GDAŃSK UNIVERSITY OF TECHNOLOGY

Ensuring compliance with the FAIR principles in access to research data by building services for data management in scientific institutions

Dr Anna Wałek
Gdańsk University of Technology Library

ALLDATA 2022 --- Barcelona 24-28.04.2022

GDAŃSK UNIVERSITY OF TECHNOLOGY

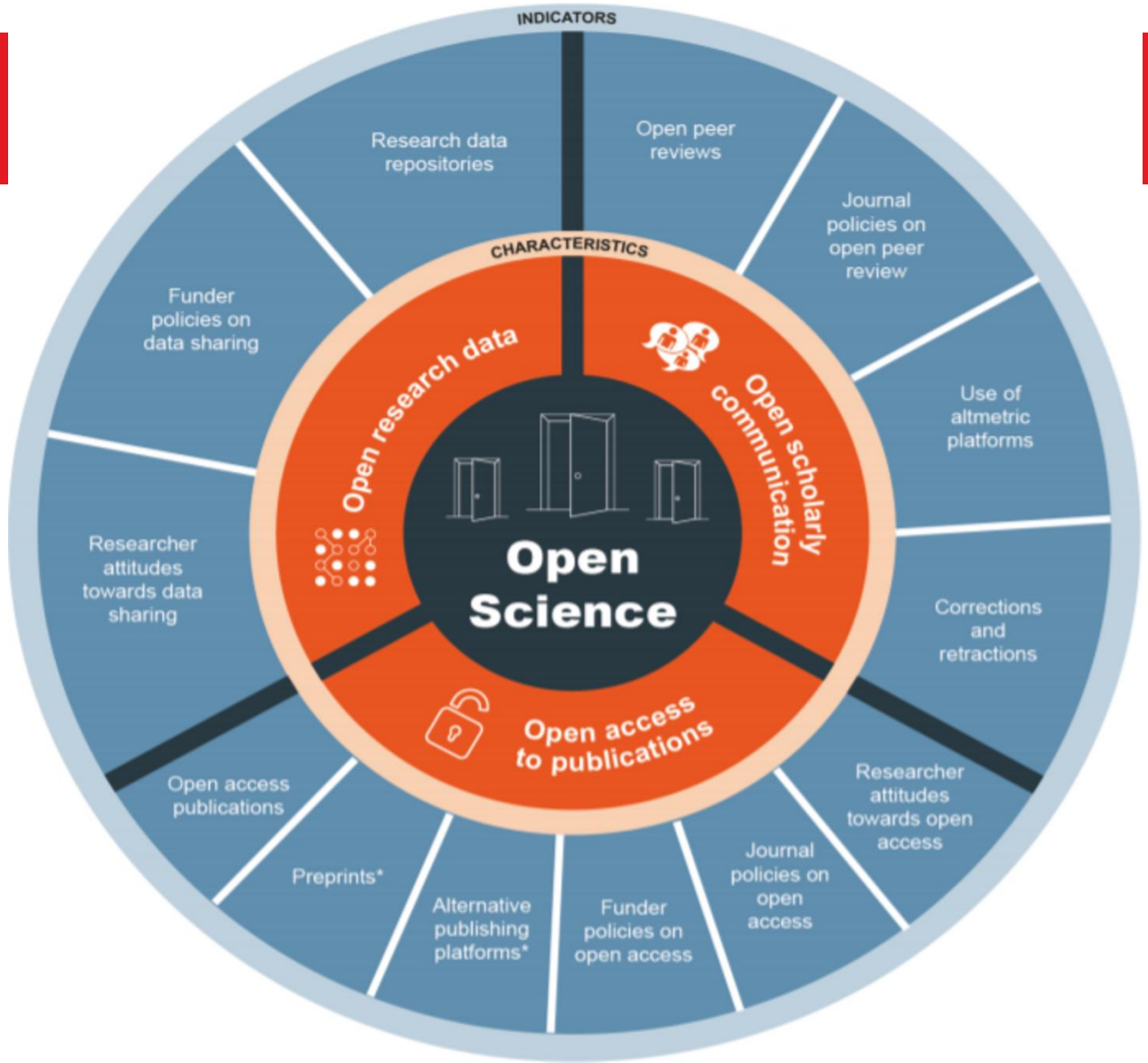
BRIDGE OF DATA

LIBRARY – THE EVOLUTION

Open Science

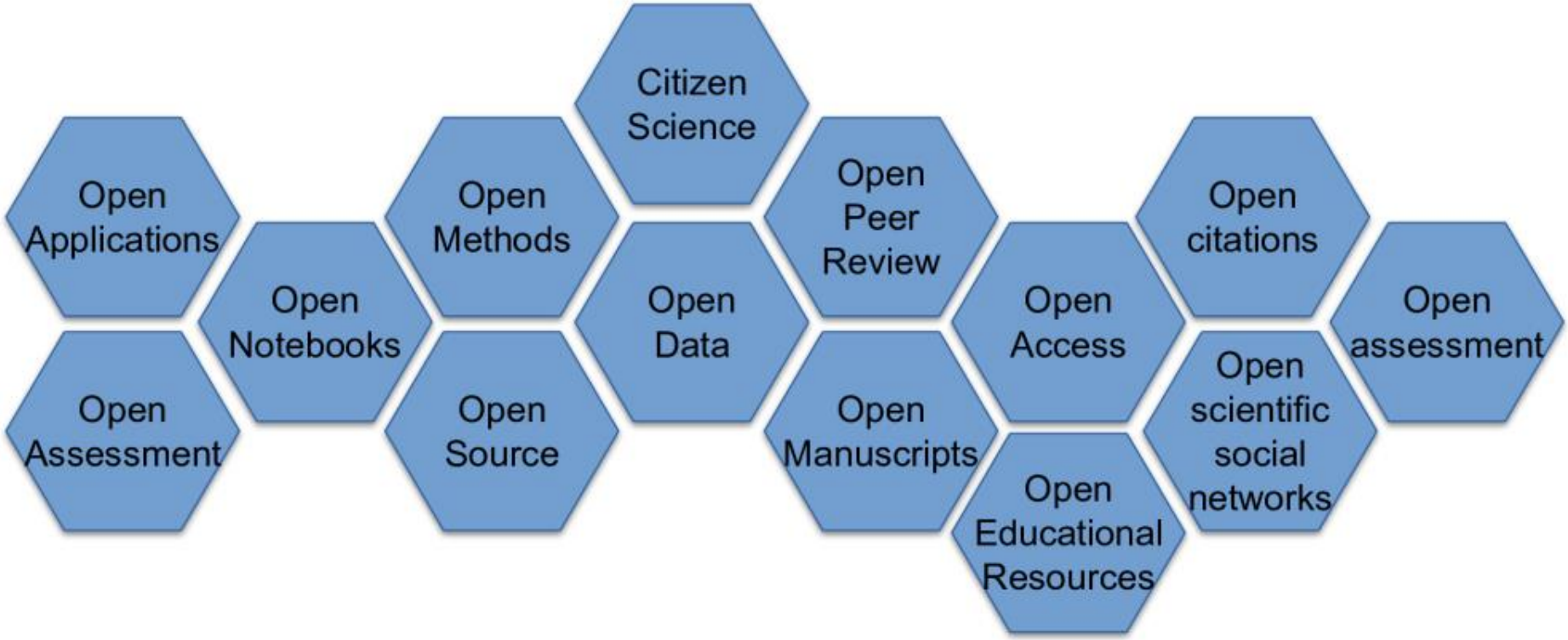
GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA



GDAŃSK UNIVERSITY OF TECHNOLOGY

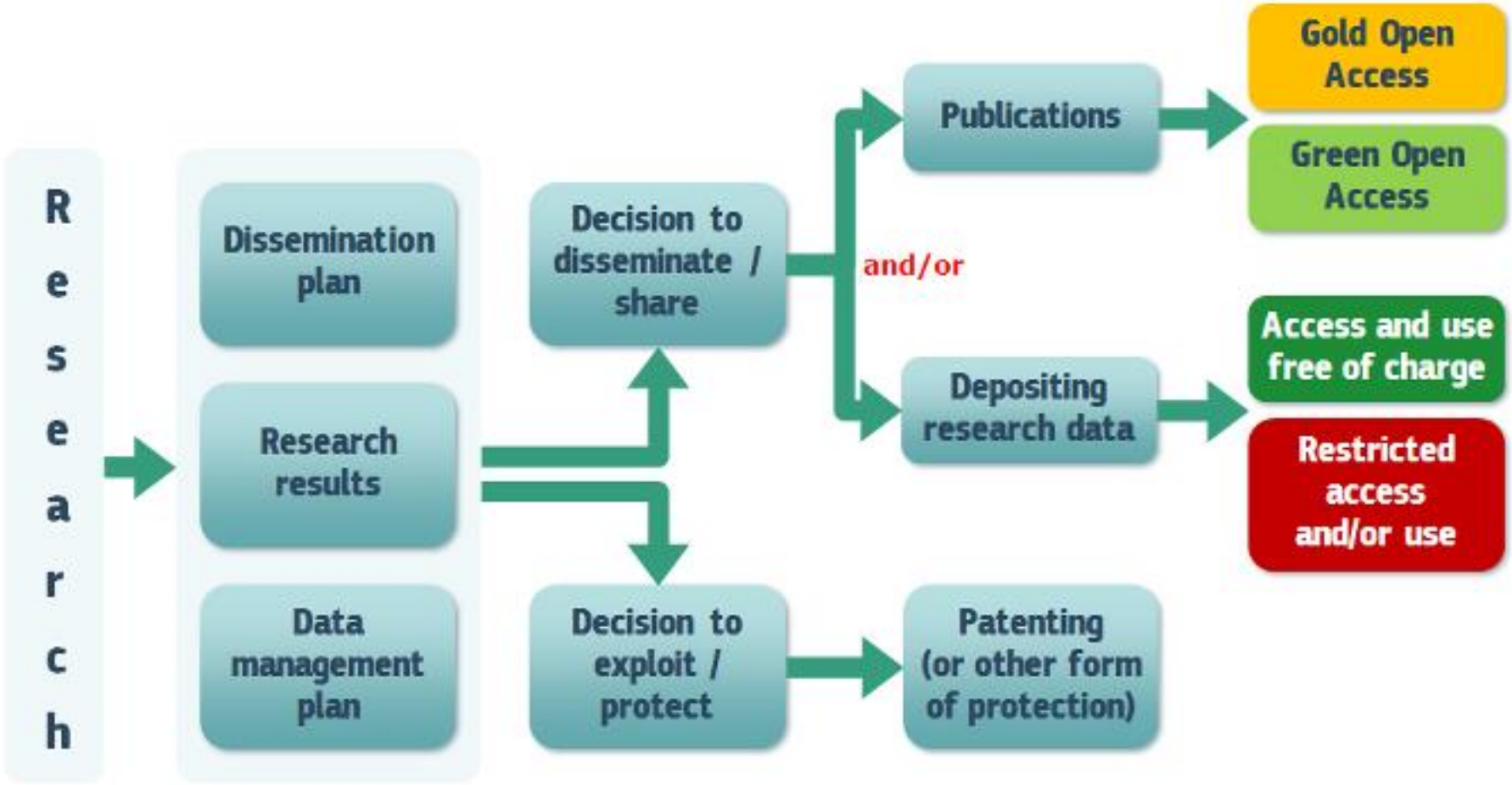
BRIDGE OF DATA



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Research results publication



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Plan S

GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

European Open Science Policy

Open Science

An approach to the scientific process that focuses on spreading knowledge as soon as it is available using digital and collaborative technology. Expert groups, publications, news and events.

PAGE CONTENTS

The EU's open science policy

Open science is a policy priority for the European Commission and the standard method of working under its research and innovation funding programmes as it improves the quality, efficiency and responsiveness of research.

When researchers share knowledge and data as early as possible in the research process with all relevant actors it helps diffuse the latest knowledge.

And when partners from across academia, industry, public authorities and citizen groups are invited to participate in the research and innovation process, creativity and trust in science increases.

That is why the Commission requires beneficiaries of research and innovation funding to make their publications available in open access and make their data as open as possible and as closed as necessary. It recognises and rewards the participation of citizens and end users.

GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

European Open Science Policy

The World Economic Forum Davos 2020 Ursula von der Leyen

"Data is a renewable resource as much as sun and wind. Every 18 months we double the amount of data we produce, 85 per cent of which is never used" .

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

National Science Centre Poland

Polish National Science Centre

One of the signatories of PlanS

DMP’s in the project proposals since 2020

Open research data
Data management plans
Open data
Open science
Research data
FAIR
FAIRest
Openess
Knowledge
DOI
DataCite
Publishing
Archiving
Repository
Preservation
Researchers
Storage
Reuse
DataDOI
Open Access
OpenAIRE
Data management
University of Tartu Library

GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

Research Data Management

What is research data:

Research data are the original sources or material that you have created or collected to conduct your research project.

They can be digital or non-digital.

The response to your research question is based on the analysis of these research data

GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

Research Data Management

types of data:

- qualitative (dealing with things numerically)
- quantitative (descriptive in nature and dealing with the quality, categorization).

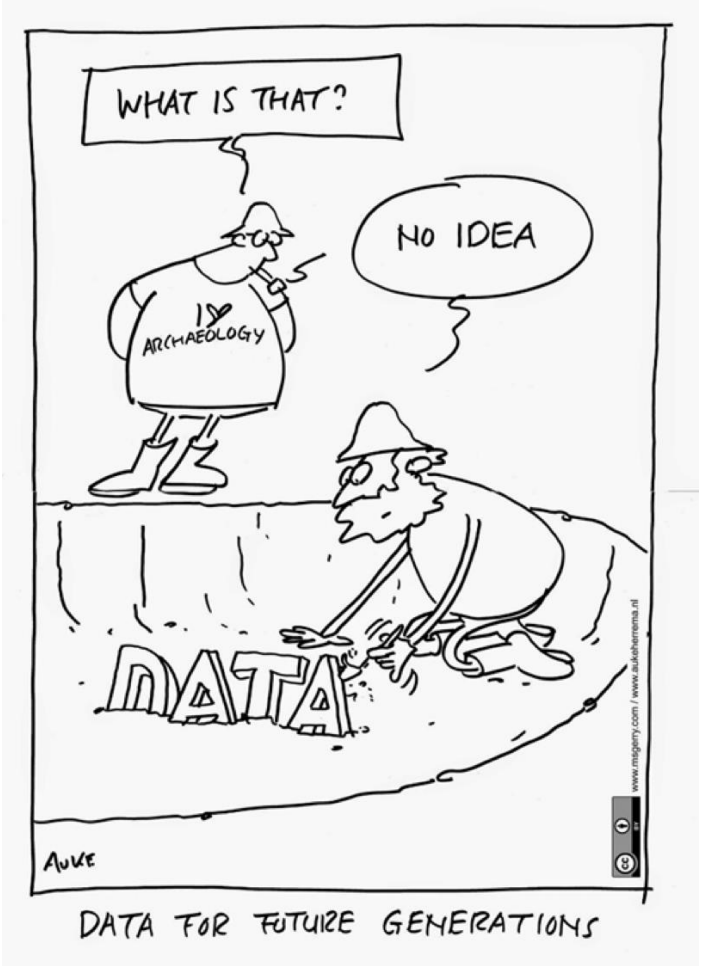
  - primary data (original data that arise from a particular experiment or observation), which are gathered and maintained by researchers
  - secondary data, often used by researchers but originally created by someone else

    - observational data – data that have been gathered from observing a particular phenomenon,
    - experimental data – in contrast – derived from controlled, randomized experiments.

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Research Data Management



DATA FOR FUTURE GENERATIONS

PUBLICATIONS AND DATA

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data Citations

Clarivate™

Who we serve | Products & Services | Resources | Contact us

Data Citation Index

Connecting data to the research it informs

Go to product | Contact us

The Data Citation Index™ provides a single point of access to quality research data from global repositories across disciplines. Descriptive records are created for data objects and linked to literature articles in the Web of Science.™

As data citation practices increase, the Data Citation Index aims to provide a clearer picture of the full impact of research output, as well as to act as a significant tool for data attribution and discovery.

As of 2 March 2022

- Data from 443 repositories
- 12,357,621 datasets
- 1,434,939 data studies
- 264,773 software

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

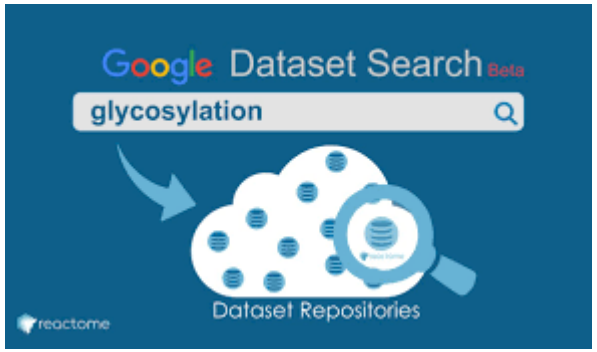
Google

Dataset Search

Wyszukaj zbiory danych

Wypróbuj koronawirus (COVID-19) lub education outcomes site:data.gov.

Learn more about Dataset Search.



Researchers perspective

Research Data Lifecycle

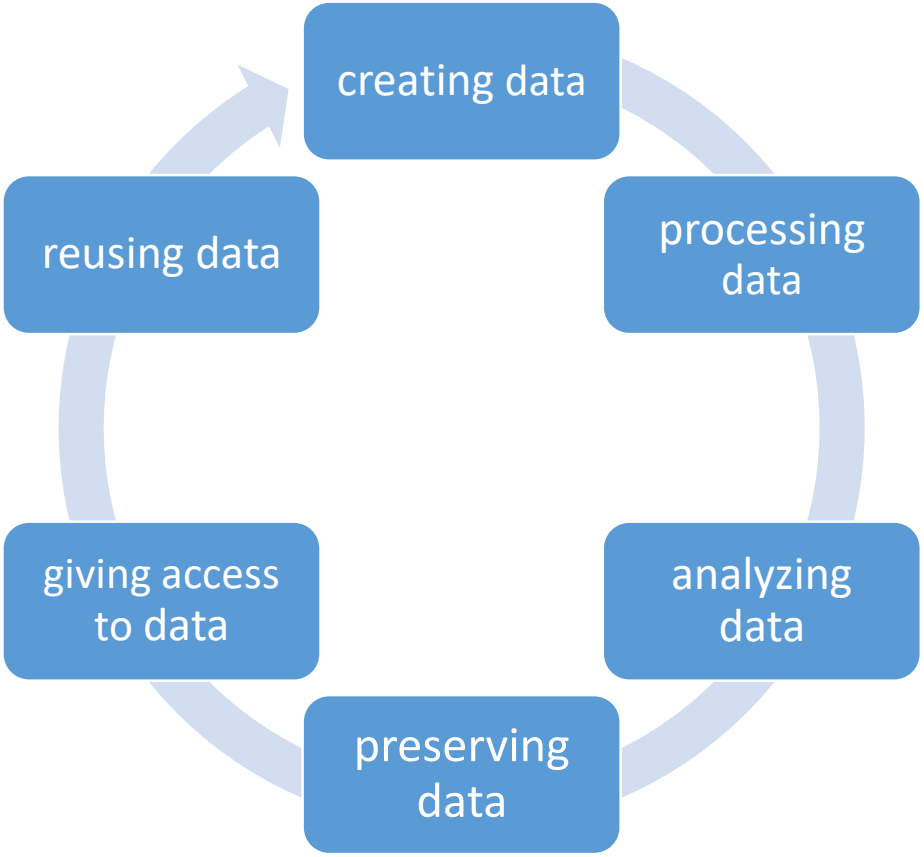
(created by the UK Data Archive)

It covers the lifespan of research data from the moment of its creation through the reuse of the data.

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Researchers perspective



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Curator's perspective

Curation lifecycle model

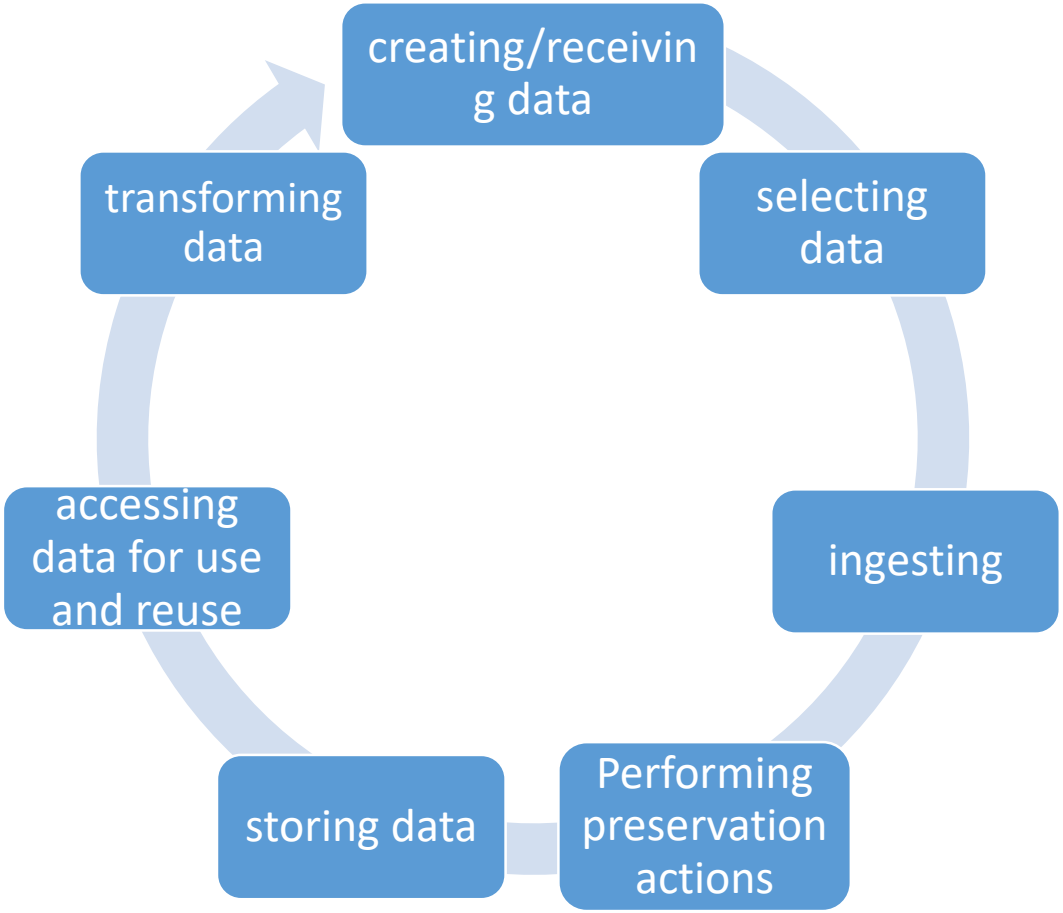
(created by The Digital Curation Centre)

all the processes and components involved in data curation from an archivist's or a curator's perspective

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Curator's perspective



Data Management Plan (DMP)

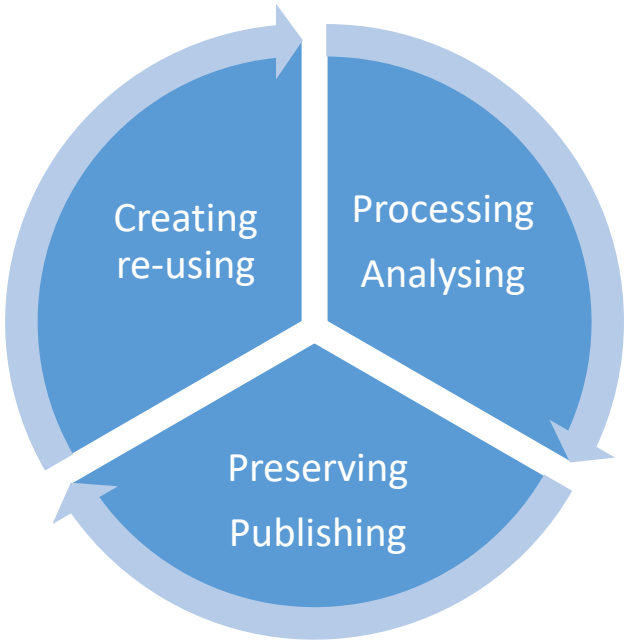
show how research data will be collected, organized, managed, and preserved during the project and after.

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

A DMP is…..?

…..a written document describing how DATA of a research project is managed during the life-cycle.



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Legal issues

PROBLEMS OCCURED DURING PREPARING DMPs IN DIFFERENT DISCIPLINES

Biology

- Ethical issues e.g. animal testing
- Replication of research
- Fear of misuse
- Not enough capacity in repositories to share data

Civil Engineering

- Fear of being scoped
- Not knowing where to share research and technical data
- Large amount of data is collected by instrumentation – authorships' concerns
- Construction data is rather complex and multiple engineering parties are often involved

Computer Science

- Eligibility to hold rights to a database
- Different practices of sharing code – e.g. via GITHUB
- Data versioning
- Hard to share ever-greater quantities of data

Chemistry

- Ethical issues with collecting research data e.g. clinical trials, patient privacy
- Concerns about data mining
- Secondary analysis of data
- Desire to protect confidential commercial information

Economics and Management

- Contractual obligations
- GDPR
- Data value cycle is very complex and might involve numerous stakeholders and different business agreements
- Complications with quality data anonymization e.g. interviews

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

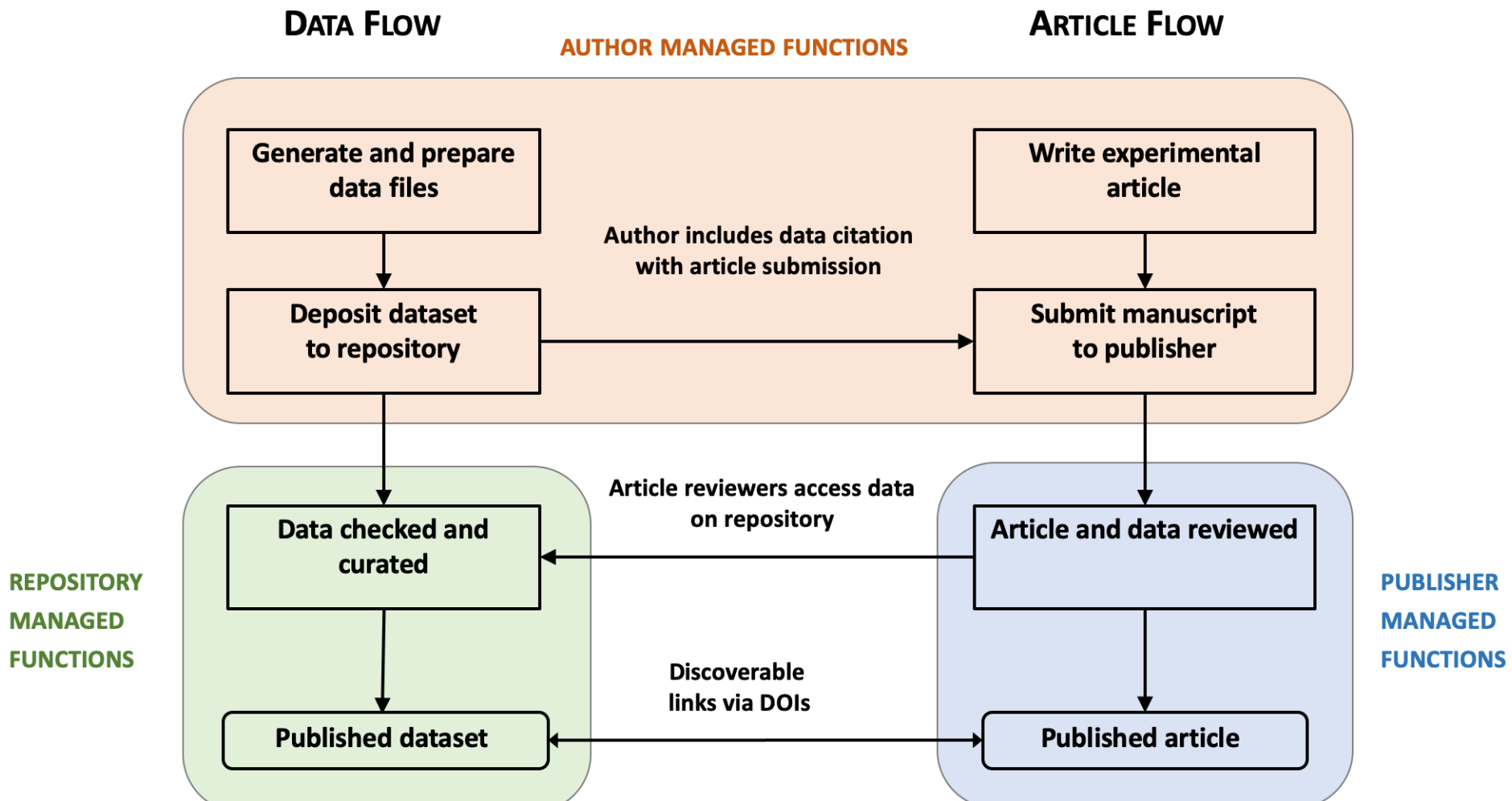
Legal arrays

Patent law	Copyright protection	Competition law	Privacy protection
• What has to be considered when research data (can) reach patentability?	• Do research data fall under copyright law at all?	• Are research data used unfairly in business transactions?	• What research data is worth protecting?
Science law	**Constitutional rights**	**International law**	**EU law**
• Can research data be licensed and published on a mandate basis?	• What constitutional limits must be observed?	• Which legal regulations exist outside Poland?	• What does the European data economy contribute for research data?
Contracts	**Labour / service law**	**Conditions of the funders**	**Policies**
• Are there any "intellectual property" agreements with research data?	• Who owns the research data collected at the universities?	• What are the conditions imposed by sponsors (NCN, industry)?	• Which legal obligations can policies develop?

based on the Katarzyna Biernacka 11.04.2019

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA



GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

Open Data

Opendatahandbook.org defines open data as:

data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike.

Opendefinition.org says that:
open data can be freely used, modified and shared by anyone for any purpose.

As matter of principle, adhere to the habit of making your **as open as possible, as closed as necessary.**

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Open Data

Can be freely used, modified & shared by anyone for any purpose

http://opendefinition.org

Limits on who can access & use data, how, or for what purpose

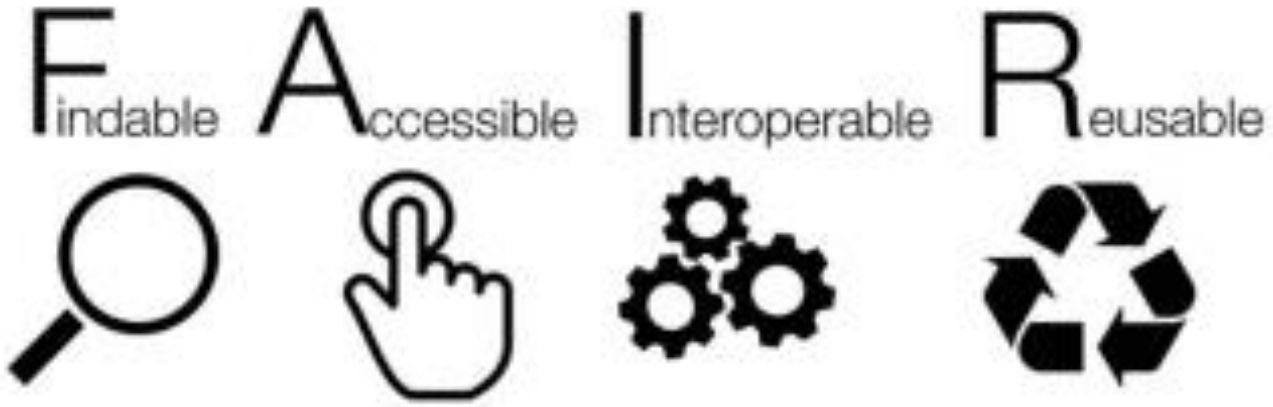
- only certain (types of) users
- only certain types of use
- ...

Under embargo
Unable to share

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

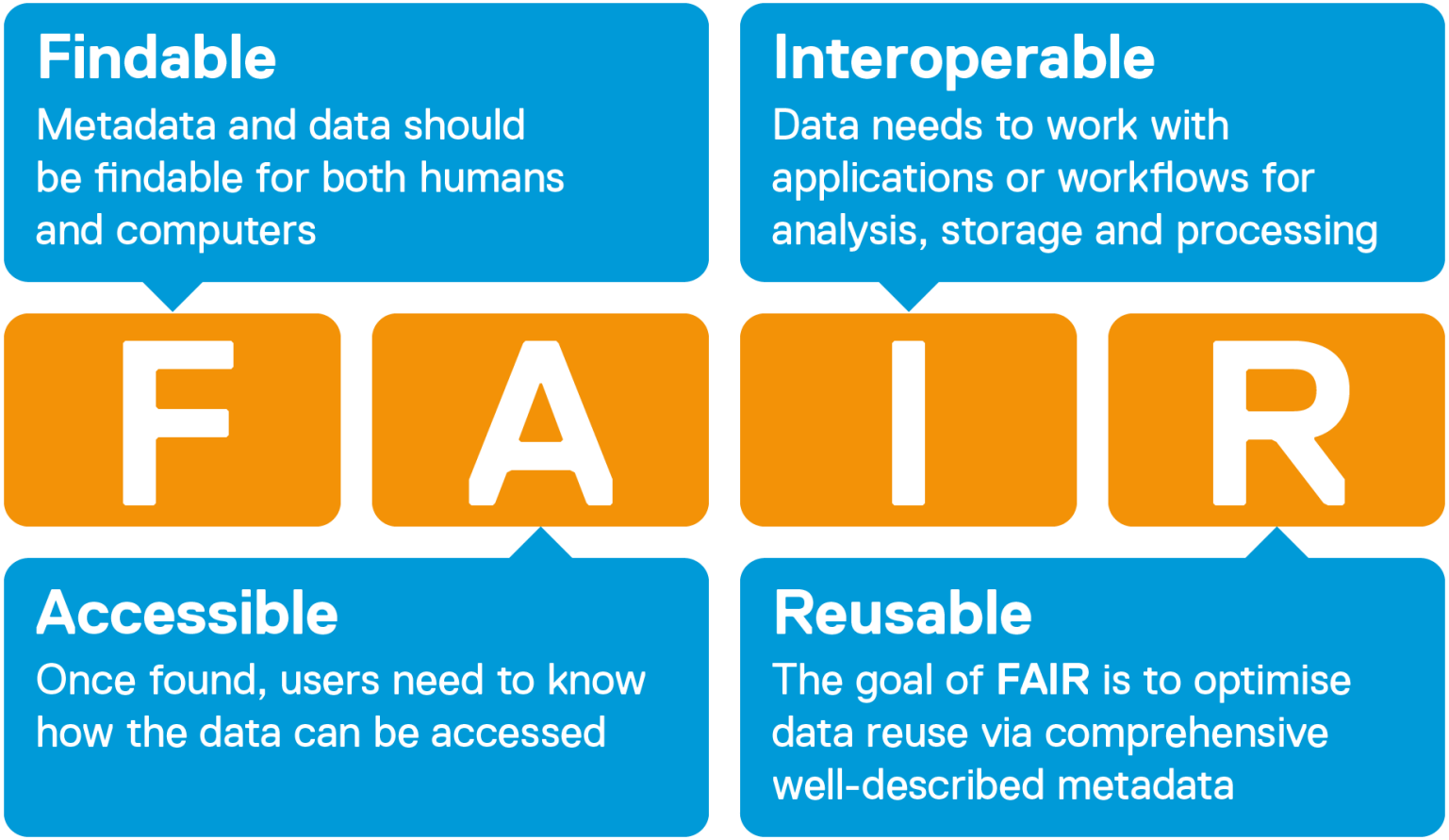
FAIR Data



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

FAIR Data



The FAIR Guiding Principles

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:

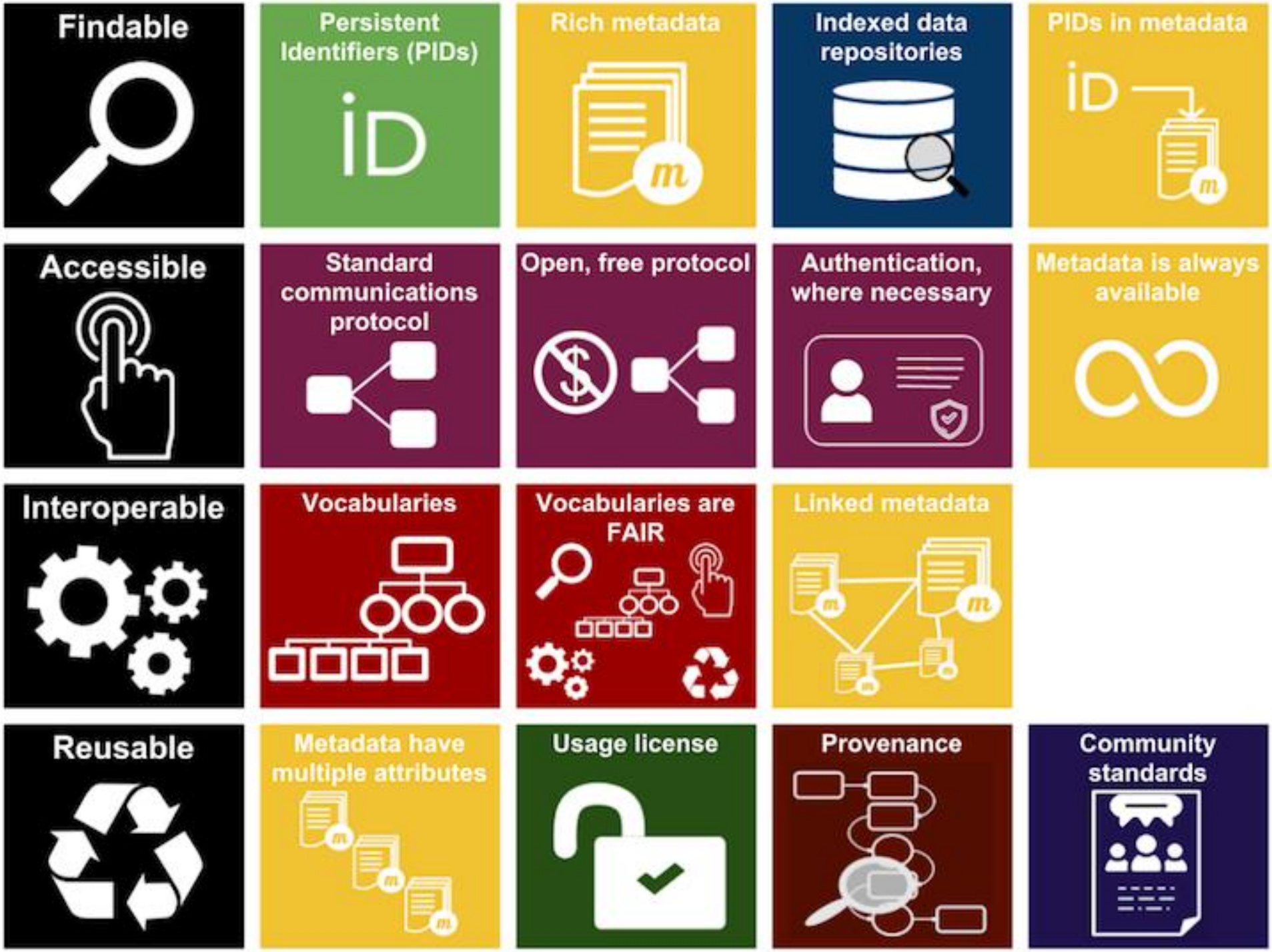
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1. the protocol is open, free and universally implementable
A1.2. the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are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

To be Interoperable:

I1. (meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
I2. (meta)data use 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 licence
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards

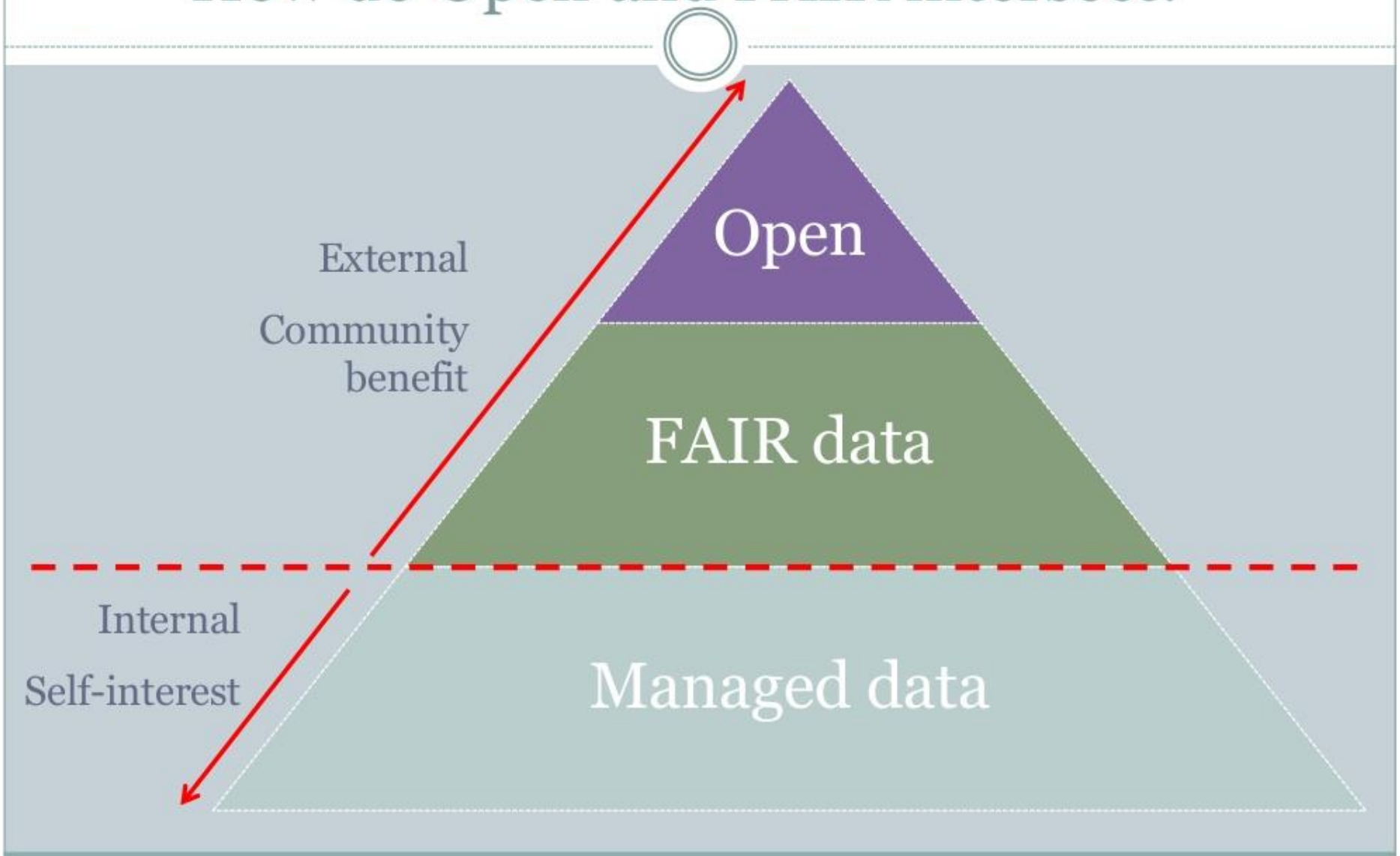

Findable
Persistent Identifiers (PIDs)
Rich metadata
Indexed data repositories
PIDs in metadata
Accessible
Standard communications protocol
Open, free protocol
Authentication, where necessary
Metadata is always available
Interoperable
Vocabularies
Vocabularies are FAIR
Linked metadata
Reusable
Metadata have multiple attributes
Usage license
Provenance
Community standards

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

≠

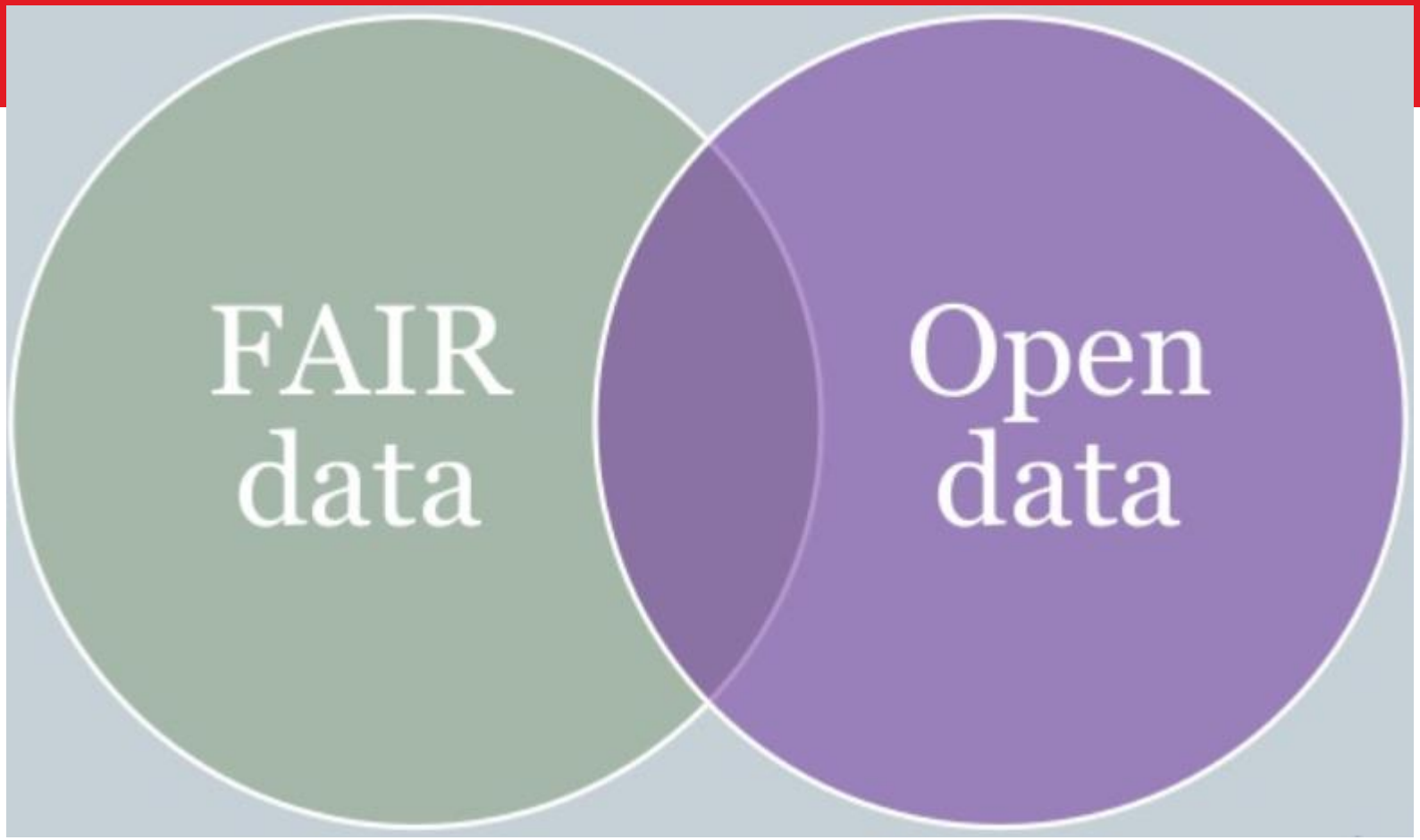
How do Open and FAIR intersect?



https://elearning.bits.vib.be/courses/writing-a-data-management-plan/lessons/fair-data/topic/fair-vs-open/

GDAŃSK UNIVERSITY OF TECHNOLOGY

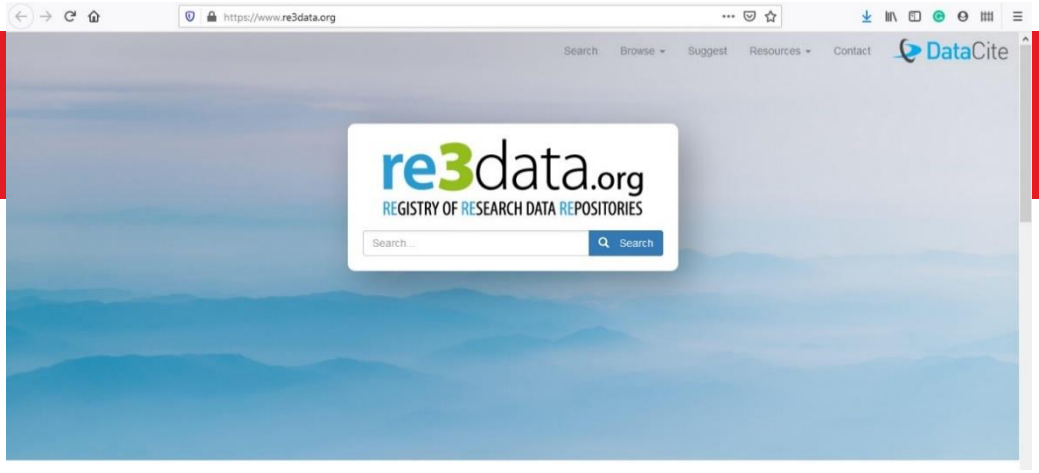
IATUL International Association of University Libraries

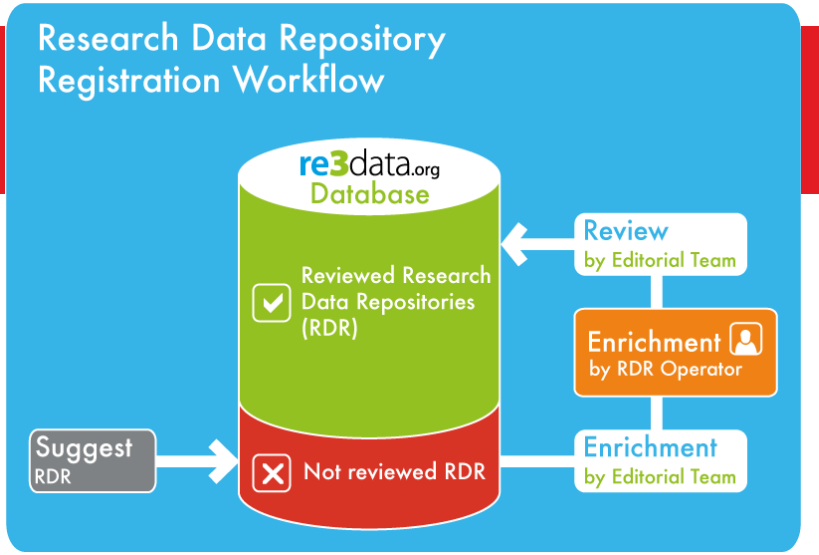


https://elearning.bits.vib.be/courses/writing-a-data-management-plan/lessons/fair-data/topic/fair-vs-open/

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA





CORE TRUST SEAL

Home | About | Certification | Certified Repositories | Apply | Contact

Core Certified Repositories

Home › Why certification › Core Certified Repositories

Applications are made public only once certification of a data repository has been approved by the CoreTrustSeal Board. Certification is against the version of the Core Trustworthy Data Repositories Requirements named in the link to the public application (e.g., 2017–2019). The CoreTrustSeal for Data Repositories is valid for three years from the certification date listed within the public application.

CORE TRUST SEAL

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Open Research Data Repository

Re3data.org

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

LIBRARY – THE MISSION

academic libraries serve the academic community as important partners in research process and information management

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Research data services

“Research data services are services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion” (Tenopir, Sandusky, Allard and Birch, 2012).

Data Librarian


team
concept
strategy
quality
experience
service
perform
professionalism
ability
goal
knowledg
competence
creativity
dy
amics
responsibility
flexibility

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

interpersonal skills

- high level communication
- advocacy
- negotiation and capability building skills

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

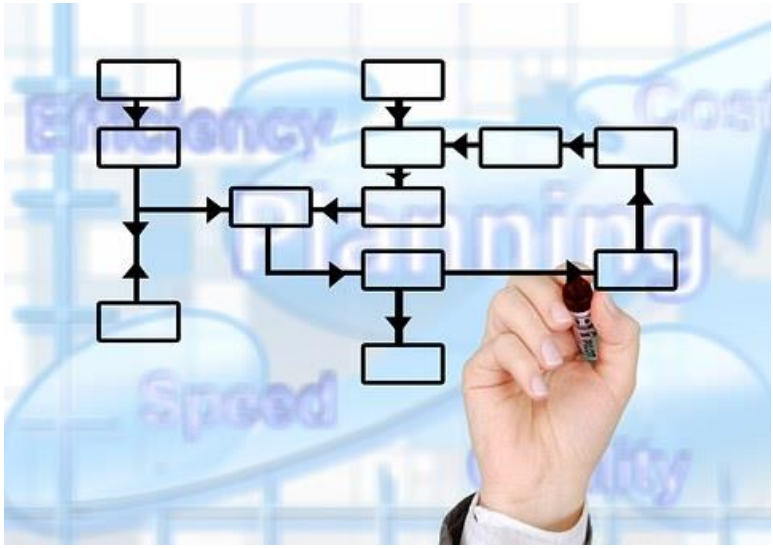
- comfortable with change,
- have a service philosophy,
- willingness to learn,
- discretion,
- “boundless curiosity”,
- be adaptable,
- assertive

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

- contextual knowledge
  - research environment
  - related funding agency policies
  - research measures
  - research evaluation activities



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

being open to

- discipline specific research life cycles
- cultures, processes,
- ethics,
- disciplinary research methods,
- intellectual property, licensing and policies,
- access norms and cultural sensitivities

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

data domain and related skills

- facilitate data sharing,
- linked data,
- data management lifecycle,
- quality control,
- data processing,
- data management planning
- ability to understand and support data storage

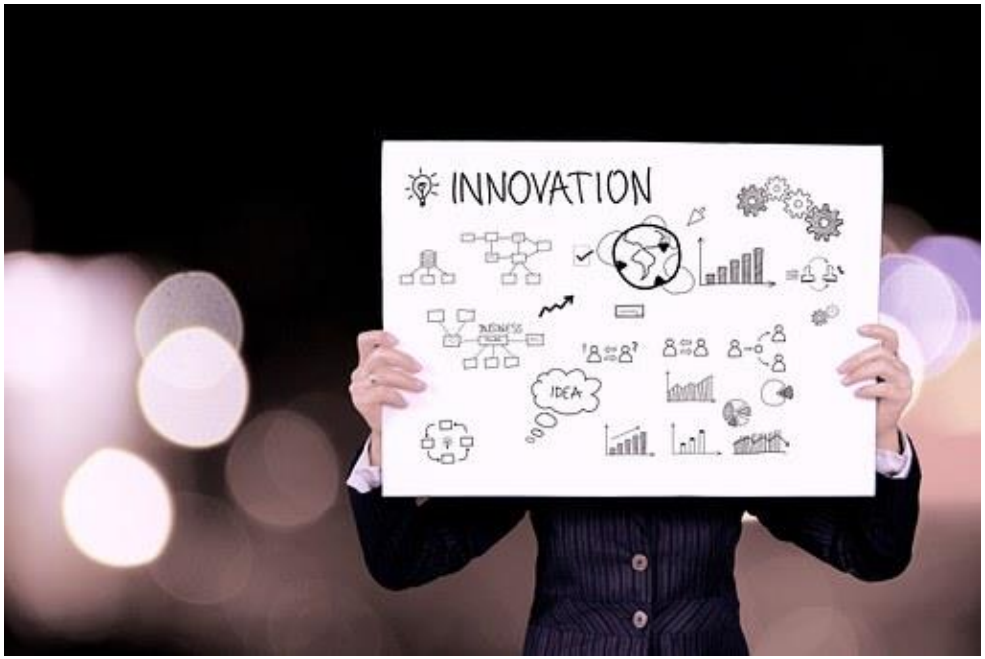
GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

skills in

- applying metadata,
- metadata standards,
- ontologies,
- structured information
- data mapping,
- cataloguing
- harvesting



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

skills in

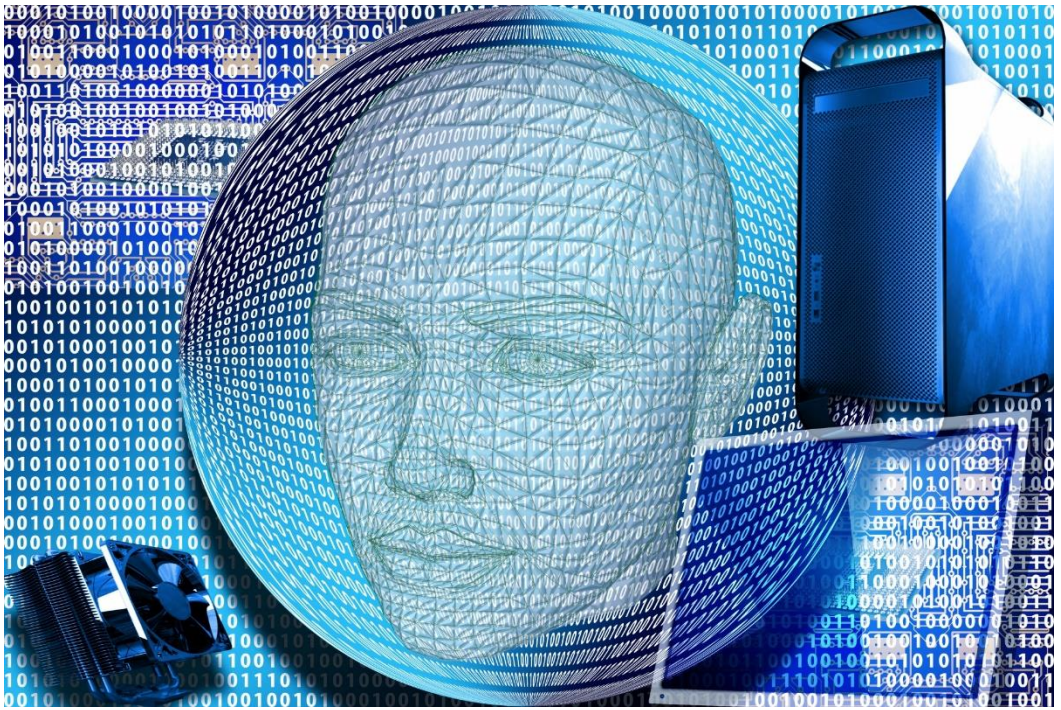
- enabling access,
- user experience,
- web interfaces,
- Graphical User Interfaces (GUI's),
- Digital Object Identifiers (DOI's),
- back up,
- version control
- naming conventions.,

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data librarian – skills

- IT skills
  - XML,
  - datatbase design,
  - APIs,
  - user centered design,
  - natural language processing tools
  - the Internet of things



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

The Bridge of Knowledge

MOST DANYCH

GDAŃSK UNIVERSITY OF TECHNOLOGY | BRIDGE OF DATA

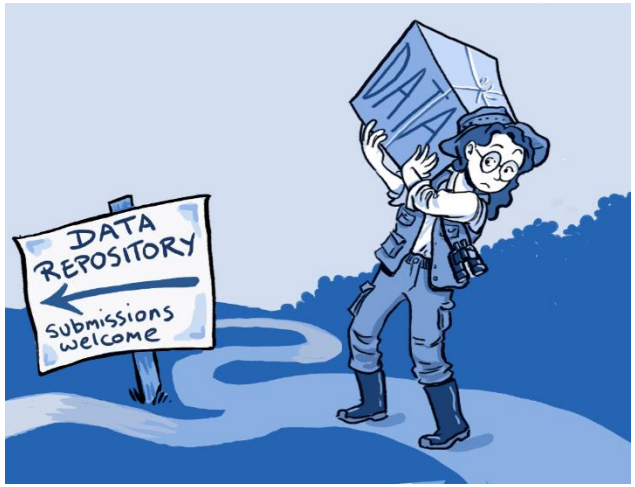
The Bridge of Data

Multidisciplinary Open System Transfering Knowledge stage II Open Research Data

The goal was to create **Competence Center** which provides expertise and support including trainings among scholars about different aspects of Open Science and face to face consultations

and to design and build a **platform dedicated to research data** generated at the **three most important universities of Pomerania**, which will increase accessibility, coherence and reuse of science, knowledge and technology resources

co-financed by the European Regional Development Fund



BRIDGE OF DATA

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Open Data Repository – mostwiedzy.pl

- Collect and store datasets from Gdańsk Tech, GUMed and UG
- Indexing datasets in Google Data, Web of Science Data Citation Index and other services
- Technological innovations such as hosting the project on the private computing cloud and storing the data on the Ceph Object Storage.
- NoSQL database – ElasticSearch.
- Repository allows researchers to perform Big Data Analysis by the Apache Zeppelin GUI on the supercomputer Tryton (40.000 cores, 1,5 PFLOPS).

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Open Science Competence Center

- Assistance and on-site tailoring training among researchers from all scientific disciplines that include Data Management Plan, open licensing or metadata standards
- Workshops regarding different aspects of Open Science as well as scholarly communication
- InfoKit regarding Open Research Data
- Metadata support
- Journals & Conference Proceedings indexing and publishing support
- Evaluation & Bibliometric support

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data steward

GDAŃSK UNIVERSITY OF TECHNOLOGY

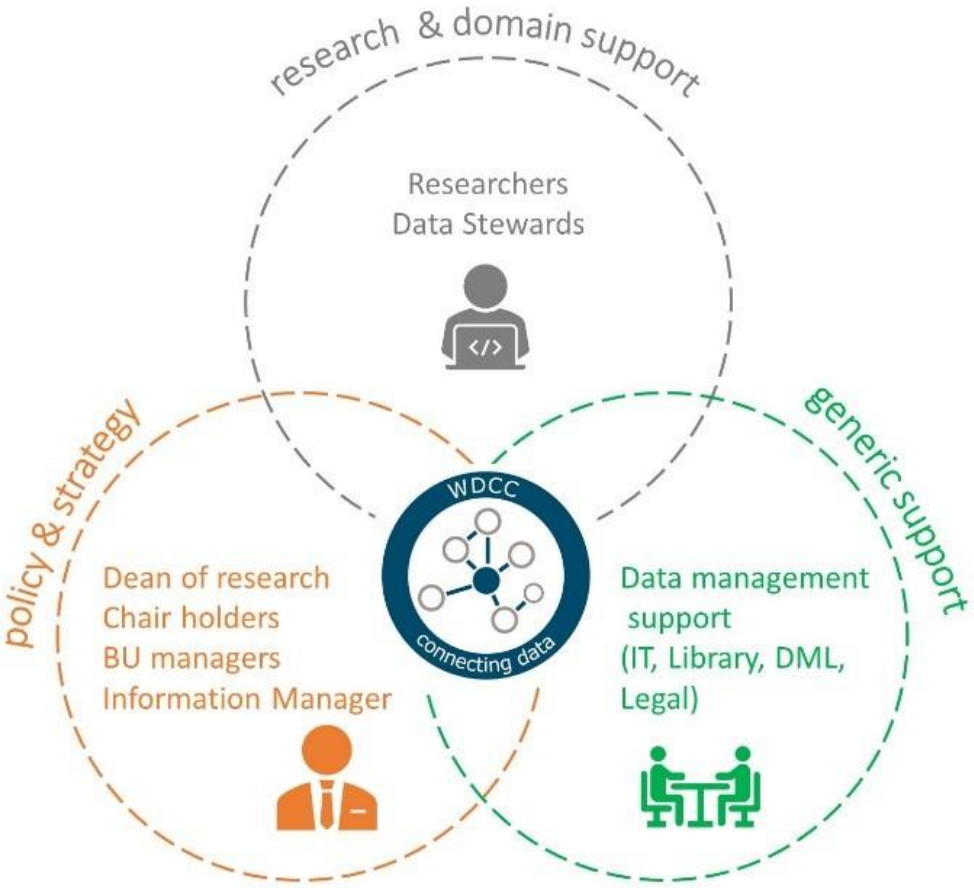
BRIDGE OF DATA

Data steward

GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Data services roles and responsibilities



GDAŃSK UNIVERSITY OF TECHNOLOGY

BRIDGE OF DATA

Open Science Competence Center - trainings

- Overview of open research data,
- Data Management Plans (national and EU grant applications),
- Legal support (data licensing, data protection, reusing data),
- Using the Bridge of Data Repository (depositing datasets),
- FAIR meta(data),
- Plan S (implications and requirements).

GDAŃSK UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GDANSK

MEDICAL UNIVERSITY OF GDAŃSK

BRIDGE OF DATA

BRIDGE OF DATA

Open Research Data

Project *"MOST DANYCH"*
is co-financed by the European Regional Development Fund
within the Operational Programme Digital Poland 2014-2020

European Funds
Digital Poland

Republic of Poland

BRIDGE OF DATA

European Union
European Regional Development Fund

Thank you

IATUL
International Association
of University Libraries

BRIDGE OF DATA