



Otwarte dane badawcze

Natalia Gruenpeter

Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego
Uniwersytetu Warszawskiego, Krajowe Biuro Otwartego Dostępu OpenAIRE

CC-BY

Treść licencji dostępna na stronie:

<https://creativecommons.org/licenses/by/4.0/legalcode.pl>



Platforma Otwartej Nauki

- wsparcie badaczy i instytucji w otwartym udostępnianiu wyników badań
- rozwijanie infrastruktury otwartej nauki
- analiza otwartej komunikacji naukowej i popularyzacja wiedzy o otwartej nauce

pon.edu.pl

[Facebook](#) | [Twitter](#)

otwartanauka.pl



RepOD

Repozytorium Otwartych Danych

Zarządzanie danymi badawczymi

2019:
15 warsztatów
~ 250 uczestników

2020:
5 szkoleń online
~ 250 uczestników



OpenAIRE, Krajowe Biuro Otwartego Dostępu

OpenAIRE's National Open Access Desks (NOADs) are gathering and updating information on open science **POLICY-INFRASTRUCTURE-TRAINING** on a periodic basis, ensuring an up-to-date feed.

- wsparcie badaczy i instytucji w otwartym udostępnianiu wyników badań
- materiały informacyjne i edukacyjne

openaire.eu/support

openaire.eu/guides



Austria	Lithuania
Belgium	Luxembourg
Bulgaria	Malta
Croatia	The Netherlands
Cyprus	Norway
Czech Republic	Poland
Denmark	Portugal
Estonia	Romania
Finland	Serbia
France	Slovakia
Germany	Slovenia
Greece	Spain
Hungary	Sweden
Iceland	Switzerland
Ireland	Turkey
Israel	United Kingdom
Italy	Latvia

Wprowadzenie

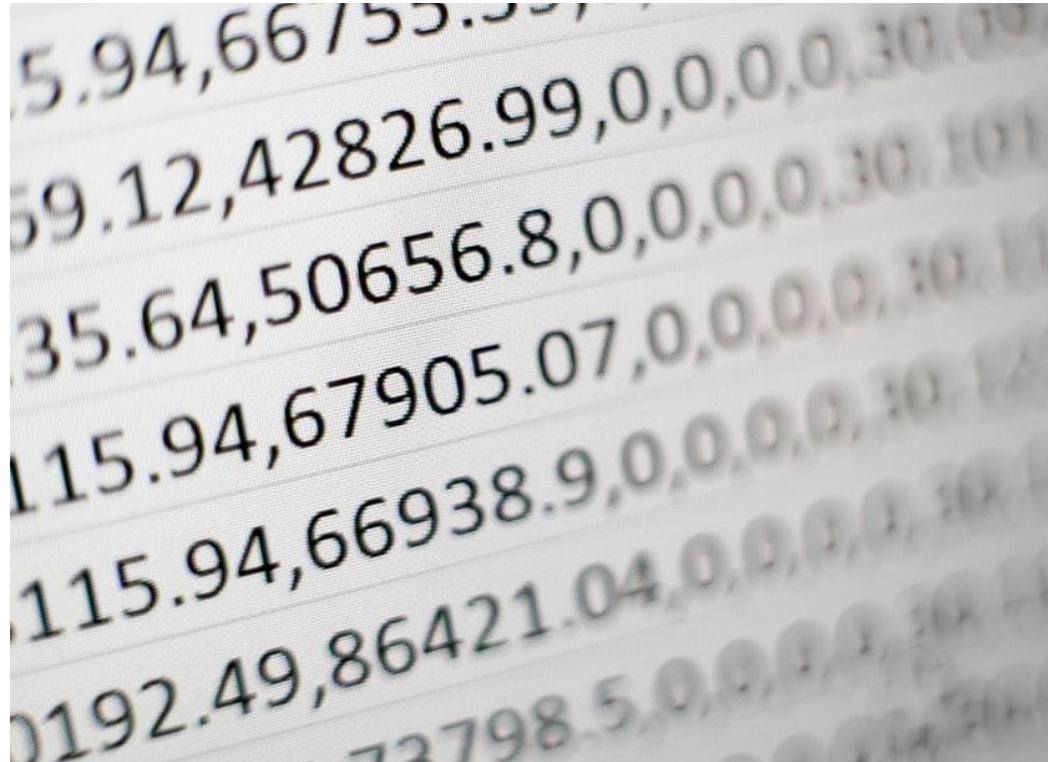
Czym są dane badawcze?



Czym są dane badawcze?

- ❖ Zarejestrowane materiały o charakterze faktograficznym, powszechnie uznawane przez społeczność naukową za niezbędne do oceny wyników badań naukowych.
- ❖ Dane zebrane, zaobserwowane lub wytworzone jako materiał do analizy, w celu uzyskania oryginalnych wyników naukowych.
- ❖ Wszystko co zostało wyprodukowane lub wytworzone w ramach prowadzonych badań.

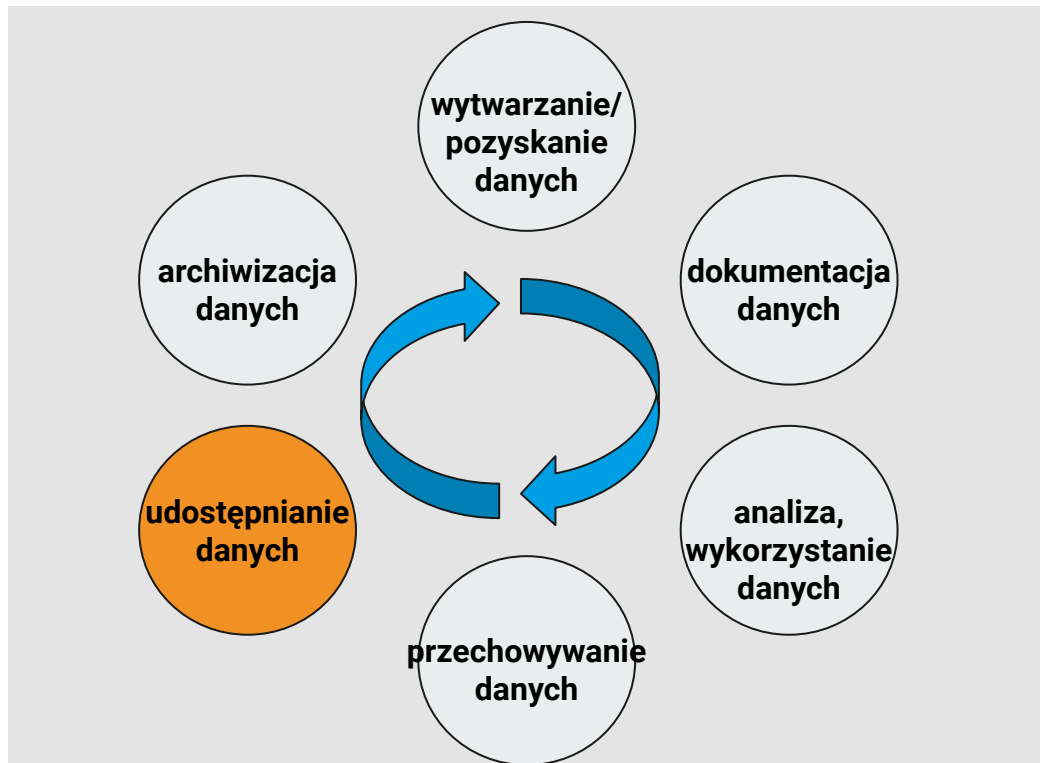
- dane liczbowe
- dokumenty tekstowe, notatki
- kwestionariusze, wyniki badań ankietowych
- fotografie, nagrania audio i wideo
- zawartość baz danych
- oprogramowanie
- wyniki symulacji komputerowych
- protokoły laboratoryjne, opisy metodologiczne
- ...



Korzyści płynące z otwartego udostępniania danych badawczych

Dane w całym cyklu projektu badawczego

Zarządzanie danymi badawczymi to sposób, w jaki organizujemy pracę z danymi badawczymi w toku całego projektu i po jego zakończeniu.





Otwarte dane badawcze

❖ właściwa organizacja danych:

→ ułatwia korzystanie z danych w przyszłości lub udostępnienie ich (np. na prośbę),

→ minimalizuje ryzyko w nieprzewidzianych sytuacjach (utrata lub zniszczenie danych/sprzętu)

❖ otwarte udostępnianie danych:

→ oszczędza czas, który trzeba poświęcić na obsługę wniosków o udostępnienie danych



Otwarte dane badawcze

❖ **promocja dorobku naukowego**

→ większa widoczność

→ wiarygodność i możliwość
weryfikacji wyników badań

❖ **ponowne wykorzystanie danych**

→ możliwość przeprowadzenia
nowych analiz z wykorzystaniem
udostępnionych danych

"A love letter to your future self": What scientists need to know about FAIR data

Following these guiding principles for sharing data can help researchers get ahead.

Jon Brock

11 FEBRUARY 2019

Nature Index 360°

The idea that scientific data should be **FAIR** — Findable, Accessible, Interoperable, and Reusable — is one increasingly endorsed by scientific institutions including the **United States National Academies of Science Engineering and Medicine**, the **European Commission**, and the Wellcome Trust. But it is yet to gain much traction among the people that ultimately matter, the scientists generating the data. The 2018 State of Open Data report, published by Digital Science, found that just 15% of researchers were “familiar with FAIR principles”. (Digital Science is operated by the Holtzbrinck Publishing Group, which also has a majority share in the publisher of the Nature Index.)

- “FAIR means thinking about the people who could benefit from your data,”
- FAIR can also be good for career advancement, particularly for early-career researchers.
- “FAIR helps you demonstrate the impact of your research when people re-use and cite your dataset,” (...) “It gets your name out there and can lead to new collaborations.”



Otwarte dane badawcze

- ❖ **ponowne wykorzystanie danych**
 - rozwój społeczny i ekonomiczny, innowacje, rozwiązanie globalnych problemów
- ❖ **zaufanie do nauki**
 - przejrzystość i wiarygodność, możliwość weryfikacji wyników badań

Scientists are unraveling the Chinese coronavirus with unprecedented speed and openness



Medical workers of the Union Hospital with the Tongji Medical College of the Huazhong University of Science and Technology in Wuhan participate in a ceremony to form an "assault team" to battle a coronavirus epidemic in China. (Cheng Min/Xinhua/AP)

By Carolyn Y. Johnson

Jan. 24, 2020 at 10:17 p.m. GMT+1

Just 10 days after a [pneumonia-like illness](#) was first reported among people who visited a seafood market in [Wuhan](#), China, scientists released the genetic sequence of [the coronavirus](#) that sickened them. That precious bit of data, freely available to any researcher who wanted to study it, unleashed a massive collaborative effort to understand the mysterious new pathogen that has been rapidly spreading in [China](#) and [beyond](#).

Ad closed by Go

Top universities ink data-sharing pact as virus spread shows need

Research-intensive university groups ask funders to provide additional support for data management

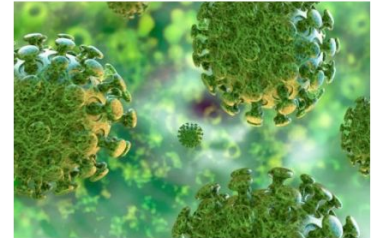
January 28, 2020

By John Ross

Twitter: @JohnRoss49

The coronavirus crisis exemplifies why researchers must systematically disclose their underpinning data, according to Australia's Group of Eight (Go8) universities.

Go8 chief executive Vicki Thomson said that research already completed could yield information capable of speeding up vaccine development, but health authorities would struggle to find it. "How do they know, if there's no database for them to access?" she asked.



Source: iStock

The coronavirus has sparked discussion at [Sorbonne University](#) in Paris, where the Go8 and counterpart organisations from eight regions – including the UK's Russell Group, the Association of American Universities, the League of European Research Universities (Leru), Japan's Research University 11 group and the African Research Universities Alliance – have gathered to thrash out a joint position on data sharing.

Deklaracja sorbońska 27.01.2020

w sprawie danych badawczych;

podpisana przez przedstawicieli 9 sieci
uniwersytetów badawczych z całego świata



Association
of American
Universities
Inquiry · Innovation · Impact



ARUA
African Research
Universities Alliance

CURIF

GERMAN U15



GROUP
OF EIGHT
AUSTRALIA

LE
RU



RUSSELL
GROUP

U¹⁵
Group of Canadian Research Universities
Regroupement des universités de recherche du Canada

Sorbonne declaration on research data rights

Based on the following principles:

- Knowledge derived from research benefits society.
- The value of research data lies on its integrity, upon which the public trust in new knowledge is founded.
- Providing access to and openly sharing data enable the development of new knowledge, accelerates discoveries for the benefit of society and economic development.
- Research data should, as much as possible, be shared openly and reused, without compromising national security, institutional autonomy, privacy, indigenous rights and the protection of intellectual property.
- The academic community is integral to identifying the complex conditions for sharing and reuse.



<https://www.leru.org/files/Sorbonne-declaration.pdf>

Dane badawcze w politykach otwartości



Polityki otwartości

→ mogą dotyczyć publikacji i/lub danych badawczych

w politykach otwartości zakres danych wymaganych do udostępnienia może być ograniczony do danych niezbędnych do weryfikacji wyników badań zaprezentowanych w publikacjach

→ wymogi grantodawcy, np. Komisja Europejska, program Horyzont 2020

→ polityka przyjęta przez wydawcę czasopisma

→ wymogi pracodawcy, uczelni lub instytutów naukowo-badawczych

→ wymogi prawne: dane badawcze jako dane sektora publicznego



Wymogi Narodowego Centrum Nauki

DANE BADAWCZE

- ❖ sporządzanie planu zarządzania danymi badawczymi,
- ❖ otwarte udostępnianie danych powiązanych z opublikowanymi artykułami

[https://www.ncn.gov.pl/finansowanie-nauki/
otwarta-nauka](https://www.ncn.gov.pl/finansowanie-nauki/otwarta-nauka)



DMP w wytycznych Narodowego Centrum Nauki

- ❖ opis danych oraz pozyskiwanie lub ponowne wykorzystanie dostępnych danych
- ❖ dokumentacja i jakość danych
- ❖ przechowywanie i tworzenie kopii zapasowych podczas badań
- ❖ wymogi prawne, kodeksy postępowania
- ❖ udostępnianie i długotrwałe przechowywanie danych
- ❖ zadania związane z zarządzaniem danymi oraz zasoby



Polityka Komisji Europejskiej

„Komisja Europejska jest przekonana, że nie powinno się płacić za dostęp lub wykorzystanie informacji, których uzyskanie zostało już sfinansowane z publicznych pieniędzy. Europejskie przedsiębiorstwa i obywatele powinni móc w pełni z nich korzystać.”

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf



Horyzont 2020: polityka otwartości

DANE BADAWCZE

- ❖ sporządzanie planu zarządzania danymi badawczymi,
- ❖ otwarte udostępnianie danych powiązanych z opublikowanymi artykułami



DMP w programie Horyzont 2020

- ❖ sposób zajmowania się danymi badawczymi podczas realizacji i po zakończeniu projektu
- ❖ jakie dane będą zbierane, wytwarzane, przetwarzane?
- ❖ jakie metodologie i standardy będą używane?
- ❖ czy dane będą udostępniane w sposób otwarty?
- ❖ jak dane będą przechowywane?
- ❖ uwzględnienie zasad FAIR

Dane badawcze

Udostępnianie co najmniej w zakresie wymaganym do weryfikacji wyników badań zaprezentowanych w publikacjach naukowych.



RESEARCH DATA - OPEN BY DEFAULT

Horizon 2020 grantees are required

take measures to ensure open access to the **data** underlying their scientific publications

provide open access to **any other research data** of their choice

Horizon 2020 grantees are **encouraged** to also **share datasets** beyond publication



<https://www.openaire.eu/how-to-comply-to-h2020-mandates-for-data>

- open by default
- as open as possible, as closed as necessary
- zasady FAIR



Ilustracja: OpenAIRE CC BY,
<https://www.exeter.ac.uk/research/openresearch/h/opendata>

Zasady FAIR

F_{indable}



A_{ccessible}



I_{nteroperable}



R_{eusable}



2.2 Definition of FAIR The FAIR guiding principles: <https://doi.org/10.1038/sdata.2016.18>

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1. the protocol is free, open and universally implementable
 - A1.2. the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (meta)data uses vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be reusable:

- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with data provenance
 - R1.3. (meta)data meet domain relevant community standards

https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

2.2 Definition of FAIR The FAIR guiding principles: <https://doi.org/10.1038/sdata.2016.18>

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their assigned information
- A1.1. the protocol is free, open, and publicly available
- A1.2. the protocol allows for third party access
- A2. metadata are accessible

To be Interoperable:

- I1. (meta)data use a formal, machine-actionable vocabulary
- I2. (meta)data uses vocabularies that are interoperable with other data
- I3. (meta)data include qualified references to other data

To be reusable:

- R1. (meta)data are richly described with machine-actionable information
- R1.1. (meta)data are released under an open license
- R1.2. (meta)data are associated with data provenance
- R1.3. (meta)data meet domain relevant community standards

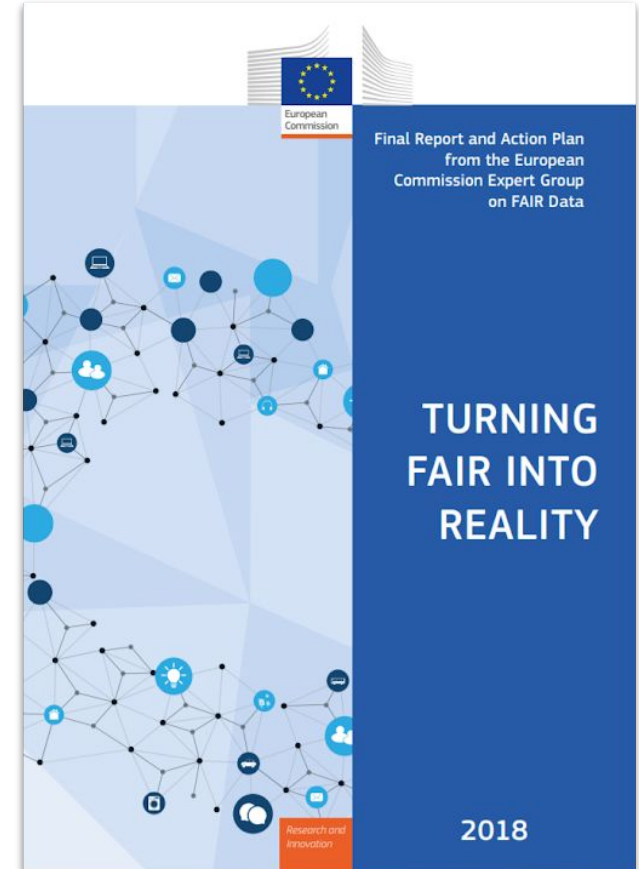
In 2016, the '**FAIR Guiding Principles for scientific data management and stewardship**' were published in *Scientific Data*. The authors intended to provide guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets. The principles emphasise machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention) because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.

<https://www.go-fair.org/fair-principles/>

https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

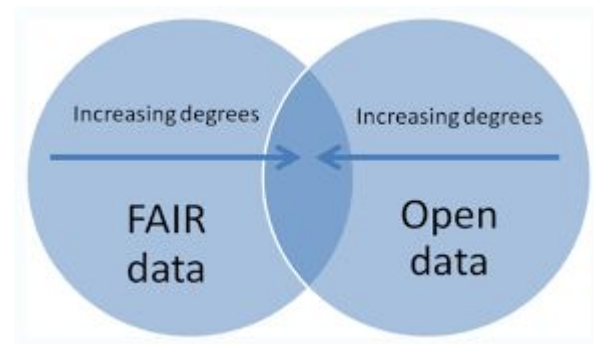
FAIR w politykach otwartości

- ❖ Polityka Komisji Europejskiej,
- ❖ Horyzont 2020 (zasady FAIR uwzględnione w DMP),
- ❖ Dyrektywa o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego,
- ❖ wytyczne Narodowego Centrum Nauki do sporządzania DMP



FAIR jako kontinuum

FAIR a otwartość



Turning FAIR into reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Union, 2018.

https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

FAIR self-assessment tool

Findable

Does the dataset have any identifiers assigned?

No identifier

Is the dataset identifier included in all metadata records/files describing the data?

No

How is the data described with metadata?

The data is not described

What type of repository or registry is the metadata record in?

The data is not described in any repository

Accessible

Interoperable

Reusable

Total across F.A.I.R

<https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/>

Findable

Czy dane opatrzone zostaną metadanymi? Czy będą opisane zgodnie z przyjętymi standardami?

Czy dane będą posiadać trwałe identyfikatory (DOI)?

Czy (meta)dane będą zamieszczone lub indeksowane w serwisie, którego zasoby można przeszukiwać?

Contact us

 because good research needs good data

Search

Home

Digital curation

About us

News

Events

Resources

Training

Projects

Community

Tailored support

Home > Resources > Metadata Standards

In this section

Briefing Papers

How-to Guides & Checklists

Developing RDM Services

Curation Lifecycle Model

Curation Reference Manual

Policy and legal

Data Management Plans

Tools

Case studies

Repository audit and assessment

Standards

Disciplinary Metadata

DIFFUSE

Publications and presentations

Roles

Curation Journals

Informatics research

External resources

Online Store

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and re-use, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Please note that a [community-maintained version of this directory](#) has been set up under the auspices of the Research Data Alliance.

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

<http://www.dcc.ac.uk/resources/metadata-standards>

Accessible

Publicly accessible

Fully accessible to persons who meet explicitly stated conditions, e.g. ethics approval for sensitive data

A de-identified / modified subset of the data is publicly accessible

Embargoed access after a specified date

Unspecified conditional access e.g. contact the data custodian for access

Access to metadata only

No access to data or metadata

Które dane zostaną udostępnione w sposób otwarty? Jeśli część danych nie może zostać udostępniona - dlaczego? Czy w takiej sytuacji udostępnione zostaną metadane?

W jaki sposób i gdzie dane zostaną udostępnione? Czy warunki dostępu będą jasno określone?

Accessible

How accessible is the data?

No access to data or metadata

Publicly accessible

Is the data available online without requiring specialised protocols or tools once access has been approved?

Will the metadata record be available even if the data is no longer available?

Interoperable

Reusable

Total across F.A.I.R.

Formaty plików

W jakich formatach zostaną zapisane dane?
Czy wymagają specjalnego, płatnego oprogramowania? Czy będzie można skorzystać z nich w przyszłości?

- ❖ rekomendowane: formaty otwarte, bezstratne,
- ❖ akceptowane: formaty powszechnie używane

Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)

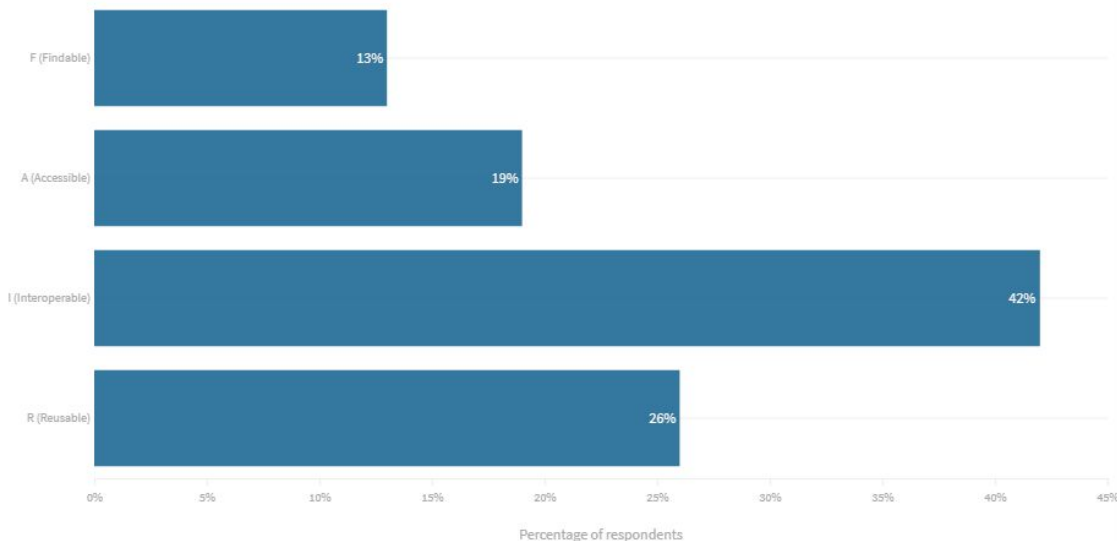
<https://www.ukdataservice.ac.uk/manage-data/formats/recommended-formats.aspx>

Interoperable

Czy możliwe będzie połączenie danych z innymi zbiorami pochodzącymi z innych źródeł?

Which of the FAIR principles do you think most needs better definition?

Interoperability is the least understood FAIR principle. Some 42% of the 187 respondents who answered this question felt that it needed further clarification.



Source: [State of Open Data](#)

Made with Flourish

https://public.flourish.studio/visualisation/213291/?utm_source=embed&utm_campaign=visualisation/213291

Reusable

Czy dane zostaną opatrzone licencją, która pozwoli na ich ponowne wykorzystanie w stopniu tak szerokim jak to możliwe?

Czy sposób pozyskiwania danych i kontekst, w jakim zostały wytworzone zostały dostatecznie dobrze opisane?

Licencje Creative Commons - trzy warstwy:

- tekst prawny,
- przystępne podsumowanie,
- dane (kod) do odczytu maszynowego.

Standard machine-readable license (e.g. Creative Commons)

Standard text based license

Non-standard machine-readable license (clearly indicating under what conditions the data may be reused)

Non-standard text-based license

No license

Reusable



Which of the following best describes the license/usage rights attached to the data?

No license



How much provenance information has been captured to facilitate data reuse?

No provenance information is recorded





Unusable data

„otwarty zbiór danych”

„dane publicznie dostępne”

„dane ogólnodostępne”

„brak ograniczeń”

„freely available”

„dane dostępne na licencji Creative Commons”



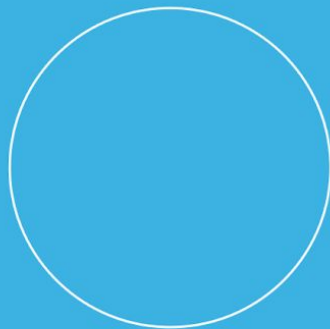
Repozytoria danych badawczych

- ❖ bezpieczne długoterminowe przechowywanie danych,
- ❖ stały adres internetowy, możliwość uzyskania trwałego identyfikatora, np. DOI,
- ❖ łatwość wyszukiwania danych,
- ❖ łatwość cytowania danych,
- ❖ podstawowe statystyki, informacje o tym, jak często dane były pobierane i oglądane.

Broszury informacyjne

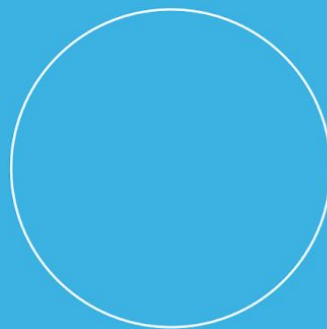
Prawne aspekty otwierania danych badawczych – poradnik

Wersja 1.0



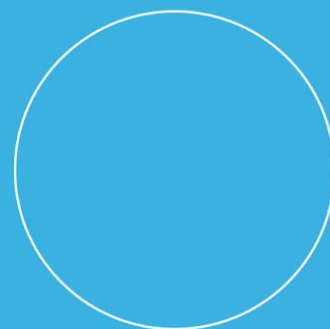
Selekcja i przygotowanie danych badawczych do udostępnienia

Wersja 1.0



Jak korzystać z zasobów w repozytoriach danych

Wersja 1.0



<https://drodb.icm.edu.pl/materialy-2/>



Fundusze
Europejskie
Polska Cyfrowa



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





Kontakt

Platforma Otwartej Nauki

pon@icm.edu.pl

Natalia Gruenpeter

n.gruenpeter@icm.edu.pl

