



RepOD

Repozytorium Otwartych Danych

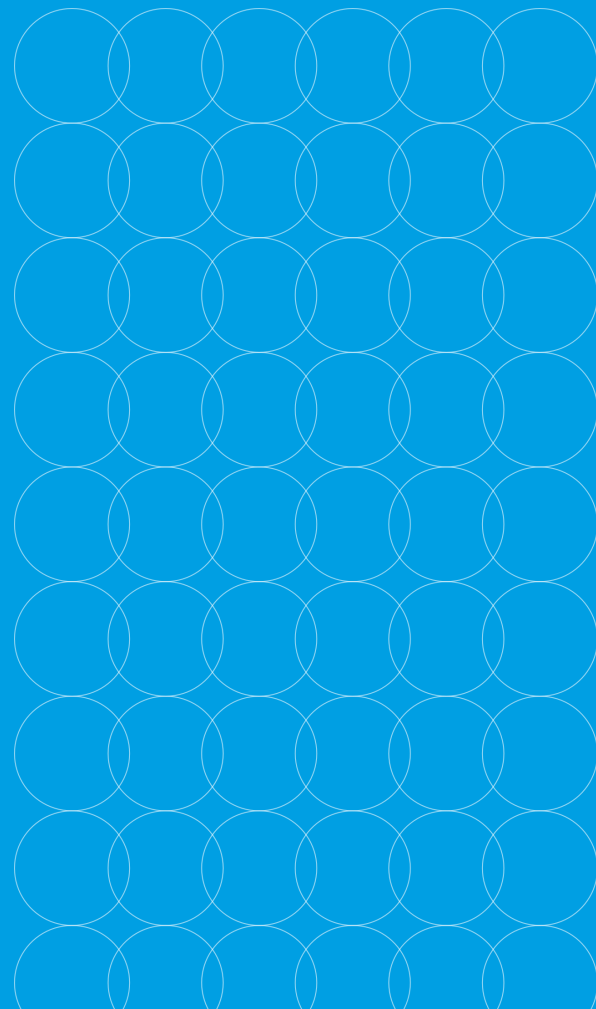
Dziedzinowe Repozytoria Otwartych Danych Badawczych

PREZENTACJA PROJEKTU

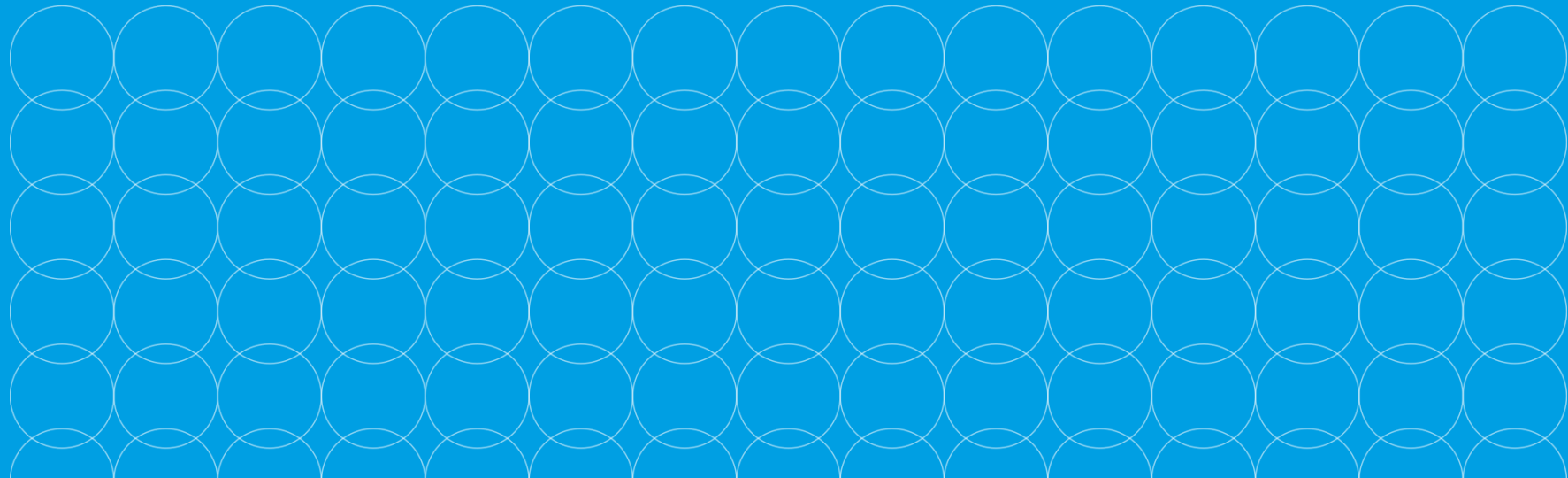


Rzeczpospolita
Polska

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 ogólnego zastosowania,
 - ❖ RDS - repozytorium danych społecznych,
 - ❖ MX-RDR - repozytorium surowych danych krystalograficznych.



RepOD

Repozytorium Otwartych Danych

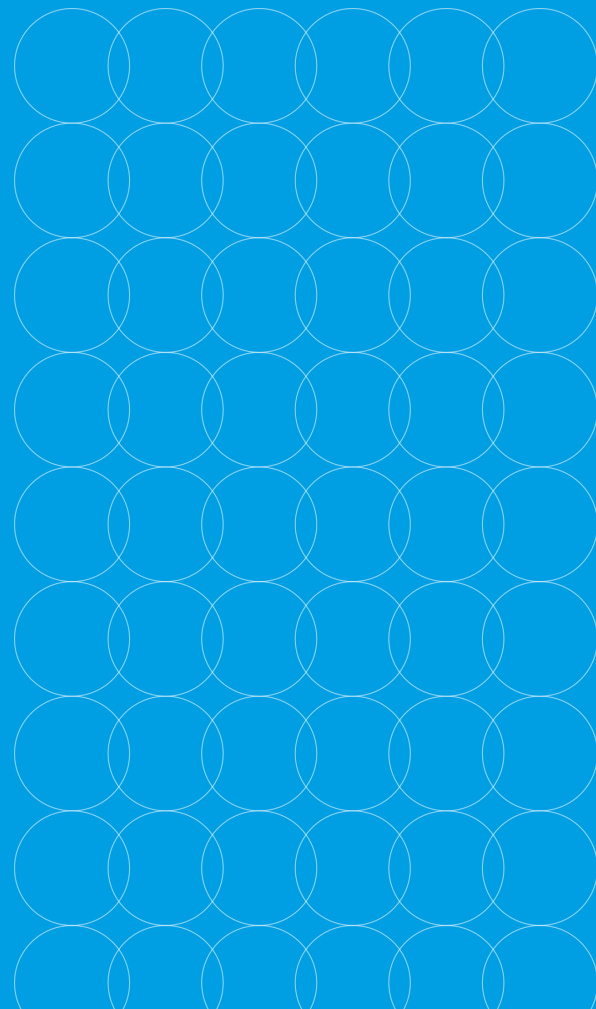
RepOD

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



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego





RDS

Repozytorium
Danych Społecznych

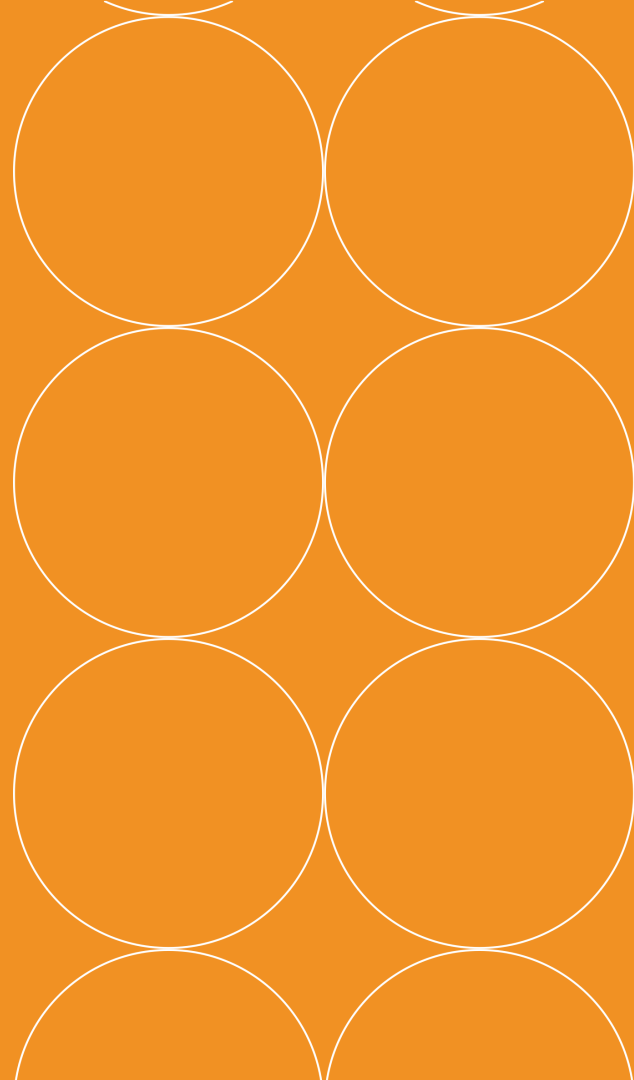
RDS: Repozytorium Danych Społecznych

- 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 POZNAŃCU



Fundusze
Europejskie
Polska Cyfrowa

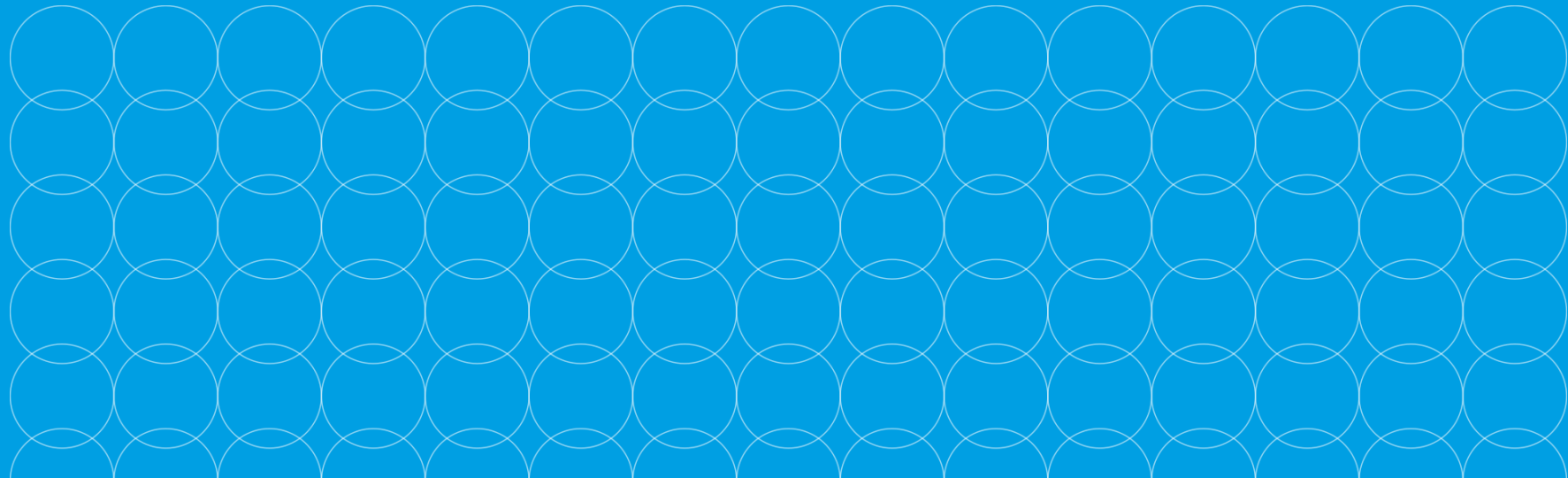


Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



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





Unia Europejska

DZIEDZINOWE REPOZYTORIA OTWARTYCH DANYCH BADAWCZYCH

Witryna informacyjna projektu

[Aktualności](#)

[O projekcie](#) ▾

[Szkolenia](#) ▾

[Pliki do pobrania](#)

[Polityka prywatności](#)

[Kontakt](#)

PLIKI DO POBRANIA

[Ulotka informacyjna](#)

[Selekcja i przygotowanie danych badawczych do udostępniania](#)

[Jak korzystać z zasobów w repozytoriach danych](#)





RepOD

Repozytorium Otwartych Danych

Zarządzanie danymi badawczymi

NATALIA GRUENPETER

CC-BY

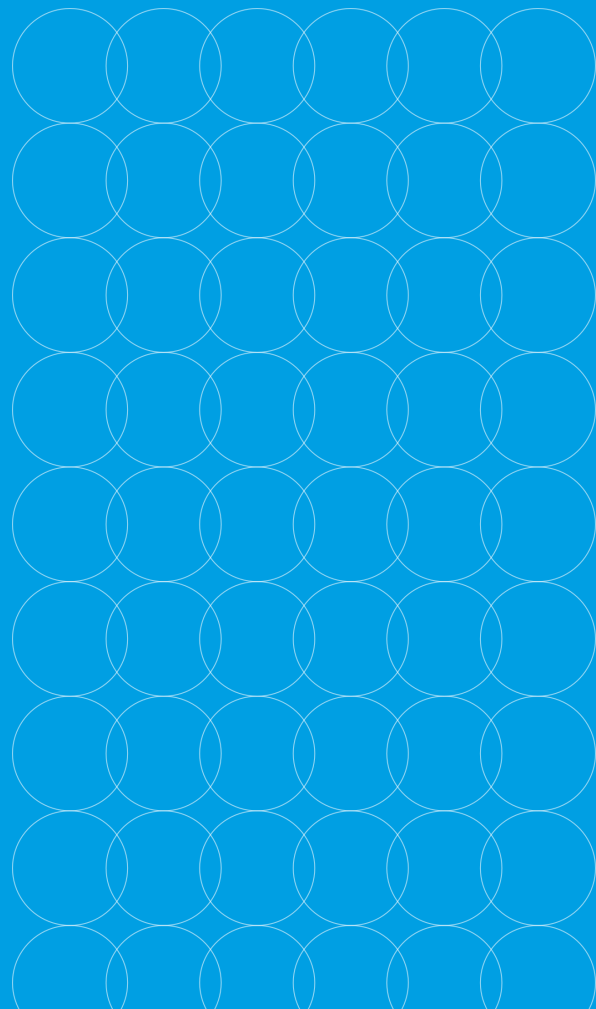
Treść licencji dostępna na stronie:

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



Rzeczpospolita
Polska

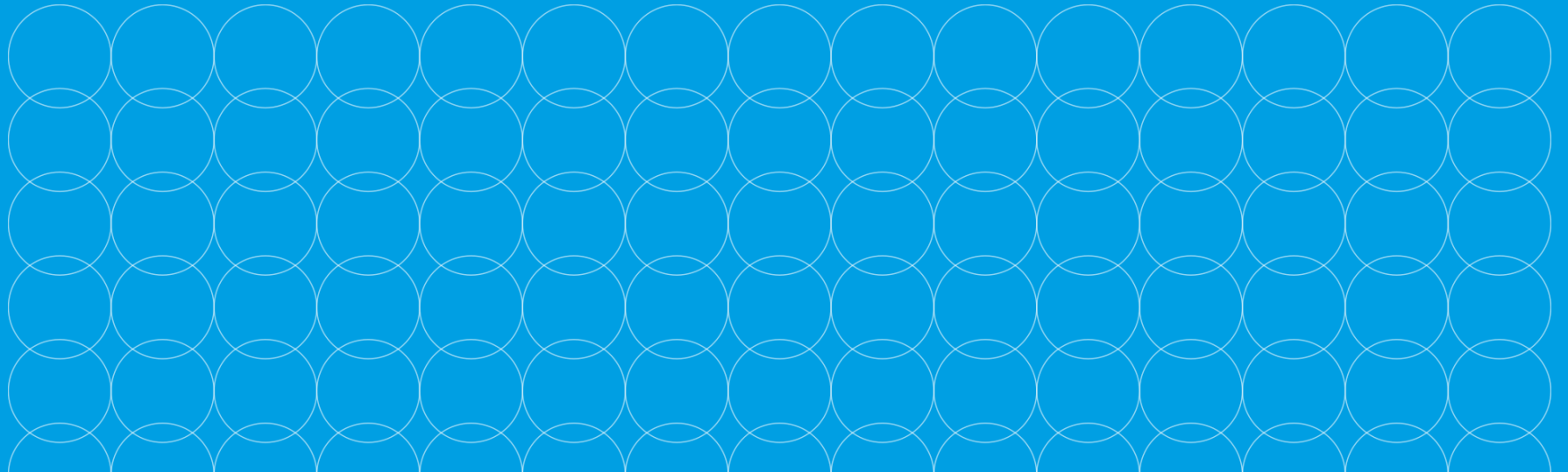
Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



Plan szkolenia część 1.

- Dane badawcze
- Zarządzanie danymi badawczymi
- Wybieranie i przygotowywanie danych do długoterminowego przechowywania
- Plan Zarządzania Danymi (Data Management Plan, DMP)
- Otwarte udostępnianie danych badawczych

Dane badawcze

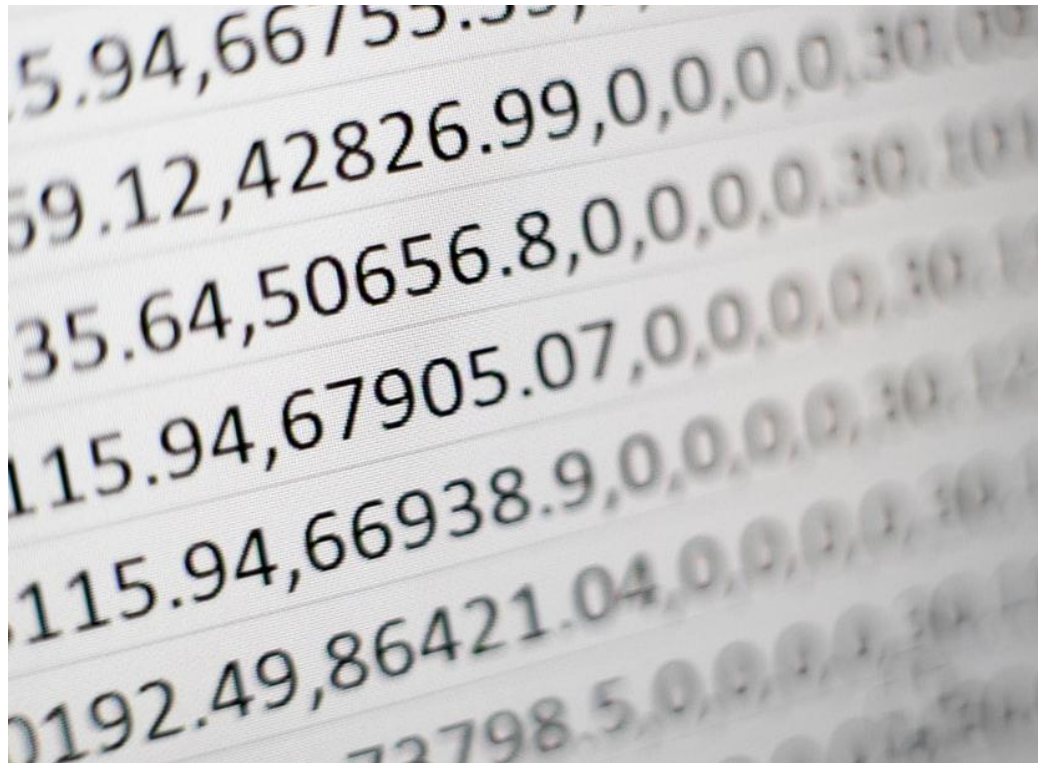


Co to są dane badawcze?

- ❖ 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.

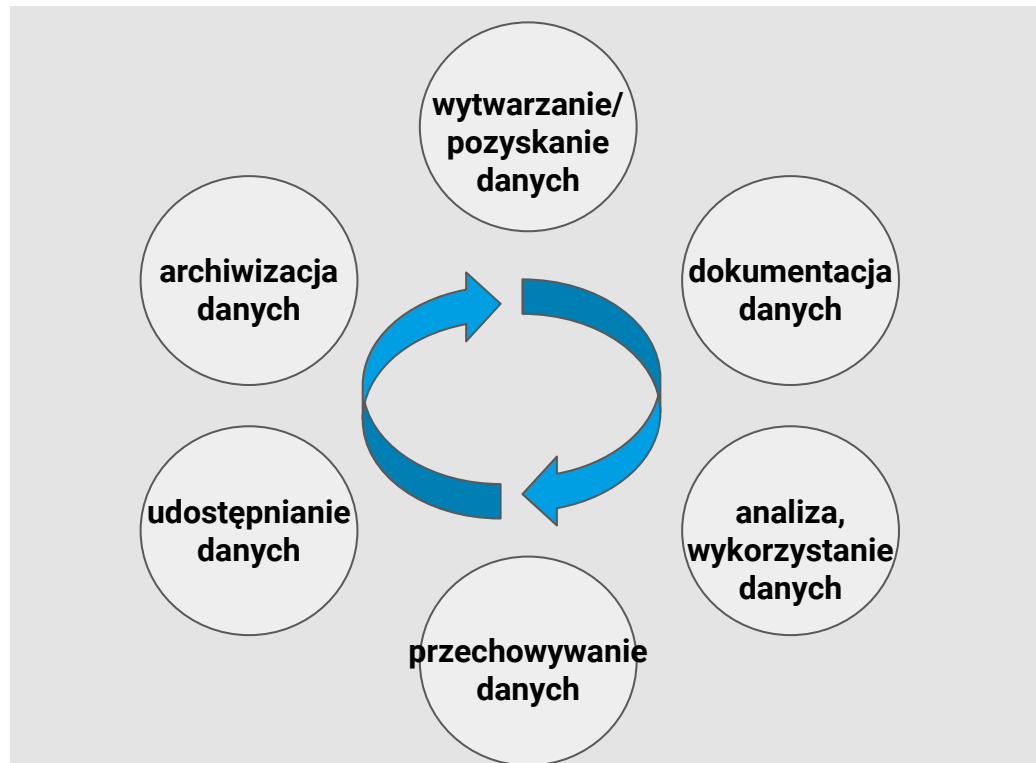
Dane w różnych aspektach

- **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**



Dane w całym cyklu projektu badawczego

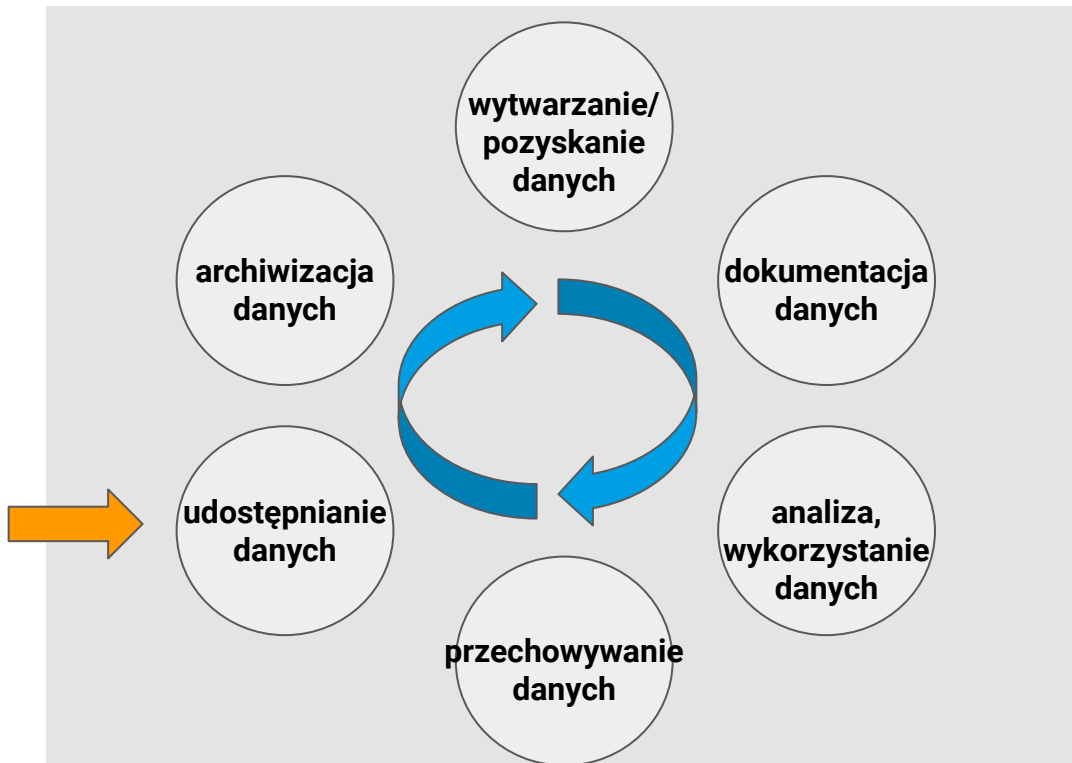
Zarządzanie danymi badawczymi sposób, w jaki organizujemy pracę z danymi badawczymi w toku całego projektu i po jego zakończeniu.



Dane w całym cyklu projektu badawczego

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

Otwarty dostęp a możliwość ponownego wykorzystania danych - w jakim zakresie? Co należy zaplanować, aby było to możliwe w jak najszerszym zakresie?



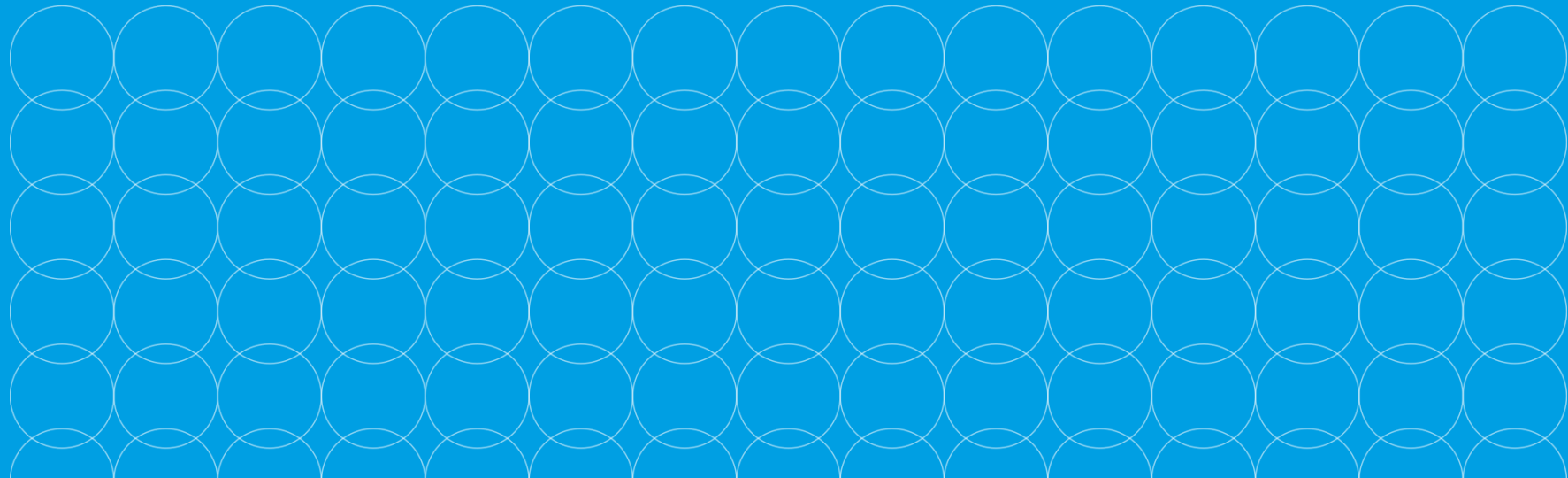
Plan Zarządzania Danymi, DMP

Formalny dokument,
który zawiera zarys
tego, co będziemy
robić z danymi
w trakcie projektu
badawczego
i po jego zakończeniu.



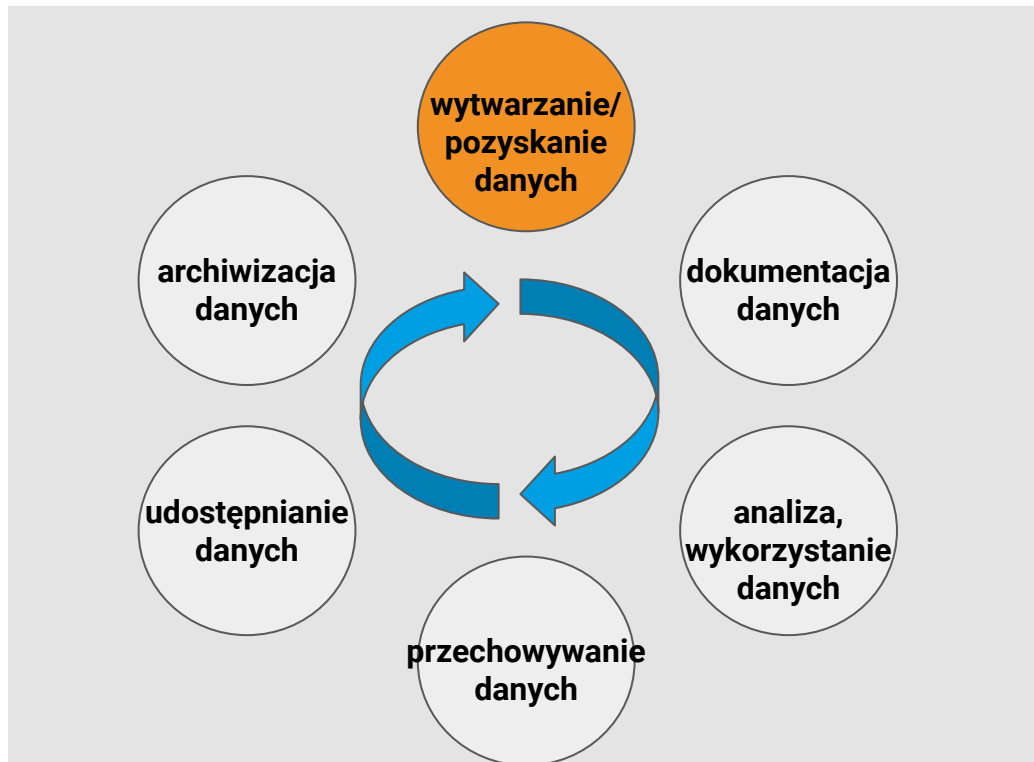
Ilustracja: [The Data Management Expert Guide](#)
by CESSDA ERIC, CC BY SA.

Zarządzanie danymi badawczymi



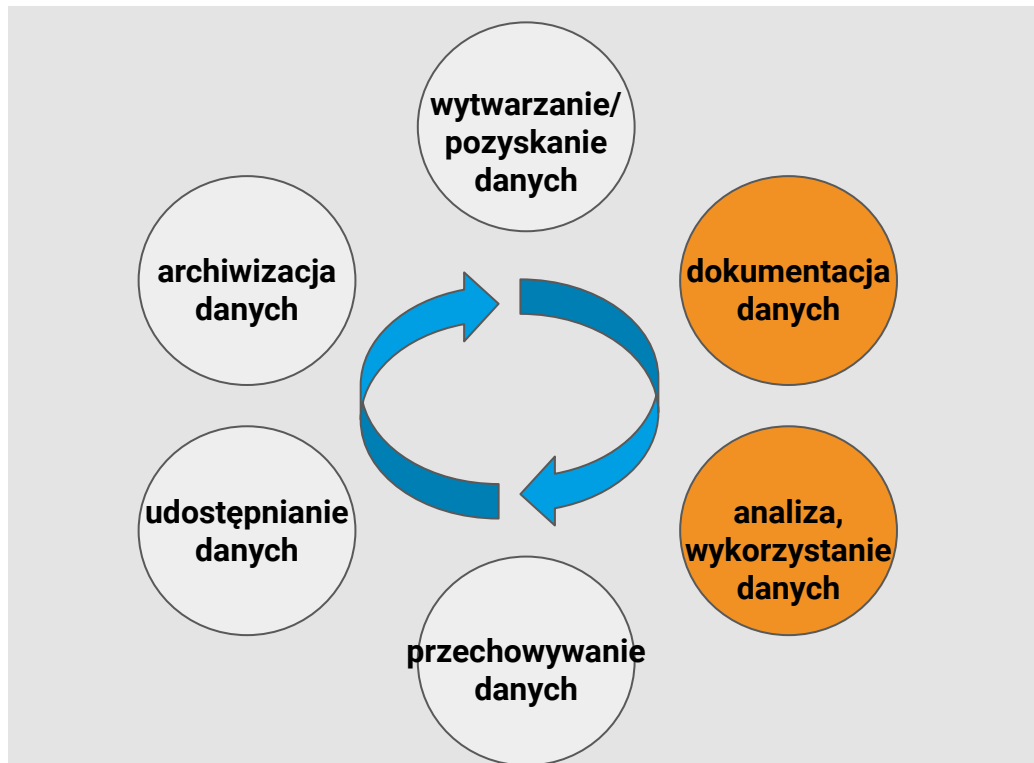
Wytwarzanie, pozyskanie danych

- 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

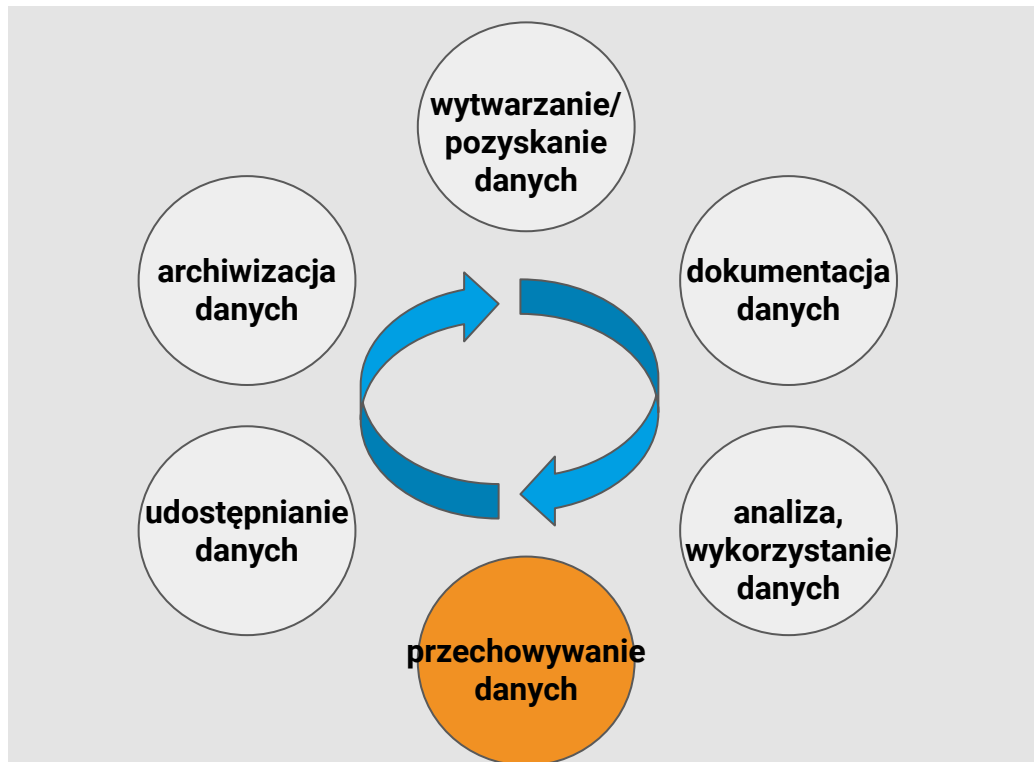
- ❖ W jakim celu/kontekście dane zostały wytworzone?
- ❖ Co zawierają zestawy danych?
- ❖ Jak dane zostały wytworzone/pozyskane?
- ❖ Kto i kiedy to zrobił?
- ❖ Jak dane były przetwarzane/opracowywane?
- ❖ Czy dane były modyfikowane?
- ❖ Jakie zastosowano metody sprawdzania jakości danych?
- ❖ W jaki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

Dokumentacja: dane

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

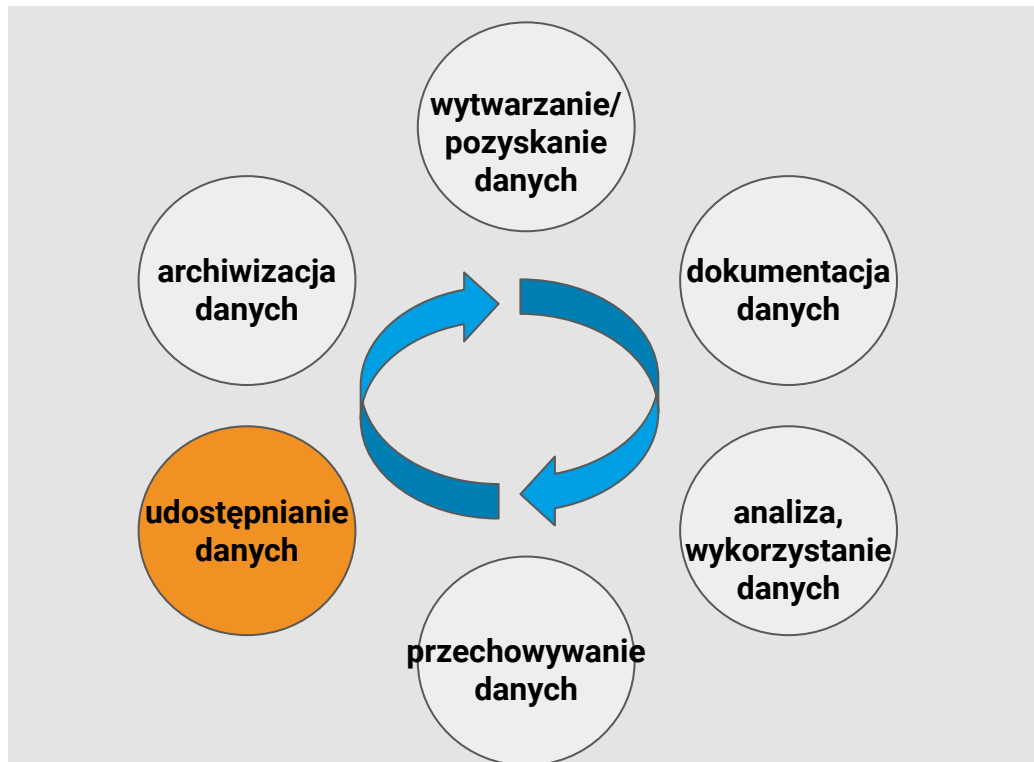
Przechowywanie danych

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



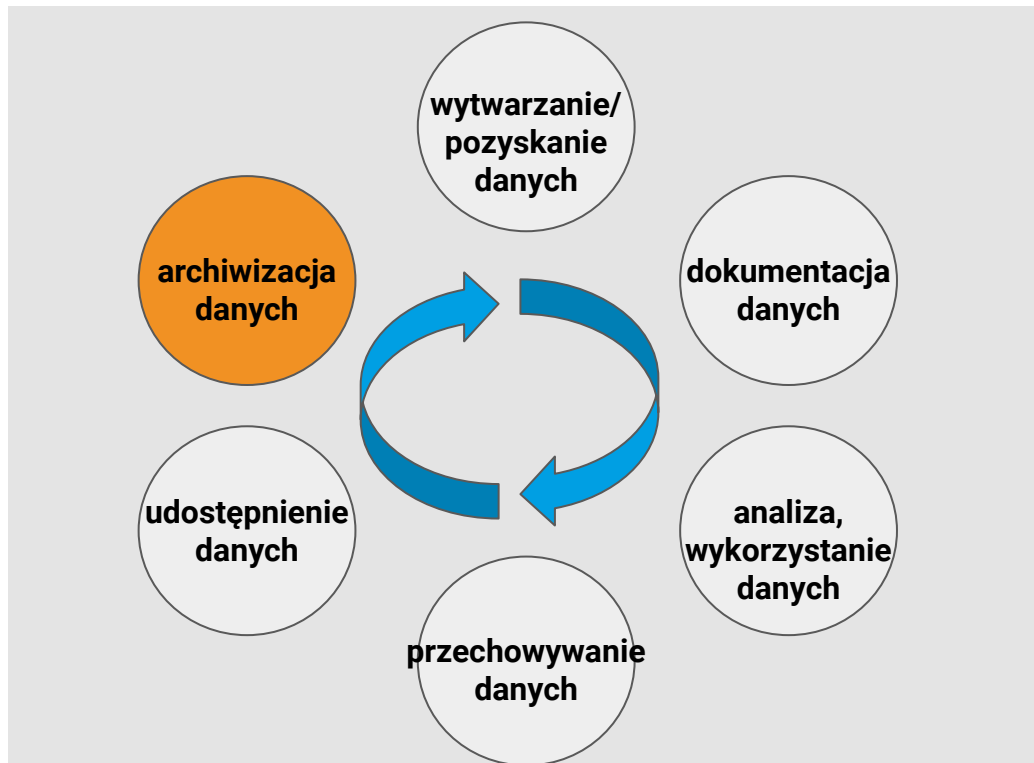
Udostępnianie danych

- 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?



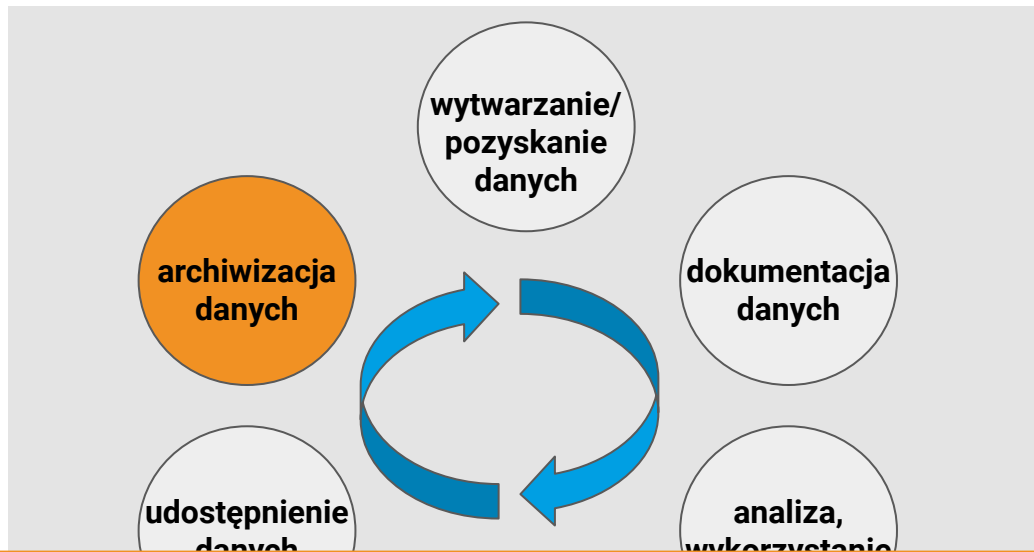
Archiwizacja danych

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



Archiwizacja danych

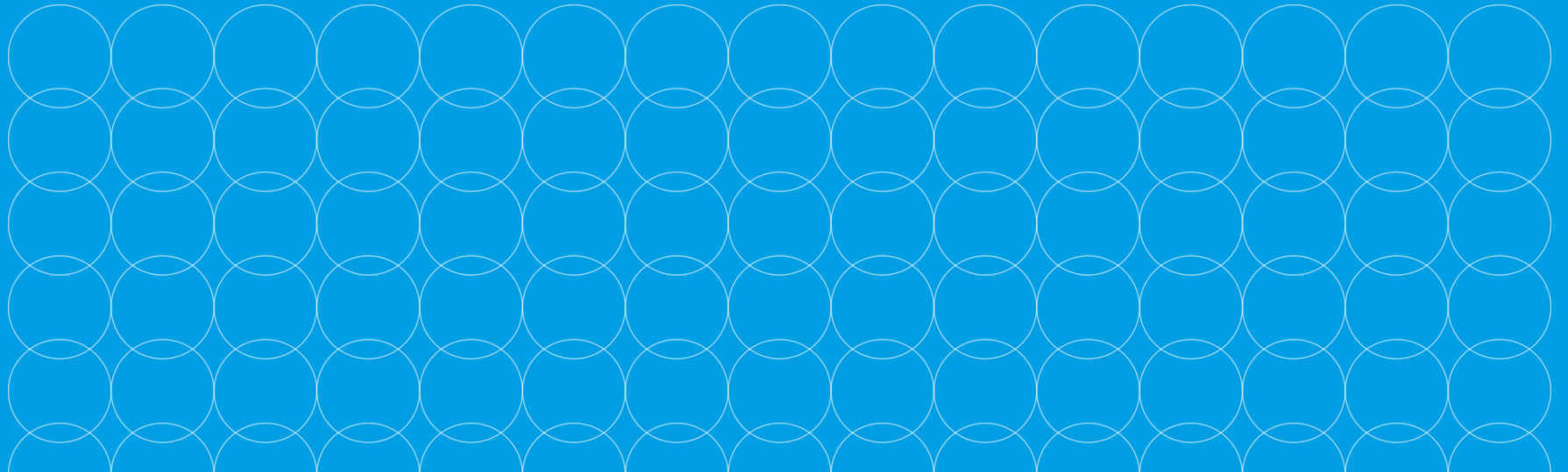
- Określenie zasad długotrwałego przechowywania danych
- Wybór danych do archiwizacji - na jakiej podstawie?
- Okres przechowywania danych



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

Wybieranie i przygotowywanie danych do długoterminowego przechowywania

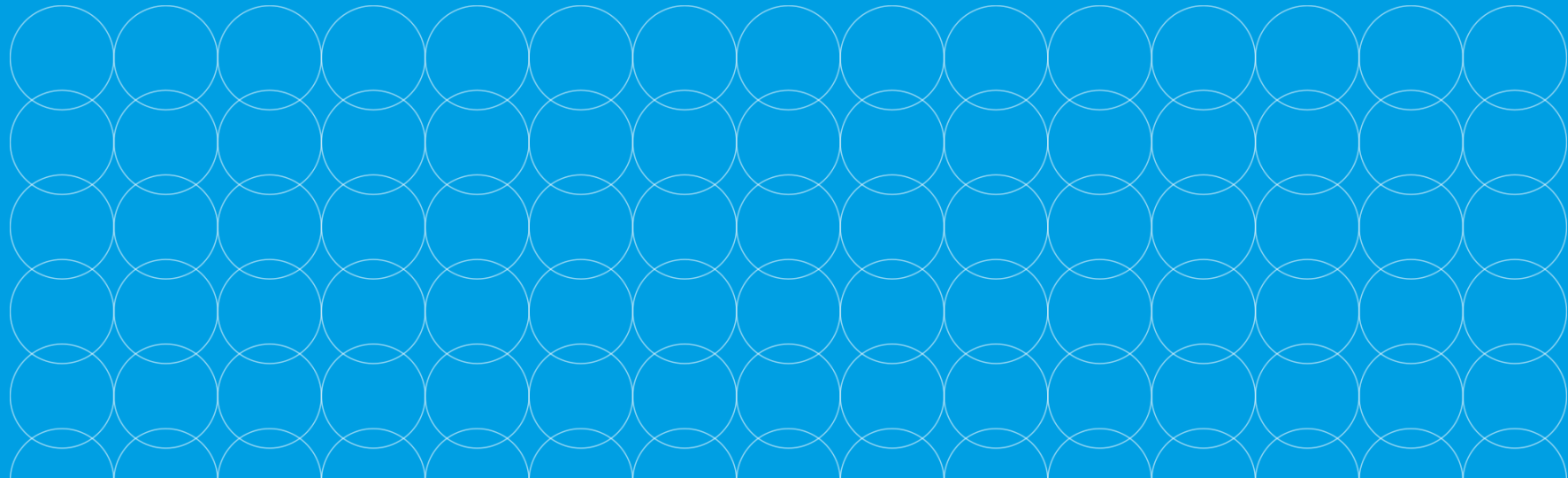


Co należy wziąć pod uwagę?

- ❖ 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?

Plan zarządzania danymi badawczymi

Data Management Plan, DMP



DMP

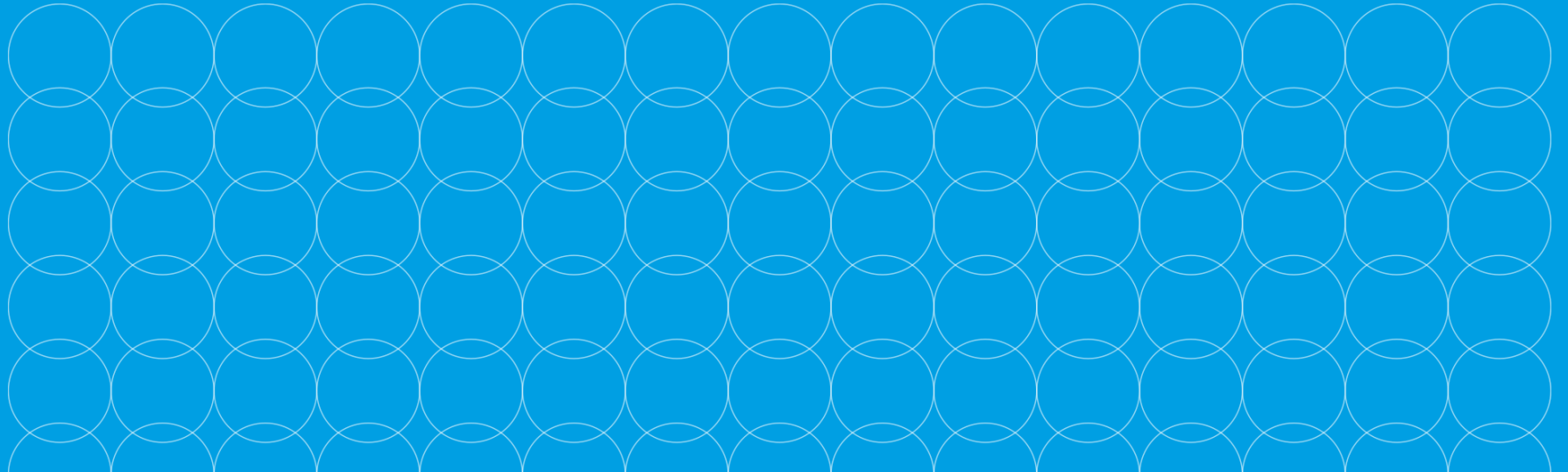
- ❖ różne wymogi → różne wzory;
- ❖ mogą być dostosowane do wymogów instytucji, grantodawcy, konkretnej dyscypliny

- ❖ żywy dokument, który może wspierać realizację projektu na każdym etapie: planowania, prowadzenia badań, upowszechniania wyników badań

Zwyczajowe elementy planu zarządzania danymi badawczymi:

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

Otwarte udostępnianie danych badawczych



Wymogi

- ❖ 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

F
indable



A
ccessible



I
nteroperable



R
eusable



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

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)

Otwarte dane - korzyści

- ❖ odpowiednie opracowanie i opisanie danych,
- ❖ możliwość zacytowania danych, a także powiązanych z danymi publikacji,
- ❖ promocja dorobku naukowego,

- ❖ możliwość weryfikacji i ponownej analizy danych,
- ❖ łatwe wyszukiwanie danych,
- ❖ rozwój nauki, możliwość przeprowadzenia nowych analiz

Repozytoria danych

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

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...

Deponowanie danych

❖ Dataverse

❖ Film

Data journals

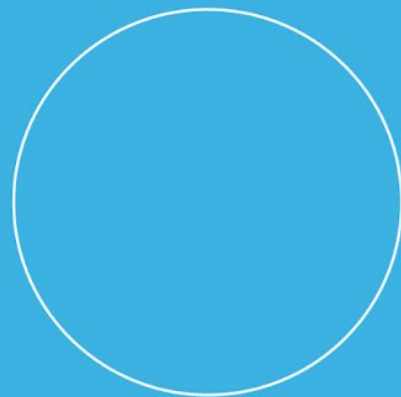
- ❖ czasopisma publikujące opisy zestawów danych badawczych,
- ❖ artykuły są recenzowane, czasopisma działają na wzór tradycyjnych czasopism publikujących artykuły,
- ❖ zestawy danych zwykle deponowane są w repozytoriach, czasami publikowane jako załączniki do artykułu

Broszury

Dostępne na stronie
drodb.icm.edu.pl

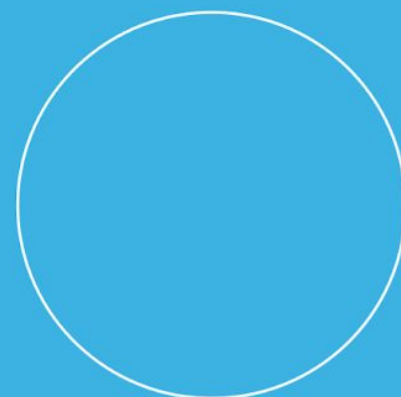
Selekcja i przygotowanie danych badawczych do udostępnienia

Wersja 1.0



Jak korzystać z zasobów w repozytoriach danych

Wersja 1.0



Dziękujemy!

Dane kontaktowe: drodb-strona@icm.edu.pl



Fundusze Europejskie
Polska Cyfrowa



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego

