



MOST DANYCH

OTWARTE DANE BADAWCZE



CENTRUM KOMPETENCJI OTWARTEJ NAUKI
BIBLIOTEKA POLITECHNIKI GDAŃSKIEJ

I N F O K I T

	strona
INFORMACJE O PROJEKCIE/WSTĘP	3
OTWARTE DANE BADAWCZE (ang. OPEN RESEARCH DATA)	4
ZASADY UDOSTĘPNIANIA DANYCH BADAWCZYCH – FAIR	6
PLAN ZARZĄDZANIA DANYMI (DMP)	10
PRZYGOTOWANIE DANYCH BADAWCZYCH DO UDOSTĘPNIENIA	12
METADANE I STANDARDY METADANYCH	14
ARCHIWIZACJA I UDOSTĘPNIANIE DANYCH BADAWCZYCH	16
OTWARTE DANE BADAWCZE – INICJATYWY	17
LICENCJONOWANIE DANYCH BADAWCZYCH	18
POPULARYZACJA DANYCH BADAWCZYCH	22
CYTOWANIE DANYCH BADAWCZYCH	24

INFORMACJE O PROJEKCIE/WSTĘP

MOST Wiedzy to platforma stworzona i uruchomiona przez Politechnikę Gdańską w 2017 roku. Ma ona na celu wspomaganie naukowców w promowaniu ich osiągnięć oraz dorobku naukowego. Jednym z elementów serwisu jest otwarte repozytorium publikacji naukowych. Udostępnianie pełnych tekstów artykułów w otwartym dostępie zwiększa ich widoczność, a także ułatwia i przyspiesza komunikację naukową oraz współpracę pomiędzy badaczami.

Otwarty dostęp do publikacji naukowych to tylko jeden z elementów Otwartej Nauki. Mając tę świadomość Politechnika Gdańska we współpracy z Uniwersytetem Gdańskim oraz Gdańskim Uniwersytetem Medycznym przystąpiła do budowy repozytorium danych badawczych – MOST DANYCH.

Repozytorium to zapewnia infrastrukturę umożliwiającą archiwizację oraz udostępnianie danych badawczych – informacji oraz materiałów zgromadzonych przez naukowców w toku procesu badawczego. Dzięki temu MOST DANYCH pozytywnie wpływa na zwiększenie dostępności, poprawę jakości badań i zwiększenie możliwości ponownego wykorzystania zasobów naukowych. Niniejszy przewodnik ma na celu zaznajomić czytelnika z najważniejszymi pojęciami związanymi z Otwartymi Danymi Badawczymi. Dzięki zwięzłej, a zarazem atrakcyjnej formie, stanowi on doskonały punkt startowy na drodze do udostępniania danych badawczych.

CENTRUM KOMPETENCJI

Drodzy naukowcy,

Doskonale zdajemy sobie sprawę, że dla wielu z Was zagadnienia związane z udostępnianiem danych badawczych stanowią jeszcze „nieznany ląd”. Aby Was zachęcić i jednocześnie pomóc w jego odkrywaniu, utworzyliśmy w Bibliotece Politechniki Gdańskiej w ramach projektu MOST DANYCH, Centrum Kompetencji Otwartej Nauki, którego głównym zadaniem jest wsparcie badaczy we wszelkich działaniach działalności związanych z „otwieraniem” nauki. Poprzez organizację szkoleń, konsultacji oraz wydarzeń, promujących ideę otwartości, pragniemy pokazać, że publikowanie w otwartym dostępie, przygotowanie planu zarządzania danymi, czy też udostępnianie danych badawczych nie są procesami skomplikowanymi, a mogą przynieść wiele korzyści i pozytywnie wpłynąć na promocję i widoczność dorobku naukowego.

Centrum Kompetencji Otwartej Nauki powstało dla Was, dlatego zapraszamy do kontaktu z nami w razie jakichkolwiek pytań lub wątpliwości dotyczących otwartości nauki.

Zespół Centrum Kompetencji Otwartej Nauki w Bibliotece PG



OTWARTE DANE BADAWCZE (ang. OPEN RESEARCH DATA)

CZYM SĄ DANE BADAWCZE?

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.

Otwarte Dane Badawcze (ORD) stanowią część szerszej idei, jaką jest Otwarta Nauka, na którą składa się również otwarty dostęp do publikacji (ang. Open Access) oraz otwarta komunikacja naukowa (ang. Open Scholarly Communication).

JAKIE KORZYŚCI PŁYNĄ Z SZEROKIEGO UDOSTĘPNIANIA DANYCH BADAWCZYCH?

- Lepsza komunikacja i wymiana informacji pomiędzy specjalistami, reprezentującymi różne dyscypliny nauki.
- Możliwość przeprowadzania analiz opartych o unikalne dane, których nie można ponownie zebrać.
- Wzrost liczby cytowań zarówno samych danych, jak i publikacji na nich opartych.
- Możliwość oceny rzetelności prowadzonych badań.
- Otwarty dostęp pozwala wykorzystać istniejące zasoby i obniżyć koszty badań.

Warto pamiętać, że Otwarte Dane Badawcze stają się coraz częściej nie tylko możliwością, a koniecznością. Agencje finansujące badania naukowe często wymagają planu zarządzania danymi (z ang. Data Management Plan), a jednym z wymogów, stawianych autorom przez redakcje czasopism, jest udostępnienie danych, na podstawie których powstała publikacja.

CZĘŚCI SKŁADOWE OTWARTEJ NAUKI WEDŁUG KOMISJI EUROPEJSKIEJ

OTWARTE
DANE
BADAWCZE

OTWARTA
KOMUNIKACJA
NAUKOWA

OTWARTY DOSTĘP
DO PUBLIKACJI

?

CZY WIESZ, ŻE...?

coraz więcej instytucji i organizacji propaguje działania na rzecz Otwartej Nauki? Unia Europejska jest liderem jeśli chodzi o Otwarte Dane Badawcze. Komisja Europejska działa aktywnie na rzecz wdrażania inicjatyw ORD przez uniwersytety i ośrodki badawcze w całej Unii Europejskiej.

Więcej informacji o etapach wdrażania Otwartej Nauki w Unii Europejskiej znajdziesz pod adresem <https://ec.europa.eu/research/openscience>

ZASADY UDOSTĘPNIANIA DANYCH BADAWCZYCH - FAIR

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: findable, accessible, interoperable, reusable.

ZASADY FAIR

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 dodatkowych narzędzi ani 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 zawierają liczne atrybuty dokładnie opisujące zbiór danych i ułatwiające użytkownikom określenie ich przydatności dla ich własnych badań.
- Zbiór danych zawiera licencję, określającą jednoznacznie warunki ponownego wykorzystania i przetwarzania danych.
- Metadane wyraźnie określają autora oraz miejsce powstania danych.
- Metadane są skonstruowane według ogólnie przyjętych standardów, specyficznych dla danej dyscypliny oraz rodzaju danych.

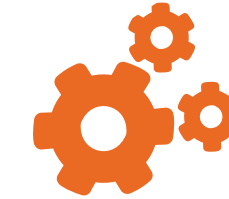
Nie są to sztywne zasady, a raczej wytyczne do właściwej pracy z danymi. Zasady FAIR zostały stworzone i są w dalszym ciągu rozwijane przede wszystkim po to, aby uczynić dane dostępnymi zarówno dla użytkowników, jak i oprogramowania komputerowego, przeszukującego bazy danych bez udziału człowieka.



FAIR



ACCESSIBLE



INTEROPERABLE



REUSABLE



CZY WIESZ, ŻE...?

zasady FAIR są wciąż rozwijane i doprecyzowywane przez międzynarodową społeczność. Inicjatywa GO FAIR organizuje spotkania i warsztaty z zakresu wprowadzania zasad FAIR. Jedną z takich inicjatyw są np. warsztaty „Metadata 4 Machines”, podczas których naukowcy pracują nad możliwością stworzenia metadanych, które będą zrozumiałe dla programów komputerowych.

Więcej publikacji i materiałów informacyjnych można znaleźć na stronie www.go-fair.org



CZAS
NA ZAGADKI!

ZAGADKA 1

Profesor Mościcki chce zachęcić swoich współpracowników do umieszczania danych badawczych w repozytorium MOST DANYCH. Uczula ich jednak, aby nie wierzyli serwisom, które podają się za to repozytorium. By wzmóc ich czujność, zadaje zagadkę:

Jedno z trzech repozytoriów to MOST DANYCH. Dwa repozytoria mówią prawdę, jedno kłamie. Które z nich to prawdziwy MOST DANYCH?

- a) Repozytorium nr 1: Jestem MOST DANYCH.
- b) Repozytorium nr 2: MOST DANYCH to repozytorium nr 3.
- c) Repozytorium nr 3: Repozytorium nr 1 to nie jest MOST DANYCH.

ZAGADKA 2

Profesor Mościcki udostępnia wytworzone przez siebie dane badawcze, zgodnie z zasadami FAIR. Jedna z nich jest dla niego szczególnie istotna, o czym w zaszyfrowanej wiadomości informuje swojego asystenta. Jaka to zasada? Rozszyfruj kod.

☂ 📎 ⚙️ 👁️ ✂️ 🏠

☀️	A	Ą	B	C	Ć	D
🕒	E	Ę	F	G	H	I
🌲	J	K	L	Ł	M	N
🎵	O	Ó	P	R	S	T
❤️	U	W	X	Y	Z	!

👁️ 🎵 ☂️ 🕒 ☂️ ❤️ ✂️ 🎵 ☂️ ☀️ ⚙️ ☀️ ⚙️ 🌲 ☂️ 🕒

odpowiedź:

PLAN ZARZĄDZANIA DANYMI (DMP)

Data Management Plan (DMP) czyli Plan Zarządzania Danymi to dokument opisujący czynności wykonywane na każdym etapie pracy z danymi badawczymi. DMP powinien powstać już na wstępnym etapie badań naukowych. Jest on coraz częściej 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.

CO POWINIEN ZAWIERAĆ DATA MANAGEMENT PLAN:

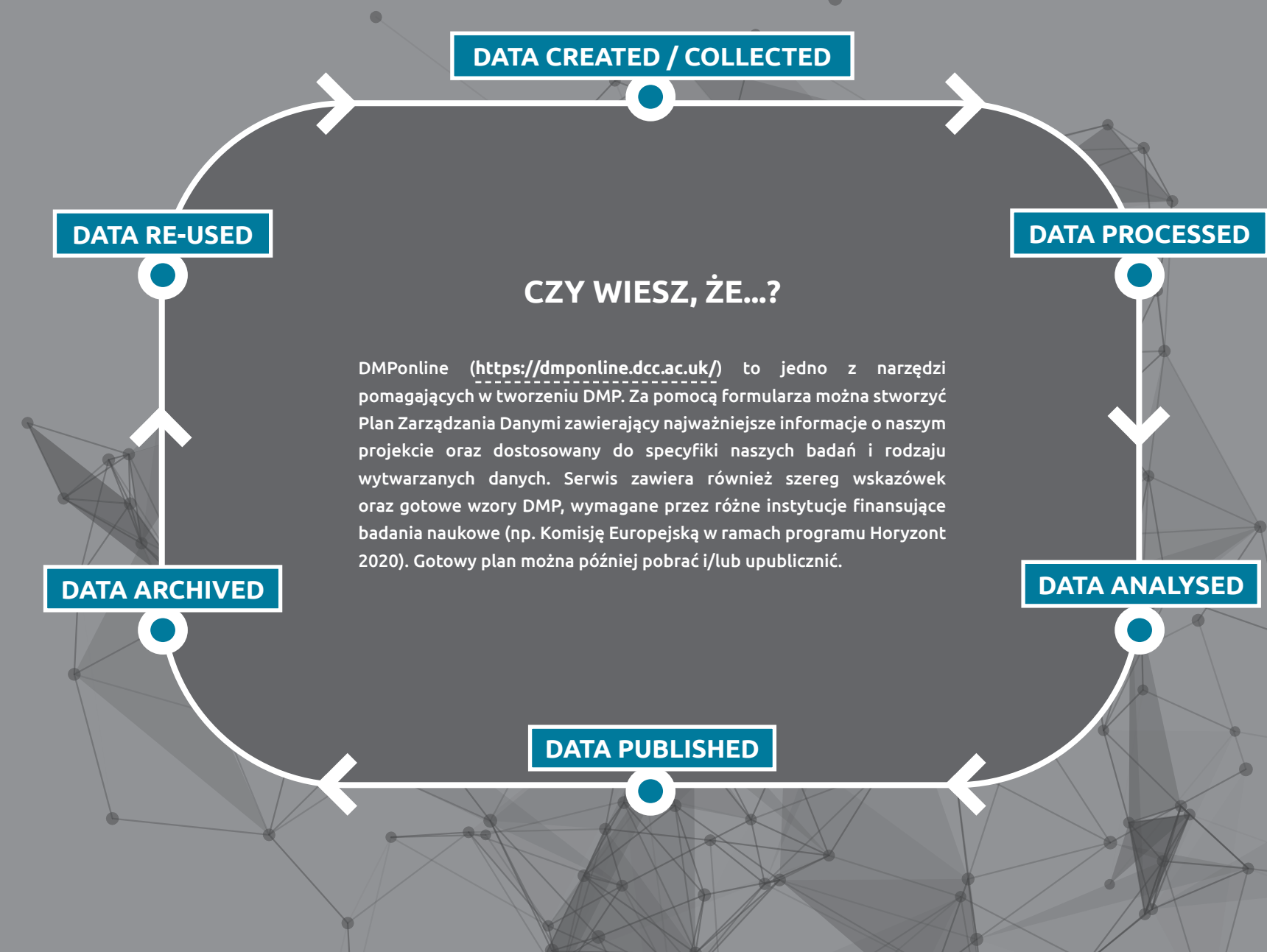
- ocenę danych już dostępnych, opis braków i potrzeb,
- opis sposobu gromadzenia danych (np. poprzez ankietę, instrument badawczy) i ich rodzaju (np. dane eksperymentalne lub obserwacyjne),
- dokumentację i standardy opisu danych (metadane),
- informacje o tym, kto będzie właścicielem praw autorskich 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 będą potrzebne do przeprowadzenia DMP.

DOBRY DMP TO LICZNE KORZYŚCI:

- 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.



CYKL ZARZĄDZANIA DANYMI



PRZYGOTOWANIE DANYCH BADAWCZYCH DO UDOSTĘPNIENIA

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 agencji 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 umożliwiające powtórzenie eksperymentu,
- kwestie ekonomiczne – jakie koszty wiążą się z zarządzaniem i przechowywaniem danych i czy są one uzasadnione.

?

CZY WIESZ, ŻE...?

dane badawcze nie muszą być idealne, mogą np. zawierać luki w pomiarach wynikające z czynników zewnętrznych. Ważne, aby zaznaczyć takie braki i opisać z czego wynikają.

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

?

CZY WIESZ, ŻE...?

niektóre zbiory danych wymagają edycji i czyszczenia. Błędy w datasetach mogą dotyczyć m.in. pisowni czy gramatyki. Mogą także zdarzyć się błędy dotyczące wartości nominalnych (terminy używane wielokrotnie). Identyfikatory dla obiektów lub kontekstów również powinny być wolne od błędów („nr katalogu” czy „miejsce #”).

Jeżeli potrzebujesz narzędzia, które pomoże Ci „oczyścić” zebrane dane, możesz korzystać np. z OpenRefine (<http://openrefine.org/>)

METADANE I STANDARDY METADANYCH

Zgodnie z zasadami FAIR istotny element każdego zbioru danych stanowią metadane. Metadane to dane o danych. Są kluczem do uzyskania dostępu do danych badawczych, ich zrozumienia i ponownego wykorzystania. Istnieją trzy główne typy metadanych:

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. Dwa z nich, często wymieniane jako oddzielne typy metadanych, to:

- metadane zarządzania prawami, które dotyczą praw własności intelektualnej,
- metadane konserwacji, które zawierają informacje potrzebne do archiwizacji i utrzymania zasobu.

Metadane powinny informować m.in. o: strukturze danych, ograniczeniach ich dotyczących (jeśli takie istnieją), o tym co dane oznaczają i w jaki sposób je cytować.

Na całym świecie powstają inicjatywy mające na celu sformalizowanie specyfikacji metadanych, aby umożliwić łatwe ponowne wykorzystanie danych. Przykładami takich inicjatyw są: Research Data Alliance (RDA, www.rd-alliance.org), OpenAire (www.openaire.eu) i Metadata 2020 (www.metadata2020.org).

Standardy metadanych mają na celu usystematyzowanie sposobu opisu danych. Metadane przygotowane według standardu posiadają stałą strukturę opisu o wyraźnie zdefiniowanych polach, dzięki czemu opis jest zawsze zrozumiały zarówno dla ludzi jak i programów komputerowych.

METADANE TO TAKŻE OPIS ZMIENNYCH, CODEBOOK I SŁOWNIK DANYCH (CONTROLLED VOCABULARY) ZAWIERAJĄCY:

- nazwy zmiennych (skrótowe i pełne, np. AGE i Age of the respondent),
- jednostki miar (np. mm),
- dozwolone wartości (np. zakres od 0 do 100),
- definicje zmiennych (np. Age=Age of the respondent in years).

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. Wybrane standardy metadanych wykorzystuje się także w różnych dziedzinach i instytucjach np.: DC (nauki o życiu), EML (ekologia), SDMX (ECB, EUROSTAT, IMF, OECD, UN), SAFE (ESA), INSPIRE ISO 19139 (nauki o ziemi), Project Open Data Metadata Schema v1.1 (Agencje federalne USA), TEI i CDWA (dyscypliny humanistyczne).

PRZYKŁADOWE SŁOWNIKI DANYCH:

- Biologia – Convention of Biological Diversity Controlled Vocabulary (CBDVoc);
- Ekonomia i nauki społeczne – Central Europe Glossary (CEG);
- Medycyna – Unified Medical Dictionary (UMD);
- Edukacja, nauki społeczne – UNESCO thesaurus.

PRZYKŁADOWE POLA METADANYCH TO:

Dataset name	Determination of the influence of green corrosion inhibitors on aluminium alloys in alkaline media
Version	1.0
Author/s	Ryl, Jacek; Wysocka, Joanna; Krakowiak, Stefan; Cieślik, Mateusz;
Description	The studies are devoted to search for green corrosion inhibitors of aluminium and its alloys, offering high corrosion inhibition efficiency in alkaline media. The project will aim at development of instantaneous impedance measurements for accurate determination of the adsorption isotherms.
Format	TA
Licence	Creative Commons Attribution 4.0 International
Funding Agency/ies	Ministry of Science and Higher Education, Republic of Poland
Keywords	corrosion inhibitor, green chemistry, aluminium alloys, instantaneous impedance measurements
DOI	<i>(wszystkie zbiory danych zdeponowane w repozytorium MOST Danych otrzymają unikalne identyfikatory DOI)</i>
Discipline	Chemical sciences
Language	English



CZY WIESZ, ŻE...?

zapisywanie metadanych możliwe jest w: pliku txt, arkuszu kalkulacyjnym, pliku XML.

Istnieje wiele przydatnych narzędzi do tworzenia metadanych. Przykładami są:

Nesstar Publisher (freeware) (www.nesstar.com) zgodny z DDI i Dublin Core, STATA, SPSS, Eenvplus dla Inspire (www.eenvplus.sinergis.it/euosmegwt/), Edytor Metadanych dla Inspire www.edytor.geoportal.gov.pl

ARCHIWIZACJA I UDOSTĘPNIANIE DANYCH BADAWCZYCH

Długoterminowa strategia przechowywania, archiwizacji i ochrony danych powinna być opisana w Planie Zarządzania Danymi. Archiwizacja ma na celu ograniczenie kosztów związanych z przechowywaniem danych oraz utrzymanie dostępności danych. Coraz częściej funkcje archiwów pełnią otwarte repozytoria danych badawczych, takie jak MOST DANYCH. W Internecie można znaleźć wiele dziedzinowych i instytucjonalnych repozytoriów danych. Czasami organizacja finansująca badania wymaga zdeponowania danych badawczych w konkretnym repozytorium.

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

- Należy dokładnie zapoznać się z warunkami korzystania z serwisu i sprawdzić czy spełnia on nasze wymagania.
- Warto również 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 (np. numeru ORCID).
- Przy wyborze 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.

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. Przykładami takich baz są: Data Citation Index, Mendeley Data czy Google Dataset Search.

OTWARTE DANE BADAWCZE – INICJATYWY

Wiele instytucji związanych z nauką (od uczelni, po instytuty badawcze, wydawców i biblioteki) wprowadza polityki oraz zasady dotyczące udostępniania, przechowywania oraz dzielenia się danymi. Poniżej znajdują się przykłady organizacji, które pomagają wprowadzać zasady FAIR.

Research Data Alliance (RDA) to organizacja społeczna, której głównym celem jest stworzenie przestrzeni do dyskusji i rozwoju infrastruktury potrzebnej do swobodnego dzielenia się danymi i prowadzenia opartych na nich badań. Powstała w 2013 r. z inicjatywy Komisji Europejskiej, Narodowej Fundacji Nauki Stanów Zjednoczonych, Narodowego Instytutu Standaryzacji i Technologii oraz Australijskiego Urzędu Innowacji.

CODATA (ang. Committee on Data for Science and Technology) to międzynarodowa organizacja z siedzibą w Paryżu. Założona została w 1966 roku jako jeden z kilkunastu stałych komitetów ICSU (Międzynarodowej Rady Stowarzyszeń Naukowych).

Digital Curation Centre (DCC) to organizacja, której głównym celem jest budowa potencjału i rozwijanie umiejętności zarządzania danymi badawczymi. DCC zapewnia fachowe porady i praktyczną pomoc organizacjom badawczym, które chcą przechowywać, zarządzać, chronić i udostępniać cyfrowe dane badawcze.



CZY WIESZ, ŻE...?

dobrze praktyki archiwizacji danych wymagają stosowania reguły 3-2-1 – tworzenie trzech kopii zapasowych, na dwóch odrębnych nośnikach, w tym jednej kopii w innej lokalizacji fizycznej np. inny budynek lub „chmura”.



CZY WIESZ, ŻE...?

politykę instytucjonalną w zakresie otwartych danych badawczych przyjęły między innymi takie uczelnie jak: Cambridge University, Manchester University oraz Melbourne University.

Wydawcy również określają swoje wymagania. Zobacz: Elsevier (Mendeley Repository oraz Elsevier Research Data) lub Wiley Data Sharing Services.

LICENCJONOWANIE DANYCH BADAWCZYCH

Zgodnie z zasadą „Reusable” dane powinny być opatrzone licencją określającą warunki korzystania z danego zbioru danych. Przed wyborem licencji należy upewnić się, jaka jest polityka naszej instytucji odnośnie udostępniania danych. Dodatkowo, agencje finansujące badania mogą wymagać stosowania określonych licencji. Dobrą praktyką jest stosowanie gotowych (standardowych) licencji, które pozwalają zaoszczędzić czas potrzebny na stworzenie własnego, nowego typu

Creative Commons to międzynarodowy projekt, w ramach którego opracowano szereg licencji gotowych do wykorzystania, co znacznie ułatwia proces udostępniania. Obecnie funkcjonują licencje CC w wersji 4.0. Przy ich stosowaniu należy pamiętać, że powstały one z myślą o utworach, a nie zbiorach danych, dlatego należy upewnić się, czy wybrana licencja pasuje do naszego zbioru.



Uznanie autorstwa - CC BY

Zezwala innym na rozpowszechnianie, zmienianie i tworzenie nowych utworów bądź zbiorów w oparciu o licencjonowany zasób, także w celach komercyjnych. Warunkiem jest oznaczenie autorstwa tego zbioru danych.



Uznanie autorstwa - Na tych samych warunkach - CC BY SA

Pozwala na kopiowanie, modyfikowanie oraz rozpowszechnianie zbioru pod warunkiem oznaczenia autorstwa i udostępniania danych zależnych na tych samych warunkach.



Uznanie autorstwa - Bez utworów zależnych - CC BY ND

Ta licencja umożliwia innym ponowne wykorzystanie danych w dowolnym celu pod warunkiem oznaczenia autorstwa. Licencja nie pozwala jednak na modyfikowanie zbioru.



Uznanie autorstwa - Użycie niekomercyjne - CC BY NC

Ta licencja pozwala innym kopiować, modyfikować i rozpowszechniać zbiór danych tak długo, jak oznaczone będzie jego autorstwo oraz wykorzystywany będzie w celach niekomercyjnych.



Uznanie autorstwa - Użycie niekomercyjne - Na tych samych warunkach - CC BY NC SA

Licencja ta pozwala innym kopiować, modyfikować i rozpowszechniać dane przy zastrzeżeniu, że zarówno oryginalne jak i zmodyfikowane zbiory zostaną użyte w celach niekomercyjnych oraz, że będą rozpowszechniane na takiej samej licencji.



Uznanie autorstwa - Użycie niekomercyjne - Bez utworów zależnych - CC BY NC ND

Ta licencja jest najbardziej restrykcyjna spośród sześciu licencji CC. Pozwala jedynie na pobieranie zbioru danych i dzielenie się nim z innymi, pod warunkiem określenia autorstwa. Zbiór nie może być modyfikowany ani wykorzystywany komercyjnie.

licencji. Określenie licencji, na jakiej udostępniane będą dane, powinno odbyć się już na etapie tworzenia Planu Zarządzania Danymi.

Przykładami otwartych licencji są licencje Creative Commons (CC) oraz licencje wolnego oprogramowania takie jak: GNU General Public License version 2 (GPLv2) i licencja Massachusetts Institute of Technology (MIT license).



CZY WIESZ, ŻE...?

w ramach projektu Open Data Commons powstały licencje, które zostały zaprojektowane specjalnie w celu licencjonowania baz danych.

Przykładami są:

Public Domain Dedication and License (PDDL) – domena publiczna dla baz danych. Zakłada nieograniczoną możliwość pobierania, udostępniania i modyfikowania baz danych.

Open Data Commons Attribution License (ODC-By) – licencja, w której jedynym warunkiem kopiowania i modyfikowania bazy danych jest uznanie autorstwa.

Open Data Commons Open Database License (ODC – ODbL) – otwarta licencja zezwalająca na kopiowanie, przetwarzanie oraz rozpowszechnianie bazy danych pod warunkiem uznania jej autorstwa oraz upowszechniania na takich samych warunkach.



CZAS
NA ZAGADKI!

ZAGADKA 3

Profesor Mościcki zapomniał, jakiego kodu używał, by dostać się do laboratorium. Kod jest jednocześnie nazwą reguły obowiązującej przy archiwizacji danych. Pomóż odzyskać kod wiedząc, że:

- trzecia cyfra to cyfra, przez którą dzielenie nie wpływa na wartość liczby
- pierwsza cyfra to liczba atomowa litu
- druga cyfra jest najmniejszą liczbą pierwszą

odpowiedź:

ZAGADKA 4

Profesor Mościcki szyfrem, jak to ma w zwyczaju, informuje dziś swojego asystenta, na jakiej licencji Creative Commons będzie chciał udostępnić wyniki swoich najnowszych badań. Wpisz zakodowane litery w odpowiednim miejscu na skali, a dowiesz się, o jakiej licencji CC myśli profesor.



- C – 17/10
- Y – 3/2
- B – 2/5
- N – 1,6

odpowiedź:

POPULARYZACJA DANYCH BADAWCZYCH

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. Wśród możliwości wyróżniamy m.in.:

DATA JOURNALS

Czasopisma recenzowane naukowo, w których artykuły opisują poszczególne zbiory danych. Publikacje zawierają bezpośredni link do zasobu, a także opisują szczegółowo założenia teoretyczne, warunki, metody oraz parametry towarzyszące powstaniu danych. Istnieją czasopisma interdyscyplinarne np. Scientific Data, Data in Brief, jak również skupione na konkretnych dyscyplinach np. Genomics Data.

NAUKOWE SERWISY SPOŁECZNOŚCIOWE NP. RESEARCHGATE, ACADEMIA.EDU CZY IMPACTSTORY

Serwisy te znakomicie wspierają komunikację pomiędzy naukowcami z całego świata. Aktualizowany na bieżąco profil naukowy może znacząco przyczynić się do wzrostu zainteresowania publikacjami oraz wykorzystania udostępnionych danych w badaniach innych.

MEDIA SPOŁECZNOŚCIOWE

Do popularyzacji danych może posłużyć również np. konto na Twitterze i regularna aktualizacja profilu. Portal ten jest uważany za najszybszy sposób wymiany informacji.

?

CZY WIESZ, ŻE...?

w opozycji do tradycyjnych wskaźników bibliometrycznych, opartych o cytowania w publikacjach naukowych, powstały alternatywne wskaźniki, z ang. altmetrics. Opierają się one na częstotliwości wzmiankowania o danej publikacji czy datasecie w mediach społecznościowych, czy serwisach dla naukowców. Więcej o alternatywnych metrykach można przeczytać na stronie <https://www.altmetric.com>

REPOZYTORIUM INSTYTUCJONALNE

Pamiętaj, 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.

Poniżej krótkie porównanie repozytorium instytucjonalnego oraz serwisu Research Gate.

	Repozytorium instytucjonalne	Research Gate
Indeksowanie (harvesting)	TAK	NIE (PRZYPADKOWE)
Archiwizacja (long-time preservation)	TAK	NIE
Model biznesowy	NON PROFIT	KOMERCYJNY
Czy spełnia wymagania wydawców i agencji finansujących badania, programów takich jak NCN, Horyzont 2020	TAK	NIE

CYTOWANIE DANYCH BADAWCZYCH

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).

PO CO CYTOWAĆ DANE BADAWCZE?

- Ma to pozytywny wpływ na transparentność i wiarygodność naukowca.
- Cytowanie zbiorów danych w powiązanych z nimi publikacjach może pozytywnie wpłynąć na liczbę cytowań tych publikacji.
- Prawidłowe cytowanie danych ułatwia oraz znacznie przyspiesza lokalizację konkretnych zbiorów danych.
- Przyczynia się to do łatwiejszego wykrywania plagiatów.
- Cytowania zbiorów danych w bazach indeksujących (np. Data Citation Index lub Google Data Search) pozytywnie wpływają na wizerunek naukowca.

To, jak opis bibliograficzny zbioru danych będzie wyglądał, zależy przede wszystkim od przyjętego w publikacji stylu cytowań (np. APA, Chicago, IEEE). Istnieją setki stylów cytowań, a to, który z nich zastosujemy najczęściej podyktowany jest wymogami stawianymi przez wydawcę. Niezależnie jednak od przyjętego stylu opis powinien zawierać następujące **podstawowe informacje: autora, rok, tytuł, miejsce udostępniania np. nazwę repozytorium, wersję, identyfikator.**

PRZYKŁAD CYTOWANIA WG STYLU APA:

Whitaker, K., Colavizza, G. (2019). Alan-turing-institute/das-public: First release with minor updates (Version v1.1) [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.3268810>

PRZYKŁAD CYTOWANIA WG STYLU IEEE

K. Whitaker and G. Colavizza. "Alan-turing-institute/das-public: First release with minor updates." (Version v1.1)" July 4, 2019. Distributed by Zenodo. <http://doi.org/10.5281/zenodo.3268810>

PRZYKŁAD CYTOWANIA WG STYLU AMA (AMERICAN MEDICAL ASSOCIATION)

Whitaker K, Colavizza G. alan-turing-institute/das-public: First release with minor updates. July 2019. <http://doi:10.5281/zenodo.3268810>



CZY WIESZ, ŻE...?

jeżeli Twój dataset ma już nadany numer DOI, aby stworzyć jego opis bibliograficzny możesz skorzystać z generatora cytowań

DOI Citation Formatter: <https://citation.crosscite.org>

TO JUŻ
OSTATNIA ZAGADKA!

ZAGADKA 5

Profesor Mościcki zachęca swoich współpracowników do większej aktywności w naukowych serwisach społecznościowych. W tym tygodniu wysłał im ten oto rebus. Jaki serwis miał tym razem na myśli profesor?



piąta litera



szósta litera



pierwsza litera



piąta litera



druga litera



ostatnia litera



piąta litera



druga litera



trzecia litera



trzecia litera



czwarta litera



czwarta litera

odpowiedź:

MOST DANYCH. Multidyscyplinary Otwarty System Tranferu Wiedzy – etap II: Open Research Data

Celem projektu jest zaprojektowanie i wybudowanie platformy pozwalającej na gromadzenie, wyszukiwanie, analizowanie i udostępnianie Otwartych Danych Badawczych oraz zasilenie jej unikalnymi danymi naukowymi z trzech najważniejszych uczelni Pomorza: Politechniki Gdańskiej, Uniwersytetu Gdańskiego i Gdańskiego Uniwersytetu Medycznego. Naukowcy będą mieli również możliwość analizy danych na superkomputerze.

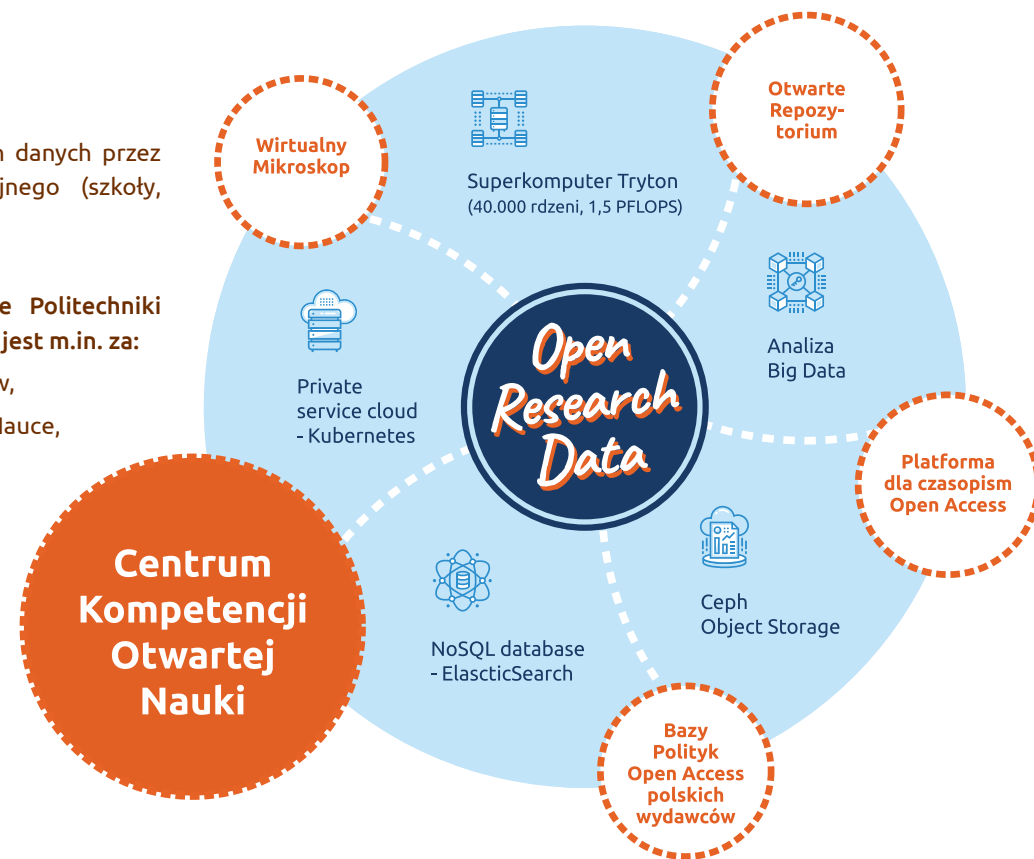


W ramach projektu powstaną również usługi i narzędzia dodatkowe, takie jak:

- baza polityk polskich wydawców naukowych względem Open Access,
- platforma do udostępniania czasopism,
- platforma do organizacji konferencji naukowych,
- wirtualny mikroskop – pozwalający na wykorzystanie zgromadzonych danych przez wszystkie grupy szeroko rozumianego społeczeństwa informacyjnego (szkoły, entuzjastów nauki, hobbystów).

Najważniejszym elementem projektu jest utworzenie w Bibliotece Politechniki Gdańskiej Centrum Kompetencji Otwartej Nauki, które odpowiedzialne jest m.in. za:

- pomoc w przygotowaniu Data Management Plan na potrzeby projektów,
- organizację szkoleń, konferencji i warsztatów poświęconych Otwartej Nauce,
- wspieranie pracowników naukowych w otwieraniu wyników badań w postaci danych badawczych i publikacji,
- wspieranie autorów w deponowaniu pełnych tekstów publikacji naukowych w instytucjonalnym repozytorium MOST Wiedzy,
- aktywną działalność na rzecz Otwartej Nauki w krajowych i międzynarodowych organizacjach, zespołach i grupach roboczych.



ROZWIĄZANIA ZAGADEK

ODPOWIEDZI

Zagadka 1:
MOST DANYCH to repozytorium nr 3. Repozytorium nr 1 kłamie.

Zagadka 2:
REUSABLE

Zagadka 3:
321

Zagadka 4:
BY NC

Zagadka 5:
Research Gate



POLITECHNIKA GDAŃSKA

kontakt

POLITECHNIKA GDAŃSKA

Gmach Główny, pokój nr 100/7
ul. Gabriela Narutowicza 11/12
80-233 Gdańsk, Polska
Telefon +48 58 3472401

MOST DANYCH

biuro.most@pg.edu.pl

<https://pg.edu.pl/most/danych>

CENTRUM KOMPETENCJI

open-data@pg.edu.pl

<https://pg.edu.pl/openscience/>

Więcej informacji znajdziesz na pg.edu.pl/openscience

Projekt „*MOST DANYCH. Multidyscyplinarny Otwarty System Transferu Wiedzy – etap II: Open Research Data*” współfinansowany jest z Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Polska Cyfrowa na lata 2014-2020



Rzeczpospolita
Polska



Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

