RepOD
Repozytorium Otwartych Danych

Fostering Institutional Collaboration through “Disciplinary Open Research Data Repositories” project

Natalia Gruenpeter, ICM UW, Poland

Open Science Fair 2021

CC-BY, https://creativecommons.org/licenses/by/4.0/

UNIWERSYTET WARSZAWSKI

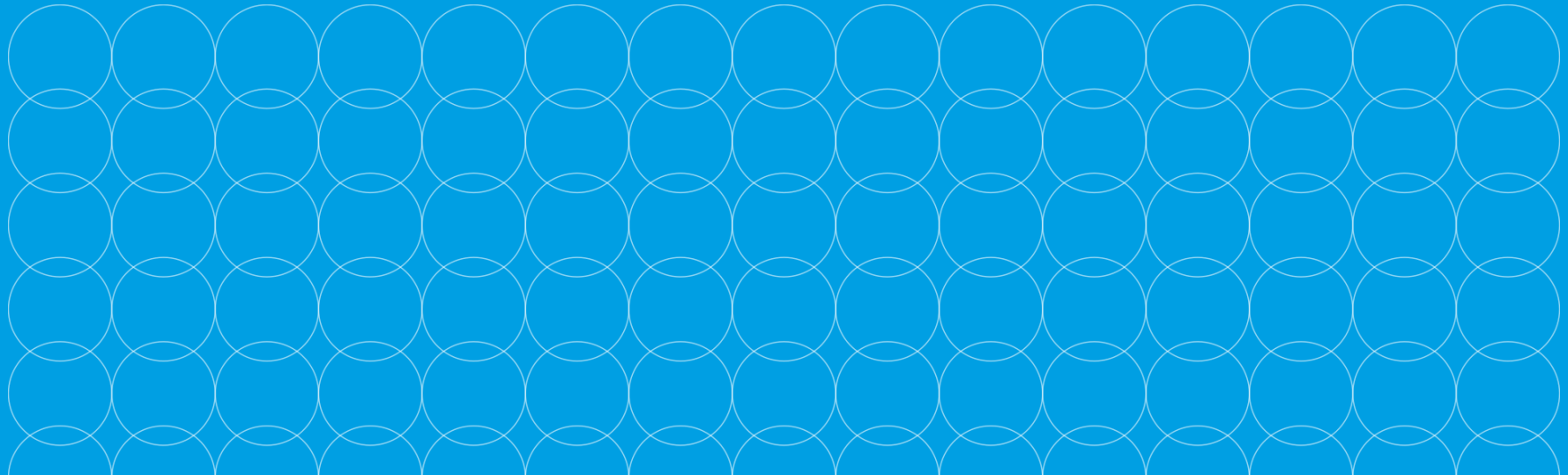
icm

European Funds Digital Poland

Republic of Poland

European Union
European Regional Development Fund

The project



Aims and objectives

- enhancing access to academic resources by making them available in open research data repositories
- improving the quality of shared data and metadata and facilitating their better use
- improving legal accessibility of already shared data

https://drodb.icm.edu.pl/english/

Institutional collaboration

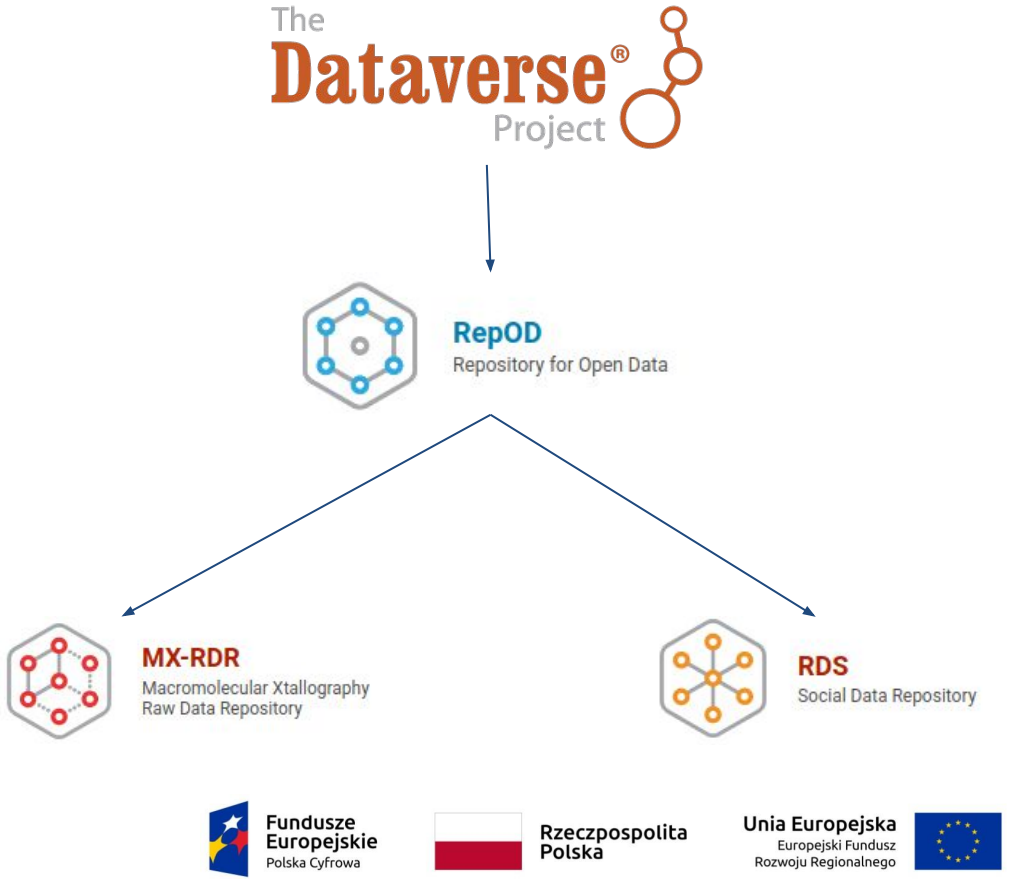
Project leader

- University of Warsaw
  - Interdisciplinary Centre for Mathematical and Computational Modelling (ICM)
  - Robert Zajonc Institute for Social Studies (ISS)

Project partners

- Institute of Philosophy and Sociology Polish Academy of Sciences (IFiS PAN),
- Adam Mickiewicz University (UAM) - responsible for crystallographic data

Data repositories



Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

RepOD

Repozytorium Otwartych Danych

Repository for Open Data (RepOD)

repod.icm.edu.pl

- a publicly accessible general data repository
- open for everyone to use
- new version (previous one was created in 2015)

UNIWERSYTET WARSZAWSKI

icm

Fundusze Europejskie Polska Cyfrowa

Rzeczpospolita Polska

Unia Europejska Europejski Fundusz Rozwoju Regionalnego

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R3JS93

RepOD
Repozytorium Otwartych Danych

Main changes:

- Polish and English interface,
- licenses assigned to files,
- WCAG improvements,
- code refactor.

UNIWERSYTET WARSZAWSKI
icm
Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

RDS
Repozytorium
Danych Społecznych

Social Data Repository (RDS)

rds.icm.edu.pl

- domain-specific repository for social science research data
- open for everyone to use

UNIWERSYTET WARSZAWSKI · icm · ISS · IFiS PAN

Fundusze Europejskie Polska Cyfrowa · Rzeczpospolita Polska · Unia Europejska Europejski Fundusz Rozwoju Regionalnego

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R31NJMYP

RDS
Repozytorium
Danych Społecznych

Main changes:

- changes in social science metadata,
- licenses assigned to files,
- Restricted Access,
- guestbooks.

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

RDS
Repozytorium
Danych Społecznych

Collections for archives

ARCHIWUM DANYCH JAKOŚCIOWYCH
Archiwum Danych Jakościowych

PADS POLSKIE ARCHIWUM DANYCH SPOŁECZNYCH
Polskie Archiwum Danych Społecznych

https://adj.ifispan.pl/index.php/en/welcome/
https://pads.org.pl/index.php/en/

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

MX-RDR
Macromolecular Xtallography
Raw Data Repository

Macromolecular Xtallography Raw Data Repository (MX-RDR)

mxrdr.icm.edu.pl

- domain-specific repository for raw diffraction data collected for macromolecular crystals
- open for everyone to use

Uniwersytet Warszawski · icm · Uniwersytet im. Adama Mickiewicza w Poznaniu

Fundusze Europejskie Polska Cyfrowa · Rzeczpospolita Polska · Unia Europejska Europejski Fundusz Rozwoju Regionalnego

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

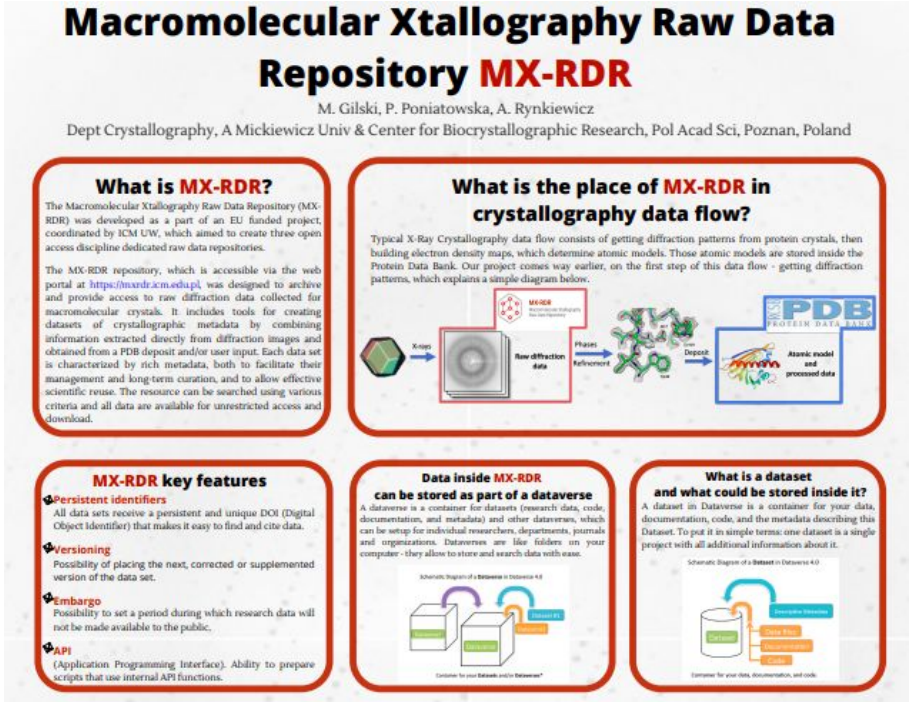
http://doi.org/10.17616/R31NJMYU

MX-RDR
Macromolecular Xtallography
Raw Data Repository

Main changes:

- macromolecular crystallography metadata,
- tools for creating datasets of crystallographic metadata by combining information extracted directly from diffraction images and obtained from a PDB deposit and/or user input.

UNIWERSYTET WARSZAWSKI
icm
UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego



Macromolecular Xtallography Raw Data Repository MX-RDR

M. Gilski, P. Poniatowska, A. Rynkiewicz
Dept Crystallography, A Mickiewicz Univ & Center for Biocrystallographic Research, Pol Acad Sci, Poznan, Poland

What is MX-RDR?

The Macromolecular Xtallography Raw Data Repository (MX-RDR) was developed as a part of an EU funded project, coordinated by ICM UW, which aimed to create three open access discipline dedicated raw data repositories.

The MX-RDR repository, which is accessible via the web portal at https://mxrdr.icm.edu.pl, was designed to archive and provide access to raw diffraction data collected for macromolecular crystals. It includes tools for creating datasets of crystallographic metadata by combining information extracted directly from diffraction images and obtained from a PDB deposit and/or user input. Each data set is characterized by rich metadata, both to facilitate their management and long-term curation, and to allow effective scientific reuse. The resource can be searched using various criteria and all data are available for unrestricted access and download.

What is the place of MX-RDR in crystallography data flow?

Typical X-Ray Crystallography data flow consists of getting diffraction patterns from protein crystals, then building electron density maps, which determine atomic models. Those atomic models are stored inside the Protein Data Bank. Our project comes way earlier, on the first step of this data flow - getting diffraction patterns, which explains a simple diagram below.

MX-RDR key features

Persistent identifiers
All data sets receive a persistent and unique DOI (Digital Object Identifier) that makes it easy to find and cite data.

Versioning
Possibility of placing the next, corrected or supplemented version of the data set.

Embargo
Possibility to set a period during which research data will not be made available to the public.

API
(Application Programming Interface). Ability to prepare scripts that use internal API functions.

Data inside MX-RDR can be stored as part of a dataverse

A dataverse is a container for datasets (research data, code, documentation, and metadata) and other dataverses, which can be setup for individual researchers, departments, journals and organizations. Dataverses are like folders on your computer - they allow to store and search data with ease.

What is a dataset and what could be stored inside it?

A dataset in Dataverse is a container for your data, documentation, code, and the metadata describing this Dataset. To put it in simple terms: one dataset is a single project with all additional information about it.

https://drodb.icm.edu.pl/wp-content/uploads/2021/07/KK62-Poster-03.pdf

UNIWERSYTET WARSZAWSKI

icm

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU

Fundusze Europejskie Polska Cyfrowa

Rzeczpospolita Polska

Unia Europejska Europejski Fundusz Rozwoju Regionalnego

RDM training

Educational materials (in Polish) concerning:

- data usage,
- data sharing,
- legal issues.

Workshops (webinars) on research data management:

- general,
- domain specific,
    - developed and co-organized with partners.

Beyond the project

Policy environment

National Science Centre

2019 - mandatory data management plans

2020 - Open Access Policy

- Plan S implementation
- Open research data

“Where possible, underlying data (basic data sets) related to the published articles should be placed in an open repository (...)”

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

Institutional collaboration

Other institutions

- institutional collections in the Repository for Open Data (general repository),
- single point of access to research data from various institutions,
- facilitates making research data FAIR,
- offered at no costs,
- currently: 11 agreements

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

Institutional collaboration

Training

- RDM training
- user guide for RepOD (webinars)

Open data repositories

- increasing the uptake of available services
- understanding research performing organisations’ needs
  - support for collection managers
  - exchange of experience and good practice

Take-home messages

- continuity and consistency builds trust
- context matters: policy environment for (open) research
- a need for flexible and agile implementation of infrastructure solutions

Fundusze Europejskie Polska Cyfrowa
Rzeczpospolita Polska
Unia Europejska Europejski Fundusz Rozwoju Regionalnego

Thank you!

Contact:

drodb@icm.edu.pl
n.gruenpeter@icm.edu.pl

RepOD
Repozytorium Otwartych Danych

UNIWERSYTET WARSZAWSKI

icm

Fundusze Europejskie Polska Cyfrowa

Rzeczpospolita Polska

Unia Europejska
Europejski Fundusz Rozwoju Regionalnego