczych należy wymienić naukowe serwisy społecznościowe (np. Researchgate, Academia.edu, Impactstory) Serwisy te znakomicie wspierają komunikację pomiędzy naukowcami z całego świata. Jeżeli profil naukowy jest na bieżąco aktualizowany to może to przyczynić się do wzrostu zainteresowania naszymi publikacjami oraz wykorzystaniem przez innych naukowców udostępnionych w repozytorium danych badawczych.

Należy również wspomnieć o mediach społecznościowych, które mogą przyczynić się do popularyzacji naszych danych badawczych. Portal Twitter jest uważany za najszybszy sposób wymiany informacji. Regularna aktualizacja profilu na tym portalu niewątpliwie może przyczynić się do wzrostu widoczności naszego dorobku naukowego.

Musimy pamiętać, że społecznościowe serwisy dla naukowców doskonale wspierają badaczy w promowaniu dorobku naukowego, jednak nie pełnią one funkcji repozytorium ani nie spełniają założeń projektowych odnośnie udostępniania efektów badań naukowych (wytyczne instytucji finansującej badania naukowe).

Podobnie jak niezbędne jest odpowiednie zacytowanie publikacji naukowej, z której zaczerpnęliśmy informacje do naszego artykułu, tak istotne jest również cytowanie danych badawczych, które zostały wykorzystane w publikacji (zarówno danych wytworzonych i udostępnionych przez nas, jak i przez innych naukowców). Jest to ważne w aspekcie transparentności i wiarygodności naukowca, oraz pozytywnego wpływu na wzrost cytowań jak również przyczynia się do łatwiejszego wykrywania plagiatów. Stylów cytowania zbiorów danych jest bardzo dużo, niezależnie jednak od przyjętego stylu opis powinien zawierać następujące podstawowe informacje: autora, rok, tytuł, miejsce udostępniania np. nazwę repozytorium, wersję, identyfikator. Cytowanie zbioru danych badawczych w bazach indeksujących (Data Citation Index) stwarza możliwość zwiększenia ich widoczności oraz pozytywnie wpływa na wizerunek naukowca.

Podsumowując należy stwierdzić, że udostępnianie danych badawczych w otwartych repozytoriach stanowi konieczność, ze względu na wymogi grantodawców, ale przyczynia się również do wzrostu widoczności naszych badań naukowych na świecie. Należy jednak zawsze pamiętać o aspektach prawnych przy pracy z danymi badawczymi.

Bibliografia

Pawłowska Maria M., Wachowicz Marta E. (2020) Wprowadzenie do zarządzania danymi naukowymi. Wydawnictwo Difin, ISBN: 9788380855199

GO FAIR International Support & Coordination Office (GFISCO) <https://www.go-fair.org/>

Materiały z broszury informacyjnej <https://mostwiedzy.pl/infokit/Infokit-pl.pdf>

Materiały <https://otwartanauka.pl/>

Otwarty dostęp - szansa czy konieczność? Jak promować otwarty dostęp?
<https://otwartanauka.pl/images/PDFs/BROSZURKA%20Warsztaty%20OA.pdf>

dr inż. Marta Piątek-Hnat
Koordynator ds. otwartego dostępu w ZUT
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Biblioteka Główna
Ośrodek Informacji i Dokumentacji Naukowej