Gromadzenie i udostępnianie danych. Trend czy konieczność?

Krzysztof Szymański
Research Intelligence

19/10/2020

ELSEVIER

Dlaczego zarządzanie danymi badawczymi staje się ważne?

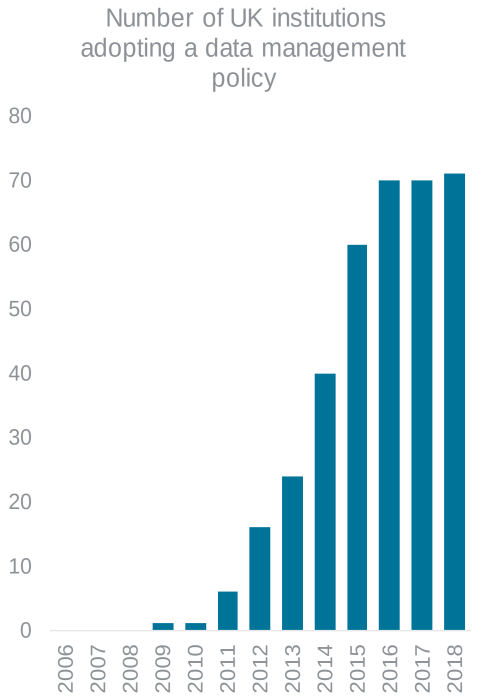
Trzy trendy łączą się, co prowadzi do przyspieszenia praktyk RDM

1) Obliczeniowe:
- Szybsze, łatwiejsze, tańsze, bardziej obliczeniowe metody prowadzenia badań
- Wraz z rozwojem analityki pojawiają się nowe możliwości wglądu w te same dane

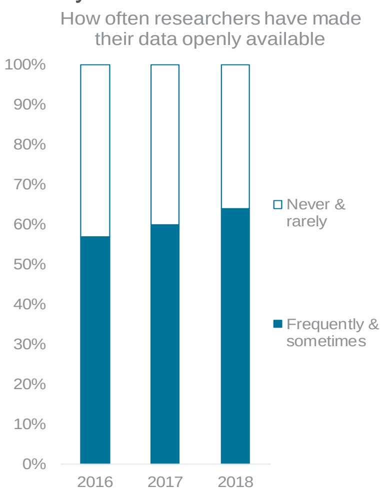
2) Fundatorzy badań:
- Agencje finansujące badania wprowadzają wymagania w zakresie udostępniania danych, mające na celu poprawę ponownego ich wykorzystania
- Następnie instytucje przyjmują polityki i plany zarządzania danymi

3) Naukowcy
- Nowa generacja naukowców, bardziej skoncentrowany na danych i ich dzieleniem się
- “Research Data is a first class citizen”

Przykład dla trendu nr 2



Przykład dla trendu nr 3:

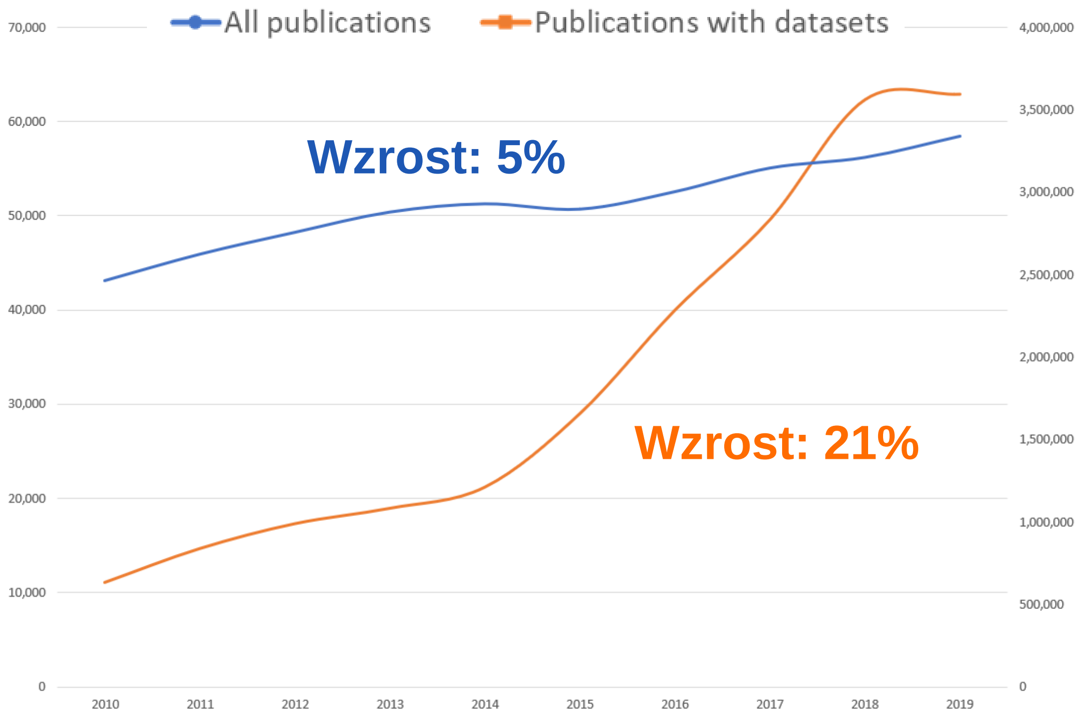


Jak naukowcy cenią sobie cytowanie danych

Data pierwszej publikacji	Tak samo lub więcej niż artykuł	Mniej niż cytowanie artykułu
<1990	64%	36%
1990s	65%	35%
2000s	60%	40%
2010s	72%	28%

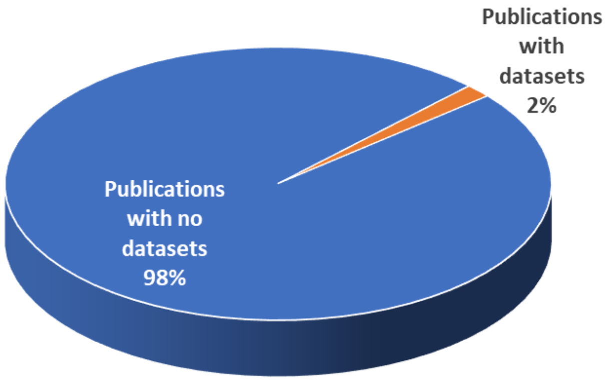
Źródła: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies
http://v2.sherpa.ac.uk/juliet/
https://figshare.com/articles/The_State_of_Open_Data_Report_2018/7195058

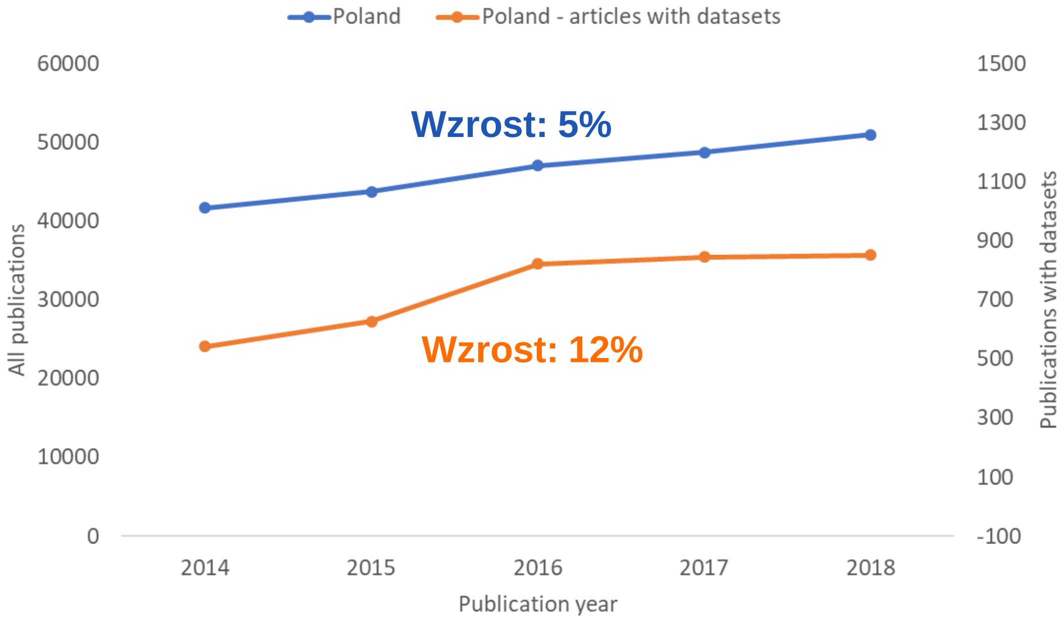
Udostępnianie danych rośnie szybko na całym świecie



Źródło: Scopus, Scholix 17/10/2020 – Wzrost = Compound Annual Growth Rate (CAGR)

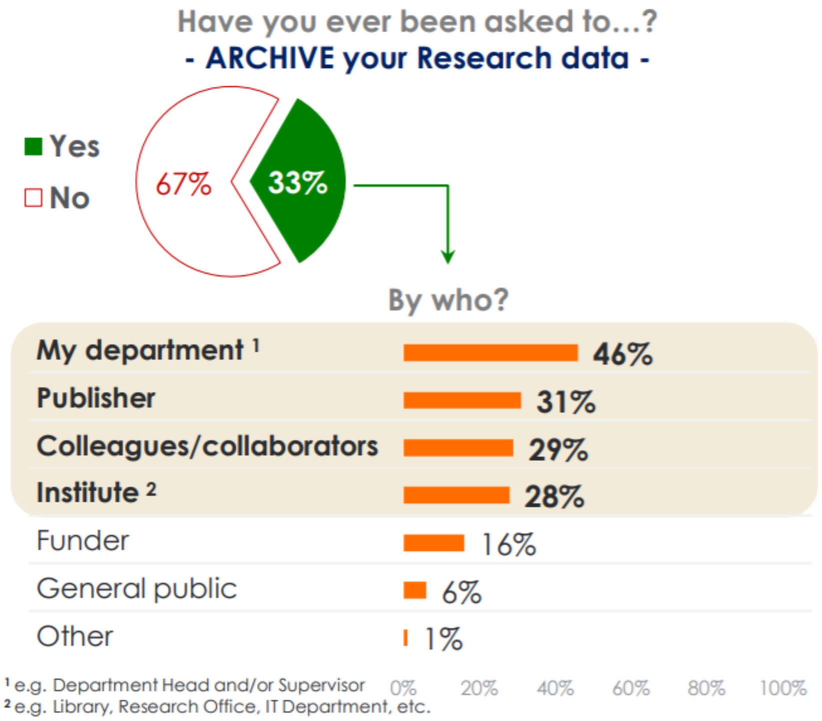
Również w Polsce

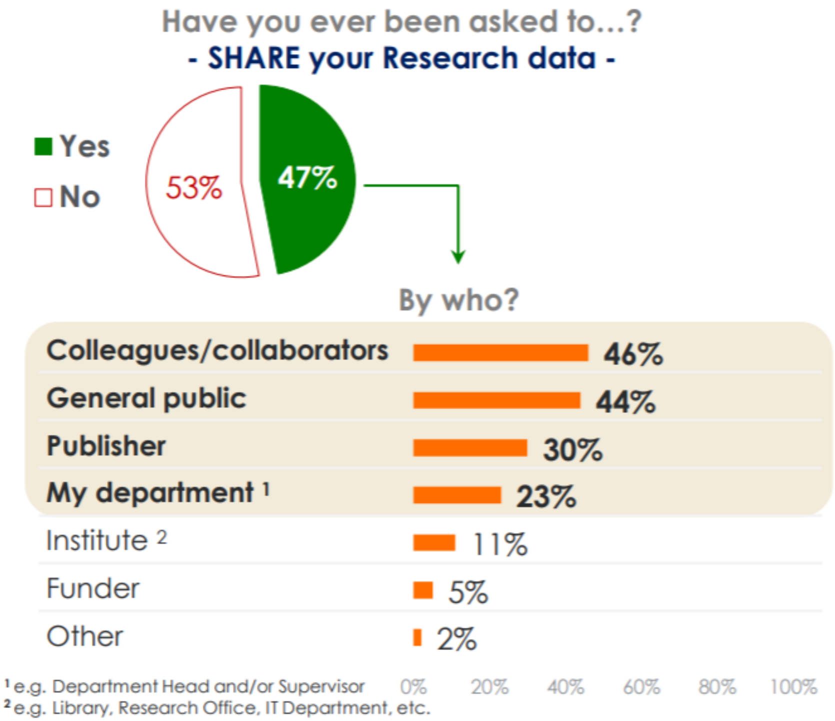




Source: SciVal, Scholix; 17/20/2020 – Wzrost = Compound Annual Growth Rate (CAGR)

Jak polscy naukowcy pracują z danymi badawczymi

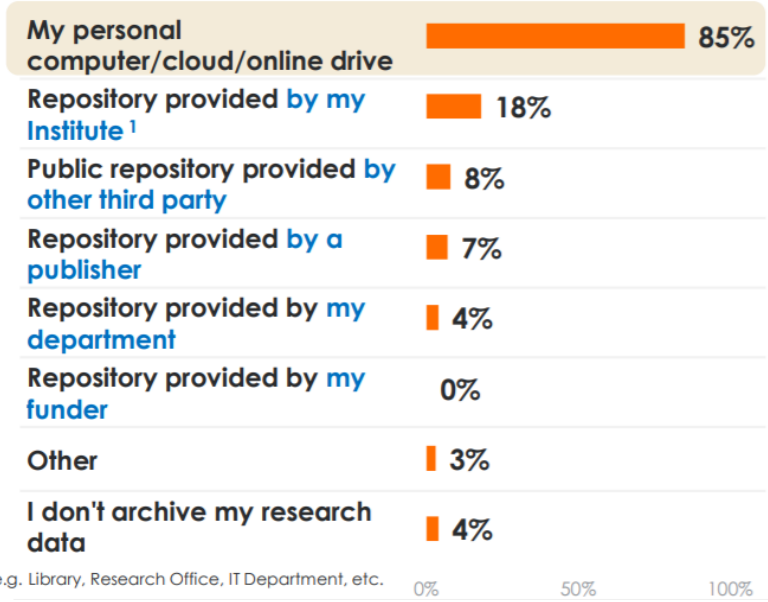




Źródło: Elsevier Online Survey, Q3 2019 – https://data.mendeley.com/datasets/nyzygp93zh/1

Jak polscy naukowcy pracują z danymi badawczymi

Where do you typically **ARCHIVE your Research data**?

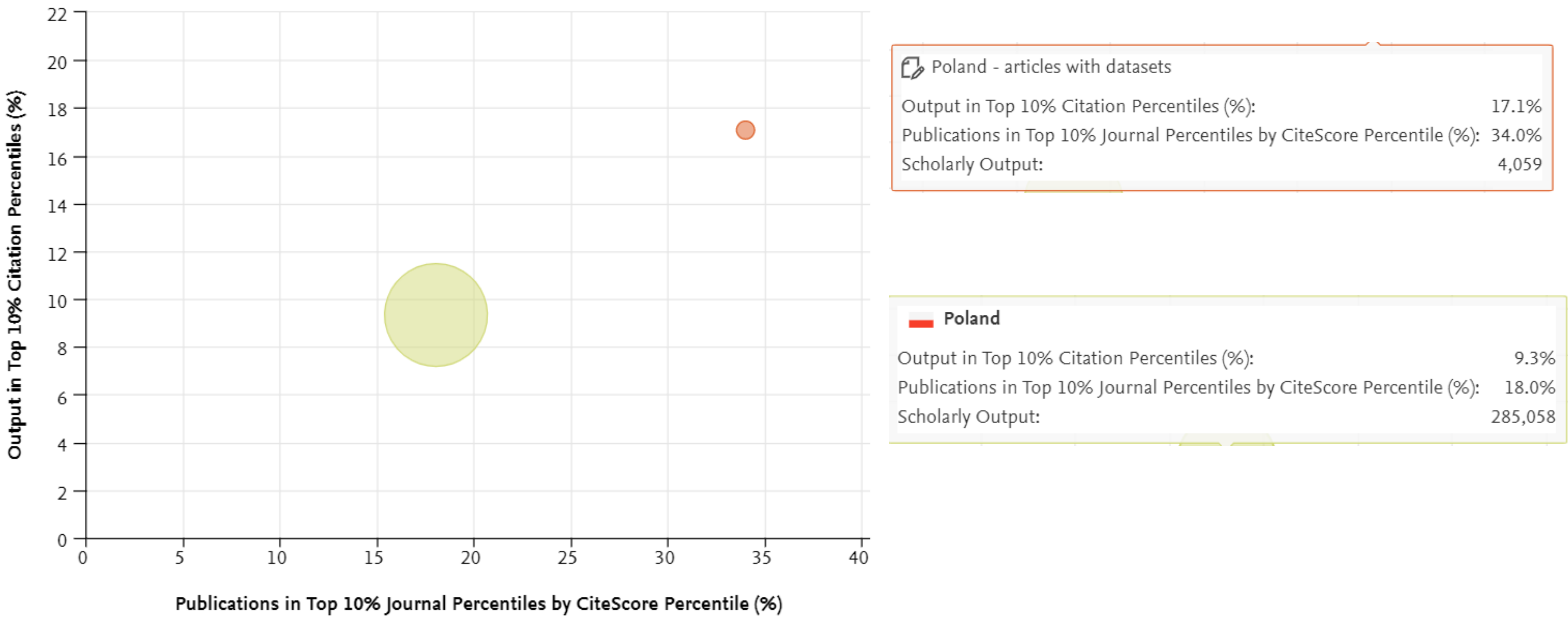


With who do you **SHARE Research data** that you used/created as part of your research project(s)

Źródło: Elsevier Online Survey, Q3 2019 – https://data.mendeley.com/datasets/nyzygp93zh/1

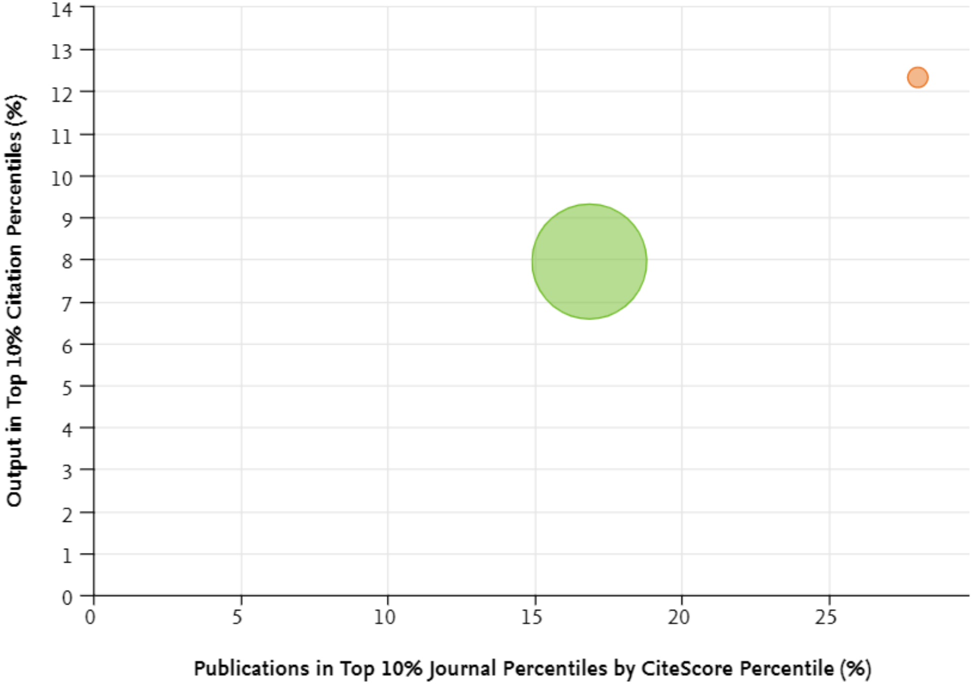
ELSEVIER

Udostępnianie danych daje większą widoczność



Source: SciVal, Scholix; 17/10/2020 – Lata 2015 - 2019

Udostępnianie danych daje większą widoczność



University of Łódź with Datasets

Output in Top 10% Citation Percentiles (%):	12.3%
Publications in Top 10% Journal Percentiles by CiteScore Percentile (%):	28.0%
Scholarly Output:	219

University of Łódź

Output in Top 10% Citation Percentiles (%):	7.9%
Publications in Top 10% Journal Percentiles by CiteScore Percentile (%):	16.8%
Scholarly Output:	8,207

Source: SciVal, Scholix; 17/10/2020 – Lata 2015 - 2019

... oraz 25% więcej cytowań

The citation advantage of linking publications to research data

Giovanni Colavizza[1,2,*], Iain Hrynaszkiewicz[3,4], Isla Staden[1,5], Kirstie Whitaker[1,6], Barbara McGillivray[1,6]

1 The Alan Turing Institute, UK.
2 University of Amsterdam, NL.
3 Springer Nature, UK.
4 Public Library of Science, UK.
5 Queen Mary University, UK.
6 University of Cambridge, UK.

Abstract

Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements in recent literature. Nevertheless, it is still unclear what proportion of these statements actually contain well-formed links to data, for example via a URL or permanent identifier, and if there is an added value in providing such links. We consider 531,889 journal articles published by PLOS and BMC, develop an automatic system for labelling their data availability statements according to four categories based on their content and the type of data availability they display, and finally analyze the citation advantage of different statement categories via regression. We find that, following mandated publisher policies, data availability statements become very common. In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository – rather than being available on request or included as supporting information files – are a fraction of the total. In 2017 and 2018, 20.8% of PLOS publications and 12.2% of BMC publications provided DAS containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% (± 1.07%) higher citation impact on average, using a citation prediction model. We discuss the potential implications of these results for authors (researchers) and journal publishers who make the effort of sharing their data in repositories. All our data and code are made available in order to reproduce and extend our results.

Źródło: https://arxiv.org/pdf/1907.02565.pdf

ELSEVIER

Wyzwanie: jak odnaleźć „nasze” dane badawcze?

W instytucji

Na przykład: dane badawcze w repozytorium instytucji

Poza instytucją

Na przykład: dane badawcze w repozytorium dziedzinowym

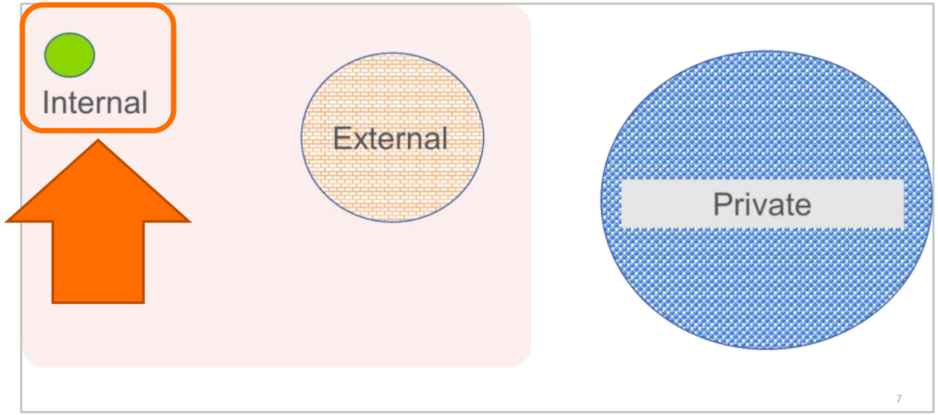
Publicznie udostępnione dane

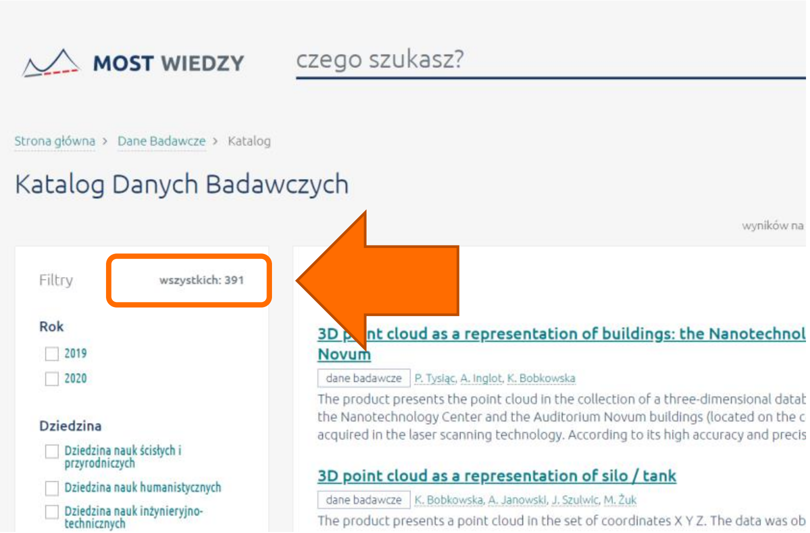
Prywatne zbiory danych

Na przykład: dane udostępniane prywatnie przez naukowców pracujących nad danym projektem

Przykład: Politechnika Gdańska

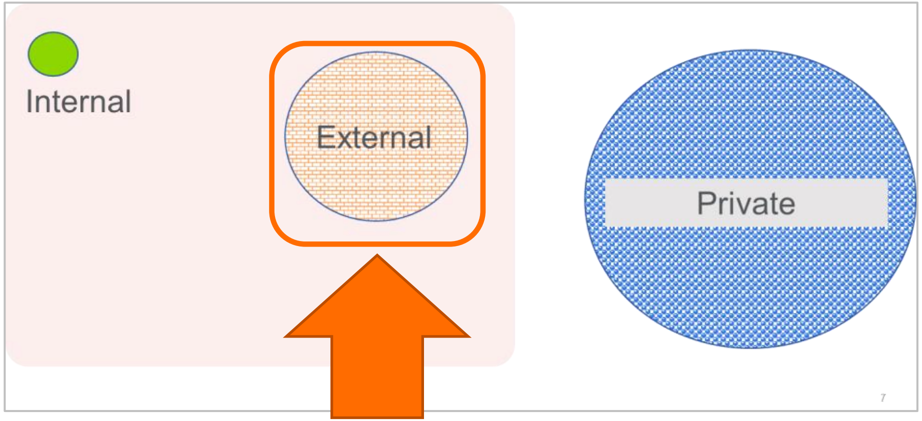
„MOST DANYCH – Multidyscyplinarny Otwarty System Transferu Wiedzy”
na dzień 17/10/2020 zawiera 391 zestawów danych

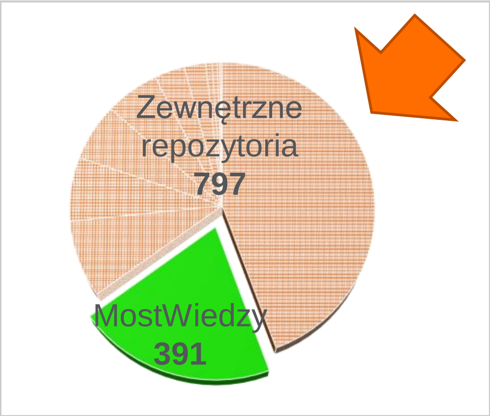




Źródło: https://mostwiedzy.pl/pl/open-research-data/catalog

W Politechnice Gdańskiej, 51% danych badawczych znajduje się poza Mostem Wiedzy

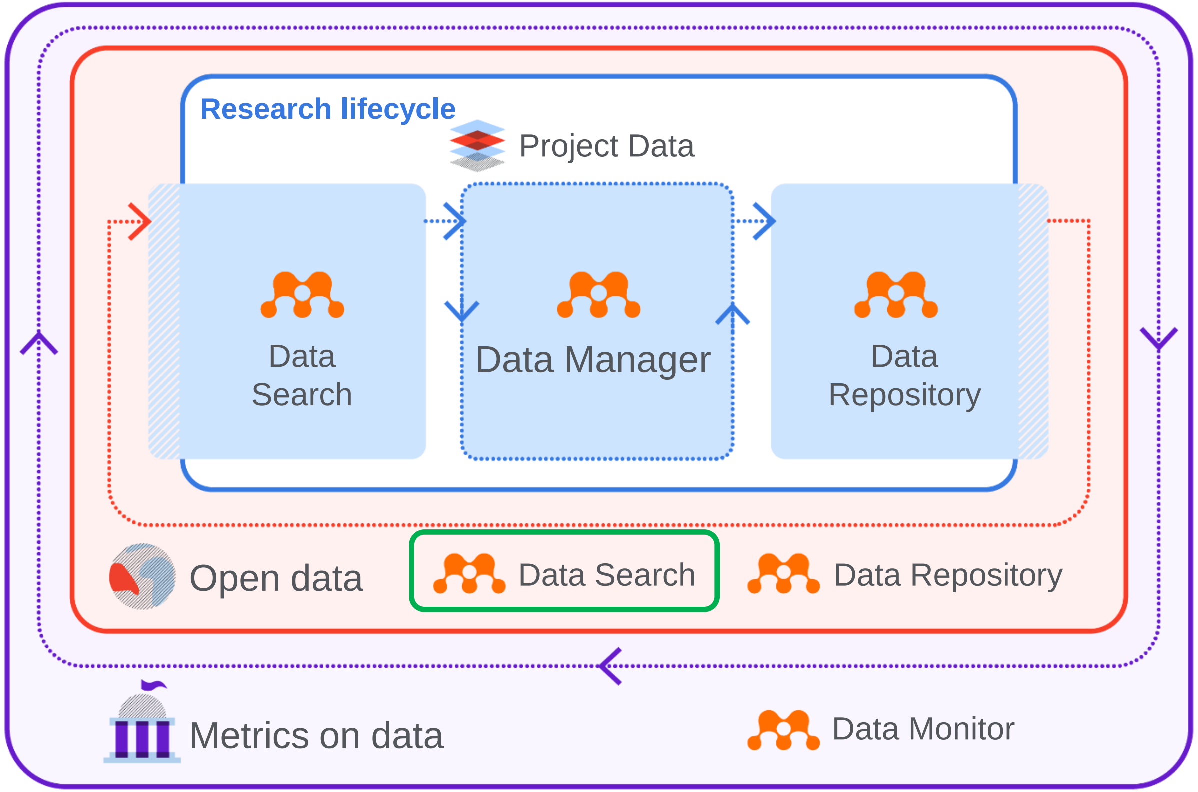




Jak znaleźć dane badawcze publikowane poza instytucjonalnym repozytorium danych?

ELSEVIER

Platforma Mendeley Data



Data Search

Odkryj otwarte dane

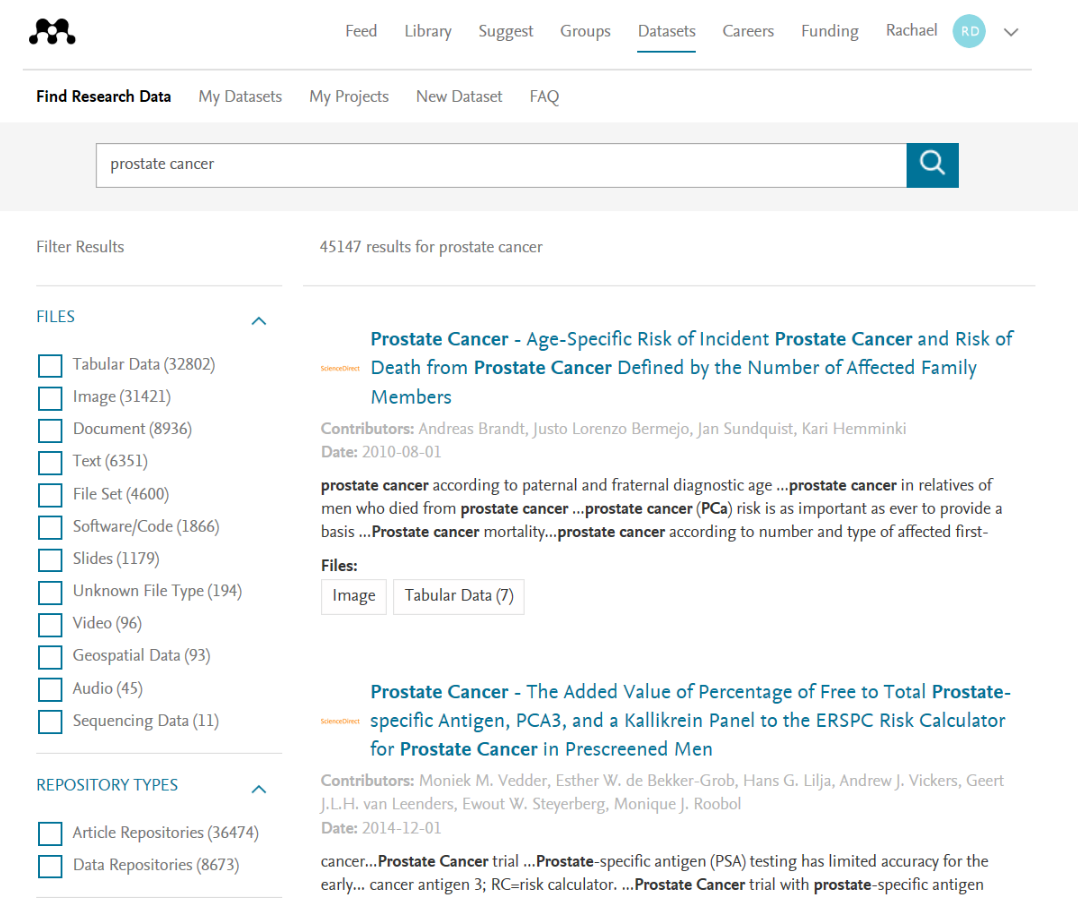
Odnaleź istotne dane

- Ponad 10 milionów zestawów danych indeksowanych w 40 największych repozytoriach dziedzinowych i ogólnych
- Głębokie indeksowanie pozwala na dotarcie do właściwych i istotnych wyników
- Przeszukiwanie nie tylko metadanych lecz całych zbiorów
- Zaawansowane filtrowanie wyników

Pobierz wybrane dane

- Linki do lokalnych repozytoriów
- Przeszukiwanie poprzez słowa kluczowe

Chcesz wiedzieć więcej?
data.mendeley.com/datasets



Jakość metadanych to podstawa

- Metadane z serwisów takich jak CrossRef czy DataCite nie wystarczą...

Pollen-based REVEALS estimates of plant cover in Europe for 36 grid-cells and the last 11700 years, supplement to: Marquer, Laurent; Gaillard, Marie-José; Sugita, Shinya; Poska, Anneli; Trondman, Anna-Kari; Mazier, Florence; Nielsen, Anne Birgitte; Fyfe, Ralph M; Jönsson, Anna Maria; Smith, Benjamin; Kaplan, Jed O; Alenius, Teja; Birks, H John B; Bjune, Anne Elisabeth; Christiansen, Jörg; Dodson, John; Edwards, Kevin J; Giesecke, Thomas; Herzschuh, Ulrike; Kangur, Mihkel; Koff, Tiiu; Latalowa, Malgorzata; Lechterbeck, Jutta; Olofsson, Jörgen; Seppä, Heikki (2017): Quantifying the effects of land use and climate on Holocene vegetation in Europe. Quaternary Science Reviews, 171, 20-37

Laurent Marquer, Marie-Jose Gaillard, Shinya Sugita, Anneli Poska, Anna-Kari Trondman, Florence Mazier, Anne Birgitte Nielsen, Ralph M Fyfe, Anna Maria Jönsson, Benjamin Smith, Jed O Kaplan, Teja Alenius, H John B Birks, Anne Elisabeth Bjune, Jörg Christiansen, John Dodson, Kevin J Edwards, Thomas Giesecke, Ulrike Herzschuh, Mihkel Kangur, Tiiu Koff, Malgorzata Latalowa, Jutta Lechterbeck, Jörgen Olofsson & Heikki Seppä

Supplementary Dataset published 2019 via PANGAEA - Data Publisher for Earth & Environmental Science

Brak informacji o afiliacji

DataCite

- A nawet dane w repozytoriach nie pomagają

Not logged in

PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH

Podobnie tutaj.

Citation: **Marquer, Laurent; Gaillard, Marie-José; Sugita, Shinya; Poska, Anneli; Trondman, Anna-Kari; Mazier, Florence; Nielsen, Anne Birgitte; Fyfe, Ralph M; Jönsson, Anna Maria; Smith, Benjamin; Kaplan, Jed O; Alenius, Teja; Birks, H John B; Bjune, Anne Elisabeth; Christiansen, Jörg; Dodson, John; Edwards, Kevin J; Giesecke, Thomas; Herzschuh, Ulrike; Kangur, Mihkel; Koff, Tiiu; Latalowa, Malgorzata; Lechterbeck, Jutta; Olofsson, Jörgen; Seppä, Heikki (2019):** Pollen-based REVEALS estimates of plant cover in Europe for 36 grid-cells and the last 11700 years. *PANGAEA*, https://doi.org/10.1594/PANGAEA.900966,

ELSEVI

www.scholix.org

SCHOLIX przychodzi z rozwiązaniem

- ✓ Wyszukiwanie danych badawczych
- ✓ Wzbogacanie metadanych

ELSEVIER

Quaternary Science Reviews
Volume 171, 1 September 2017, Pages 20-37

Quantifying the effects of land use and climate on Holocene vegetation in Europe

10.1016/j.quascirev.2017.07.001

Show more

https://doi.org/10.1016/j.quascirev.2017.07.001 Get rights and content

Jest wymieniony na

PANGAEA.
Data Publisher for Earth & Environmental Science

Marquer, Laurent; Gaillard, Marie-José; Sugita, Shinya; Poska, Anneli; Trondman, Anna-Kari; Mazier, Florence; Nielsen, Anne Birgitte; Fyfe, Ralph M; Jönsson, Anna Maria; Smith, Benjamin; Kaplan, Jed O; Alenius, Teja; Birks, H John B; Bjune, Anne Elisabeth; Christiansen, Jörg; Dodson, John; Edwards, Kevin J; Giesecke, Thomas;

10.1594/PANGAEA.900966

Scopus

Quaternary Science Reviews
Volume 171, 1 September 2017, Pages 20-37

Quantifying the effects of land use and climate on Holocene vegetation in Europe (Article)

Marquer, L.[a,b,e], Gaillard, M.-J.[b], Sugita, S.[c], Poska, A.[a,d], Trondman, A.-K.[b], Mazier, F.[e], Nielsen, A.B.[a,f], Fyfe, R.M.[g], Jönsson, A.M.[a], Smith, B.[a], Kaplan, J.O.[h], Alenius, T.[i,j], Birks, H.J.B.[k,l], Bjune, A.E.[k,m], Christiansen, J.[n], Dodson, J.[o,p], Edwards, K.J.[q,r], Giesecke, T.[n], Herzschuh, U.[s], Kangur, M.[c], Koff, T.[c], Latałowa, M.[t], Lechterbeck, J.[u], Olofsson, J.[a], Seppä, H.[v]

Hide additional authors Save all to author list

[a]Department of Physical Geography and Ecosystem Science, Lund University, Sweden
[b]Department of Biology and Environmental Science, Linnaeus University, Sweden
[c]Institute of Ecology, Tallinn University, Estonia
[d]Institute of Geology, Tallinn University of Technology, Estonia
[e]GEODE, UMR-CNRS 5602, Université Toulouse Jean Jaurès, France
[f]Department of Geology, Lund University, Sweden
[g]School of Geography, Earth and Environmental Sciences, University of Plymouth, United Kingdom
[h]ARVE, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[i]Department of Philosophy, History, Culture and Art Studies, University of Helsinki, Finland
[j]Department of Archaeology, University of Turku, Finland
[k]Department of Biology and Bjerknes Centre for Climate Research, University of Bergen, Norway
[l]Environmental Change Research Centre, University College London, United Kingdom
[m]Uni Research Climate, Bergen, Norway
[n]Department of Palynology and Climate Dynamics, Albrecht-von-Haller-Institute for Plant Sciences, University of Göttingen, Germany
[o]School of Earth and Environmental Sciences, University of Wollongong, Wollongong, Australia
[p]State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, Shaanxi, China
[q]Departments of Geography & Environment and Archaeology, School of Geosciences, University of Aberdeen, United Kingdom
[r]Clare Hall, University of Cambridge, United Kingdom
[s]Alfred-Wegener-Institut Potsdam and Institut für Geowissenschaften, Universität Potsdam, Germany
[t]Laboratory of Palaeoecology and Archaebotany, Department of Plant Ecology, University of Gdańsk, Poland
[u]Arkeologisk Museum, Universitetet i Stavanger, Stavanger, Norway

ELSEVIER

Author IDs

Mendeley Data

PANGAEA.

Pollen-based REVEALS estimates of plant cover in Europe for 36 grid-cells and the last 11700 years

Contributors: Marquer, Laurent, Gaillard, Marie-José, Sugita, Shinya, Poska, Anneli, Trondman, Anna-Kari, Mazier, Florence, Nielsen, Anne Birgitte, Fyfe, Ralph M, Jönsson, Anna Maria, Smith, Benjamin

Date: 2019-05-02

327027 ... Abstract: This dataset corresponds to the pollen-based REVEALS estimates of 25 plant taxa for Europe and associated standard errors, published in Marquer et al. (2017). This is part of the results from the Swedish project LandClim I (Gaillard et al., 2010; Trondman et al., 2015; Marquer et

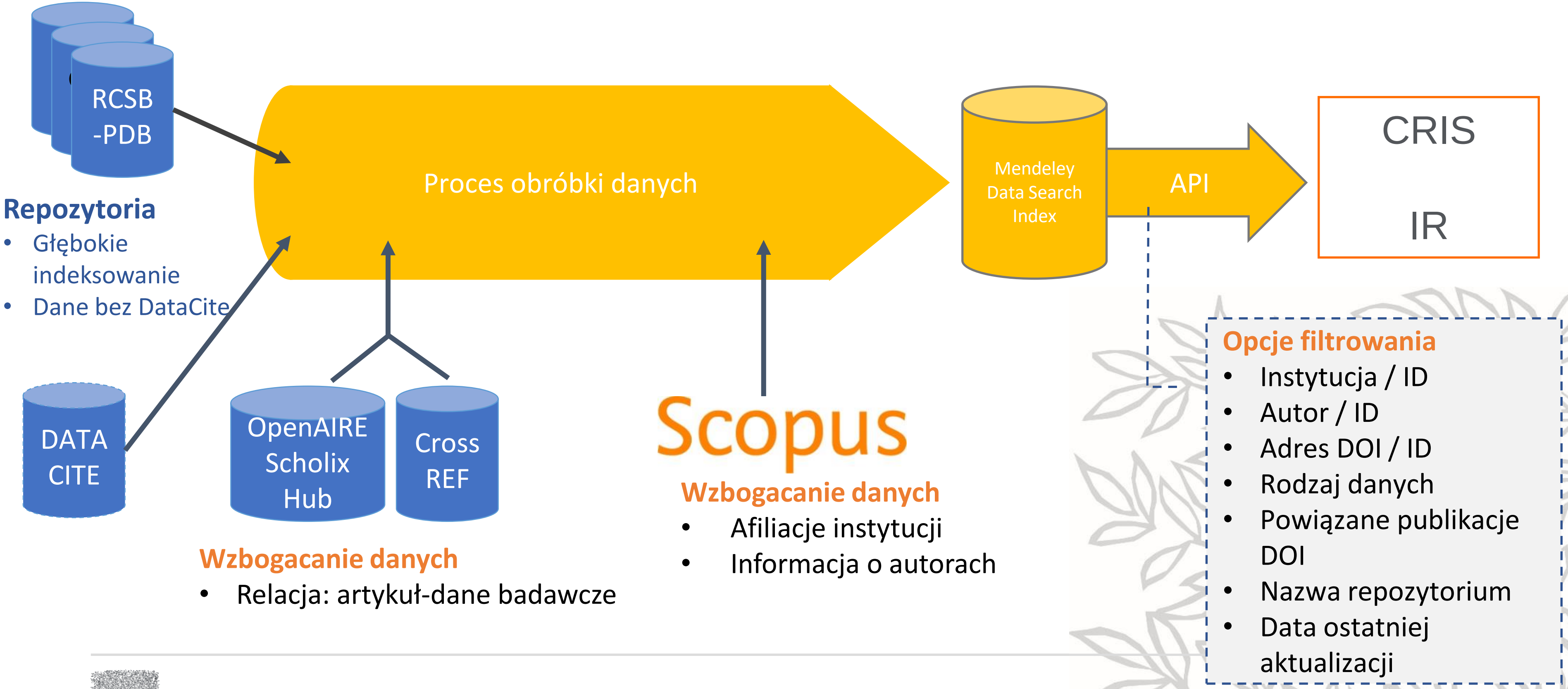
?

Afiliacje, IDs

Files:

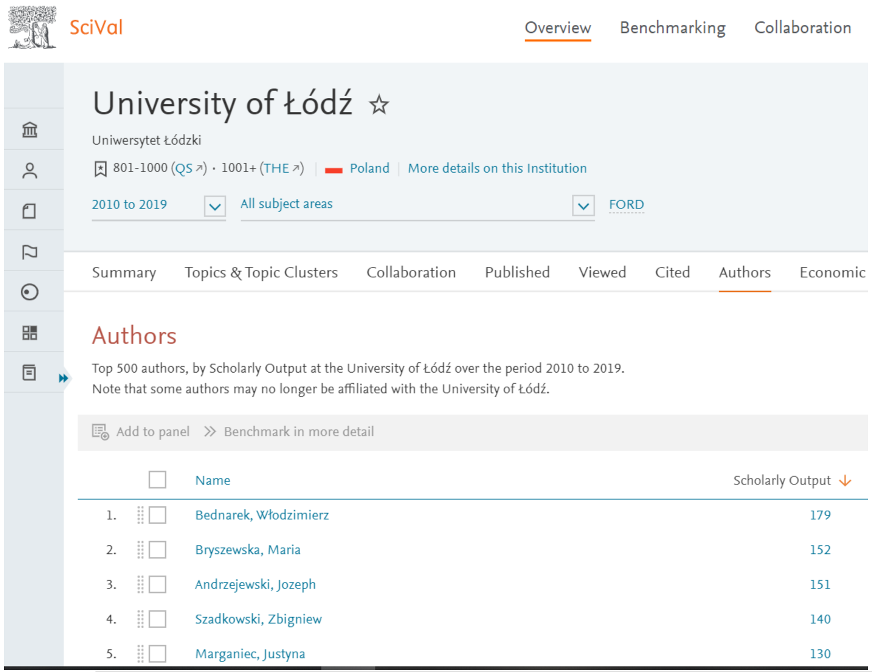
Tabular Data

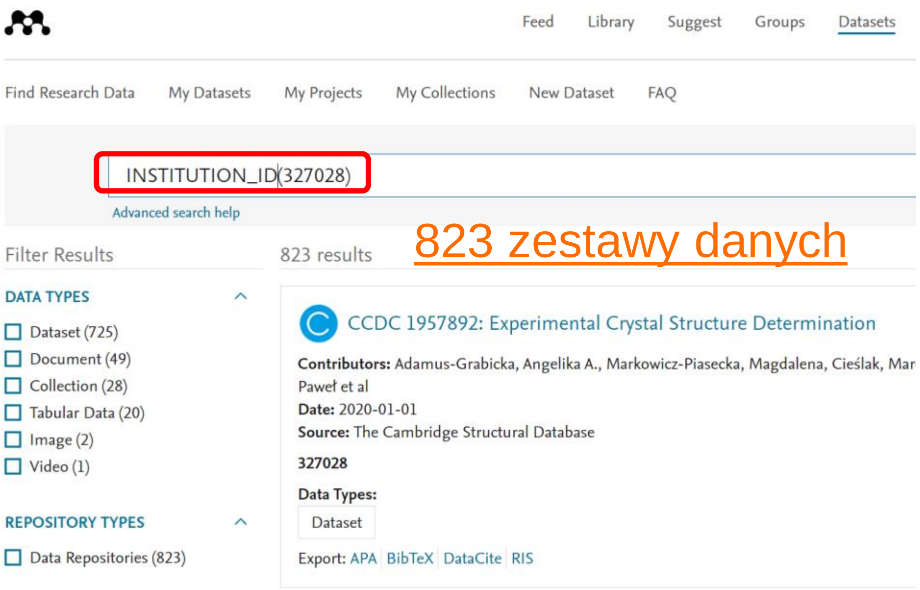
Jak działa nasze wsparcie: zbieranie danych+ API



Przykład: Uniwersytet Łódzki: SciVal ID 327028

scival.com/overview/summary?uri=Institution/327028

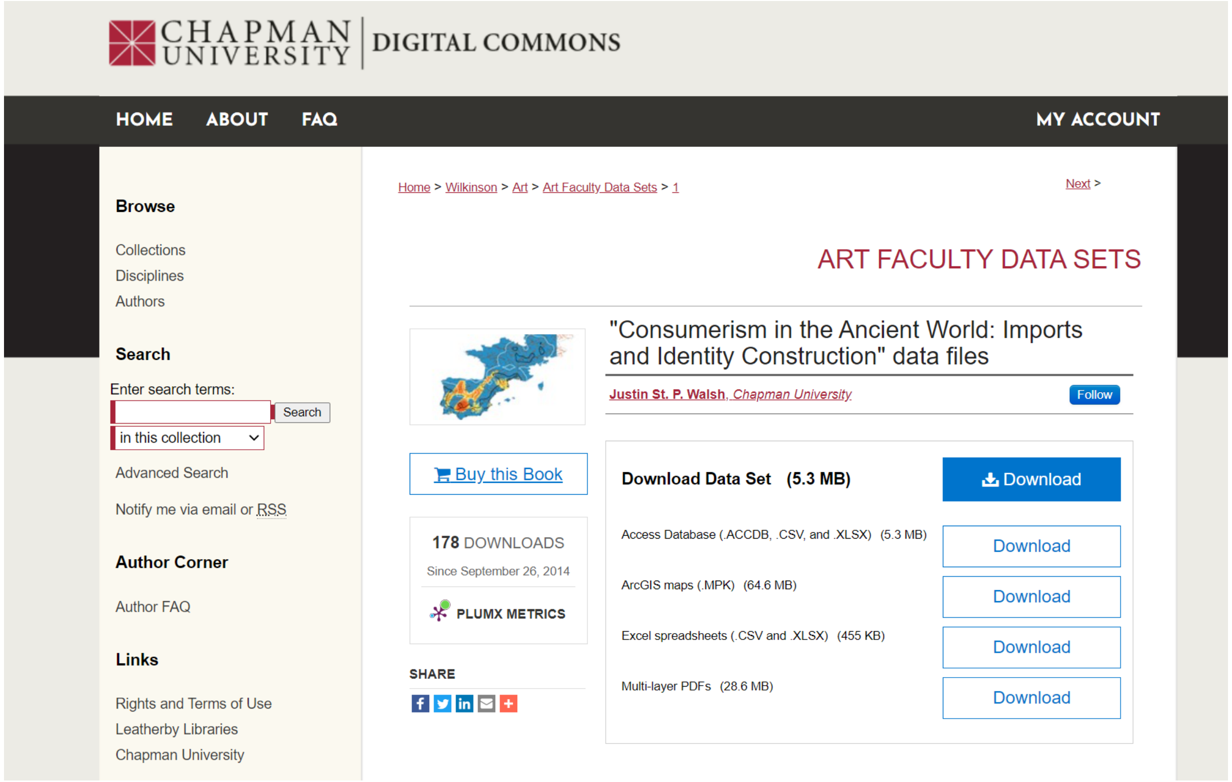




823 zestawy danych

ELSEVIER

Przykład wykorzystania repozytorium dla danych badawczych



Statystyki wykorzystania danych badawczych - przykłady

"Consumerism in the Ancient World: Imports and Identity Construction" data files

1,122 Usage | 2 Social Media

Citation Data: Art Faculty Datasets
Publication Year: 2014

Home
Overview
Highlights
Twitter

Metrics Details

USAGE	1,122
Abstract Views	944
Digital Commons	944
Downloads	178
Digital Commons	178
SOCIAL MEDIA	2
Tweets	2
Twitter	2

Most Recent Tweet

See all tweets

Justin Walsh
@jstpwalsh
Replying to @jstpwalsh
If the eResources ever go away from Routledge again, they are permanently available on the Chapman Digital Commons

"Consumerism in the Ancient World: Imports and Identity ...
digitalcommons.chapman....

Data Description

These files were created for the book Consumerism in the Ancient World: Imports and Identity Construction (Routledge 2014). They include almost 24,000 Greek vases or fragments of vases dating ca. 800-300 BCE from 233 sites in Portugal, Spain, France, Germany, and Switzerland. The maps illustrate the results of a geostatistical technique called kriging, which predicts distributions of vases on the basis of samples. The maps are available ... ArcGIS format or in reduced ver...

Show more

Name	Type	Downloads
Public Assignment	Commercial	96
The University of Cape Coast	Education	76
University of Kwazulu-Natal	Education	60
Iowa State University	Education	59
Clayton County Public Schools	Education	58
Ministry of Information and Communications Government of Sierra Leone	Government	55
Backbone Malabo	Commercial	46
Super iMax	Commercial	42
Halliburton Company	Commercial	37
Makerere University Main Campus	Education	35

Readership distribution

Referrer	Downloads
www.google.com/	15735
www.google.co.uk/	1417
www.google.com.gh/	692
lib.dr.iastate.edu/	606
www.google.com/url	606
www.google.ca/	546
www.google.com/search	387
www.google.co.za/	380
com.google.android.googlequicksearchbox	356
www.google.co.in/	332

ELSEVIER

Inne przykłady z Polski (lata 2010-2019)

		Publikacje z danymi	Publikacje ogółem	Udział w całości	Dane badawcze
1	Polska Akademia Nauk	1,554	57101	3%	3760
2	Uniwersytet Wrocławski	910	9699	9%	3974
3	Uniwersytet Adama Mickiewicza	845	12802	7%	4094
4	Uniwersytet Jagielloński	689	27806	2%	1996
5	Uniwersytet Warszawski	648	22826	3%	2615
6	Uniwersytet Śląski	323	8296	4%	1350
7	Politechnika Wrocławska	277	16228	2%	1093
8	Poliechnika Łódzka	267	9496	3%	1105
9	Poliechnika Warszawska	258	19940	1%	1137
10	Uniwersytet Mikołaja Kopernika	247	9959	2%	688

ELSEVIER

Integracja z CRIS czy IR (przykład z Pure)

New Reporting BETA

Compliance report

Elsevier (demo e

Create new workspace

Expand mode Save Save as Share Exp

Projects | Awards | Research outputs | Datasets

	A	B	C	D	E	F
	Project Title	H2020 Contract ID	Article DOI	Article Title	Associated Dataset DOI	Associated Dataset Title
1	EUROFusion	633053	10.1080/10420151003729870	Single-crystal diamond MIS diode for deep UV d...		
2	EUROFusion	633053	10.1088/1361-6587/aac60f	Feasibility of a far infrared laser based polarimet...		
3	EUROFusion	633053	10.1016/j.fusengdes.2018.05.065	Progress in the design of the superconducting ...	10.17632/gy7bfzcgyd.3	Datasets for Potential of Robotic Lab Methods ...
4	EUROFusion	633053	10.1016/j.ssi.2018.06.008	Carbon powder material obtained from an innov...	10.6084/m9.figshare.5900077.v1	Additional file 2: of Agrobacterium-mediated a...
5	EUROFusion	633053	10.1016/j.fusengdes.2018.05.053	Shutdown dose rate neutronics experiment duri...		
6	EUROFusion	633053	10.1016/j.fusengdes.2018.04.001	DEMO design activity in Europe: Progress and u...	10.17632/g75gcpp49k.1	Citation Contexts for Nobel Prize Winning Pap...
7	EUROFusion	633053	10.1088/1361-6587/aaaa49	Reconstruction of flux coordinates from discretiz...		
8	EUROFusion	633053	10.1051/epjconf/201715307028	Neutron streaming studies along JET shielding ...	10.17632/btchxktzyw.1	Supplementary data tables for "A standardized ...
9	EUROFusion	633053	10.1088/1741-4326/aa6412	Overview of the TCV tokamak program: Scientifi...		
10	EUROFusion	633053	10.1016/j.nme.2017.05.010	LIBS experiments for quantitative detection of re...	10.5517/ccxzls7	CCDC 863560: Experimental Crystal Structure ...
11	EUROFusion	633053	10.1016/j.esr.2016.11.002	Fusion power in a future low carbon global elect...		
12	EUROFusion	633053	10.1063/1.4940785	Structure of wave-particle resonances and Alfvén...		

Zamiast podsumowania

Plany Zarządzania Danym Badawczymi
https://dcc.ac.uk/guidance/how-guides/develop-data-plan

Różne polityki dotyczące zarządzania danymi
https://dcc.ac.uk/guidance/policy

Kurs – Zarządzanie Danymi Badawczymy – jęz. angielski
https://rdmla.github.io/

Czy warto mieć „data stewards”?
https://www.openaire.eu/item/the-role-and-value-of-data-stewards-in-universities-a-tu-delft-case-study-on-data-stewardship
https://www.youtube.com/watch?v=YwFkOK-n4fc&feature=youtu.be

Od czego zacząć:
https://dcc.ac.uk/guidance/curation-lifecycle-model]

Pytania?

Więcej szkoleń:
https://www.elsevier.com/pl-pl/promo/szkoleniaonline

Kontakt:
Krzysztof Szymański
k.szymanski@elsevier.com