



RepOD

Repozytorium Otwartych Danych

Dziedzinowe Repozytoria Otwartych Danych Badawczych

PREZENTACJA PROJEKTU



UNIWERSYTET
WARSZAWSKI

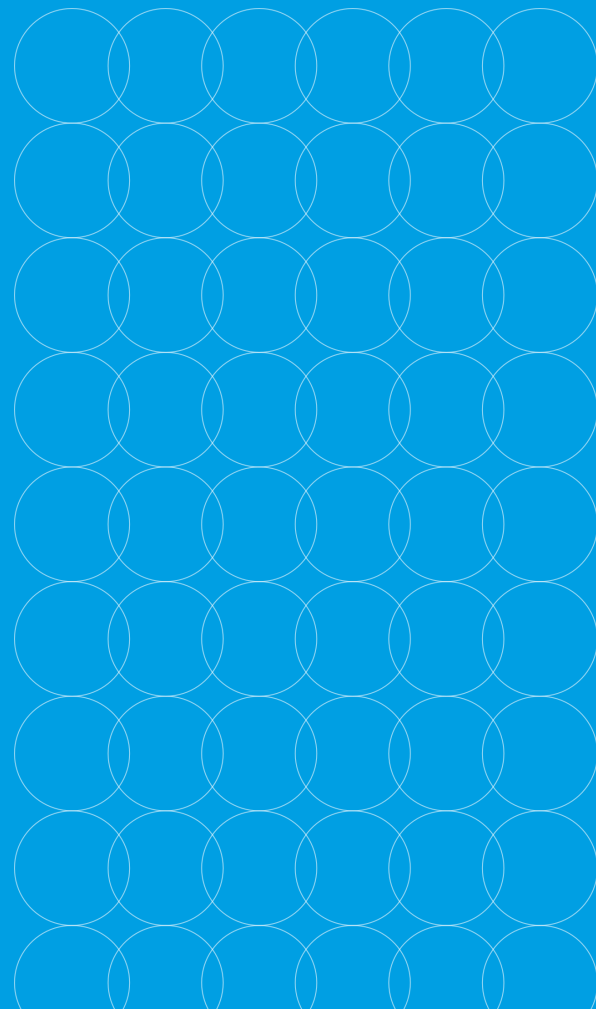


Fundusze
Europejskie
Polska Cyfrowa

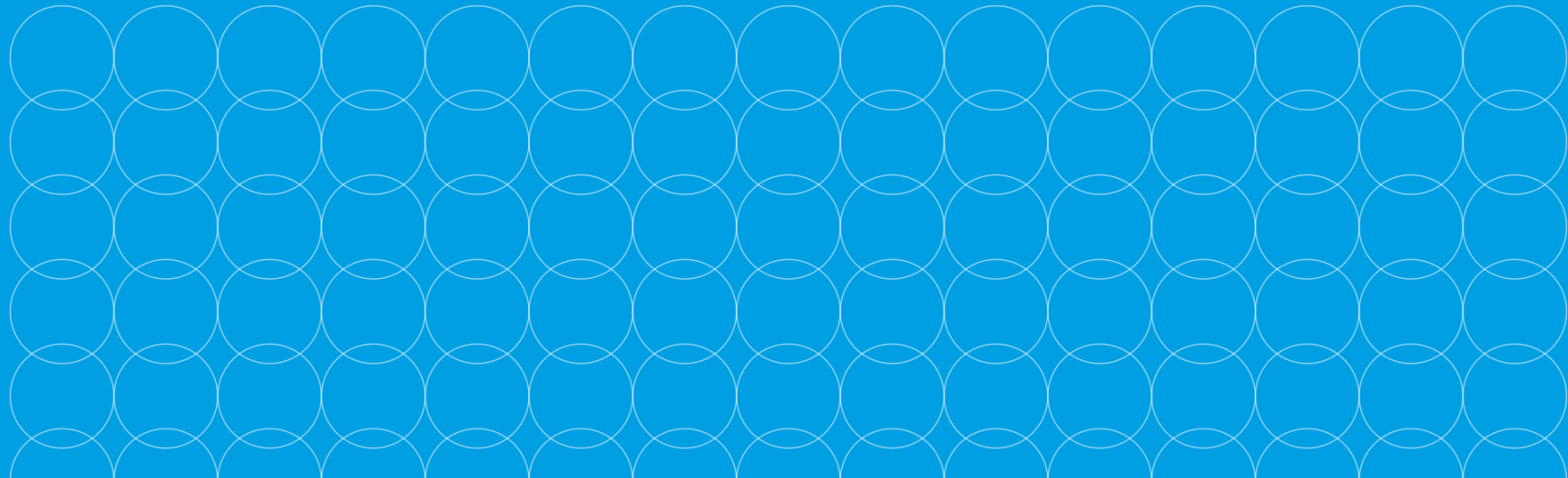


Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Informacje o projekcie



Ogólne informacje

Beneficjent projektu: Uniwersytet Warszawski,
jednostki realizujące: ICM UW, ISS UW

Partnerzy projektu:
Instytut Filozofii i Socjologii PAN,
Uniwersytet Adama Mickiewicza w Poznaniu

Okres realizacji: 1 sierpnia 2018 r. - 31 lipca 2021 r.

Kwota dofinansowania: 4 998 889 PLN
(w tym UE 4 230 559,76 PLN,
budżet państwa 768 329,24 PLN).

Cele projektu

- Udostępnienie danych badawczych: społecznych, krystalograficznych i innych.
- Stworzenie oprogramowania dla trzech repozytoriów (Dataverse jako punkt wyjścia, możliwość dalszych modyfikacji w przyszłości).
- Uruchomienie trzech repozytoriów



RepOD

Repozytorium Otwartych Danych

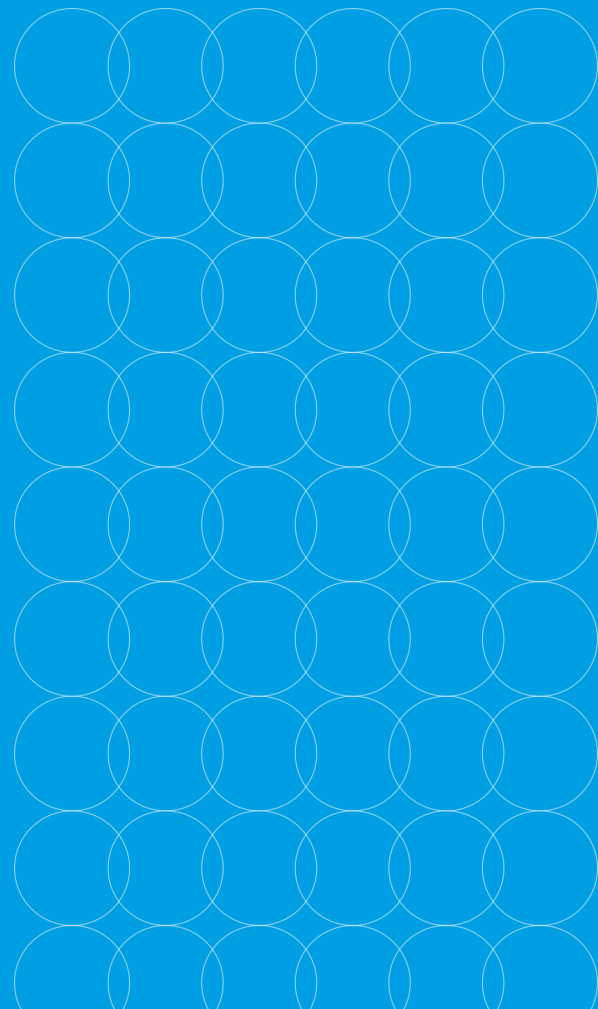
RepOD

- Repozytorium otwarte dla wszystkich zainteresowanych użytkowników.
- Zastąpi działającą pilotażową wersję repozytorium.



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





RDS

Repozytorium
Danych Społecznych

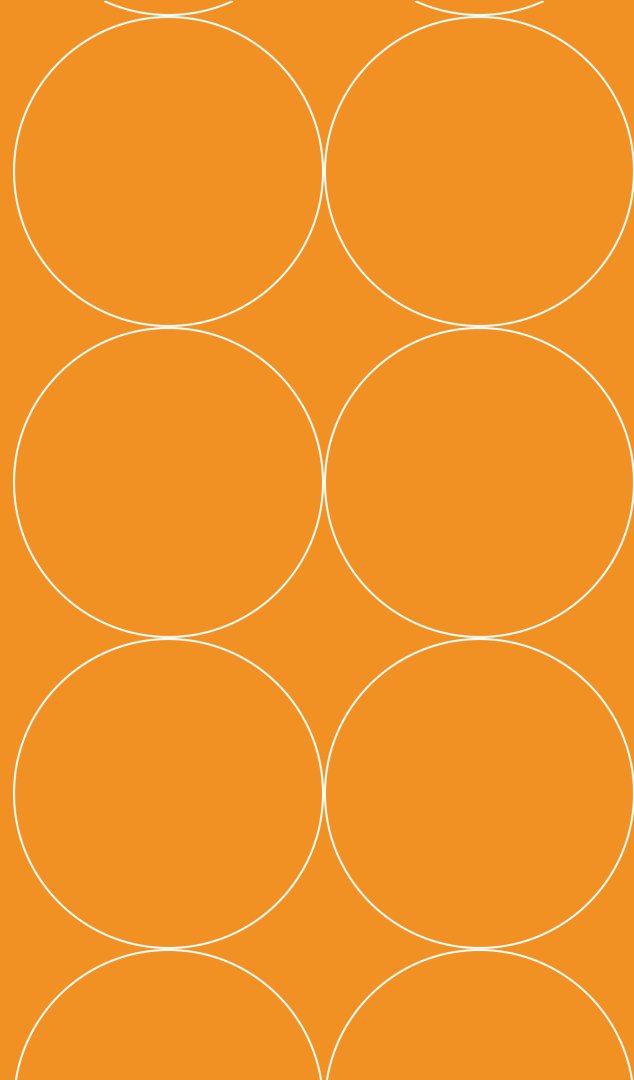
RDS: Repozytorium Danych Społecznych

- Repozytorium otwarte dla wszystkich zainteresowanych użytkowników.
- 400 zbiorów danych (jakościowych i ilościowych) udostępnionych w ramach projektu.



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





MX-RDR

Macromolecular Xtallography
Raw Data Repository

MX-RDR: Repozytorium Danych Krystalograficznych

- Repozytorium otwarte dla wszystkich zainteresowanych użytkowników.
- 200 zbiorów surowych danych krystalograficznych udostępnionych w ramach projektu.



UNIWERSYTET
WARSZAWSKI



UNIWERSYTET
IM. ADAMA MICKIEWICZA
W POZNANIU



Fundusze
Europejskie
Polska Cyfrowa

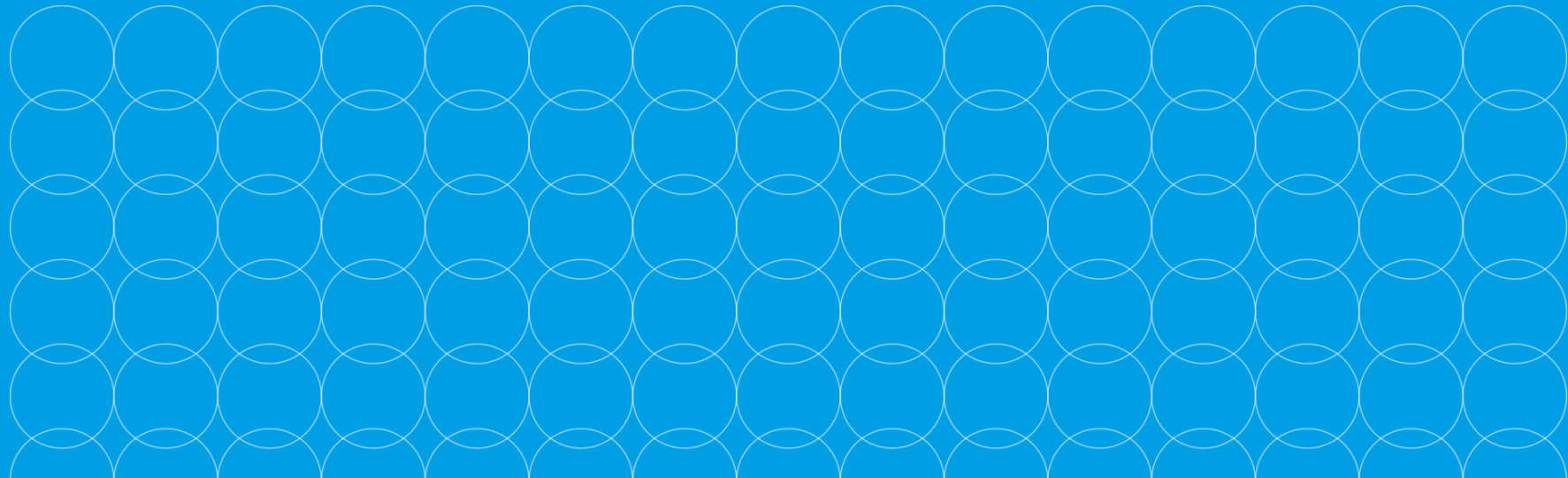


Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



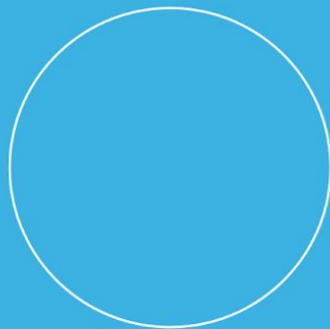
**Planowany start wersji produkcyjnych
repozytoriów: połowa 2020 r.**



Broszury informacyjne

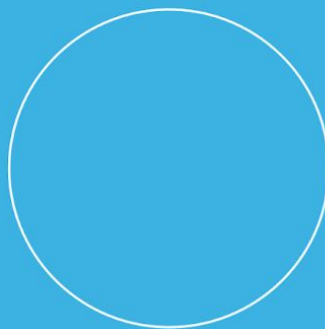
Prawne aspekty otwierania danych badawczych – poradnik

Wersja 1.0



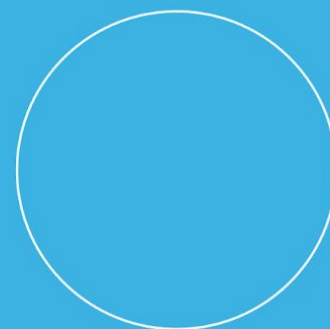
Selekcja i przygotowanie danych badawczych do udostępnienia

Wersja 1.0



Jak korzystać z zasobów w repozytoriach danych

Wersja 1.0



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



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





RepOD

Repozytorium Otwartych Danych

Zarządzanie danymi badawczymi

NATALIA GRUENPETER

CC-BY

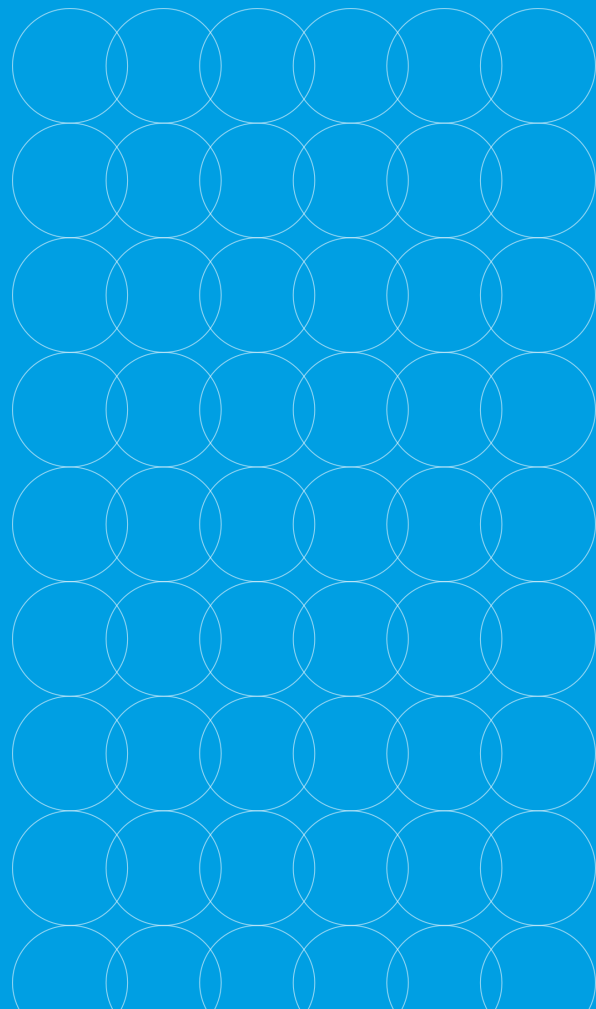
Treść licencji dostępna na stronie:

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



Rzeczpospolita
Polska

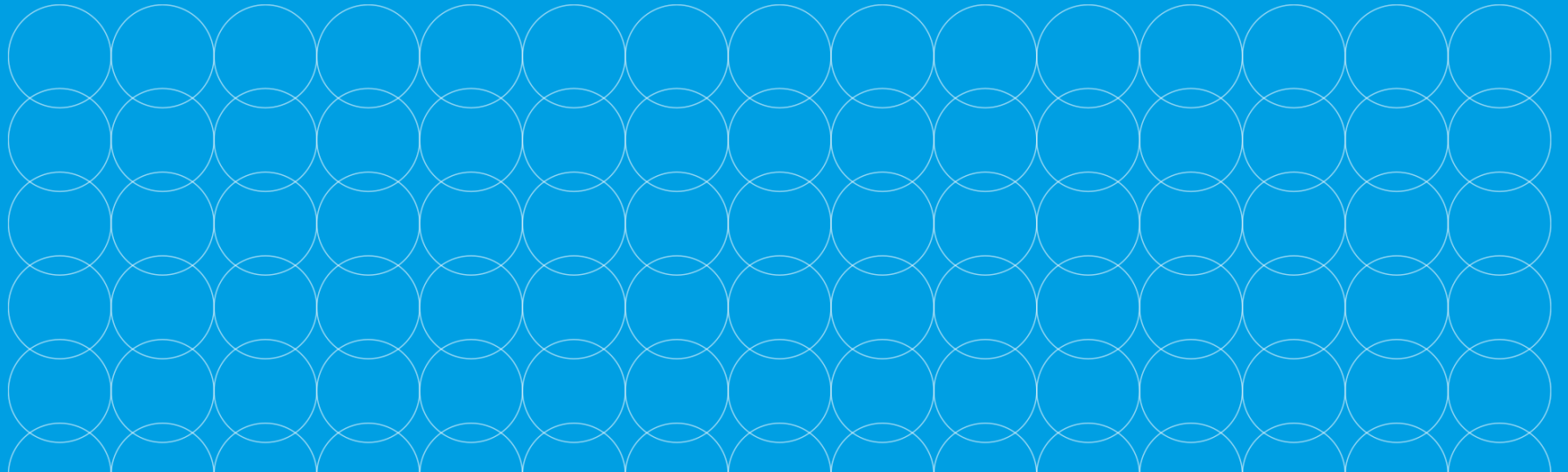
Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Plan szkolenia część 1.

- Dane badawcze - definicje**
- Korzyści płynące z prawidłowego zarządzania danymi badawczymi**
- Plan Zarządzania Danymi (Data Management Plan, DMP)**
- Dane w Horyzoncie 2020 i grantach NCN**
- Selekcja i przygotowanie danych do udostępnienia**

Dane badawcze - definicje

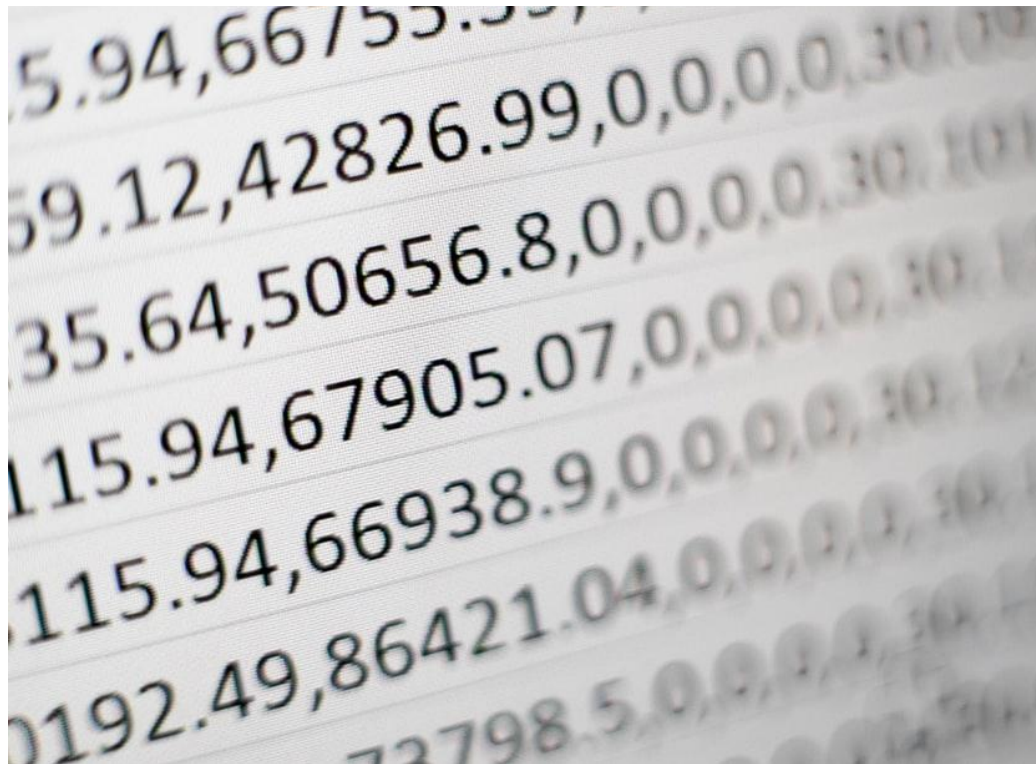


Co to są dane badawcze?

- ❖ 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.

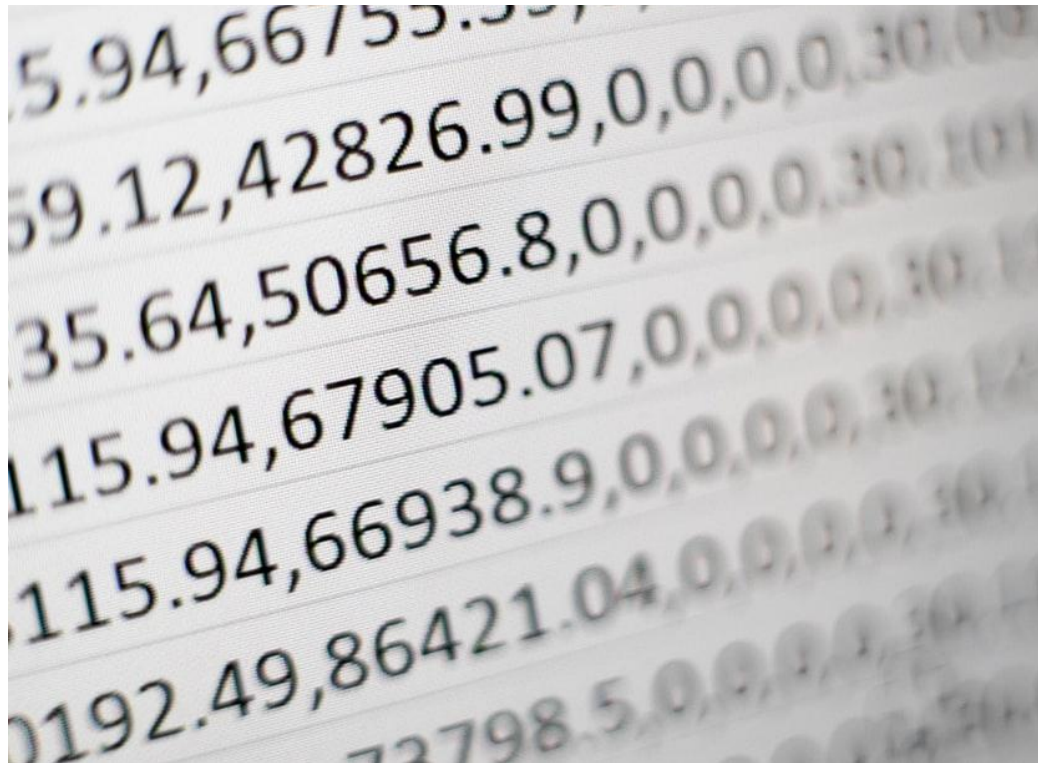
Dane badawcze

- 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 ...

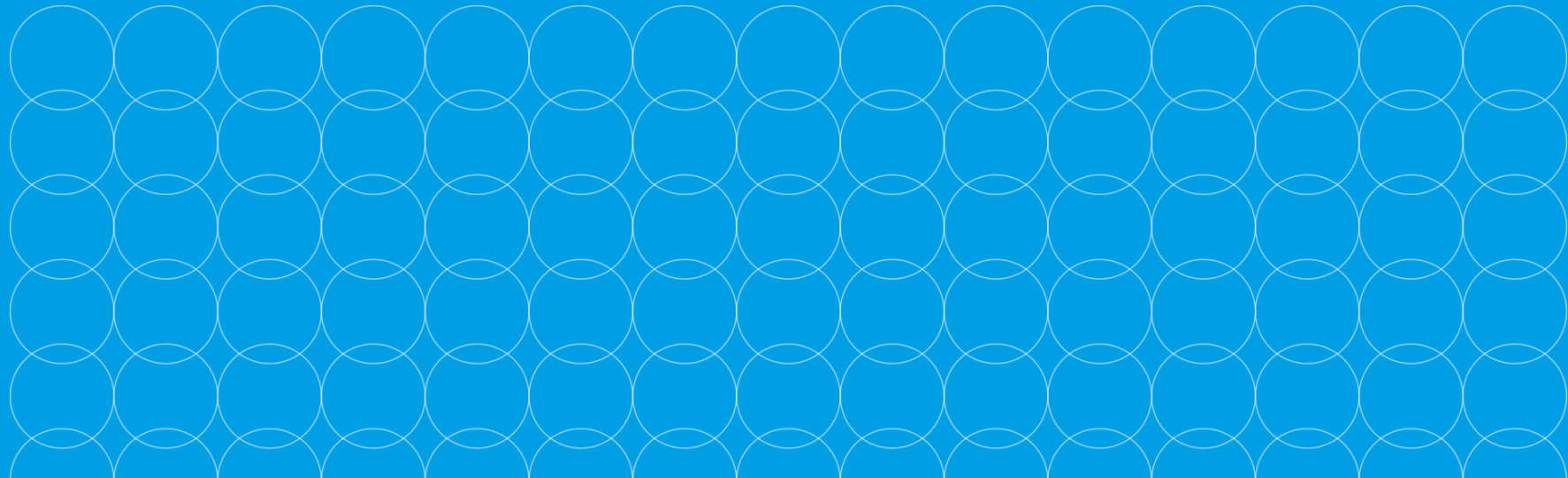


Dane w różnych aspektach

- **Rodzaj danych**
dane liczbowe, tekstowe, wizualne, audio, wideo, geolokalizacyjne...
- **Źródło danych**
sposób pozyskania/wytworzenia danych: eksperyment, obserwacja, symulacja...
- **Forma i format danych**
- **Rozmiar i złożoność danych**



Zarządzanie danymi badawczymi - korzyści



Korzyści dla badaczy

- 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

Table 2. “How much of your data do you currently store or deposit in the following locations?”

Storage location	None	Some	Most	All	n*
On my institution’s server	469 (27.2%)	468 (27.1%)	333 (19.3%)	407 (23.6%)	1727
On the principal investigator’s server	655 (41.0%)	342 (21.4%)	240 (15.0%)	282 (17.6%)	1598
On a departmental server	809 (52.0%)	342 (22.0%)	192 (12.3%)	146 (9.4%)	1556
On my personal computer	161 (9.2%)	503 (28.7%)	356 (20.3%)	718 (41.0%)	1753
On paper in my office	698 (44.4%)	654 (41.6%)	116 (7.4%)	81 (5.1%)	1573
USB/external drive	557 (35.6%)	496 (31.7%)	184 (11.8%)	281 (18.0%)	1565
In a discipline-based repository, (e.g. NEON or LTER)	1074 (49.2%)	218 (10.0%)	92 (4.2%)	31 (1.4%)	1536
In a publisher or publisher-related repository (e.g., specific publisher or Dryad)	1030 (66.8%)	334 (21.7%)	53 (3.4%)	19 (1.2%)	1541
Other data repository or archive (e.g.					
In my institution’s repository					
Cloud storage					
Other					

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide

Carol Tenopir, Natalie M. Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, Robert J. Sandusky

Published: March 11, 2020 • <https://doi.org/10.1371/journal.pone.0229003>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0229003>

Table 2. “How much of your data do you currently store or deposit in the following locations?”

Storage location	None	Some	Most	All	n*
On my institution’s server	469 (27.2%)	468 (27.1%)	333 (19.3%)	407 (23.6%)	1727
On the principal investigator’s server	655 (41.0%)	342 (21.4%)	240 (15.0%)	282 (17.6%)	1598
On a departmental server	809 (52.0%)	342 (22.0%)	192 (12.3%)	146 (9.4%)	1556
On my personal computer	161 (9.2%)	503 (28.7%)	356 (20.3%)	718 (41.0%)	1753
On paper in my office					1573
USB/external drive					1565
In a discipline-based repository, (e.g.					1536
In a publisher or publisher-related re					1541
Other data repository or archive (e.g.					1584
In my institution’s repository					1553
Cloud storage					1593
Other					979

Good data practice

- discipline-based repository – 5.6%
- publisher or publisher-related repository – 4.6%
- other data repository or archive – 16.6%
- institution’s repository – 16.5%

Mediocre data practice

- cloud – 23.6%
- institutional server – 42.9%
- departmental server – 21.7%
- PI’s server – 32.6%

Bad data practice

- personal computer – 61.3%
- paper in my office – 12.5%
- thumb/external drive – 29.8%

Repozytoria

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

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

Following these guiding principles for sharing c

Jon Brock

11 FEBRUARY 2018

Nature Index 360°

The idea that scientific data should be FAIR – and Reusable – is one increasingly endorsed by the **United States National Academies of Sciences**, the **European Commission**, and the Wellcome Trust among the people that ultimately matter, the scientists. The State of Open Data report, published by Digital Science researchers were “familiar with FAIR principles”. (Digital Science is a part of the Holtzbrinck Publishing Group, which also has a presence on Nature Index.)

But as LeMay argues, there are both altruistic and selfish reasons for researchers to take the next step and make their data FAIR. “Most people get into research because they want to make a difference,” she says. “That includes making your data as useful as possible.” FAIR can also be good for career advancement, particularly for early-career researchers. “FAIR helps you demonstrate the impact of your research when people reuse and cite your dataset,” LeMay says. “It gets your name out there and can lead to new collaborations.”

Heller agrees. However, the true benefit of FAIR, he argues, is in providing a framework for researchers to manage their own data so they can themselves find it, understand it, and reuse it. “As a scientist, you should treat your data like a love letter to your future self,” he says.

Korzyści dla społeczeństwa

- większe zaufanie do nauki
 - ◆ większa przejrzystość i wiarygodność nauki, możliwość weryfikacji wyników badań
- rozwój społeczny i ekonomiczny
 - ◆ możliwość wykorzystania danych w innych poza naukowych kontekstach, np. biznes, innowacje
 - ◆ szybsza komunikacja naukowa i współpraca sprzyja rozwiązywaniu globalnych problemów

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

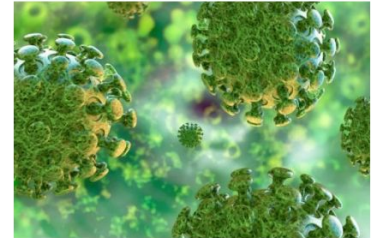
January 28, 2020

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

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

U¹⁵

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.



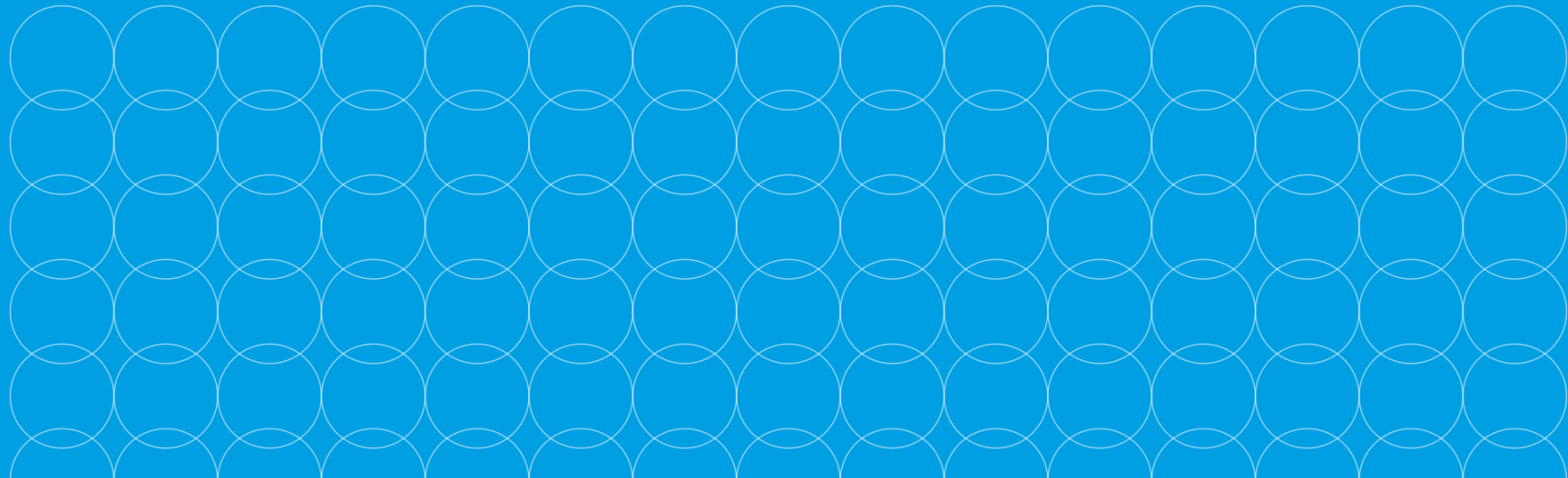
<https://www.leru.org/files/Sorbonne-declaration.pdf>

Dlaczego nas to dotyczy?

- polityki otwartości
 - ◆ wymogi grantodawcy, np. Horyzont 2020,
 - ◆ polityka przyjęta przez wydawcę czasopisma, np. udostępnienie danych w zakresie wymaganym do weryfikacji wyników zaprezentowanych w artykule,
 - ◆ wymogi pracodawcy, instytucjonalna polityka otwartości

Plan zarządzania danymi badawczymi

Data Management Plan, DMP



DMP

- formalny dokument, który zawiera zarys tego, co będziemy robić z danymi w trakcie projektu badawczego i po jego zakończeniu,
- żywy dokument, który ma wspierać realizację projektu na każdym etapie,
- różne wymogi → różne wzory;

Zwyczajowe elementy planu zarządzania danymi badawczymi:

- ◆ wytworzenie, pozyskanie danych,
- ◆ formaty i nazewnictwo plików,
- ◆ przechowywanie danych,
- ◆ zasady dostępu,
- ◆ prawne i etyczne aspekty
- ◆ niezbędne zasoby

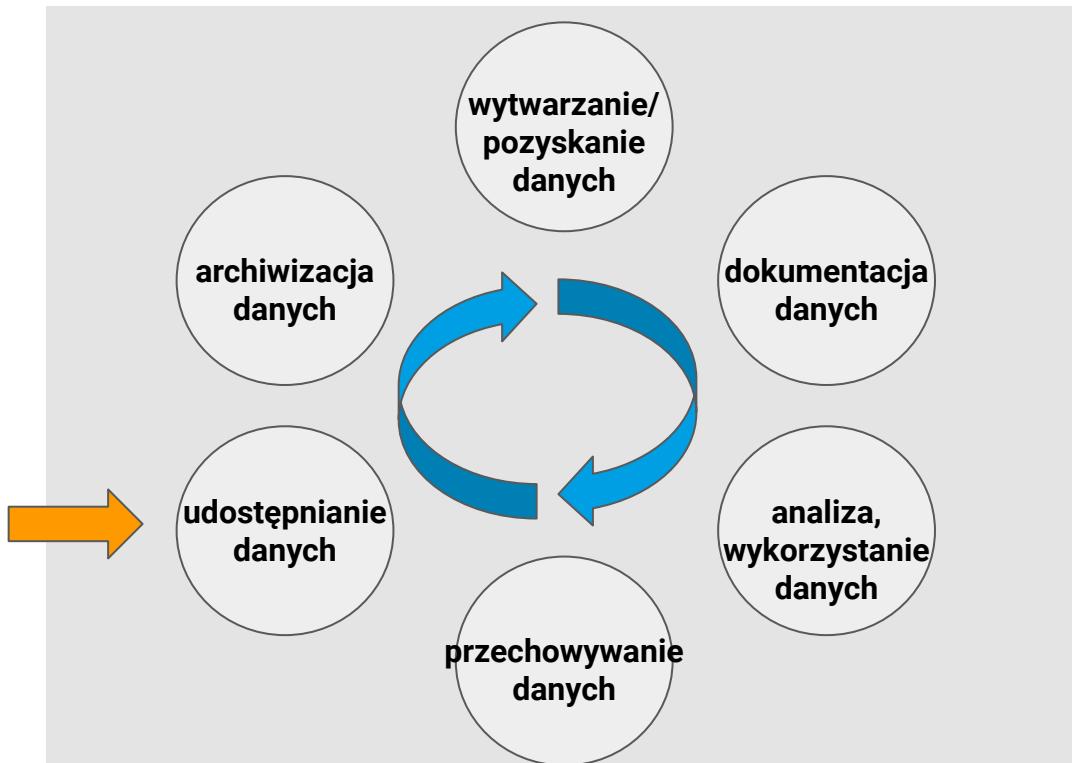
Zalety sporządzania DMP



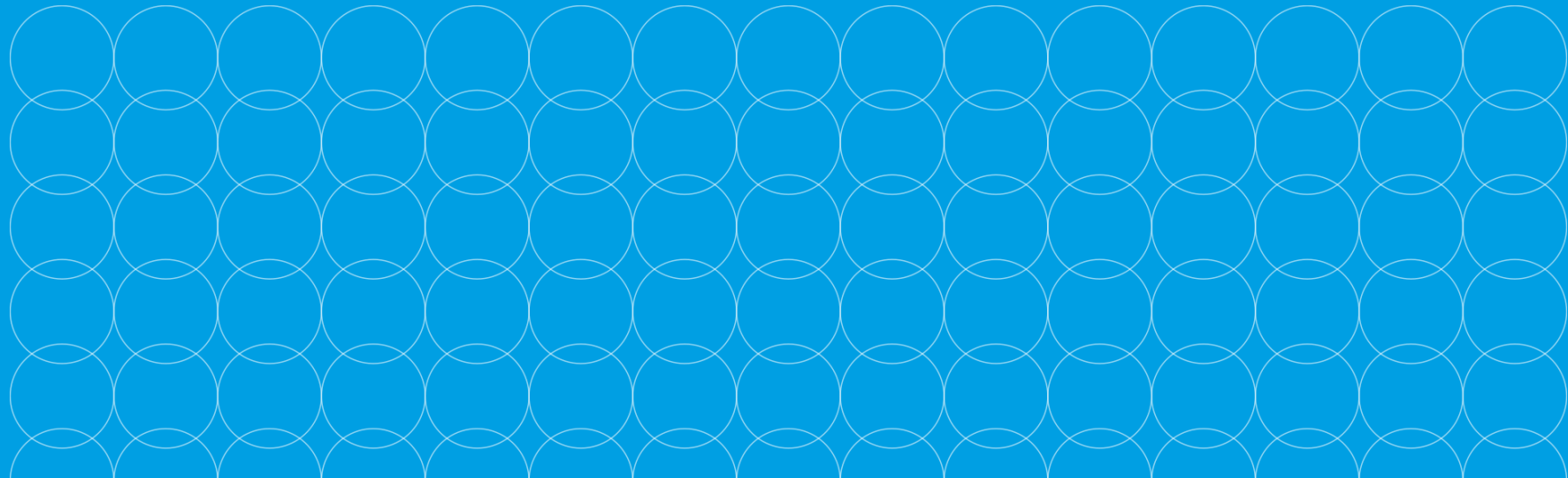
Ilustracja: [The Data Management Expert Guide](#)
by CESSDA ERIC, CC BY SA.

Dane w całym cyklu projektu badawczego

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



Dane w Horyzoncie 2020 i grantach NCN



Stanowisko KE

„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

W ramach programu Horyzontu 2020 kwestie związane z zarządzaniem i udostępnianiem danych badawczych reguluje Open Research Data Pilot, którym od 2017 roku objęte są domyślnie wszystkie obszary tematyczne (ale nie wszystkie instrumenty).

Dwa kluczowe elementy:

- opracowanie DMP
- otwarte udostępnianie danych

Otwarte udostępnianie danych badawczych

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>

Otwarte udostępnianie danych badawczych

- open by default
- as open as possible, as closed as necessary
- zasady FAIR
- opt-out w uzasadnionych sytuacjach



Ilustracja: OpenAIRE CC BY,

<https://www.exeter.ac.uk/research/openresearch/opendata>

Opt-out

W uzasadnionych przypadkach możliwa jest rezygnacja (opt-out) z udziału w Open Research Data Pilot, która skutkuje częściowym lub całkowitym uchyleniem obowiązku udostępniania danych badawczych; Można z niej skorzystać na każdym etapie.

Opting out – partially or entirely

By extending the pilot, open access becomes the default setting for research data generated in Horizon 2020.

However, not all data can be open. Projects can therefore opt out at any stage (either before or after signing the grant) and so free themselves retroactively from the obligations associated with the conditions – if:

- participation is incompatible with the obligation to protect results that can reasonably be expected to be commercially or industrially exploited
- participation is incompatible with the need for confidentiality in connection with security issues
- participation is incompatible with rules on protecting personal data
- participation would mean that the project's main aim might not be achieved
- the project will not generate / collect any research data or
- there are other legitimate reasons (you can enter these in a free-text box at the proposal stage).

The Commission's approach can therefore be described as "as open as possible, as closed as necessary".

F
indable



A
ccessible



I
nteroperable



R
eusable



Ilustracja: Sungya Pundir, Wikimedia Commons,
CC BY-SA 4.0

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



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 identifier using a standardized protocol
- A1.1. the protocol is free, open, simple, robust and scalable
- A1.2. the protocol allows for the use of authentication and authorization where appropriate
- A2. metadata are accessible, either directly or through a standard query service

To be Interoperable:

- I1. (meta)data use a formal, machine-actionable vocabulary (e.g., ontology, controlled vocabulary)
- I2. (meta)data uses vocabularies that are interoperable with other standards
- I3. (meta)data include qualified relationships to other data

To be reusable:

- R1. (meta)data are richly described with machine-actionable terms
- 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

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?

The screenshot shows the DCC website's 'Metadata Standards' page. At the top, the DCC logo is followed by the tagline 'because good research needs good data' and a search bar. A navigation menu includes links for Home, Digital curation, About us, News, Events, Resources, Training, Projects, Community, and Tailored support. The breadcrumb trail reads 'Home > Resources > Metadata Standards'. The main content area is titled 'Disciplinary Metadata' and contains a paragraph explaining the importance of metadata for research data access and re-use, mentioning that many academic disciplines have supported initiatives to formalise metadata specifications. It also notes that for those disciplines that have not yet settled on a metadata standard, the General Research Data section provides links to broader metadata standards. A note mentions a community-maintained version of the directory. On the left, a sidebar lists various resources under the heading 'In this section', including 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 (with 'Disciplinary Metadata' selected), Publications and presentations, Roles, Curation Journals, Informatics research, External resources, and Online Store. Below the main text, a 'Search by Discipline' section features five image-based links: Biology (microscope), Earth Science (coastline), General Research Data (circuit board), Physical Science (molecular model), and Social Science & Humanities (hands raised).

<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?

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.

No access to data or metadata

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

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)

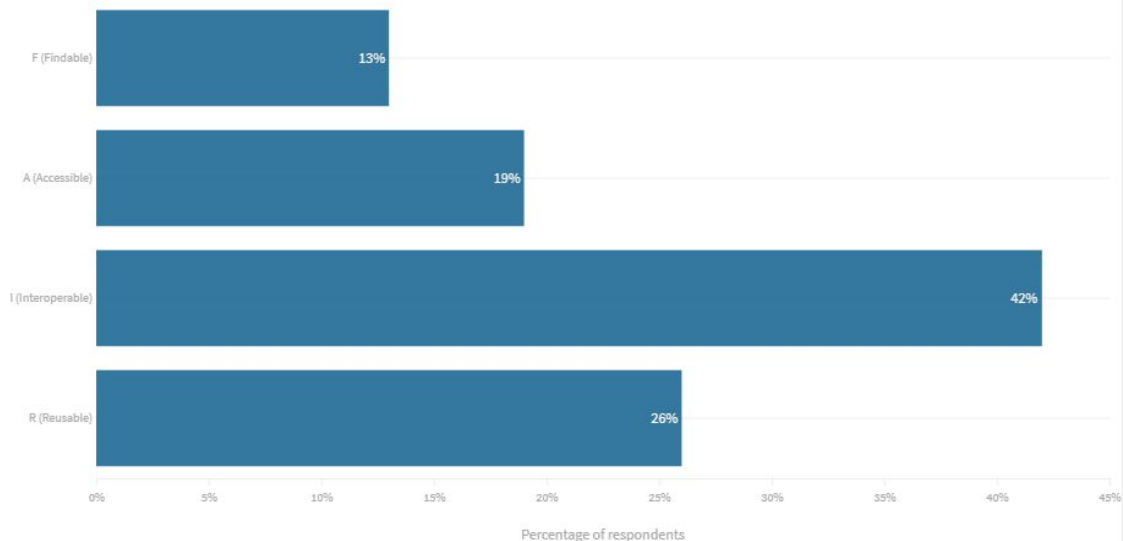
<https://www.ukdataservice.ac.uk/manage-data/formats/recommended-formats.aspx>

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

Rola dokumentacji



Czy do danych dołączona zostanie dokumentacja? Czy wskazane zostanie źródło (pochodzenie) danych: kto i w jaki sposób je wytworzył? Jak były przetwarzane? Czy zawierają dane z innych źródeł?

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”

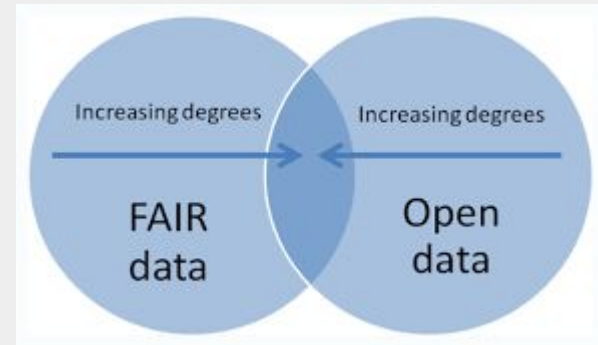
„brak ograniczeń”

„freely available”

„dane dostępne na licencji Creative Commons”

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

<https://www.ands-nectar-rds.org.au/fair-tool>

FAIR self-assessment tool

Welcome to the ARDC FAIR Data self-assessment tool. Using this tool you will be able to assess the 'FAIRness' of a dataset and determine how to enhance its FAIRness (where applicable).

This self-assessment tool has been designed predominantly for data librarians and IT staff, but could be used by software engineers developing FAIR Data tools and services, and researchers provided they have assistance from research support staff.

You will be asked questions related to the principles underpinning Findable, Accessible, Interoperable and Reusable. Once you have answered all the questions in each section you will be given a 'green bar' indicator based on your answers in that section, and when all sections are completed, an overall 'FAIRness' indicator is provided.

Please be aware that additional explanatory information is provided within the tool. The (i) information button provides an overview of each of the FAIR high-level elements (Findable, Accessible, Interoperable and Reusable). Additionally, each question is hyperlinked, leading users to explanatory information and links to wider resources on related topics.

Findable (i)

Does the dataset have any identifiers assigned?

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

How is the data described with metadata?

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

Accessible (i)

Interoperable (i)

Reusable (i)

Szanowni Państwo,

mając na uwadze fakt, że Komisja Europejska w swoich *Zaleceniach z dnia 17 lipca 2012 r. w sprawie dostępu do informacji naukowej i jej ochrony* wezwwała kraje członkowskie, aby jasno określiły strategię rozpowszechniania i otwartego dostępu do publikacji naukowych, Ministerstwo Nauki i Szkolnictwa Wyższego w przyjętym w 2015 roku dokumencie zatytułowanym *Kierunki rozwoju otwartego dostępu do publikacji i wyników badań naukowych w Polsce*¹ sformułowało wytyczne dotyczące wprowadzenia otwartego dostępu przez jednostki naukowe i uczelnie poprzez:

1. przyjęcie instytucjonalnej polityki otwartego dostępu;
2. wyznaczenie przez kierowników jednostek naukowych i uczelni pełnomocników ds. *Open Access* (OA);
3. wspieranie modelu otwartego czasopism oraz
4. korzystanie z doświadczeń i potencjału bibliotek prowadzących otwarte repozytoria.

Mając powyższe na uwadze, wraz z innymi agencjami finansującymi badania naukowe zrzeszonymi w Science Europe, podjęliśmy działania mające na celu wypracowanie wspólnych europejskich wytycznych dotyczących zarządzania danymi naukowymi i sposobu ich udostępniania w tzw. otwartym dostępie. Praktyczny przewodnik dot. ujednoczonych europejskich praktyk związanych z zarządzaniem danymi naukowymi został opublikowany w styczniu br.²

W oparciu o wytyczne zawarte w tym przewodniku planujemy jeszcze w tym roku wprowadzić do formularza wniosku o finansowanie projektu załącznik, w którym wnioskodawca przedstawi skrócony plan zarządzania danymi badawczymi (z ang. *Data Management Plan*), które powstaną w ramach realizacji projektu. Plan ten będzie uzupełniany na etapie składania raportu końcowego i oceniany po zakończeniu realizacji projektu.

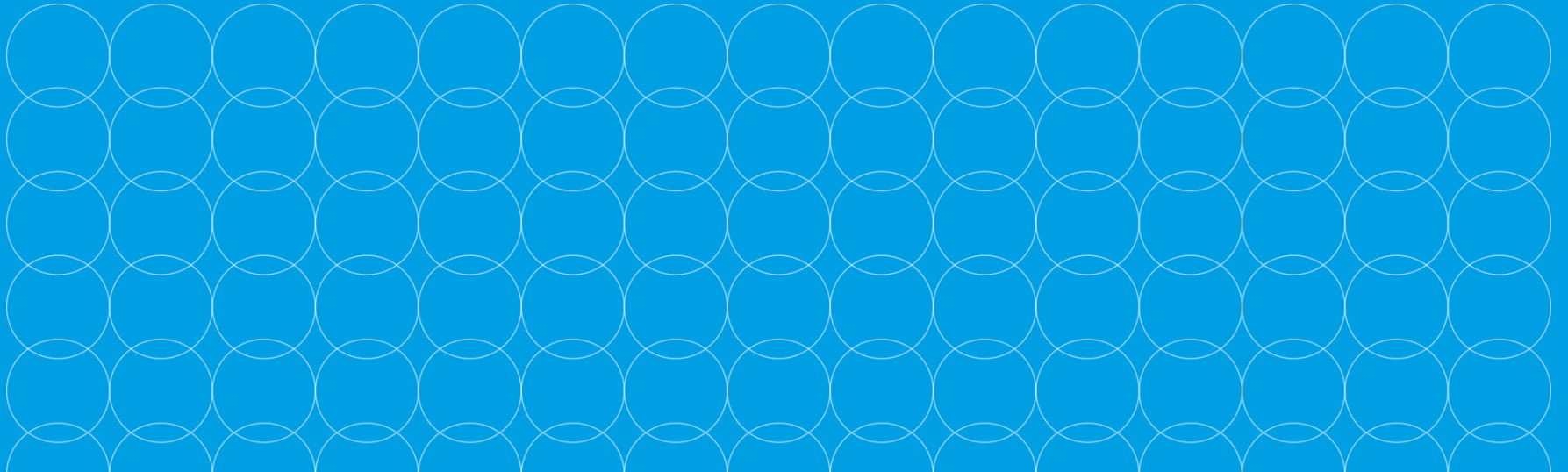
DMP jako
załącznik
do wniosku

- ❖ 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

<https://www.ncn.gov.pl/finansowanie-nauki/otwarta-nauka>

<https://www.ncn.gov.pl/aktualnosci/2020-03-06-plan-zarzadzania-danymi-pytania>

Selekcja i przygotowywanie danych do udostępnienia

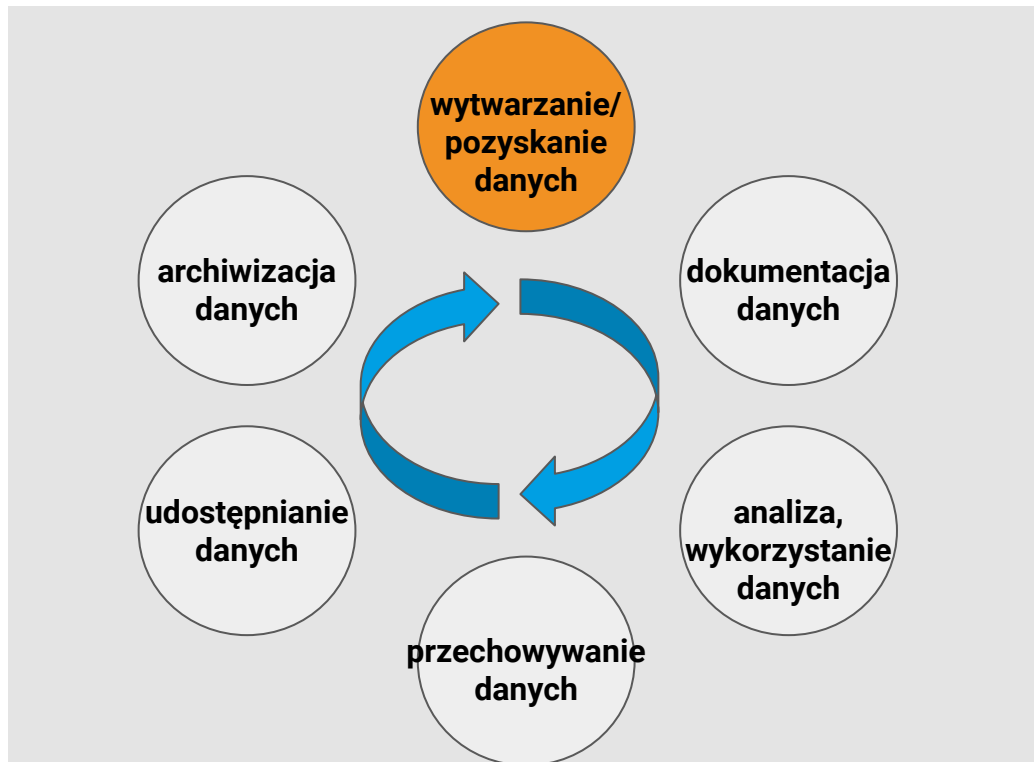


Co należy wziąć pod uwagę?

- ❖ Do czego jesteśmy zobowiązani?
- ❖ Jaka jest wartość naukowa lub historyczna danych?
- ❖ Jak unikalne są nasze dane? Czy istnieje możliwość ich ponownego zebrania/wytworzenia?
- ❖ Czy dane mogą wykorzystać inni?
- ❖ Jakie koszty wiążą się z zarządzaniem i przechowywaniem danych?

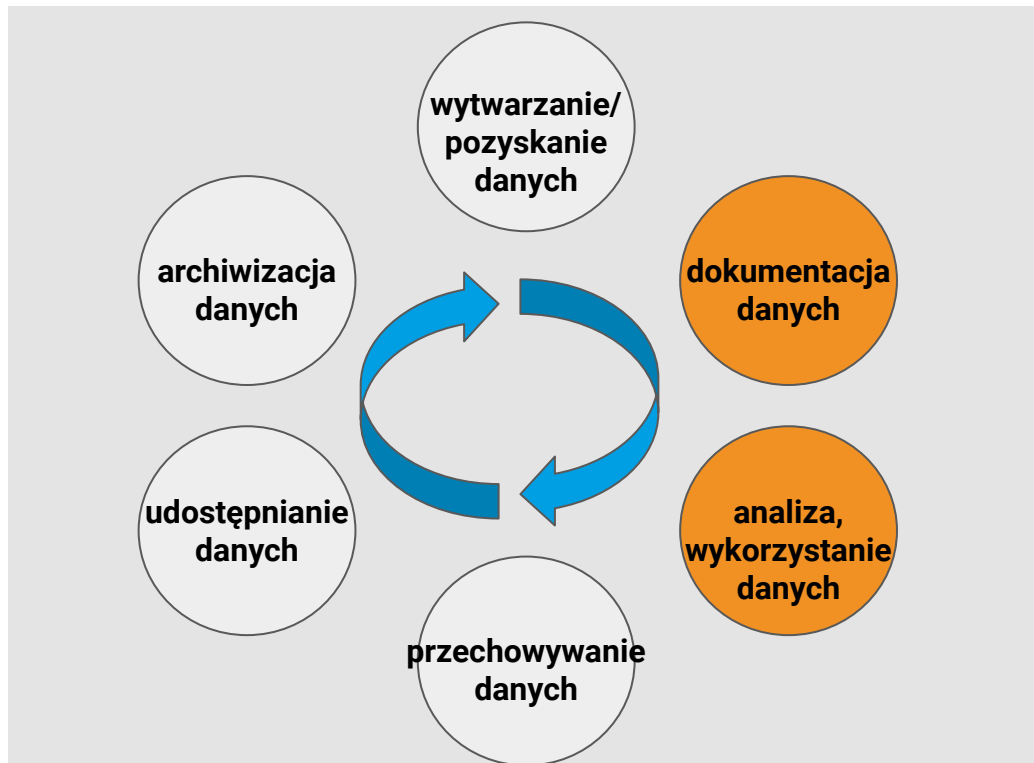
Wytwarzanie, pozyskanie danych

- Przemyslenie kwestii prawnych na jak najwcześniejszym etapie
- Przyjęcie jednolitego i spójnego sposobu organizacji i nazewnictwa danych
- Wybór odpowiednich formatów zapisu danych
- Tworzenie dokumentacji



Dokumentacja i analiza danych

- Co jest niezbędne do właściwego zrozumienia danych?
- dokumentacja ułatwia odpowiednie zacytowanie i ponowne wykorzystanie danych
- dokumentacja na dwóch poziomach:
 - ◆ projektu,
 - ◆ danych;
- osobne pliki z dokumentacją: readme.txt



Dokumentacja: projekt

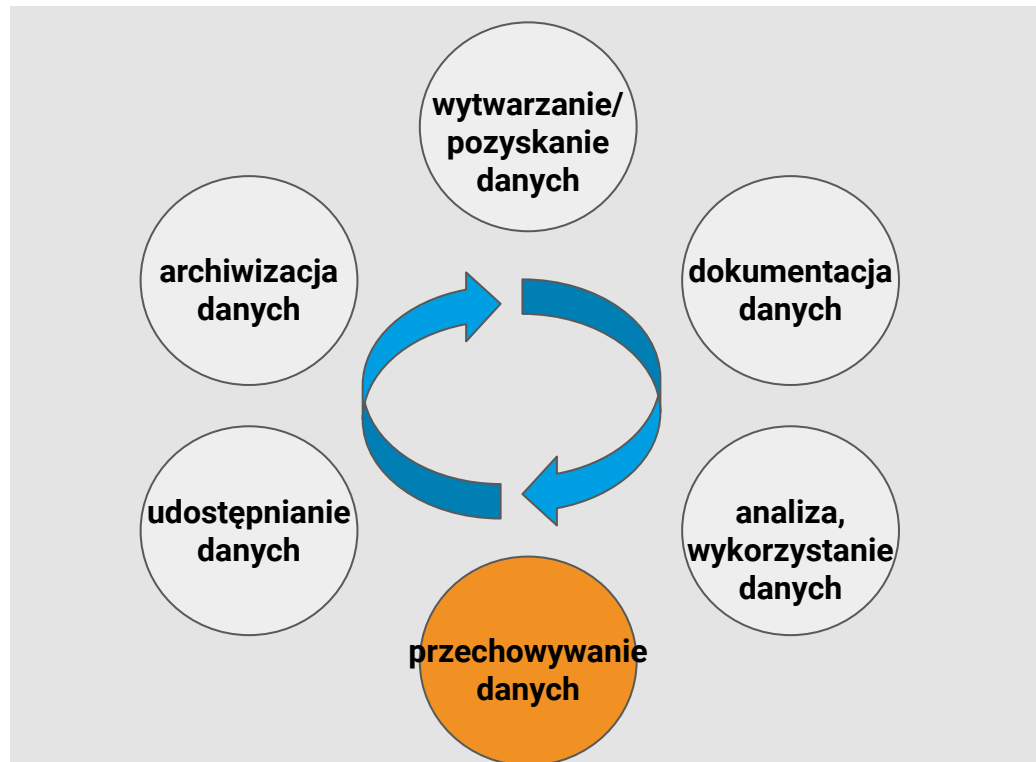
- ❖ W jakim celu/kontekście dane zostały wytworzone?
- ❖ Co zawierają zestawy danych?
- ❖ Jak dane zostały wytworzone/pozyskane?
- ❖ Kto i kiedy to zrobił?
- ❖ Jak dane były przetwarzane/opracowywane?
- ❖ Czy dane były modyfikowane?
- ❖ Jakie zastosowano metody sprawdzania jakości danych?

Dokumentacja: dane

- ❖ nazwy i opisy zmiennych, skrótów, schematów klasyfikacyjnych,
- ❖ informacje o urządzeniach pomiarowych, ich ustawieniach, parametrach itp.
- ❖ testy jakości danych

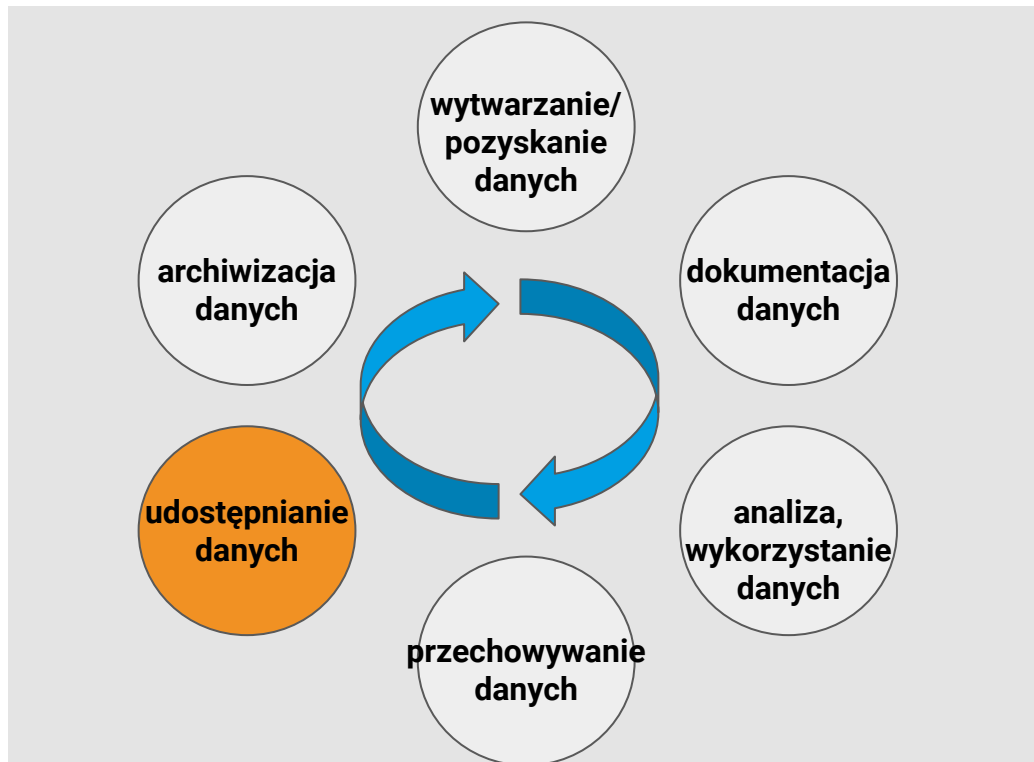
Przechowywanie danych

- Jak bezpiecznie przechowywać dane w czasie realizacji projektu?
- Jak zapewnić dostęp do danych dla zespołu, osób uprawnionych?
- Tworzenie kopii zapasowych i procedur kontroli danych.



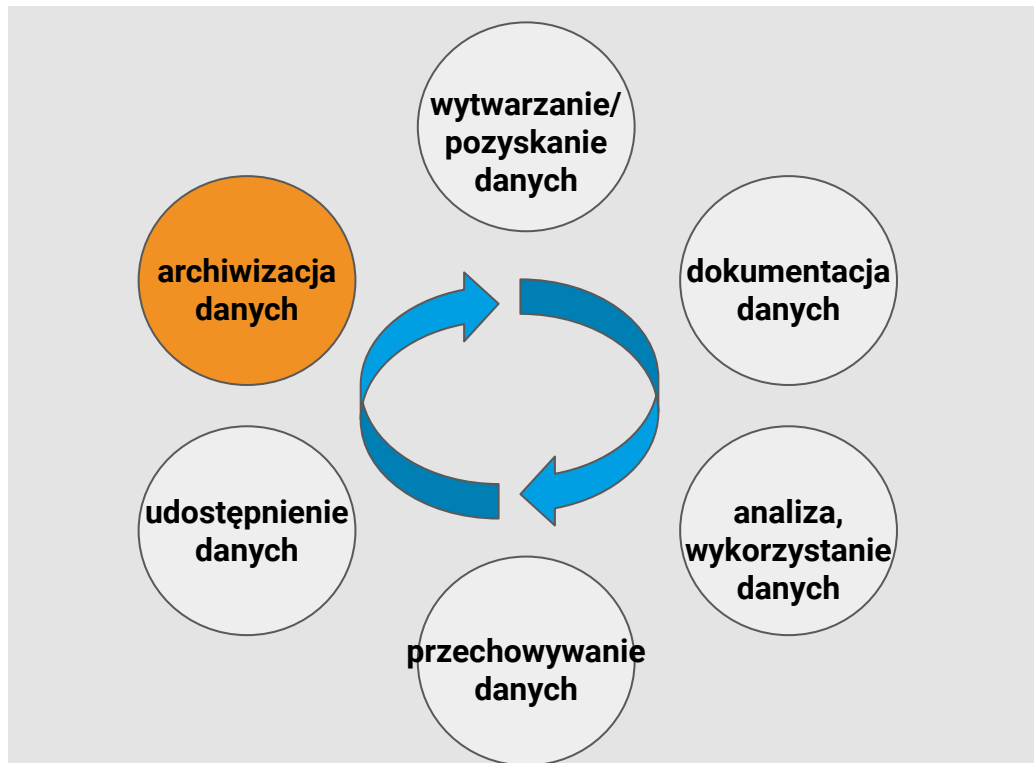
Udostępnianie danych

- Jakie dane jesteśmy zobowiązani udostępnić?
- Jakie dane chcemy i mamy możliwość udostępnić niezależnie od zobowiązań?
- Na jakich zasadach udostępniamy dane badawcze?
- Gdzie udostępnimy dane badawcze?



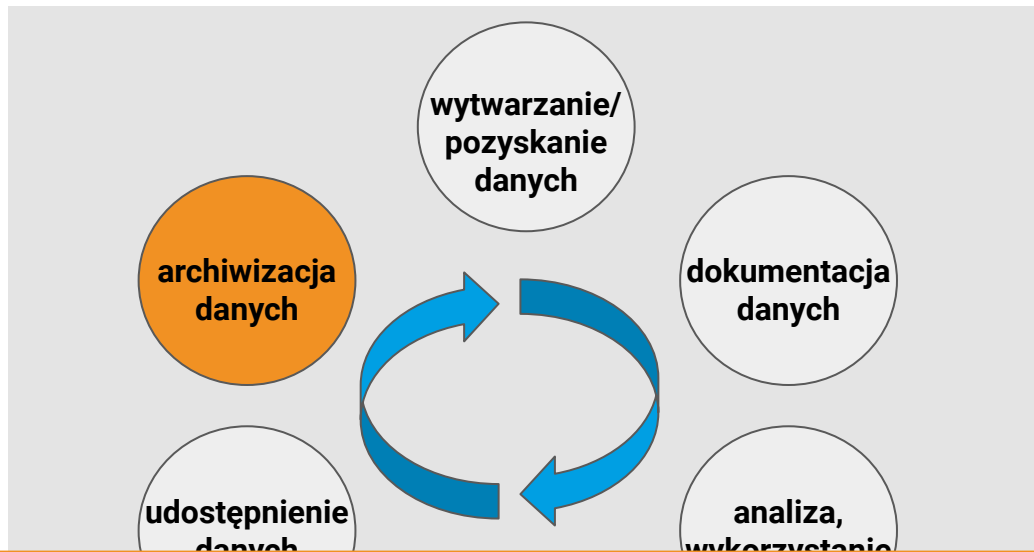
Archiwizacja danych

- Określenie zasad długotrwałego przechowywania danych
- Wybór danych do archiwizacji - na jakiej podstawie?
- Okres przechowywania danych



Archiwizacja danych

- Określenie zasad długotrwałego przechowywania danych
- Wybór danych do archiwizacji - na jakiej podstawie?
- Okres przechowywania danych



² Dane surowe i przetworzone powinny być przechowywane przez okres odpowiedni dla danej dyscypliny i zastosowanej metodologii. W rozumieniu NCN uzasadniony okres przechowywania danych to minimum 10 lat.

źródło: https://www.ncn.gov.pl/sites/default/files/pliki/regulaminy/wytyczne_zarzadzanie_danymi.pdf

Jak wybrać repozytorium?

1.
Repozytoria
dziedziczne



RDS
Repozytorium
Danych Społecznych



MX-RDR
Macromolecular Xtallography
Raw Data Repository

2.
Repozytoria
instytucjonalne

3.
Repozytoria ogólnego
przeznaczenia



RepOD
Repozytorium Otwartych Danych

re3data.org

→ katalog repozytoriów
danych badawczych

→ ponad 2 400
repozytoriów

→ możliwość przeszukiwania
katalogu wg różnych
kryteriów



Filter

- Subjects ⊞
- Content Types ⊞
- Countries ⊞
- AID systems ⊞
- API ⊞
- Certificates ⊞
- Data access ⊞
- Data access restrictions ⊞
- Database access ⊞
- Database access restrictions ⊞
- Database licenses ⊞
- Data licenses ⊞
- Data upload ⊞
- Data upload restrictions ⊞
- Enhanced publication ⊞
- Institution responsibility type ⊞
- Institution type ⊞
- Keywords ⊞
- Metadata standards ⊞
- PID systems ⊞
- Provider types ⊞
- Quality management ⊞
- Repository languages ⊞
- Software ⊞
- Syndications ⊞
- Repository types ⊞
- Versioning ⊞

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Search...

Search



Data journals

- ❖ czasopisma publikujące opisy zestawów danych badawczych,
- ❖ artykuły są recenzowane, czasopisma działają na wzór tradycyjnych czasopism publikujących artykuły,
- ❖ zestawy danych zwykle deponowane są w repozytoriach, czasami publikowane jako załączniki do artykułu

Strony

drodb.icm.edu.pl

pon.edu.pl/dane

Dziękujemy!

Kontakt: drodb@icm.edu.pl



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

