

RepOD

– Repozytorium Otwartych Danych Badawczych

Marta Hoffman-Sommer

Platforma Otwartej Nauki, ICM, Uniwersytet Warszawski

Seminarium Open Access, Biblioteka Uniwersytetu Łódzkiego, 20.10.2015



UNIWERSYTET WARSZAWSKI
Interdyscyplinarne Centrum Modelowania
Matematycznego i Komputerowego
www.icm.edu.pl



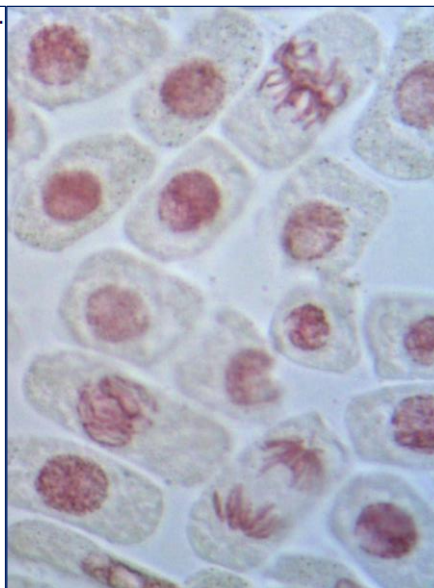
Co jest przedmiotem otwartego dostępu?

KTH Biblioteket, CC-BY-SA
<https://www.flickr.com/photos/kthbiblioteket/4472640423/>



Artykuły i książki

Channel	Raw	Int.	Intensity	Avg.
11481	61,73	69	186	60
42142	181,65	447	232	73
37539	151,37	403	248	67
26707	127,18	302	210	69
33831	145,82	329	232	71
30312	135,32	310	224	73
20118	83,82	125	240	71
16894	83,22	140	203	71
16143	82,36	115	196	71
19950	95	159	210	73
24331	98,11	174	248	75
21530	106,06	222	203	71
11831	67,99	77	174	67
46601	194,17	428	240	79
52345	180,5	468	290	81
43917	177,08	428	248	77
43813	208,63	478	210	83
39835	177,83	422	224	81
20207	103,1	170	196	77
17899	91,32	136	196	75
15462	88,86	136	174	73
18585	94,82	155	196	74
21416	109,27	197	196	79
26097	112,49	212	232	77
11463	63,68	73	180	65
36909	144,18	277	256	77
40585	145,47	293	279	75
32514	140,15	256	232	79
38101	127	283	300	77
29338	104,78	203	280	73
26193	93,88	144	279	77



```
<TEI version="5.0" xmlns="http://
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>TEI中文指引</title>
    </titleStmt>
    <publicationStmt>
      <p>將與TEI 中文在地化計劃等文件一
    </publicationStmt>
    <sourceDesc>
      <p>譯自TEI P5 英文指引</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
<text>
  <body>
    <p>這是TEI P5的中文指引...</p>
  </body>
</text>
</TEI>
```

Dane badawcze

„Zarejestrowane materiały o charakterze faktograficznym powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.”

Co zaliczamy do danych badawczych?

Dane liczbowe

Dokumenty tekstowe, notatki laboratoryjne

Kwestionariusze, ankiety, wyniki badań ankietowych

Nagrania audio i video, zdjęcia

Próbki, artefakty, obiekty

Zawartość baz danych (video, audio, teksty, obrazy)

Modele matematyczne, algorytmy

Oprogramowanie (skrypty, pliki wejściowe...)

Wyniki symulacji komputerowych

Protokoły laboratoryjne, opisy metodologiczne

Co to znaczy, że dane są otwarte?

1. Można się dowiedzieć, że istnieją
2. Można je obejrzeć
3. Można z nich korzystać

Otwarte dane



Dane publiczne



Dane badawcze

Dane statystyczne

Finanse

Kultura

Klimat i środowisko

Transport

Otwarte dane

Dane publiczne

Dane badawcze

Dane statystyczne

Finanse

Kultura

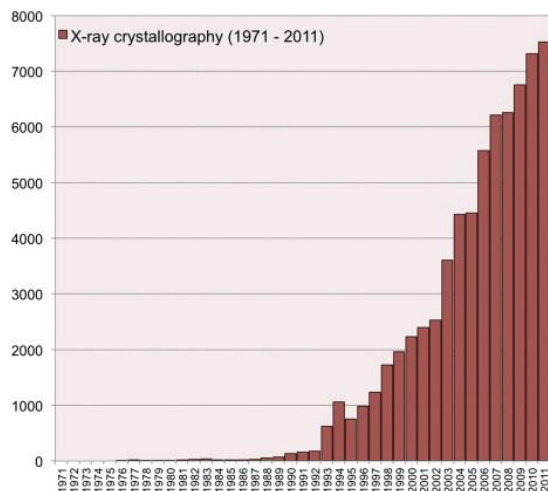
Klimat i środowisko

Transport

Jak otwierać dane badawcze?

Otwarte dane – repozytoria specjalistyczne

Protein Data Bank – od 1971 roku



Berman, Kleywegt, Nakamura, Markley (2012)
<http://dx.doi.org/10.1016/j.str.2012.01.010>



Oxford Text Archive – od 1976 roku

University of Oxford Text Archive

University of Oxford Text Archive: [Home](#) | [About](#) | [Catalogue](#) | [TCP](#) | [Contact](#) | [Help and FAQ](#) | [Search OTA](#)

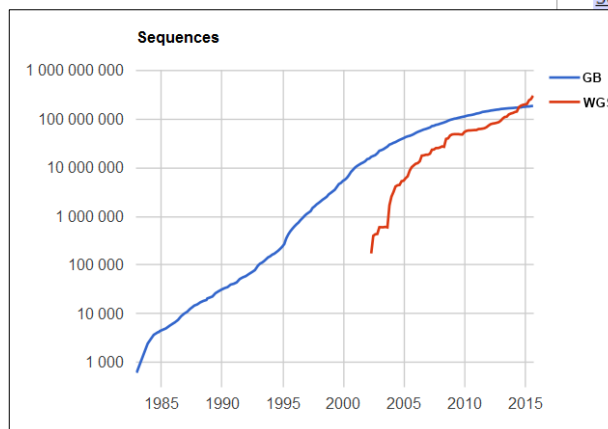
TEI Texts | Corpora | Legacy formats

Search: Show 10 entries

ID	Title	Author	Date	Language	Availability
5700	As you Like it.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
5699	ALL'S Well, that Ends Well.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
5698	The third Part of Henry the Sixt, with the death of the Duke of YORKE.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
57	The second Part of Henry the Sixt, with the death of the Good Duke HVMFREY.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
56	The Second Part of Henry the Fourth, Containing his Death: and the Coronation of King Henry the Fift.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
55	The first Part of Henry the Sixt.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
54	The First Part of Henry the Fourth, with the Life and Death of HENRY Sirnamed HOT-SPVRRE.	Shakespeare, William, 1564-1616	1623	English	CC BY-SA
53	Plain directions for the treatment of	Ogden, Bernard, b. 1767. 1797.	eng	eng	CC BY-SA

GenBank – od 1982 roku

<http://www.ncbi.nlm.nih.gov/genbank/statistics>



Co z danymi, dla których brakuje specjalistycznych repozytoriów?

→ Szeroko zakrojone lub ogólne repozytoria danych



→ Czasopisma publikujące dane (*data journals*)



Data in Brief



RepOD - serwis dla polskiej społeczności akademickiej

Dane:

(1) badawcze

(2) otwarte

→ ze wszystkich dziedzin nauki

→ wszystkie formaty plików



www.repod.pon.pl

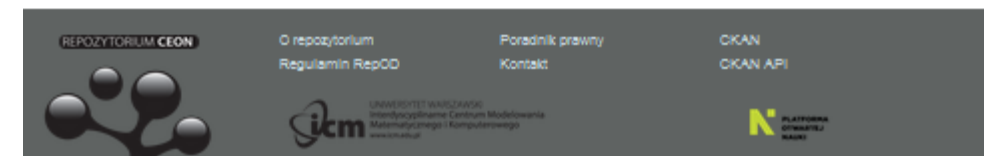


Po co udostępniać dane badawcze w otwartym repozytorium?

Otwarte udostępnianie danych badawczych może przynieść korzyści zarówno rozwojowi nauki, jak i karierze badacza:

- udostępnienie danych umożliwi ich ponowną analizę i zachęca do nowych interpretacji,
- otwarte dane można wykorzystywać do prowadzenia nowych badań, a także łączyć je ze sobą, tworząc nowe zestawienia,
- z otwartych danych mogą korzystać zarówno inni naukowcy, jak i osoby spoza środowiska akademickiego,
- udostępnienie danych ułatwia sprawdzenie, czy opublikowane już prace naukowe opierają się na powtarzalnych wynikach,
- dane zdeponowane w repozytorium są bezpiecznie, długoterminowo przechowywane,
- przygotowanie danych do udostępnienia wymaga ich odpowiedniego opracowania i opisanie, dzięki czemu łatwiej z nich skorzystać w przyszłości,
- dane zdeponowane w repozytorium posiadają stały URL i uzyskują numer DOI (digital object identifier), co ułatwia ich prawidłowe cytowanie oraz umożliwia umieszczenie listy opublikowanych zbiorów danych w CV,
- dane w repozytorium są opatrzone zstandaryzowanym zestawem metadanych, dzięki czemu są łatwe do wyszukania,
- repozytorium zapewnia badaczom informacje o tym, jak często ich dane były oglądane i pobierane.

Na całym świecie coraz więcej instytucji finansujących badania naukowe wprowadza politykę otwartych danych. Niektóre z nich nakładają na swoich grantobiorców obowiązek otwartego udostępniania wytworzonych danych. Także Komisja Europejska uruchomiła Pilotat Otwartych Danych Badawczych w programie Horyzont 2020.





ckan

The open source data portal software

- Otwarte oprogramowanie
- Stworzone przez brytyjską Open Knowledge Foundation, obecnie rozwijane przez CKAN Association
- Przeznaczone przede wszystkim dla danych publicznych

Zbiór danych

(1) metadane

🏠 / Foldery użytkowników / Marta Hoffman-Sommer / Time-course data of the ...

Time-course data of the phosphorylation of MAP kinases in yeast cells treated with alpha-factor and/or salt

Obserwujący
0

Folder użytkownika

Zbiór danych Grupy Strumień aktywności Powiązane

Time-course data of the phosphorylation of MAP kinases in yeast cells treated with alpha-factor and/or salt

This dataset contains the results of an experiment meant to investigate crosstalk between Ste11-dependent mitogen-activated protein kinase (MAPK) pathways in cells of the yeast *Saccharomyces cerevisiae*. Yeast cells were stimulated with the hormone alpha-factor and with salt (NaCl) in three different scenarios: (1) only alpha-factor, (2) alpha-factor and salt simultaneously, (3) only salt. Samples were drawn from the cultures at defined time points after stimulation, total proteins were isolated and analyzed by Western blotting with antibodies against the phosphorylated forms of three yeast MAPKs:

(2) pliki

Pliki w tym zbiorze

 TXT	ReadMe_forMAPKdataset This is a text file containing all information necessary to understand and... 6 wejść 1 pobranie	↗ Eksploruj
 ZIP	Reports This archive contains text files (tab-separated values) containing the raw... 1 wejście 2 pobrania	↗ Eksploruj
 ZIP	Pictures This archive contains jpeg files presenting scans of all Western blots... 0 wejść 3 pobrania	↗ Eksploruj
 XLS	MAPK-PP-normalized_data This is an xls spreadsheet containing processed quantitative data. Data... 4 wejścia 1 pobranie	↗ Eksploruj
 PDF	MAPK-PP-plots This file contains images of graphs obtained by plotting the processed data... 6 wejść 9 pobrań	↗ Eksploruj

Metadane

- Tytuł, autor(autorzy), rok wydania, krótki opis...
- Link do powiązanej publikacji naukowej
- Informacje o finansowaniu badań
- Licencja prawna

→ zgodność z wymogami OpenAIRE

→ zgodność z wymogami DataCite



DOI – *digital object identifier*



Zbiór danych to zbiór plików (zasobów), wraz z ich opisami i innymi informacjami, dostępny pod stałym adresem URL. Zbiór danych jest podstawową jednostką wyszukiwania danych przez innych użytkowników.

* Tytuł:

* URL: <dataset> [Edytuj](#)

Autor:

Imię:

* Nazwisko:

Email:

Afiliacja:

* Rok wydania:

Powiązana publikacja:

Opis:

Możesz tu używać formatowania Markdown

Słowa kluczowe:

Typ zasobu:

Obszar badań:

Instytucja finansująca badania:

Program finansowania badań:

Numer grantu:

Licencja dla plików: ☒ Udzielam następującej licencji dla wszystkich plików tego zbioru:

☒ Osobno wybiorę licencje dla poszczególnych plików tego zbioru.

Wybierana powyżej Licencja dla plików będzie dotyczyć jedynie zawartości plików, które dodasz do tego zbioru danych. Wysyłając ten formularz, zrzekasz się - w granicach dopuszczonych prawem - wszelkich praw autorskich i pokrewnych zapewniających ochronę prawną zbioru danych jako zestawienia, jak również wprowadzonych do formularza wartości metadanych, na mocy licencji CC0 1.0.

[Usuń](#)

* Wymagane pole

[Dalej: Dodaj dane](#)

Pliki

- Dowolnie wiele plików w zbiorze
- Różnorodność formatów

 / Zbiory danych / Utwórz zbiór danych

? Czym jest zasób?

Zasobem może być jakikolwiek plik zawierający pożyteczne dane.

1 Utwórz zbiór danych

2 Dodaj dane

Plik:  Załaduj plik

Nazwa:

Opis:

Opis zawartości tego konkretnego pliku.

Możesz tu używać formatowania Markdown

Format:

Licencja: [Poradnik prawny RepOD](#)

Poprzednia strona

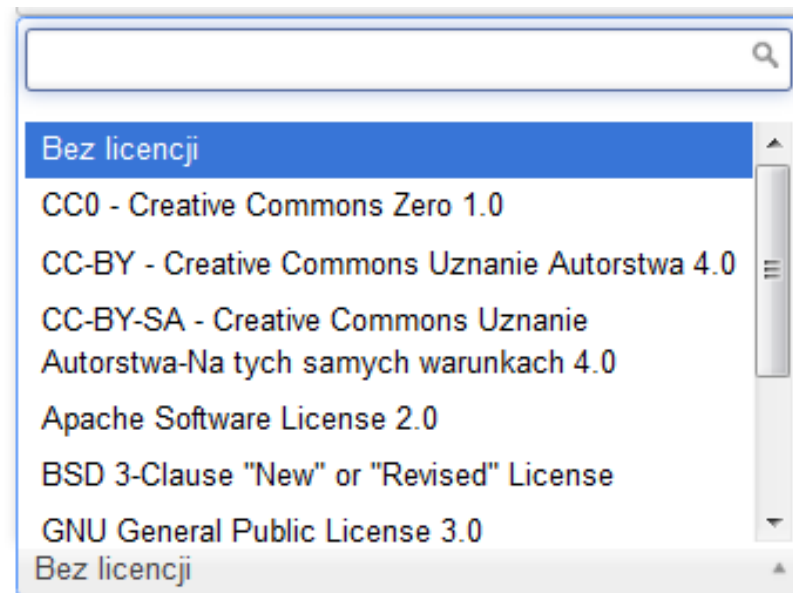
Zapisz jako kopię roboczą

Zapisz i dodaj kolejny

Publikuj zbiór danych

Licencje prawne

- bez licencji: dozwolony użytek
- oświadczenie CC0: domena publiczna
- otwarte licencje CC
- licencje otwartego oprogramowania



→ metadane oraz zbiór danych jako zestawienie plików są
uwolnione do domeny publicznej na mocy oświadczenia CC0

Po co umieszczać dane badawcze w repozytorium?

- bezpieczna archiwizacja
- dane są opisane i udokumentowane – łatwiej z nich skorzystać w przyszłości
- łatwość wyszukiwania przez wszystkich zainteresowanych
- dane uzyskują stabilny URL i/lub identyfikator – mogą być jednoznacznie cytowane w pracach naukowych

Dlaczego warto udostępniać dane w sposób **otwarty**?

The Importance of Sharing

Sens udostępniania

Po co otwierać dane badawcze?

- udostępnienie danych umożliwia **nowe analizy i interpretacje**
- otwarte dane można łączyć ze sobą, tworząc **nowe zestawienia**
- z otwartych danych mogą korzystać **naukowcy z innych dziedzin** oraz **osoby spoza środowiska akademickiego**, co daje szansę na ich **nowatorskie wykorzystanie**
- dzięki otwartym danym badania naukowe posuwają się **szybciej**, jest **więcej współpracy** naukowej
- łatwiej sprawdzić, czy opublikowane już prace naukowe opierają się na **powtarzalnych wynikach**, łatwiej także przeciwdziałać oszustwom i nierzetelności naukowej

Zespół RepOD w ICM UW

Programiści: Mateusz Neumann, Tomasz Błażejczyk

Webmaster: Monika Maksymiuk, Łukasz Pawełczak

Projekt graficzny: Jakub Rakusa-Suszczewski

Kwestie prawne: Krzysztof Siewicz

Koordinacja: Marta Hoffman-Sommer, Przemysław Kwiatkowski i Jakub Szprot

Dziękuję!

Kontakt:

msommer@icm.edu.pl

Serwis:

www.repod.pon.pl

Prezentacja jest dostępna na licencji Creative Commons - Uznanie autorstwa 3.0 Polska
(<http://creativecommons.org/licenses/by/3.0/pl/legalcode>)

