# The Bridge of Data Project Objectives

Anna Wałek[1*], Michał Nowacki[2], Paweł Lubomski[2]

[1] Gdańsk University of Technology Library, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[2] IT Services Center (CUI), Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: anna.walek@pg.edu.pl; ORCID: 0000-0001-8782-013X

## Abstract

Open Research Data (ORD) is one of the emerging trends for researchers across the globe. However, it has to be stressed that the level of implementation and awareness of ORD varies between countries. Many initiatives have been created in Polish scientific institutions to support the process of opening publications. These are mainly Open Access (OA) repositories, implementing the so-called green road of OA. However, only a few universities in Poland offer their researchers essential tools and substantive support for opening research data.

The Gdańsk University of Technology has created several tools and services supporting the open sharing of scientific research results, including scientific publications and research data.

The Bridge of Data project was established to support researchers in their activities regarding different Open Science layers. The project is unique in this part of Central and Eastern Europe. It involves three Pomeranian universities: Gdańsk University of Technology (leader), the University of Gdańsk, and the Medical University of Gdańsk.

The chapter aims to present the Bridge of Data project's assumptions and results, which provides both an open data repository and a range of additional services for researchers who want to share their research results openly.

Project assumptions will be presented along with their genesis and the result of their implementation. The technical aspects of creating IT tools (data repository, data analysis on a supercomputer, platforms for scientific journals and conferences) will be highlighted. The Open Science Competence Center, its tasks and the manner of their implementation will also be discussed.

An important aspect discussed in the chapter will also be the cooperation of various university teams to create tailored tools and services. These groups include librarians

who design technical solutions and support services, IT specialists building IT platforms and tools, and researchers representing various scientific disciplines. The last group enters data into the repository and helps adapt the functionality of the tools to users' requirements.

Based on an analysis of the functionality of the IT tools and based on reports on the activities of the Competence Center, a solution model will be drawn up that can be compared with other implementation cases of similar tools and services.

**Keywords:** Open Research Data, data repository, Open Science, Bridge of Data Project, Bridge of Knowledge, data steward, data librarian, data management

https://doi.org/10.34808/x55q-sz53_dyr_roz1

## The genesis of the project – The Bridge of Knowledge

In 2016, the Gdańsk University of Technology (Gdańsk Tech) launched the project titled "Multidisciplinary Open System Transferring Knowledge" (in Polish: "Multidyscyplinarny Otwarty System Transferu Wiedzy – MOST Wiedzy"). The project acronym (MOST Wiedzy) means "the Bridge of Knowledge". The project's main aim was to build an IT platform (mostwiedzy.pl) that supports sharing information about universities' achievements and the research potential of their researchers. In addition, the idea of the project was to facilitate establishing cooperation between researchers and between researchers and business (Krawczyk and Lubomski, 2017; Wałek and Lubomski, 2017).


**Fig. 1.1.** Scheme of the MOST Wiedzy (Bridge of Knowledge) project

The project also started introducing the idea of Open Access (OA) and Open Science (OS) among university employees. The enormous success of this project and new trends in opening science prompted the project team to launch the next project being a continuation of the previous. The "Bridge of Data" project (Bridge of Data – Multidisciplinary Open System Transferring Knowledge. Stage II Open Research Data) started in 2018 and focuses on implementing the Open Research Data (ORD) repository. It also provides

system modules that support organisers of scientific conferences and editorial offices of scientific journals in their processes. The project's key objective was to create substantive support in the Open Science Competence Center (OSCC) at the Gdańsk University of Technology Library (Gdańsk Tech Library).

Both projects are co-financed by the European Regional Development Fund within the Digital Poland Operational Program for 2014–2020.

## The Bridge of Knowledge portal

As a result of the Bridge of Knowledge project, the mostwiedzy.pl portal was created (MOST Wiedzy – your knowledge portal, 2021). It contains databases of related data and repositories and is divided into five modules.

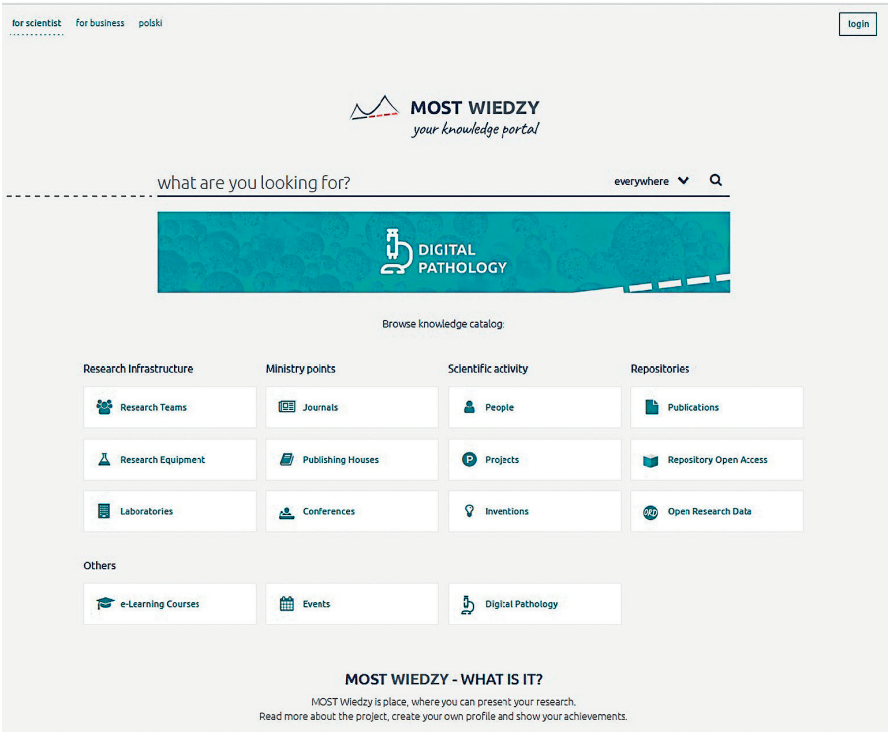



**Fig. 1.2.** The main page of the mostwiedzy.pl portal

The first one, "Research infrastructure", includes databases containing:

- research teams – the record contains a list of team members, their research interests, research offer and service offer (e.g. for the purposes of cooperation with business);
- research equipment – the details of the record contain the name of the manufacturer, parameters, laboratory and the maintainer of the equipment;

- laboratories – the record contains detailed information about the laboratory equipment, people associated with it, research topics being pursued, commercial offerings, related research data, and other resources.

The second module is named "Ministry points" and refers to the Ministry of Education and Science scores on which the scientific evaluation of researchers and research units is based. These points are awarded to journals, conference materials and monographs and are published on the list of publishers and journals announced by the Ministry. The "scientific journals" modules create a separate extensive database of journals and publishing policies developed by the Competence Center and will be discussed later in this chapter.

"Scientific activity", the third module, contains people, projects and inventions databases. One of the critical elements is the scientific profile. It has information about researchers, publications, ongoing projects, scientific achievements, organisations, didactic activities, and shared research data. The profile is completed with a bio and can be linked to profiles on social media.

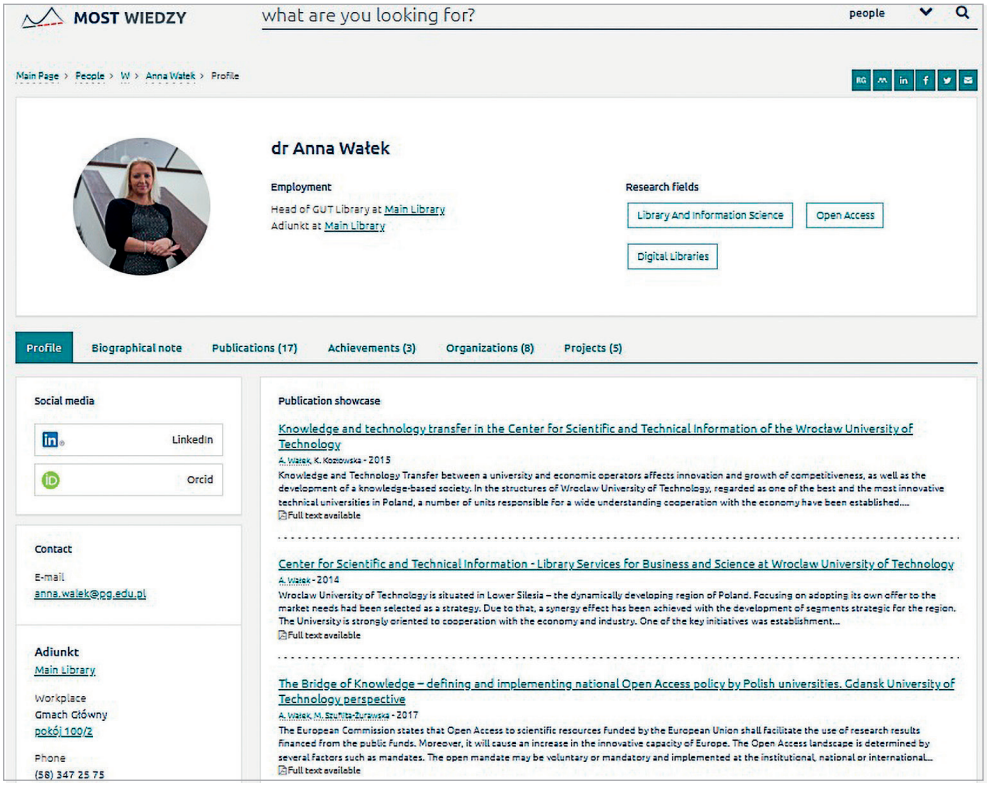



**Fig. 1.3.** Scientific profile

The fourth module includes three repositories behind the implementation of the Open Science policy. These are, in turn:

- publications repository – containing bibliographic records of all publications authored by people with a profile on the portal;
- Open Access Repository, containing publications available in full text, following the Open Access green route;
- Open Research Data repository containing data sets.

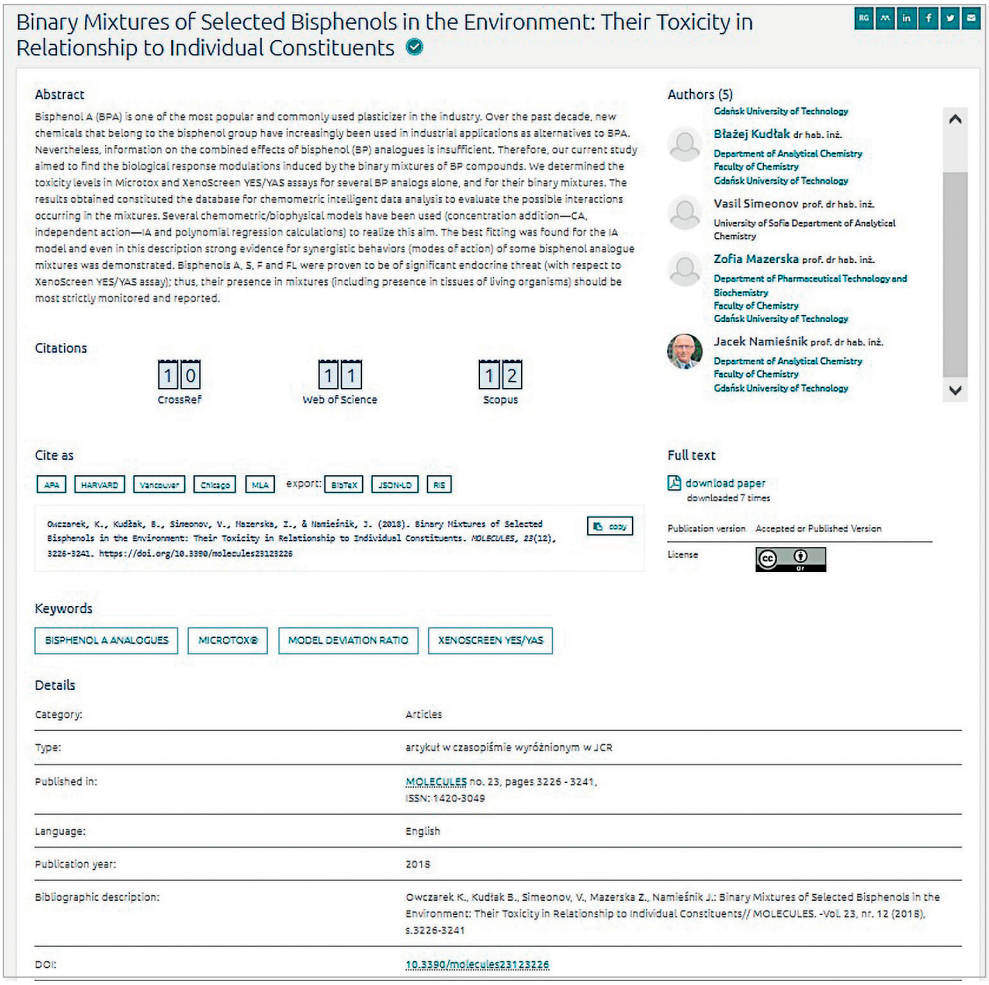
The last fifth module includes additional databases, including events (both scientific and university life events), descriptions of online courses available on the university's e-learning platform and a virtual microscope (digital pathology).

## Open Access Repository

The Open Access Repository contains records of all publications in full text. The database enables searching and sorting results and refining the search according to specific criteria. A publication record contains bibliographic data, citation information, altimetric data, export to various citation styles, license information, and a full-text button. It also provides publication statistics such as the number of views and downloads and an intelligent mechanism for suggesting similar publications.

The process of depositing publications in the repository is carried out using the system for registering scientific achievements on the MojaPG portal (My Gdańsk Tech). When entering bibliographic data, the author can add a publication file, select the license under which the publication is made available and specify the type of publication (preprint, postprint, author's accepted manuscript, version of record, etc.). The main principle of the authors of the repository project was the idea of a "one-stop-shop", allowing authors to enter all the necessary data during one login to one system. Authors can additionally select an embargo period for the publication and also add a preprint, which, after the embargo expires, is replaced with the correct version of the article. Finally, after entering all the data, the author sends them for verification.

Regarding several steps that need to be made by academic staff to deposit publications in the repository, the Library has already formed a Library Repository Services (LRS) Team (currently a part of the Open Science Competence Center). Due to the university's administrative structure, the initial phase of depositing the documents is supervised by the Department of Scientific Matters (DSM). This unit is responsible for research output registration and data transfer to the national POLON system that the Ministry of Education and Science supports. Additionally, the DSM team must verify and validate the metadata and bibliometric analysis. In the process of document depositing, the Library's team's primary role is to check for publishers' policy, file formatting and editing. In addition, the Library Repository Services Team checks the publisher's policy and embargo and makes sure that the uploaded file's type and format are correct (Wałek and Szuflita-Żurawska, 2017).



Binary Mixtures of Selected Bisphenols in the Environment: Their Toxicity in Relationship to Individual Constituents

Abstract

Bisphenol A (BPA) is one of the most popular and commonly used plasticizer in the industry. Over the past decade, new chemicals that belong to the bisphenol group have increasingly been used in industrial applications as alternatives to BPA. Nevertheless, information on the combined effects of bisphenol (BP) analogues is insufficient. Therefore, our current study aimed to find the biological response modulations induced by the binary mixtures of BP compounds. We determined the toxicity levels in Microtox and XenoScreen YES/YAS assays for several BP analogs alone, and for their binary mixtures. The results obtained constituted the database for chemometric intelligent data analysis to evaluate the possible interactions occurring in the mixtures. Several chemometric/biophysical models have been used (concentration addition—CA, independent action—IA and polynomial regression calculations) to realize this aim. The best fitting was found for the IA model and even in this description strong evidence for synergistic behaviors (modes of action) of some bisphenol analogue mixtures was demonstrated. Bisphenols A, S, F and FL were proven to be of significant endocrine threat (with respect to XenoScreen YES/YAS assay); thus, their presence in mixtures (including presence in tissues of living organisms) should be most strictly monitored and reported.

Authors (5)

Gdańsk University of Technology

Błażej Kudłak dr hab. inż.
Department of Analytical Chemistry
Faculty of Chemistry
Gdańsk University of Technology

Vasil Simeonov prof. dr hab. inż.
University of Sofia Department of Analytical Chemistry

Zofia Mazerska prof. dr hab. inż.
Department of Pharmaceutical Technology and Biochemistry
Faculty of Chemistry
Gdańsk University of Technology

Jacek Namieśnik prof. dr hab. inż.
Department of Analytical Chemistry
Faculty of Chemistry
Gdańsk University of Technology

Citations

CrossRef 10 Web of Science 11 Scopus 12

Cite as

APA HARVARD Vancouver Chicago MLA export: BibTeX JSON-LD RIS

Owczarek, K., Kudłak, B., Simeonov, V., Mazerska, Z., & Namieśnik, J. (2018). Binary Mixtures of Selected Bisphenols in the Environment: Their Toxicity in Relationship to Individual Constituents. MOLECULES, 23(12), 3226-3241. https://doi.org/10.3390/molecules23123226

copy

Full text

download paper
downloaded 7 times

Publication version Accepted or Published Version

License

Keywords

BISPHENOL A ANALOGUES MICROTOX® MODEL DEVIATION RATIO XENOSCREEN YES/YAS

Details

| | |
|---|---|
| Category: | Articles |
| Type: | artykuł w czasopiśmie wyróżnionym w JCR |
| Published in: | MOLECULES no. 23, pages 3226 - 3241, ISSN: 1420-3049 |
| Language: | English |
| Publication year: | 2018 |
| Bibliographic description: | Owczarek K., Kudłak B., Simeonov, V., Mazerska Z., Namieśnik J.: Binary Mixtures of Selected Bisphenols in the Environment: Their Toxicity in Relationship to Individual Constituents// MOLECULES. -Vol. 23, nr. 12 (2018), s.3226-3241 |
| DOI: | 10.3390/molecules23123226 |

**Fig. 1.4.** Publication record in the Open Access Repository

Authors with a profile on the portal mostwiedzy.pl who are not employees of the Gdańsk University of Technology may log in to the portal using a Google account or ORCID identifier (soon also other identifiers). Then they enter publication data and files directly into the portal via a dedicated interface, and the workflow differs from the one presented above.

Based on the experience of the Competence Center and comments from users, the repository is constantly being improved. Recently added, among other citation styles and their export, a test version of the bibliography, the possibility of thanking the author for sharing the file with the publication and linking the publication with other resources (funding source, project, research data, etc.).

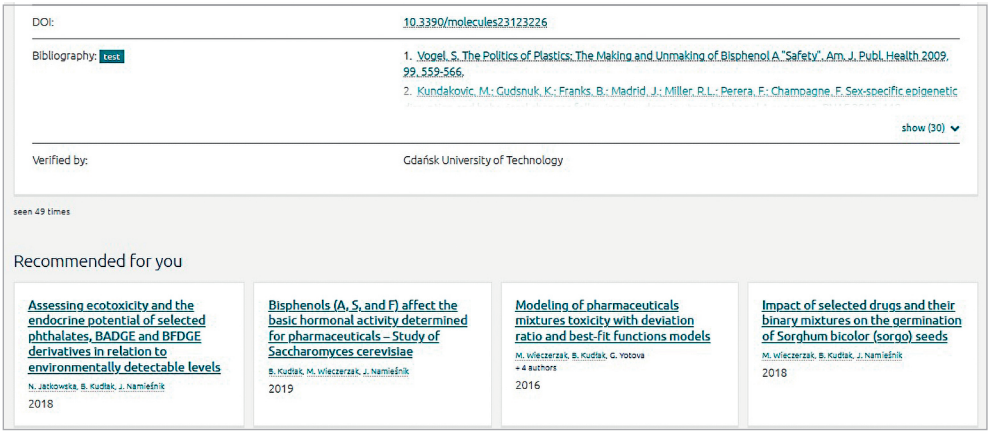



**Fig. 1.5.** Recommendations based on the current publication and the search history

## Open Research Data Catalogue

The main task of the Open Research Data repository is to provide data from three universities that make up the Bridge of Data consortium. Designing the platform and its elements, from the metadata schema to the content management tools, was long and tedious. However, it resulted from various teams' close and effective cooperation (Library, IT, researchers) at all three universities.

The first stage was the preparation by the substantive team of a list of requirements and functionalities that the new system should provide. This list was based on the best practices of similar repositories with a worldwide reputation and some shortcomings that the team saw in other such repositories. Examples of good practices were, for instance, Figshare, Zenodo, and some domain repositories. However, the team was also aware that creating a tool was only the first step. For researchers to use it and see its benefits, it should provide full functionality and indexing in search engines, indexes, and lists of repositories recommended by publishers and funding agencies.

Since the repository collected data from several disciplines, the challenge was to provide appropriate functions and a universal metadata description.

The metadata should be human- and machine-readable and compatible with commonly used metadata standards.

Selecting the standards that will be appropriate for dataset collections and fulfil the FAIR principles (Wilkinson et al. 2016) is a weighty and challenging decision. For scientific publications, mostwiedzy.pl already supported Dublin Core and Highwire Press tags. Additionally, to ensure the project's compatibility with five stars Open Data, each object is described by schema.org with JSON-LD formatting. The creators of the datasets are encouraged to link the datasets to other objects existing in the Bridge of Knowledge Portal, such as publications, scientific projects, teams, laboratories or other datasets. Moreover, research data may be grouped into series. All the mentioned links are also represented in the descriptive, JSON-LD based metadata. These links to other data sources, combined

with different features and rules, such as using non-proprietary file format and assigning unchangeable DOI and URLs to each dataset, fulfil the requirements of 5-star Open Data in terms of metadata descriptions of datasets.

As metadata is essential for efficiently storing, sorting, retrieving, sharing, and linking scientific data, the team decided to use the set of atributes of DDI (Data Documentation Initiative) and DataCite standards for the first level to ensure the description of granular levels of resources of the metadata. Due to the wide range of disciplines covered by the project, from humanities, social sciences, technical and engineering to medical science, the team was looking for a standard that best reflects the needs and assumptions. The chosen standard is quite general, flexible, and more accessible for all disciplines and broader communities than others. Besides this it is more interoperable than other standards, which will result in better indexing of the provided datasets in various search engines and data hubs, increasing awareness of their presence and availability. The second level of metadata was subject-specific and more constrained to ensure that scientific objects were more findable and reusable (for example, the INSPIRE standard for GIS data). The second stage of the development of the metadata schema for the Data Repository will be implemented in the next phase of the repository development. Then the metadata with a higher level of detail will be defined.

The data repository has a hierarchical structure that allows, e.g. research teams to assign a specific collection of datasets to particular projects and then a sub-collection to different research objects such as individual scholars, publications, software or images.

The process of depositing datasets is based on a dedicated interface. It is intuitive and equipped with additional tools to facilitate navigation and fill in the individual description fields.

The first screen of the repository contains a window with hints of what the author of the deposited dataset can expect from the repository. For example, it briefly describes such functionalities as backup, broadcasting DOI, versioning of datasets, links to other resources, graphical abstracts or private links. In addition, it contains contact details of the Competence Center with a hint on what the Center's employees can help with. Finally, a button also allows you to easily add a new data set and a list of the latest added datasets.

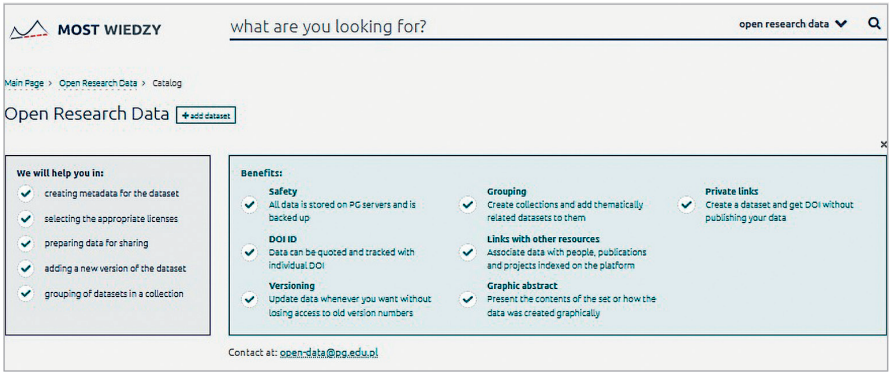



**Fig. 1.6.** The first screen of the Open Research Data Catalogue with hints for authors

After starting the new dataset wizard, the author receives a form in which it is required to enter individual metadata. It is where you enter all the metadata that allows you to search and index the data set correctly. First of all, the authorship of the dataset or the participation of individual people and organisations in its creation should be specified. For this, there is an extensive "authors" field with the possibility of choosing different roles. In this and several other places, the author can count on a hint about what options he can choose. It is crucial for those authors who do not have much experience in data depositing and may not understand the difference between authorship of a publication and participation in data production. The tooltips expand after hovering the mouse over the appropriate marker.


**Fig. 1.7.** The first stage of the description of the dataset with tooltips

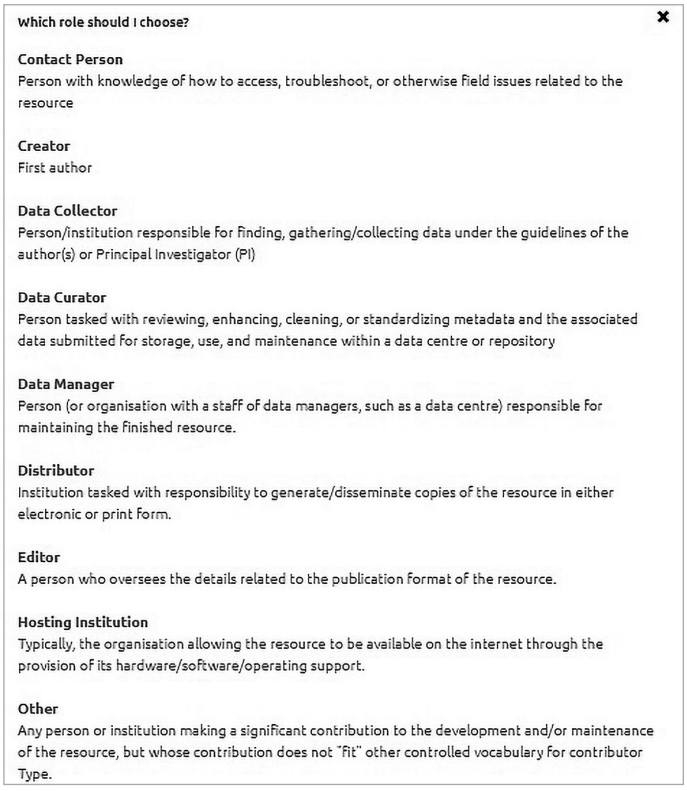



**Fig. 1.8.** An example of the tooltips for the roles in data creation.

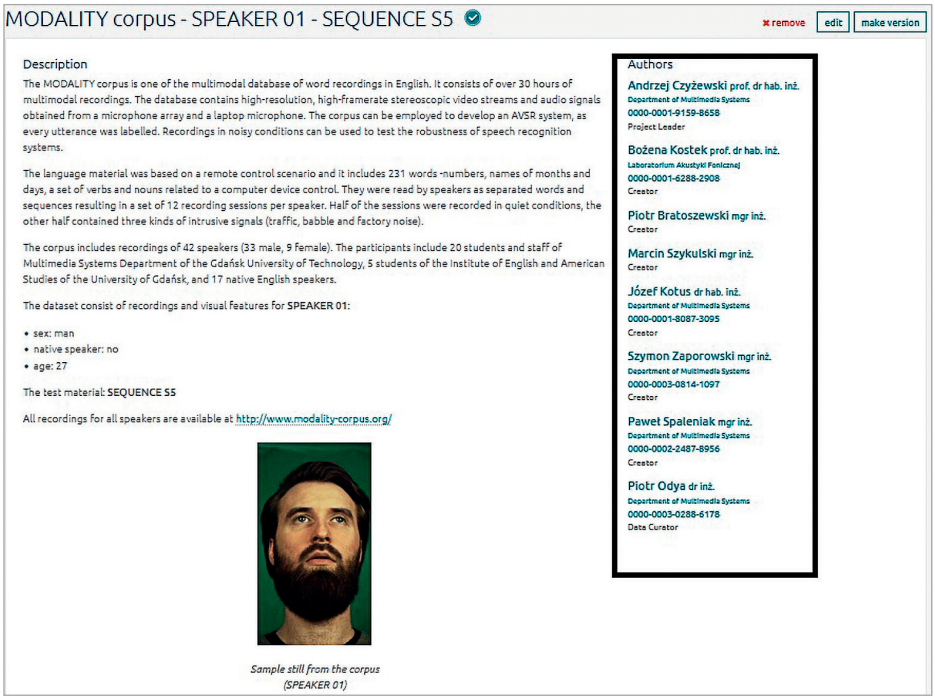



**Fig. 1.9.** An example of a dataset with different roles for the authors of the dataset

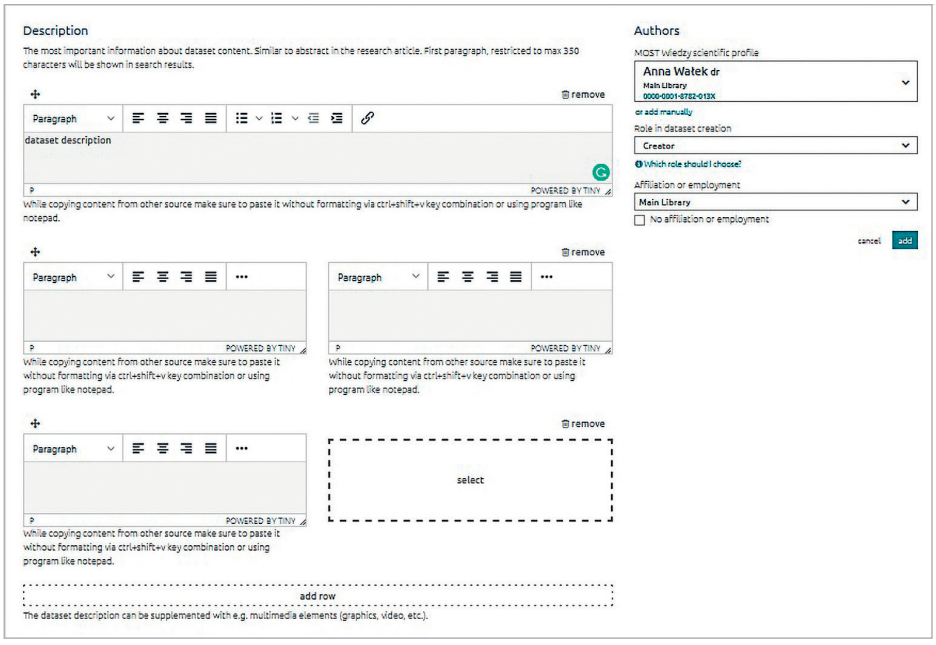



**Fig. 1.10.** An extensive option to add a dataset description, including a graphic description

The next step is to fill in the metadata about the dataset, starting with a unique title and description of the dataset. An additional option is to introduce an expanded description and a graphic description, which allows you to present the dataset's contents in a detailed or visual way. It is beneficial for datasets containing video material or showing complex processes.

As research data may result from various types of projects, be linked into groups or series, and be related to publications, all functionalities have been added to the repository that allows for linking and relating individual resources to each other. It is also possible to create dataset versions. These functionalities make it possible to adjust the dataset collection to individual needs and improve their searchability.

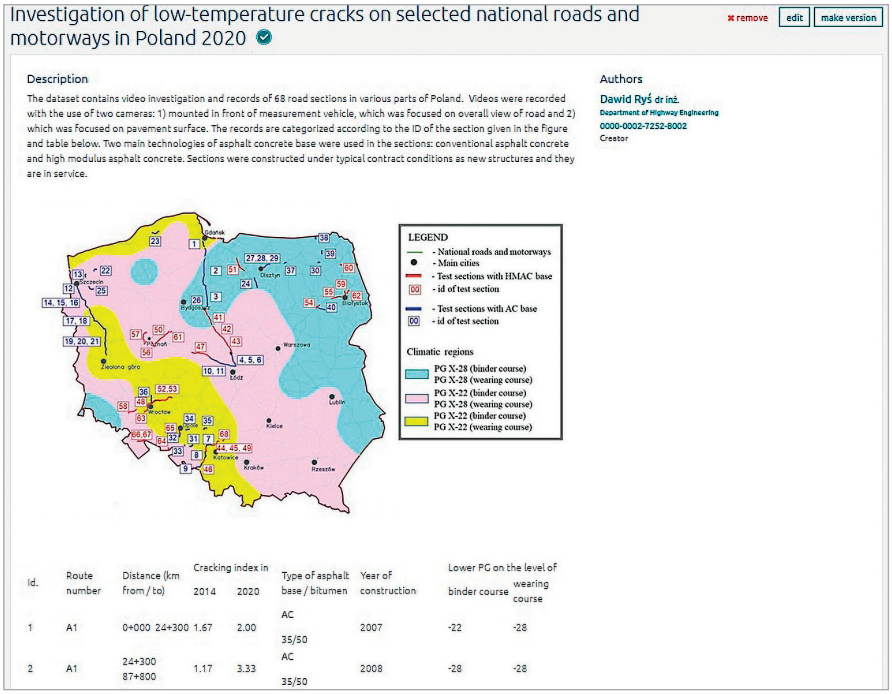



**Fig. 1.11.** An example of a dataset with an extensive text and graphic description

In the following steps, the complex metadata information is added: dataset creation date, publication date (if it is, for example, earlier than the date of deposit in the repository), the language of the data itself, research areas represented, information on funding, keywords and ethical papers approval along with its number. In addition, the author can specify whether a new DOI is to be assigned and could add links to other resources such as other datasets or publications (data relationship).

In the next step, a data file (or files) is added, a license is selected, and a possible embargo is determined. The author also specifies whether the data is raw or processed and whether additional software is required to open it. If so, it is necessary to provide the name of the software and a possible link to download it.

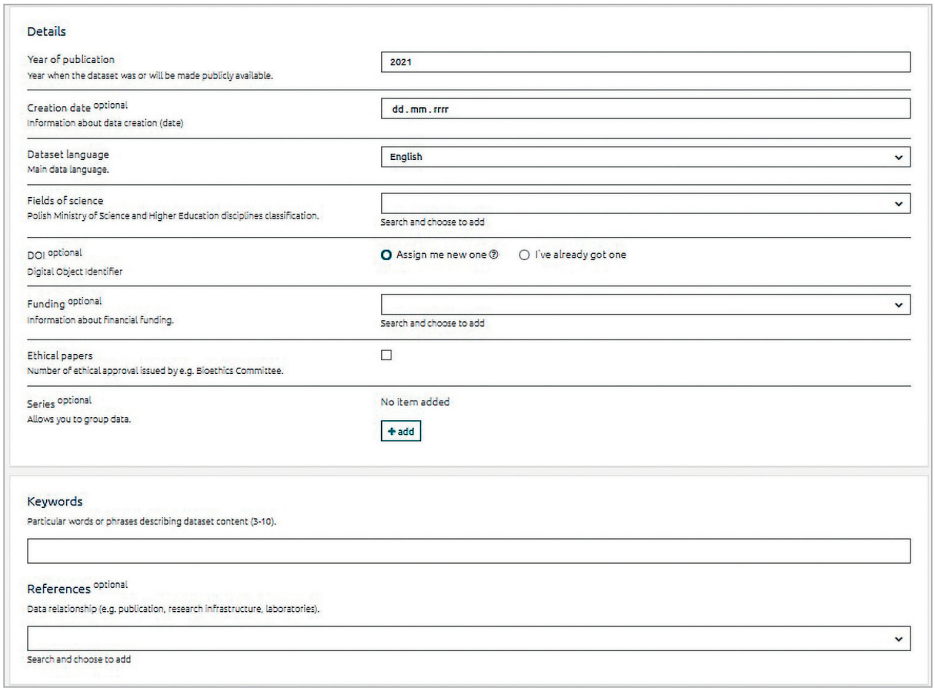



**Fig. 1.12.** A form for filling in metadata for the dataset

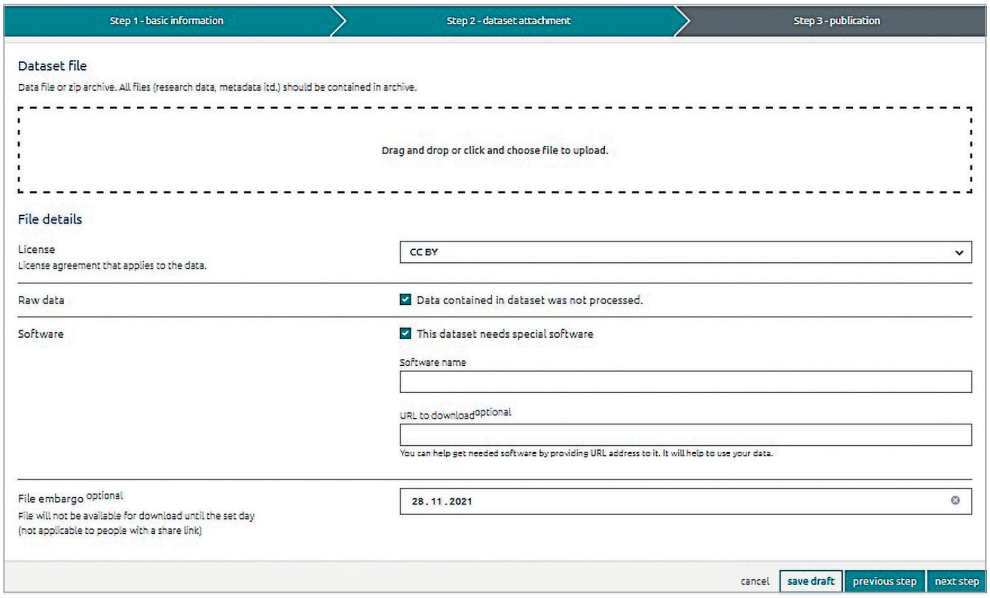



**Fig. 1.13.** A form for adding a data file and setting access rules

Access to data is often required to review scientific publications, but before it is made publicly available. Then the solution often used by the authors is the so-called private link that can be generated for data that has not yet been made available but is, for example, in a draft, invisible to the public. Each link owner, i.e. the reviewer, has access to the deposited dataset, but it does not involve publication. It is also possible to deposit data with the limited embargo and restricted access. Then there is an option to send a request for sharing directly to the author.

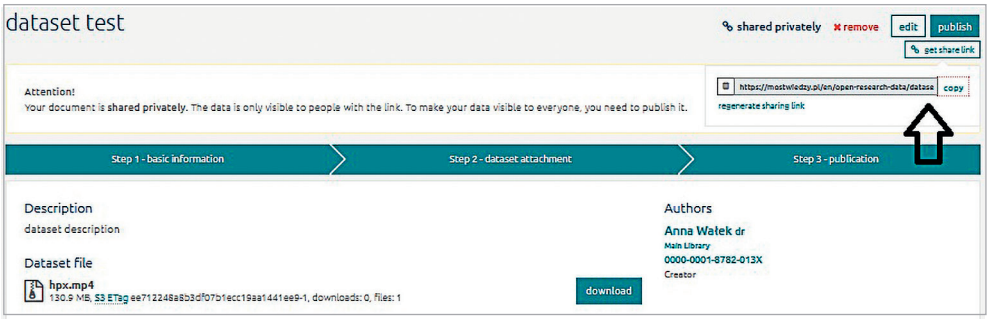



**Fig. 1.14.** A private link option for the dataset in draft

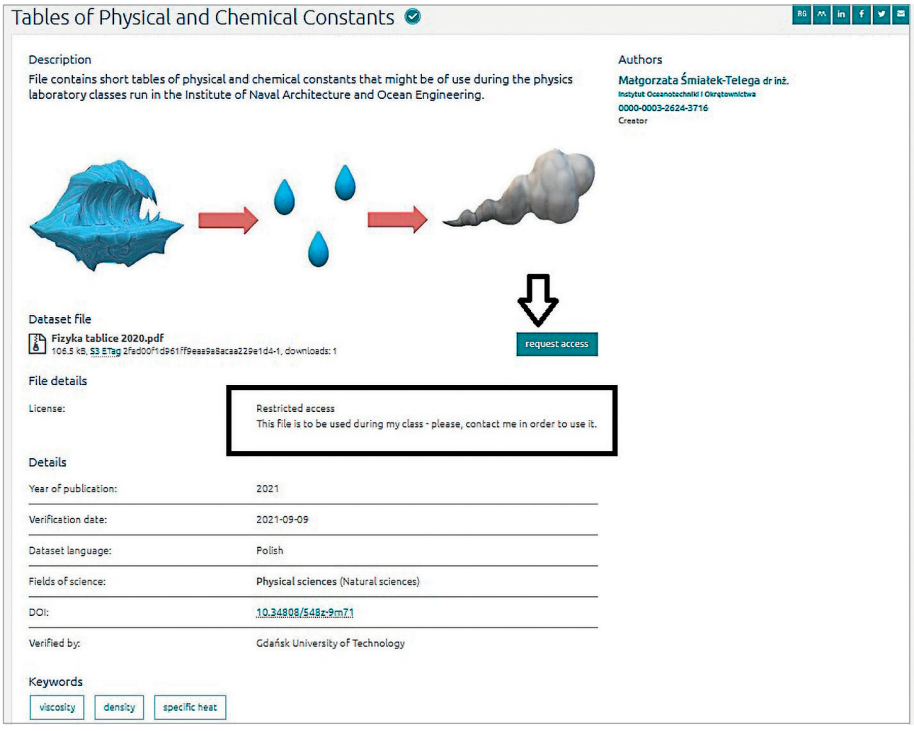



**Fig. 1.15.** An example of the dataset with a restricted access

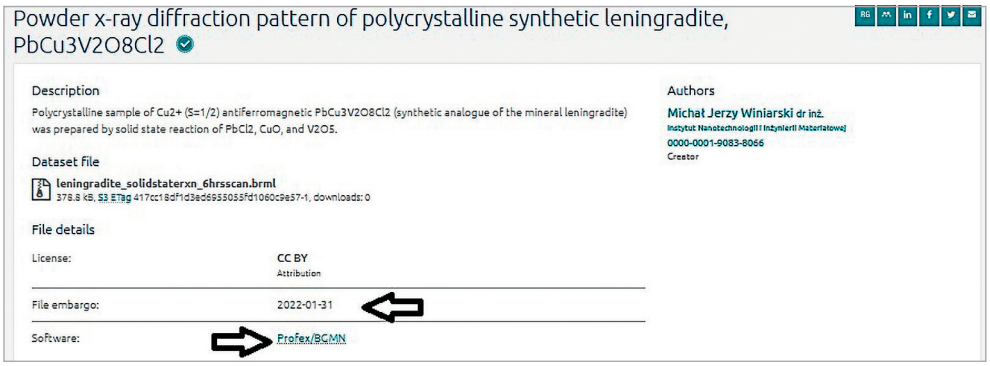



**Fig. 1.16.** An example of the dataset with the embargo and external software usage

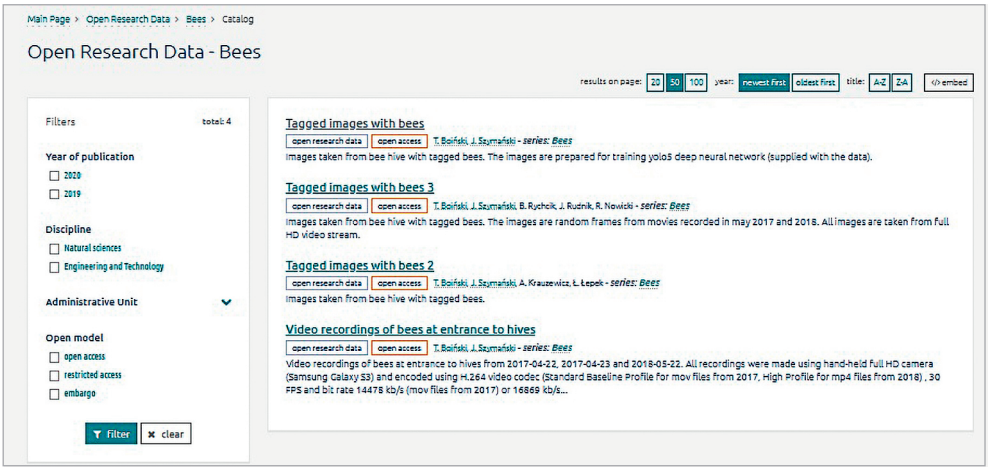



**Fig. 1.17.** An example of datasets grouped into series

The data and metadata prepared in this way are then sent to an employee of the Competence Center for checking and approval. Then, they become visible in the Open Research Data Catalog depending on the granted access rights. In addition to presenting data and metadata, a dataset record has critical features such as citation support, file list preview, and a download button.

## Open Science Competence Center

Currently, there is no national policy regarding Open Research Data in Poland. However, in 2019, the National Science Centre (NSC), which has already signed up to Plan S, set up the obligation to attach the short Data Management Plan to all grant applications starting from September 2019. It was the direct cause of the project team's work and activities intensifying. The Competence Center was created immediately after the project was launched – in the autumn of 2018. Being the only university in Poland with a team of professionals who can support research teams in the preparation of DMPs for the

needs of project applications, Gdańsk University of Technology has grown into the role of a leader.

Based on extensive experience and activity in various international teams and expert groups, members of the substantive team, already at the stage of preparing the substantive assumptions of the project, were aware that it would be unique and future-proof. In 2017, when preparing the grant application, the team relied solely on foreign experience and research because data repositories or dedicated services did not exist in Poland. Nevertheless, it was known that this trend would also enter the Polish environment, and the team wanted to prepare for this fact.

The basis of the project, the idea of which was created in 2017, was the awareness of the lack of solutions in the Polish environment, and above all in Gdańsk, that would help deal with the impending need for data management and their sharing system.

In February 2020, representatives of the Gdańsk University of Technology Library were invited as experts by the National Science Centre to conduct a series of training courses for librarians and university administration employees from all over Poland. In the following months, the employees of the Centre trained several hundred people from all over Poland – researchers, librarians and administration employees. In addition, special individual training was also prepared at the request of the authorities of several Polish universities.

These and subsequent initiatives, training and consultations for many Polish research centres, and participation in international initiatives such as GO FAIR confirmed that the project's implementation and assumptions came at the best possible time.

Research data are becoming increasingly important for researchers who appreciate the benefits of sharing data or are obliged by funding providers to provide open access to their research results. In addition, scientists are beginning to see the advantage of reusing data sets. It has been confirmed by the results of the research conducted by Digital Science-Figshare. In addition, "The State of Open Data Report", published in 2016, 2017, 2018 and 2019, examined global attitudes toward Open Research Data (Science Digital, 2020).

The Bridge of Data project participants were aware that building only the technical infrastructure without sufficient support for the researchers would not be successful. That is why an essential element of the project is substantive support for academics provided by the Open Science Competence Center (OSCC) at the Gdańsk University of Technology Library. The OSCC was established to fill the gap in the area of scholarly communication support at scientific institutions.

The idea of creating the OSCC arose from the global trend of data stewardship and experiences from the previous Bridge of Knowledge Project. It has been revealed that researchers' knowledge of Open Science has some gaps, especially those related to copyright and research data management issues. To resolve these practical difficulties faced by researchers, a support team that includes data specialists and librarians has been organised and is managed at the Gdańsk Tech Library. Part of the team is recognised as data stewards – a relatively new position in the Polish academic landscape. A data steward is usually seen as a disciplinary expert with diverse knowledge and experience in research

data management practices. Another critical role played by OSCC members is that of data support librarians. Their essential characteristic is supporting researchers at multiple stages of the data life cycle, both during the research process and during the curation process (Wałek, 2018).

The Open Science Competence Center is a particular contact point that offers help and supporting services for "sharing" research culture. The Center provides various training, consultancies, and other events promoting the idea of opening science and trying to popularise its benefits such as broader collaboration, increasing usage and citations of scientific articles and data, faster impact, and greater public engagement.

Currently, different types of training are offered to scientific staff and students, such as tailored training (e.g. concerning the scientific discipline), face-to-face consultations, and online webinars divided into thematic blocks:

- 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).

Center members also validate the metadata descriptions and data formats entered into the repository.

Another main challenge for the OSCC was considering the differences between scientific disciplines and their different scholarly communication practices regarding sharing scientific output to provide complex support with RDM. Different approaches are necessary when dealing, e.g. with humanities data and other problems that occur within medical data. Each field has its scholarly communication practices and should be treated individually concerning maintaining the regulations and procedures following the research data issues.

The Research Data Management Services team is not the only team within the Center. In addition to the LRS mentioned above, another team consists of specialists in publishing policies and journal models of scientific publishers (Open Access policies team).

The employees of the Center also support the development of additional services, such as a platform for publishing scientific journals based on OJS and a platform supporting the organisation of conferences based on Indico. It makes the Competence Center team a multitasking team that provides comprehensive support for all the Open Science processes.

The Competence Center team comprises employees of various sections of the Library, specialising in a specific issue. Organizationally, it is located in the Scientific and Technical Information Services, but it is under the task of the Center leader, who is under the substantive project coordinator.

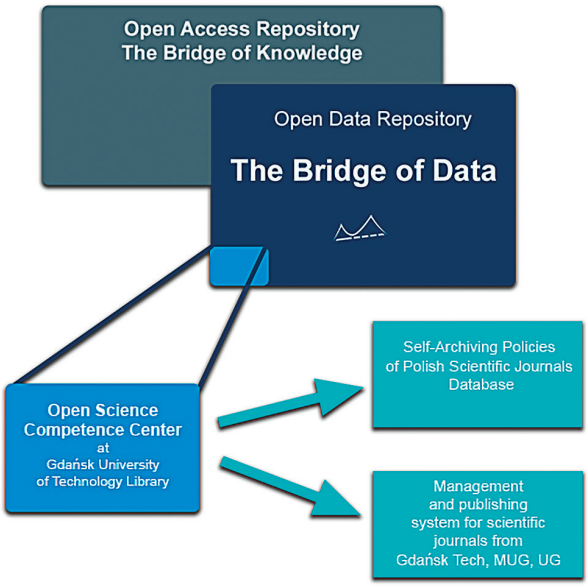



**Fig. 1.18.** The scheme of the Open Science Competence Center

## Research teams

A project assumption was to involve representatives of various scientific disciplines represented at all three universities initially engaged in the project. The process of shaping the team took several months and was very dynamic. In the beginning, three team leaders were appointed, whose task was to identify people at their universities who have and would like to share their research data. Then, teams were formed. The Competence Center conducted a series of training to familiarise the team members with Open Science principles, the basics of data management, and complex issues such as licensing, metadata standards, and the data formats that can be shared in the repository. The team members significantly influenced, among other things, the primary metadata description in the repository. It turned out that the diversity of data disciplines, types and formats forced the creation of one universal form, allowing for trouble-free entry of all possible kinds of data. Many functions such as private links and restricted access were also introduced at the researchers' request. The future role of team members is also to play the role of data champion in their units – the ambassador of the project and the idea of opening research data. The scientific team members took on the implementation of the assumed indicator of the project, which is the provision of approx. thirty thousand datasets only during the project's duration (until the end of 2021). The teams' activities are coordinated, supervised and accounted for by the substantive coordinator of the project. It is worth mentioning that all research teams described their experiences and examples of their activities in the field of data opening in the other part of this monograph.

## Database of Policies

The Database of Copyright and Open Access Self-Archiving Policies of Polish Scientific Journals collects and analyses the publishing policy in Open Access, copyright management, and the content of publications by both authors and users. According to the creators' concept, the database was to fulfil a function similar to SHERPA / RoMEO. Besides this, the database of publishing policies substantially complements the information in SHERPA / RoMEO, comprehensively collecting information about Polish journals, both those indexed in it and those not yet registered there. Moreover, the database extends the info relevant to Polish users by providing data on the current scoring and publication model. Unfortunately, information about the journal model has not been registered by the available services so far, and it is necessary to verify compliance with the funders' requirements.

The collected information is presented in the form of a journal profile, the description of which consists of several sections. The first part presents primary data about the journal, i.e. ISSN and eISSN, website address, publisher name, and assigned scientific disciplines, according to the Ministry of Education and Science's current list. The following sections provide information on the journal's scores (current and previous years), the journal's model, and the CiteScore scores. In the next section, the user will find a comprehensive set of data on the publishing policy of the selected journal, i.e. the license used and the conditions of self-archiving – i.e. the possibility of using the article's content after its publication by the author. When determining the conditions of self-archiving, particular emphasis is placed on whether authors have the right to deposit the text (at least in their institutional repositories), what types of text are subject to this consent (submitted, accepted, and published versions) and whether there is a time embargo (i.e. whether the editorial office requires a periodic abstention from self-archiving from the author). The data is collected based on publicly available information and direct contact with the editorial office or publisher. The communication aims to obtain and supplement the information required to accurately present the journal's profile in the database (Wałek and Kokot, 2020; Kokot and Szymik, 2020).

The database recipients are primarily members of the research community: researchers, doctoral students, and students. However, a database is also a convenient tool that allows academic librarians to define the so-called Green Open Access rules, especially those involved in developing institutional repositories.

There are over 3,500 records in the database. Each of them was assigned a score according to the criteria of the Ministry, a publication model consistent with the definition included, among others in Plan S, and a policy that defines the principles of auto-archiving.

The data was collected, verified, and entered by a team of about 20 Gdańsk Tech Library employees. It should be emphasised that the database was designed and implemented by a group of librarians and is the result of experience gained through cooperation with the academic community. The technical IT team has perfectly executed all these assumptions. Furthermore, the need to respond to the scientific community's changing needs made it possible to adjust new librarianship skills to those already existing.

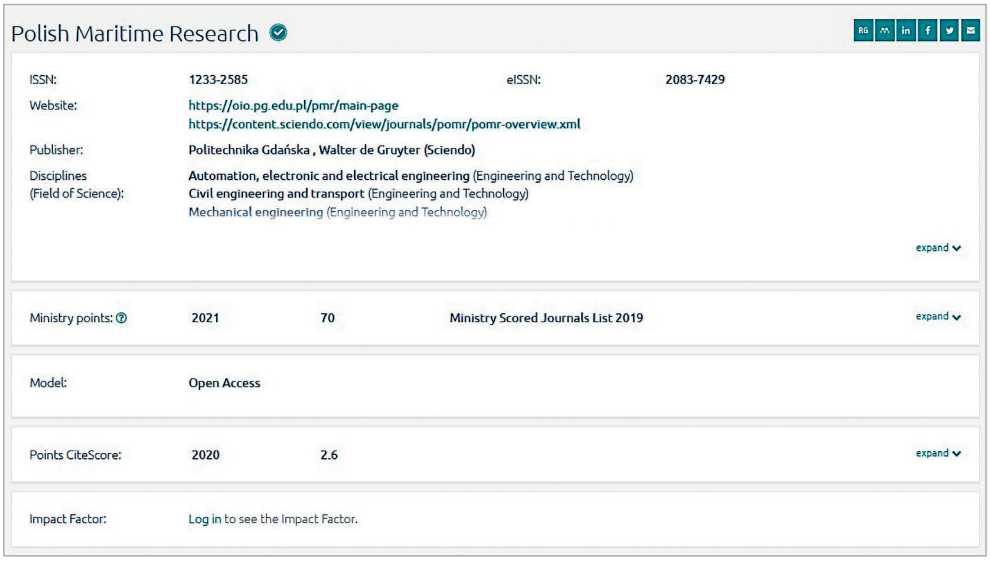



**Fig. 1.19.** An example of the journal's record in the policies database

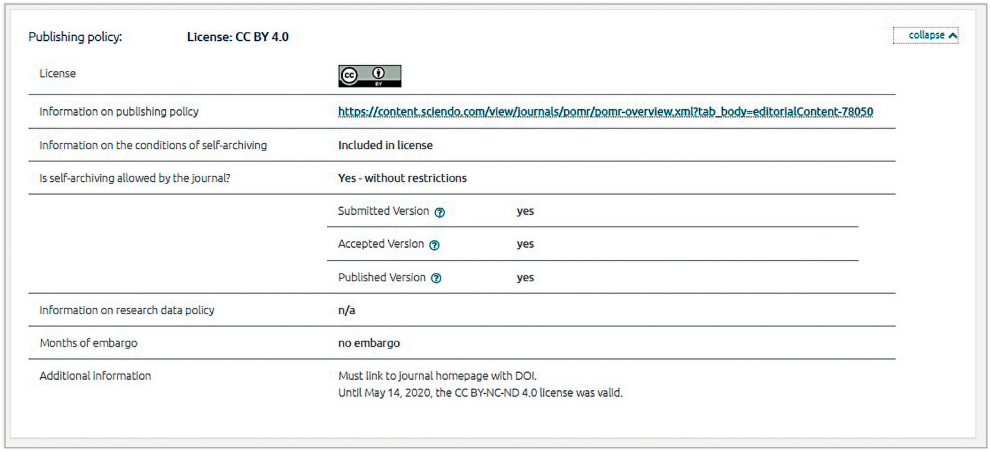



**Fig. 1.20.** An example of the journal publishing policy in details

## Virtual Microscope

Under the Bridge of Data project, the team from the Medical University of Gdańsk (MUG) undertook the creation of the Digital Tissue and Cell Atlas and the Virtual Microscope with the cooperation of the Gdańsk University of Technology and CI TASK (Center of lnformatics Tri-Citry Academic Supercomputer and network). Experts in pathomorphology gathered carefully selected histological and cytological specimens. Samples were processed, stained and scanned in the MUG's Department of Medical Laboratory Diagnostics (DMLD) (see the chapter dedicated to the Virtual Microscope tool).

Virtual Microscope is an application made available as part of the platform. Its role is to provide comfortable and quick access to a collection of digital microscopic images stored in a data repository. The shared images are characterised by very high quality and resolution.

The principle of its operation is very similar to popular map-sharing services such as Google Maps. However, as the data to be displayed to the user are very large, it is impossible to send them in full. Instead, the data is divided into small fragments and transmitted only when there is a need to display a specific area. Thanks to this, the user can start viewing an image, the total size of which is several dozen gigabytes, practically without any noticeable delay, even on a mobile device.

To enter the virtual microscope page, you can use the mostwiedzy.pl portal page or select a direct address (https://wirtualnymikroskop.mostwiedzy.pl/ and www.digitalpathology.pl).

Being on the main page, we have to choose from 3 options:

- go to the main page,
- a list of images along with the search engine – slide list,
- organ list with graphic preview – select organ.

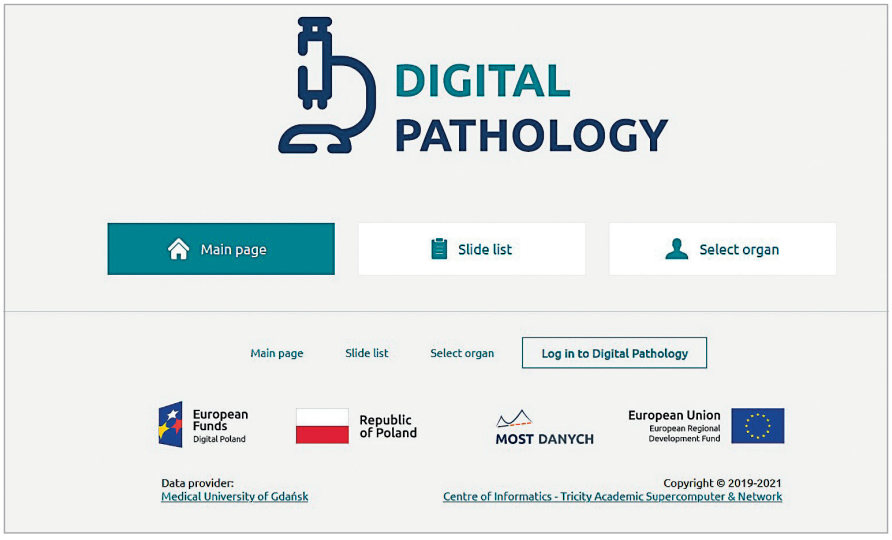



**Fig. 1.21.** The main page of the Virtual Microscope (digital pathology) tool

On the page with a list of images (slide list), we have the opportunity to search for us interesting research through the search engine. If we are interested in a graphical preview of the search organs, we can use the organ selection option. When we click on the selected organ, there will be a list with the appropriate filter set. The sample preview field contains the navigation buttons and option selection.

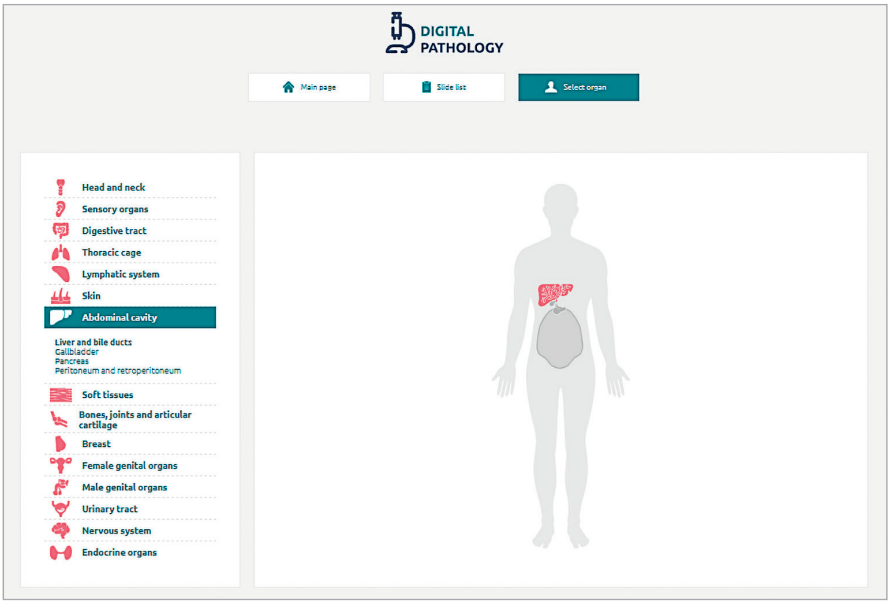



**Fig. 1.22.** The view of the selected organ

With the help of a virtual microscope, you can view the deposited images, which were created due to the preparation and digitisation of thousands of tissues and cells. It is an innovative way of presenting and broader use of research data collected in the repository.

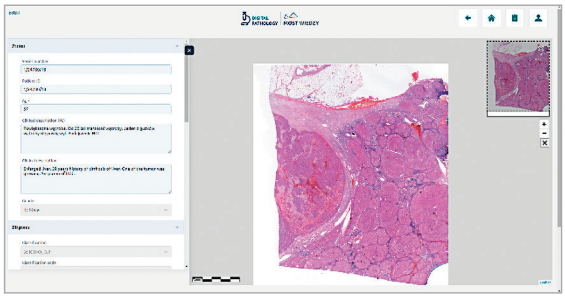


**Fig. 1.23.** The preview and navigation

The general scheme of creating resources for the Virtual Microscope is presented in the diagram below. However, the entire process will be described in more detail in the third chapter of this book.

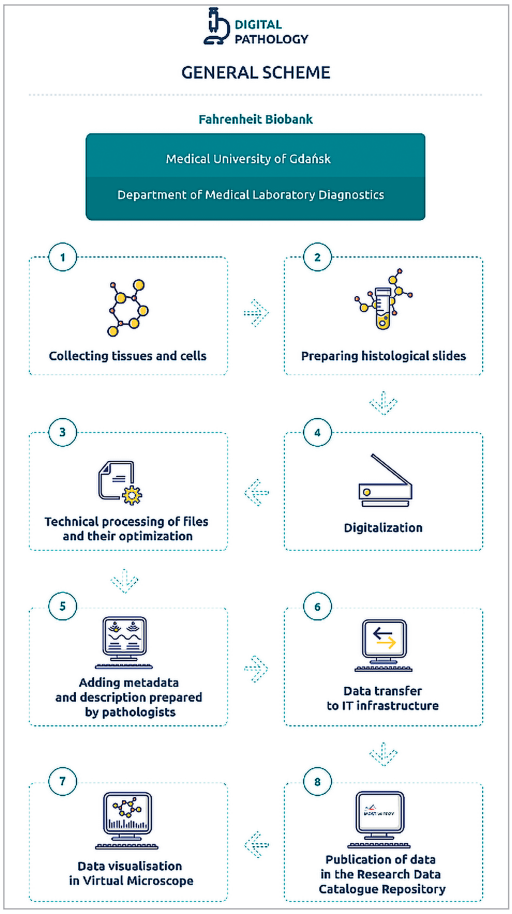



**Fig. 1.24.** General scheme of the Virtual Microscope – Digital Patology database workflow

## Services supporting the editing of scientific journals

As part of handling the publishing process of scientific journals by the mostwiedzy.pl platform, it was decided to adopt the Open Journal Systems (OJS) solution. It is open-source software for managing and publishing scientific journals developed and published by PKP (Public Knowledge Project). It is currently the most widely used open-source journal publishing platform, with over 10,000 journal titles worldwide.

OJS is a comprehensive tool for managing the entire editorial process to publish articles and issues online. The system has implemented a typical journal structure with its periodicity. It marks subsequent numbers and encapsulates texts into identifying data (metadata), such as the author, title, abstract, keywords, or DOI (digital object identifier). The available indexes and the ability to browse and search the journal's content additionally make OJS a highly functional tool for readers.

Supervision over the circulation of the text, from the moment of its submission to publication (or rejection), enables the journal's editors to meet the deadlines for the publication of subsequent issues of the journals.

The OJS system has been integrated with the mostwiedzy.pl portal. The integration was carried out at the level of authentication of system users – to use the system, you must have an account in the Bridge of Knowledge system – and visually adjusted (in the information layer and the journal reader) to the graphic design of the mostwiedzy.pl portal.

In the information layer, the OJS system functions as a Content Management System (CMS), under which the editorial team of the journal could launch the journal's website, publishing information about the journal itself, guidelines for authors of articles, guidelines for article reviewers and providing access to the current and archival issues of the journal.

## Services supporting the organisation of scientific conferences

When preparing services supporting the organisation of scientific conferences, it was decided to adapt the Indico system, created by CERN (European Organization for Nuclear Research). It is open-source software for organising events, supporting the archiving of materials related to the event and supporting the cooperation of the organising team.

Indico is a web application that facilitates the organisation of events of any size, from short meetings and lectures to large conferences, including scientific conferences, the necessary stage of obtaining and reviewing conference materials (articles, lectures, workshops, posters, etc.).

Indico offers a wide range of features that include presenting the main page of the event, workflows for scientific articles and their abstracts, and a full-fledged user registration system. Additional system functions are:

- a multi-level authorisation system for the event team,
- uploading and downloading articles, presentations and other documents,
- archiving event materials and event metadata,
- reviewing conference materials.

An additional advantage of the system is an active community (Indico Community) that shares knowledge and experience related to the system's implementation, configuration, and development.

The Indico system provides full support throughout the life cycle of a scientific conference, from the implementation of the processes of submitting and reviewing abstracts of presentations through submitting articles to the possibility of preparing materials for the post-conference publication with organisational support for the entire event – registration of participants, communication with participants, sharing materials with them.

## E-services overview

From a technical perspective, the Bridge of Knowledge platform integrates multidisciplinary and multidimensional data from many databases. The metadata appropriately describes each piece of data to connect with other objects and to machines that understand

the metadata format. In addition, there are e-services dedicated to sharing information about research output and its metrics and potential. The Open Access repository is one of the biggest in Poland. The unique journals catalogue gathers information about the model and open access policy of, especially, Polish journals.

More and more scientists are interested in depositing their datasets in the repository, undergoing the Core Trust Seal certification process.

Newly introduced modules dedicated to scientific conferences and editors of scientific journals facilitate their work significantly. The entire process of registration, reporting, reviewing and open access publishing and assigning DOIs is supported.

Every object located on the platform is easily findable by an elastic and intelligent search engine. It analyses the search context and history and returns the results that best suit the user's interests. The platform also suggests other objects that the user might be interested in. This process is ready for machine use. The whole platform is accessible by an open REST API so that it is easy to integrate with and easy to embed data on different websites.

All e-services offered by the platform are available free of charge for all scientists.

## High-level architecture of the platform

The Bridge of Knowledge platform is a proprietary solution developed by Gdańsk University of Technology teams using open-source technologies and components. It is based on the Elasticsearch indexing and search engine, which supports full-text search (Free and Open Search: The Creators of Elasticsearch, ELK & Kibana, 2021). Metadata and data are located in a NoSQL database that provides object storage compatible with the S3 API (Amazon S3 REST API Introduction, 2006). It is located in two private clouds – the main engine works on the IT Services Center cloud. In contrast, the open research data repository is located in the TASK (Center of lnformatics Tri-Citry Academic Supercomputer and network) and is served directly by dedicated services from their cloud. Both clouds are managed by Kubernetes (The Kubernetes Authors, 2021).

All services that run inside the clouds can be scaled to the appropriate level of performance. The individual elements are:

- Front-end in PHP technology – a service based on the Front-end of mostwiedzy.pl responsible for the presentation layer
- BigDataAnalysis using Apache Spark technology – searching for data based on metadata and performing advanced BigData analyses
- StorageProxy using the NGINX server – data access authorisation service realised using JSON Web Token
- ImageServer – DICOM file server – serving DICOM files in a format that can be presented in a web browser, also with JSON WebToken checking
- BusinessLogic JAVA and Spring technologies – support for the business logic of the portal, authorisation control, generating JSON Web Token, searching and saving metadata

- ElasticSearch database – NoSQL database of the mostwiedzy.pl portal, including ORD file metadata
- Object data warehouse – a warehouse with a size of up to 150 TB responsible for ORD data storage, based on Ceph software

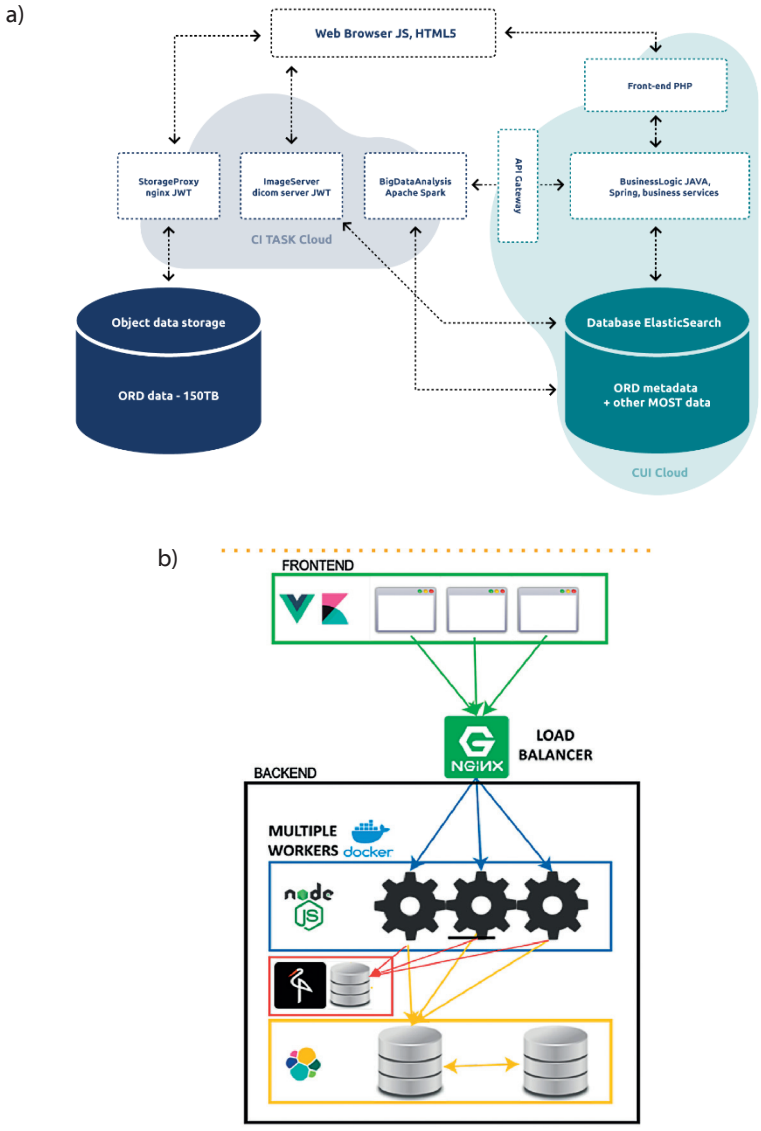



**Fig. 1.25.** High-level architecture of the platform

The platform and its e-services are available via every modern web browser. The user interface was designed with the needs of people with disabilities in mind. It also flexibly

adapts to the device being used, making it easy to use on PCs, laptops, tablets, and smartphones (Lubomski, Pszczoliński and Kalinowski, 2017).

## Integration, network and storage challenges

Gathering and exchanging such a large amount of data implies extensive integrations with many external services, such as ORCiD, CrossRef, DataCite, Clarivate Analytics, Scopus, etc. (Wałek and Lubomski, 2017).

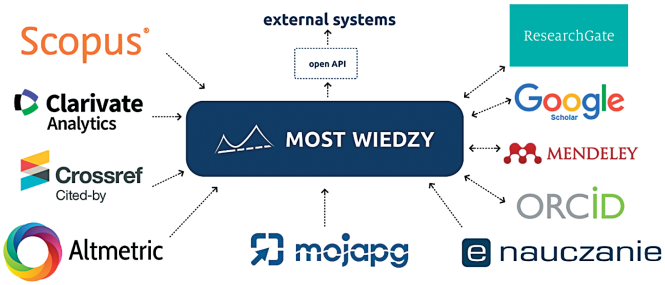



**Fig. 1.26.** Interoperability of Bridge of Knowledge platform

The Open Research Repository is expected to include over 150 TB of data by the end of the project (June 2022). It introduces many challenges related to network and storage that had to be taken into account during the design and implementation process. It is essential in the e-service that implements big-data analysis of the open datasets from the repository on the supercomputer located in the TASK (Center of Informatics Tri-Citry Academic Supercomputer and network). The TASK provides a high-speed internet network with direct connections to the PIONIER and GEANT networks. Moreover, both clouds are connected via dedicated, duplicated fibre optic connections. Every dataset is located in fast object storage. A dedicated custom backup process that is reliable and cost-effective was designed and implemented.

## Quality, security and reliability assurance

One of the principles of the project was to serve high quality, verified data in an efficient, secure and reliable way. That is why most of the data are published due to multi-step review workflows. Data are verified by highly qualified data stewards and specialists in open access and bibliometrics.

From a technical point of view, private clouds ensure easy horizontal scaling and adaptation to the current load. The search engine uses numerous indexes that serve results quickly regardless of the repository and database size, which should be standard in modern systems. Thanks to applying a rolling release process, maintenance works are done without downtime and inaccessibility, which are inconvenient for users (Lubomski,

Pszczoliński and Nowacki, 2017). Moreover, many monitoring services that control the health of each service and react if something goes wrong have been implemented.

## Ways of measuring traffic and scale of success

It is crucial to have real-time monitoring of network traffic in the system and historical data for analysis. We use Google Tag Manager and Google Analytics to track user behaviour and flow on our platform. In addition, some counters and sensors serve us valuable information about how the platform is used and what is worth expanding.

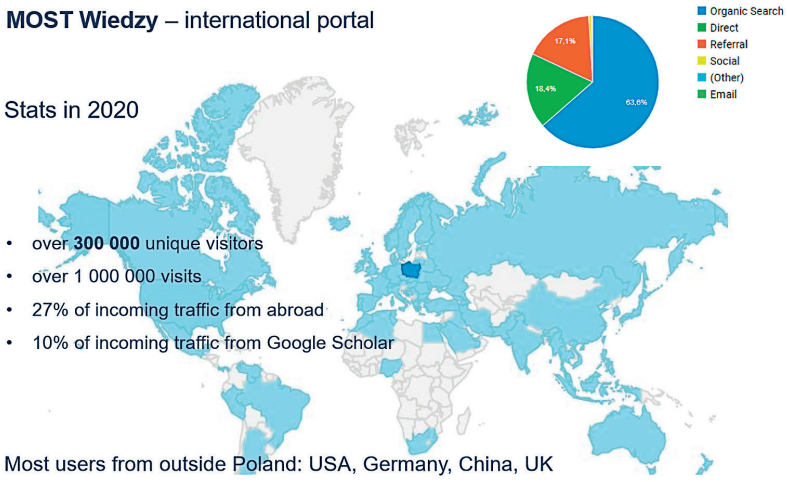



**Fig. 1.27.** Traffic on the Bridge of Knowledge platform

Based on Google Analytics stats, we can see that the Bridge of Knowledge platform is an international platform (nearly 30% of traffic is from outside of Poland) and is doubling the number of users year on year (over 450,000 unique users in 2021). It is worth emphasising that over 60% of traffic comes from organic search.

## Data searching and binding

According to the RDF standard (Resource Description Framework (RDF), 2014), all data gathered on the platform is organised as connected objects with a specific type with descriptive metadata defined (according to the RDF standard). Therefore, it does not matter how a user gets to a resource (internal or external search engine, a direct link, or an internal or external catalogue) – it is possible to see the connected objects and follow the bi-directional links between the objects. It is very useful for users as well as machines.

This approach to data organisation on the platform makes it compliant with the 5 Star Open Data standard (Hausenblas, 2015).

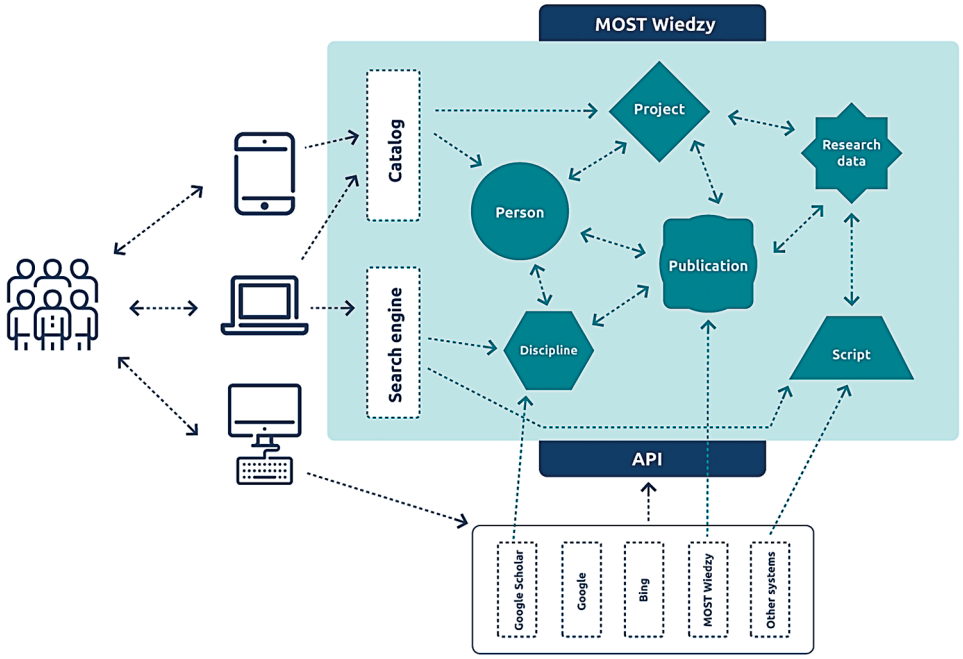



**Fig. 1.28.** The organisation of data on the Bridge of Knowledge platform

## Indexing by external search engines and SEO

The Open Research Data Catalogue created at the Bridge of Knowledge platform, by design, implements all of the principles of FAIR Data. To promote datasets among international researchers, the project emphasises the high indexability of the gathered metadata. To accomplish this goal, three crucial steps have been taken:

- decision on metadata format for the description of multidisciplinary datasets,
- implementation of a user-friendly dataset upload form,
- providing the gathered metadata for indexing robots and registration in searchable resources.

Ensuring a proper metadata description for each dataset begins when scientists upload data to the repository. It was decided to base the data description model on the popular schema provided by DataCite. The Bridge of Knowledge dataset upload form requires a minimum scope of data that the author needs to provide. Moreover, it encourages the indication of recognised identifiers for researchers (ORCID) or institutions (ror.org). The data is then presented on the website in a user-friendly and transparent form. The form of presentation meets the standard of accessibility for the blind and visually impaired – WCAG 2.1. Apart from a human-readable presentation, an equally strong emphasis was placed on a detailed, unambiguous description for indexing robots. For this purpose, it was decided to describe the data in JSON-LD format based on Schema.org. This method of describing the datasets stored in the repository resulted in a very high indexation rate

in Google Data Search. Each dataset published on the Bridge of Knowledge platform is indexed directly on DataCite.org.

The Bridge of Data Open Research Data Repository has been indexed by the Web of Science, which confirms the high standards of the metadata description of datasets.

In summary, all principles of FAIR Data are achieved in The Bridge of Knowledge Data Repository. The deposited data are:

- **Findable –** All datasets available on the portal have a unique URL address and DOI identifiers. If the datasets don't have their identifier at the time of publication, they receive a new DOI number generated by the repository. URLs and identifiers do not change over time and allow users to reach every version of a given dataset. The datasets are described with metadata in JSON-LD format based on Schema.org and are indexed in the Web of Science, Google or Google Dataset Search, making them easy to find.
- **Accessible** – The repository has a dedicated catalogue of Open Research Data, fully accessible by a standard web browser. The ability to search/read metadata and download datasets is open and does not require performing activities such as registration or providing contact details. Over 98% of datasets deposited in the repository are available for download without registration. Whether the author chooses to deposit data within the restricted access, there is an option for the user viewing the dataset description to send a request for accessing the dataset data. The request is transferred directly to the author.
- **Interoperable –** Providing metadata of deposited datasets is mandatory for creators. The Open Science Competence Center verifies the quality of metadata upon approval. Moreover, the creators are encouraged to create metadata connections to other resources available in the Bridge of Knowledge Portal, such as publications, projects (funding), laboratories, teams or other datasets. The feature of automatically generated citation has been implemented to improve the visibility and reusability of data. Every record can be saved as a citation formatted in the most popular citation styles (APA, Harvard, Vancouver, Chicago, and MLA) or exported to format files native to citation manager software (RIS, BIBTEX). Metadata can also be downloaded as a JSON-LD file or DataCite schema.
- **Reusable** – To ensure the possibility of data reuse, providing licenses for each record is mandatory for all data depositors. Each dataset contains a license that specifies the conditions for the reuse of the dataset. They are established when the dataset is uploaded and are guaranteed not to change over time. The repository allows depositors to choose from creative commons licences or upload a separate file containing their license (custom license). When discovering a dataset, information about the licence is always displayed. The data depositors are encouraged to use open formats whenever possible without losing information/quality. In the case of non-standard formats, we recommend that authors generate and attach a descriptive file (e.g. Readme.txt) containing all the information allowing the correct interpretation of the data. Whether the data needs dedicated software to analyse it, data depositors are asked to specify it – this is one of the metadata values describing the dataset.

## Summary

The Gdańsk University of Technology has created several tools and services supporting the open sharing of scientific research results, including scientific publications and research data. The Bridge of Data project was established to support researchers in their activities regarding different Open Science layers. Thanks to the involvement of specialists in various fields and the scientists themselves, representing the three most prominent universities in Gdańsk, the tools and services produced are at the world level in terms of their quality and functionality. The project's success results from cooperation between all the teams that make up the project team.

## References

Amazon Web Services, Inc. (2006) '*Amazon S3 REST API Introduction*'. Available at: https://docs.aws.amazon.com/AmazonS3/latest/API/Welcome.html (Accessed: 8th March 2021).

DDI (2021) '*Document, Discover and Interoperate*'. Available at: https://ddialliance.org/ (Accessed: 20th April 2021).

Elasticsearch, B.V. (2021) 'Free and Open Search: The Creators of Elasticsearch', *ELK & Kibana | Elastic*. Available at: https://www.elastic.co/ (Accessed: 8th March 2021).

Gdańsk University of Technology (2021) '*MOST Wiedzy – your knowledge portal*'. Available at: https://mostwiedzy.pl/en/ (Accessed: 8th March 2021).

Kokot-Kanikuła, K. and Szymik, A. (2021) 'Zadania i kompetencje bibliotekarzy w zespole ds. polityk wydawniczych polskich czasopism naukowych', *Biuletyn EBIB*, 3(198), pp. 1–8. Available at: http://ebibojs.pl/index.php/ebib/article/view/738 (Accessed: 25th May 2022).

Krawczyk, H. and Lubomski, P. (2017) 'Multidisciplinary Open System Transferring Knowledge for R2B Development', *TASK Quarterly*, 21(4), pp. 299–308. DOI: 10.17466/tq2017/21.4/c.

Lubomski, P., Pszczoliński, P. and Kalinowski, A. (2017) 'Techniczne aspekty niezawodnej chmury obliczeniowej', *INFOBAZY 2017 – Bazy Danych dla Nauki – dane, wiedza, e-usługi*, pp. 1–6. Available at: https://mostwiedzy.pl/pl/publication/download/1/techniczne-aspekty-niezawodnej-chmury-obliczeniowej_17758.pdf (Accessed: 25th May 2022).

Lubomski, P., Pszczoliński, P. and Nowacki, M. (2017) 'Dobre praktyki i narzędzia open source w zwinnym wytwarzaniu oprogramowania'. *INFOBAZY 2017 – Bazy Danych dla Nauki – dane, wiedza, e-usługi*, pp. 1–6. Available at: https://mostwiedzy.pl/pl/publication/download/1/dobre-praktyki-i-narzedzia-open-source-w-zwinnym-wytwarzaniu-oprogramowania_17756.pdf (Accessed: 25th May 2022).

Hausenblas, M. (2015) '*5-star Open Data*'. Available at: https://5stardata.info/en/ https://5stardata. info/en/ (Accessed: 8th March 2021).

RDF Working Group (2014) '*Resource Description Framework (RDF)*'. Available at: https://www.w3.org/RDF/ (Accessed: 8th March 2021).

Science, D. et al. (2020) 'The State of Open Data 2020'. *Digital Science*. DOI: 10.6084/m9.figshare.13227875.v2.

The Kubernetes Authors (2021) '*What is Kubernetes*?' Available at: https://kubernetes.io/docs/concepts/overview/what-is-kubernetes/ (Accessed: 8th March 2021).

Wałek, A. (2018) 'Is data management a new 'digitisation'? A change of the role of librarians in the context of changing academic libraries' tasks, *IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies*, pp. 1–11. Available at: https://mostwiedzy.pl/pl/publication/download/1/is-data-management-a-new-digitisation-a-change-of-the-role-of-librarians-in-the-context-of-changing-_30244.pdf (Accessed: 25th May 2022).

Wałek, A. and Kokot, K. (2020) 'Untying a tight knot – Database of Copyright and Open Access Self-Archiving Policies of Polish Scientific Journals', *Open-Access-Tage 2020 (OAT2020)*, Bielefeld, Germany. Zenodo. DOI: 10.5281/zenodo.4018894.

Wałek, A. and Lubomski, P. (2017) 'The Bridge to Knowledge – Open Access to Scientific Research Results on Multidisciplinary Open System Transferring Knowledge Platform', *TASK Quarterly*, 21(4), pp. 333–342. DOI: 10.17466/tq2017/21.4/e.

Wałek, A. and Szuflita-Żurawska, M. (2017) 'The Bridge of Knowledge – defining and implementing national Open Access policy by Polish universities'. Gdansk University of Technology perspective. *IFLA World Library and Information Congress, 83rd IFLA General Conference and Assembly: Libraries. Solidarity. Society*, pp. 1-8 Available at: https://mostwiedzy.pl/pl/publication/download/1/the-bridge-of-knowledge-defining-and-implementing-national-open-access-policy-by-polish-universities_16116.pdf (Accessed: 25th May 2022).

Wilkinson, M. et al. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data,* 3, 160018. DOI: 10.1038/sdata.2016.18.