Sharing research data across disciplines

Edited by Anna Wałek

Gdańsk Tech Publishing House

GDAŃSK UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE
CHAIRMAN OF EDITORIAL BOARD
Dariusz Mikielewicz

EDITOR OF SCIENTIFIC PUBLICATIONS
Michał Szydłowski

REVIEWERS
Henryk Krawczyk
Ignasi Labastida

ASSOCIATE EDITORS
ASSOCIATE EDITORS
Beata Adamczak
Ewa Cypukow
Aleksander Mroziński

LAYOUT
Wioleta Lipska-Kamińska

COVER DESIGN
Łukasz Zgraja

Gdańsk University of Technology publications may be purchased at
https://www.sklep.pg.edu.pl

DOI
https://doi.org/10.34808/x55q-sz53_dyr

ISBN 978-83-7348-864-9

Table of Contents

Introduction

Research data are all data collected, observed or produced during a research process to obtain original scientific results. Depending on how or for what purposes they were created, we can distinguish, among other things, observation, experimental, reference, compilation data or simulation data.

Each discipline of science creates its research data specific to it. For example, biology may include photos made with a microscope or films documenting animal behaviour, civil engineering and environmental protection – geographical or spatial data and archival documents in the case of history. The research data will also include descriptions of procedures, laboratory logbooks or notes from experiments.

There are many types of data, standards for describing them and how they can be made available. The authors of this monograph intended to present practices, standards and experiences developed by a team of scientists, librarians and IT specialists implementing the "The Bridge of Data" project (Bridge of Data – Multidisciplinary Open System Transferring Knowledge. Stage II Open Research Data).

This project, resulting from the work of three Pomeranian universities – Gdańsk University of Technology, University of Gdańsk and Medical University of Gdańsk – aims to gather and make available in the repository more than 30 000 data sets from various scientific disciplines (until the end of June 2022). Therefore, interdisciplinary teams have been created at each partner university. They produce, describe, adapt and make available research data on the Bridge of Knowledge platform and act as ambassadors of openness at their universities, departments, institutes, etc.

The dissemination of practices related to the management and sharing of research data has led to new occupational and professional specialisations. They include the data steward and data librarian. A data steward is usually seen as a disciplinary expert with diverse knowledge and experience in research data management practices. In this case, the knowledge of research practice and active involvement in research activities is essential. The role of the data support librarians is to support researchers at multiple stages of the data life cycle, both during the research process and during the curation process. Specialists of this type work at the Open Science Competence Center (OSCC), established in the Library of the Gdańsk University of Technology (Gdańsk Tech Library). Their task is, inter alia, to support research teams and scientists themselves in the creation and implementation of a Data Management Plan; to conduct training, briefings and consultations

on broadly understood Open Science; as well as to verify and validate the data sets and their metadata descriptions in the research data repository.

As a result of the work of the teams of scientists, with the substantive support of the Open Science Competence Center in the GdańskTech Library and with the technical support of IT centres of the Gdańsk Tech, data were created and made available, which are the starting point for this study. Moreover, this work results in a repository platform dedicated to data – the Open Research Data Catalogue on the Bridge of Knowledge platform, data analysis, storage and backup tools, and support services provided by the team of experts of the Competence Center employed in the project.

The monograph consists of three essential parts, which are then divided into chapters. The first part of the monograph includes chapters detailing the activities directly related to the Bridge of Data project assumptions.

It starts with the introductory chapter, "The Bridge of Data Project Objectives". It clarifies the genesis of the Bridge of Data project and the previous Bridge of Knowledge project. The progress of their implementation has been described, and the most critical assumptions, effects, tools, and services created. This chapter mainly describes the establishment of the Bridge of Knowledge repository and the Open Research Data Catalogue as an integral part of the platform. It approximates the technical and organisational solutions applied and organises knowledge of guidelines for the creation of open data repositories, as well as, among other things, their indexation and certification. In addition, this chapter approximates the genesis, tasks and functioning principles of the Competence Center established in the Gdańsk Tech Library as a unit supporting the research data management process and providing many services to the scientific community directly related to Open Science and Open Research Data. Other tools and services developed as part of the project, including a virtual microscope, data analysis service and policy base, are also signalled.

The chapter entitled "Data Analysis in Bridge of Data" presents issues in analysing research data on a supercomputer. It describes the architecture of this solution and the infrastructure, together with the Tryton supercomputer. Finally, it indicates how scientists use the solutions developed to perform complex calculations and simulations.

Subsequent chapters describe data management practices in several selected scientific disciplines.

This part opens a chapter entitled "The Digital Tissue and Cell Atlas and the Virtual Microscope", whose authors describe the creation of a digital cell and tissue atlas and the Virtual Microscope tools within the Bridge of Data project. The tool described, resulting from the digitisation and accessibility of digital images of tissues collected from patients, responds to contemporary didactics and education challenges, shaping digital skills and promoting the use of modern technologies. The main idea behind creating such modern digital tools is to use them to develop new distance teaching methods.

The chapter entitled "Towards Open Research Data in the Economics Discipline" describes synthetically the issues related to sharing research data in management and economics sciences. It presents, among other things, opportunities and possible doubts

related to the provision of data in these specific disciplines, where research is often based not only on public funds but also on cooperation with business and industry.

The next chapter, "A text as a set of research data. A number of aspects of data acquisition and creation of datasets in Neo-Latin studies", presents an approach to humanistic data resulting from historical and neo-Latin works. The authors of this chapter, specialising in neo-Latin studies and the history of early-modern education, share experiences related to the creation and sharing of research data created based on Latin texts and inscriptions.

"Surf zone currents in the coastal zone of the Southern Baltic Sea – a modelling approach" is the title of the chapter, which introduces issues related to the use of data collected in a dataset. The authors present the possibilities for re-use the data collected based on modelling data.

The last chapter in this section – "Using synchronously registered biosignals dataset for teaching basics of medical data analysis – case study", describes the possibilities of using research data in teaching based on an example of medical data.

The third part of the monograph consists of chapters taking the form of data descriptors. There are a total of 40, and they represent all scientific disciplines, the representatives of which are involved in the Bridge of Data project.

The data descriptor is a revised article describing and referring to specifically selected data sets. It is quoted (so that researchers receive greater recognition for their work) and aims at facilitating the tracing, interpretation and re-use of data sets. The descriptor shall provide the information needed to interpret data contained in a data set; it is linked to one or more trusted data resources in which data files, code or other resources are stored. Many publishers, including Springer/BioMed Central, Elsevier, Wiley, Faculty of 1000 and Ubiquity Press, offer some form of data-based publications – sometimes referred to as data documents or notes rather than data descriptors but all, to a large extent, have similar objectives to increase the visibility of data sets in the reviewed literature.

All chapters of the data descriptor type have been created according to the standard. It is based on the best practices used in the reviewed scientific journals of the largest publishers. It contains the following essential elements: abstract, keywords, specification table containing primary data describing a dataset (subject area – research areas represented by this dataset described; More specific subject area – specification of research areas; Type of data; How the data were acquired – explaining how the data were obtained; Data format; Data source location; Data accessibility), Background (constituting a type of introduction and explanation), Methods, Data quality and availability (including a description of the method of obtaining data and maintaining quality thereof, as well as DOI number of the datasets described and the licence on which they are available). All data sets described were deposited in the Open Research Data Catalogue repository on the Bridge of Knowledge platform (mostwiedzy.pl).

The reference to specific data in the repository shall allow the reader to find the data immediately and to verify and deepen the information provided in the given chapter.

Although the data descriptors represent different scientific disciplines, the types of data and ways of obtaining them have been grouped in such a way as to highlight further

the spectrum of research carried out at three partner universities and to illustrate the diversity and usefulness of the data produced in the research processes. For example, here are datasets used in medicine, diagnostics, biomedical engineering (including mechanical properties of human stomach tissue, immunological data, body temperature measurements, measurements of physiological parameters, thermal imaging, use of lasers in medicine, hydrogel properties), biological data used in ecology (e.g. describing microphytobenthos or ecosystem in the Gulf of Puck, cyanobacteria in the Baltic Sea – their review and extent of occurrence, presence of mercury in the environment or occurrence of Baltic herring), in addition datasets containing oceanographic data (including within the scope of physical and chemical oceanography, describing wind parameters, water temperature, water level and salinity in the case of e.g. the Baltic Sea, as well as ice cover level in the Arctic), meteorological data (e.g. weather simulations, troposphere parameters in tropics, image of the Gulf of Puck using an echo probe), road traffic monitoring data, data concerning acoustic engineering, spectroscopy, metrology, testing of steel for construction of vessels or thermoplastic materials used by 3D printers have been described. The data represented are also related to areas and phenomena such as physics and mathematical theories. Moreover, some data show lego bricks recordings used in the teaching of neurological networks (AI), data on architecture and spatial planning, classics, legal data on EU law, as well as data on social sciences – including survey data on graduates' careers or social behaviour.

The aim of an attempt to describe as much data as possible and the diversity thereof was to approximate to the reader of this publication both the complexity of research processes and data acquisition methods in different disciplines. Thus, it also shows various data types, formats, and methods – collection, description, and sharing.

Some of the names appearing in individual authors' studies may occasionally be interchangeable, although an attempt has been made to standardise this issue as far as possible. The most important names in the text include the institutions, projects, and initiatives that, either thematically or by authors' affiliation, are related to this monograph. The authors of the chapters represent three universities. The first one is the Gdańsk University of Technology (the acronyms of GUT and Gdańsk Tech are generally used. Gdańsk Tech is the abbreviated name of the university in English, adopted by the University Senate in 2021, which will ultimately replace the acronym previously used). The Gdańsk Tech is represented by employees and doctoral students of individual Departments, Library, IT Services Center and TASK Computer Center (Center of Informatics, Tri-City Academic Supercomputer and networK – under the TASK CC acronym). In addition, the University of Gdańsk (UG) and Gdańsk Medical University (GUMed) are represented. The staff members of each department indicate their full affiliation and contact details at the beginning of each chapter. Whenever the text does not indicate the name of the university but only the name of the department, faculty, unit or institute, it should be assumed that the author refers to the entity at which he/she is affiliated or to another entity indicated, for example, in the abstract or the initial part of the chapter. This principle applies to the authors representing all three universities.

The text repeatedly includes names related to the Bridge of Data project. "MOST Wiedzy" is the Polish name of the project implemented by the Gdańsk University of Technology in 2016–2020. The Bridge of Knowledge is the English equivalent of this name. The Polish and English versions of the name may occur interchangeably in the text. As part of this project, mostwiedzy.pl platform was created, the functionalities of which are described in chapter 1. Alternatively, the name of the Bridge of Knowledge platform may be used with the name MOST Wiedzy, but it does not have its Internet address in this language version. The only address version is the Polish version (mostwiedzy.pl). MOST DANYCH project (The Bridge of Data) is an extension – continuation of the Bridge of Knowledge project. Within the framework of the project, the tools and services described in Chapter 1. have been created. The Open Research Data Catalogue is also referred to in the text as the data repository or Bridge of Data repository. It is a functional extension of the mostwiedzy.pl platform to allow depositing research data.

The monograph is accompanied by a list of contents and an index to facilitate the navigation of the publication's content.

We hope that the content gathered and made available by us will prove helpful to you, both in implementing your practices related to the management of research data and designing tools and services supporting these processes.

This publication could only be created thanks to the involvement and support of the entire project team implementing the Bridge of Data project. Special thanks are due to the authors of the various chapters. Without whose expertise and effort to select and describe the subject matter, the monograph could not have gained a practical dimension. I would also like to thank the reviewers, Prof. Henryk Krawczyk and Dr Ignasi Labastida i Juan, who have given us many valuable comments and suggestions. The editorial team in the Gdańsk Tech Library – in particular Aleksander Mroziński, Beata Adamczak and Ewa Cypukow – watched over the collection, correction and technical preparation of the texts and many other organisational aspects related, among others, to communicate with the authors and proofreaders. On behalf of myself and the entire team, I hope that the monograph handed over to you will prove interesting and valuable.

Anna Wałek, PhD
Scientific editor

The monograph was created under the Bridge of Data project (Bridge of Data – Multidisciplinary Open System Transferring Knowledge. Stage II Open Research Data) co-financed by the European Regional Development Fund within the Digital Poland Operational Program for 2014–2020.

1
Part

The Bridge of Data Project Assumptions

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.

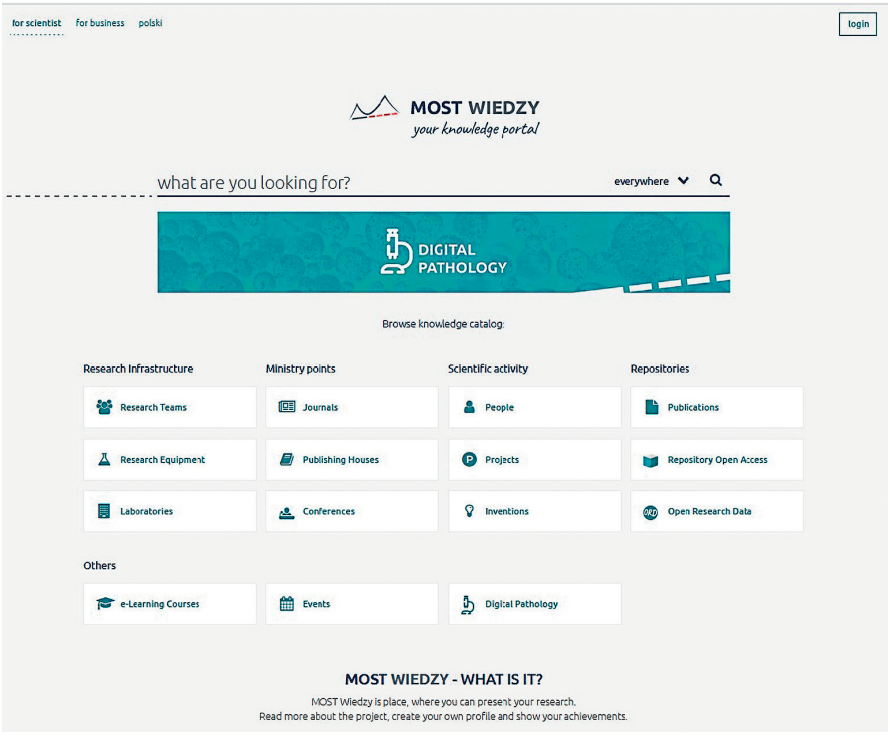


Fig. 1.2. The main page of the mostwiedzy.pl portal

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.

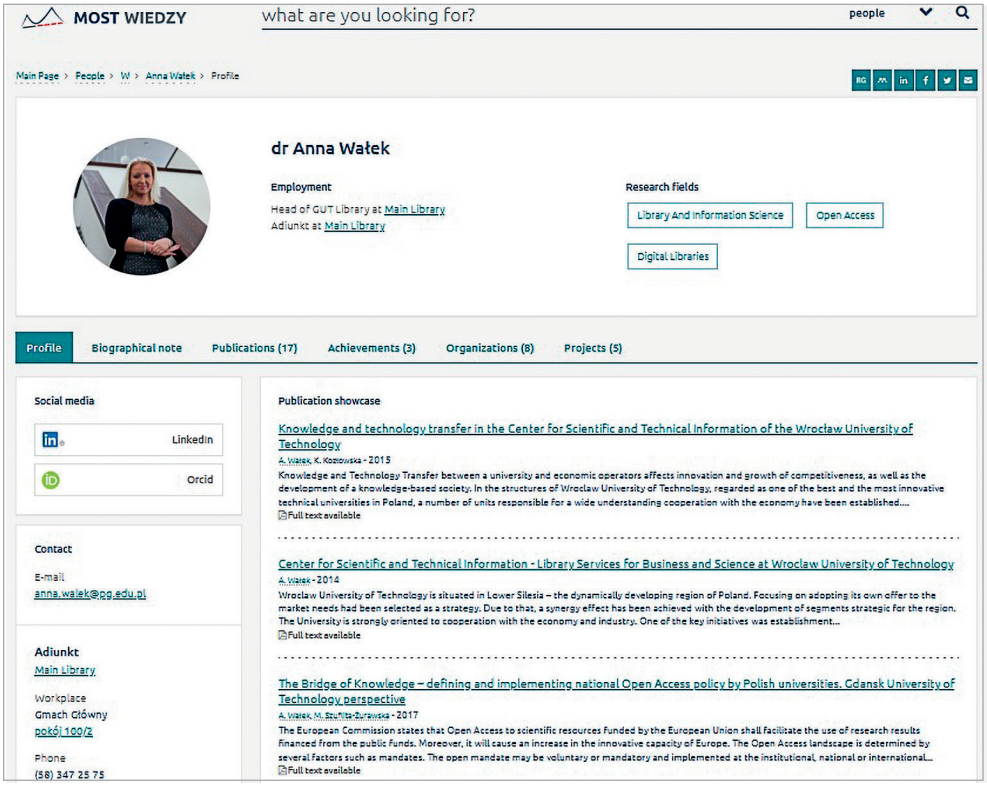


Fig. 1.3. Scientific profile

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).

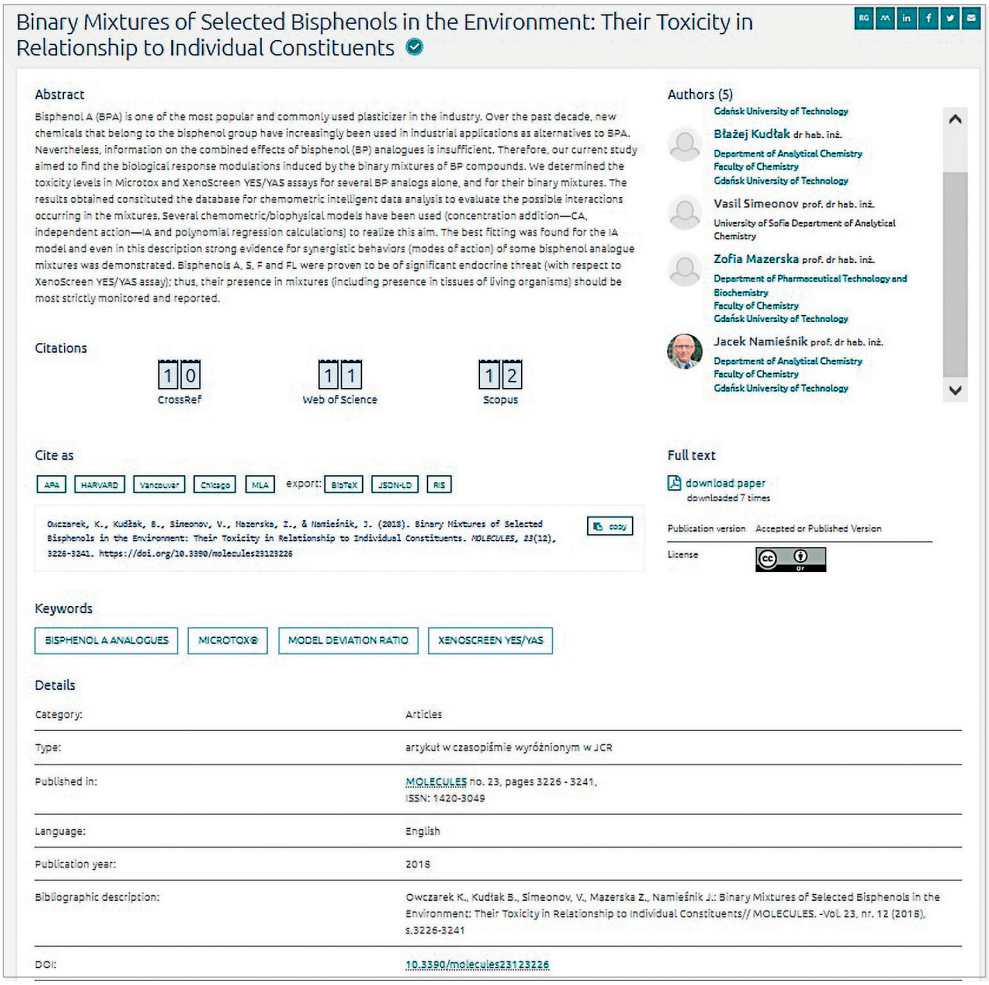


Fig. 1.4. Publication record in the Open Access Repository

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.).

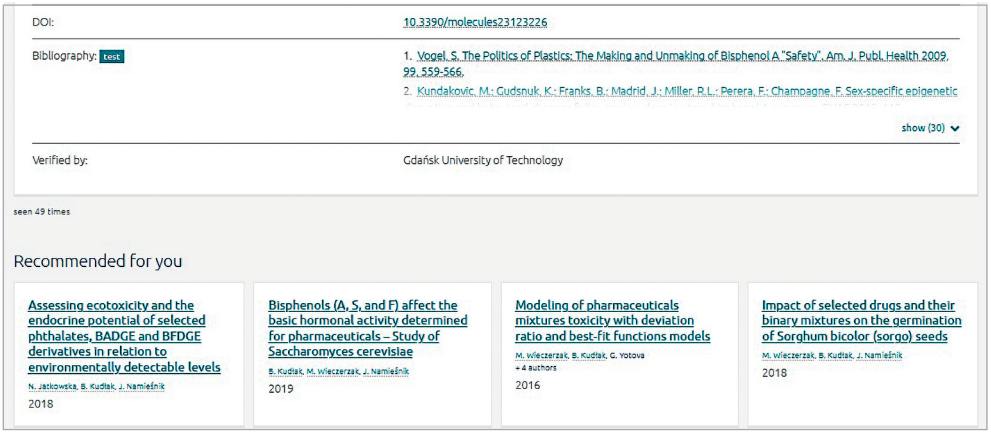


Fig. 1.5. Recommendations based on the current publication and the search history

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.

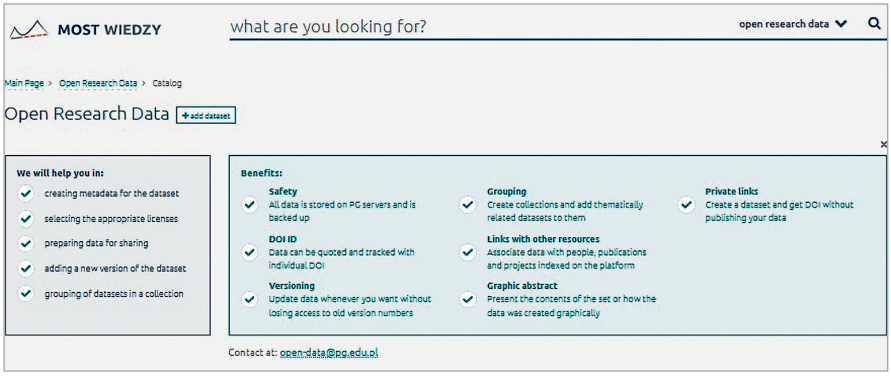


Fig. 1.6. The first screen of the Open Research Data Catalogue with hints for authors

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

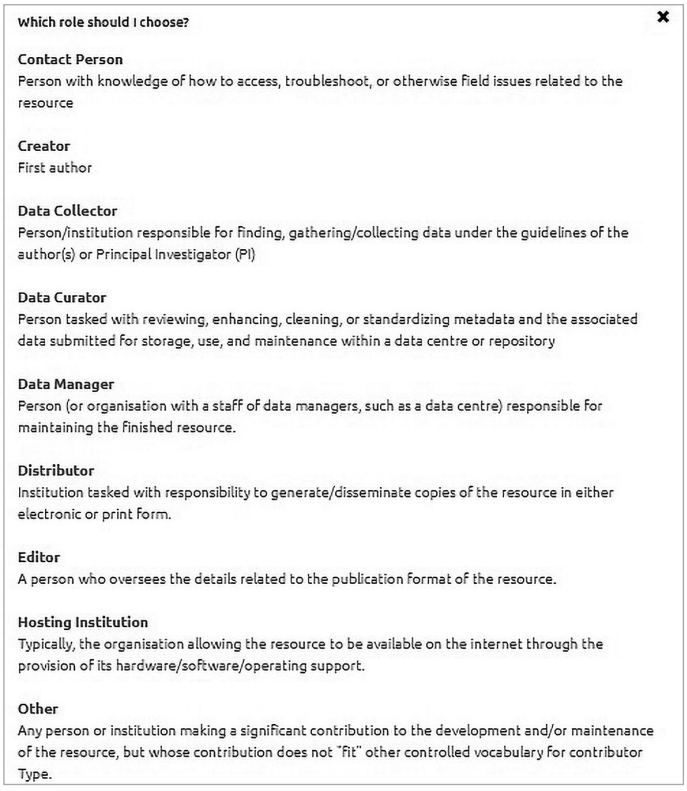


Fig. 1.8. An example of the tooltips for the roles in data creation.

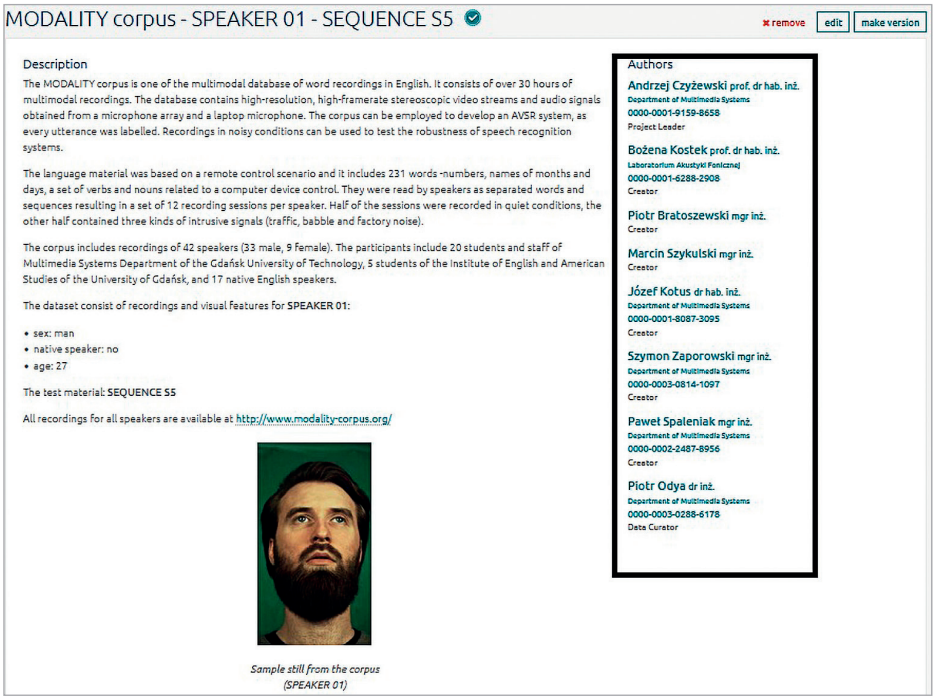


Fig. 1.9. An example of a dataset with different roles for the authors of the dataset

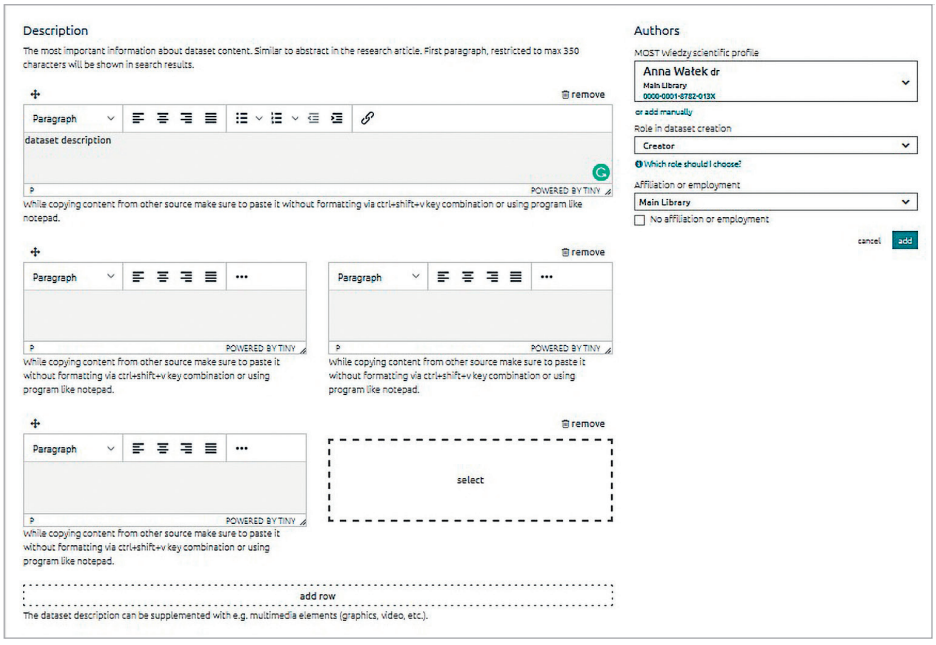


Fig. 1.10. An extensive option to add a dataset description, including a graphic description

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.

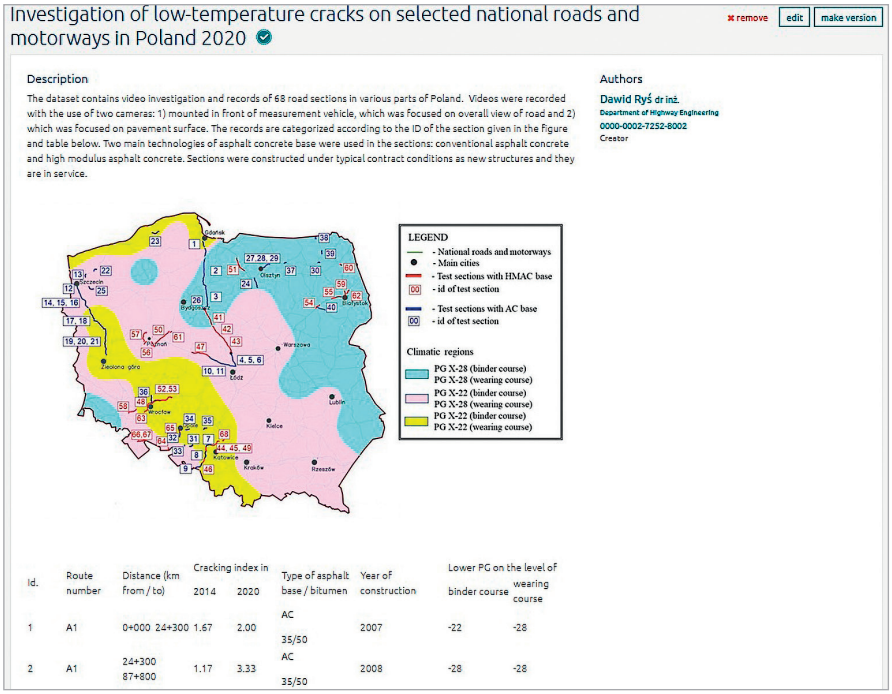


Fig. 1.11. An example of a dataset with an extensive text and graphic description

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.

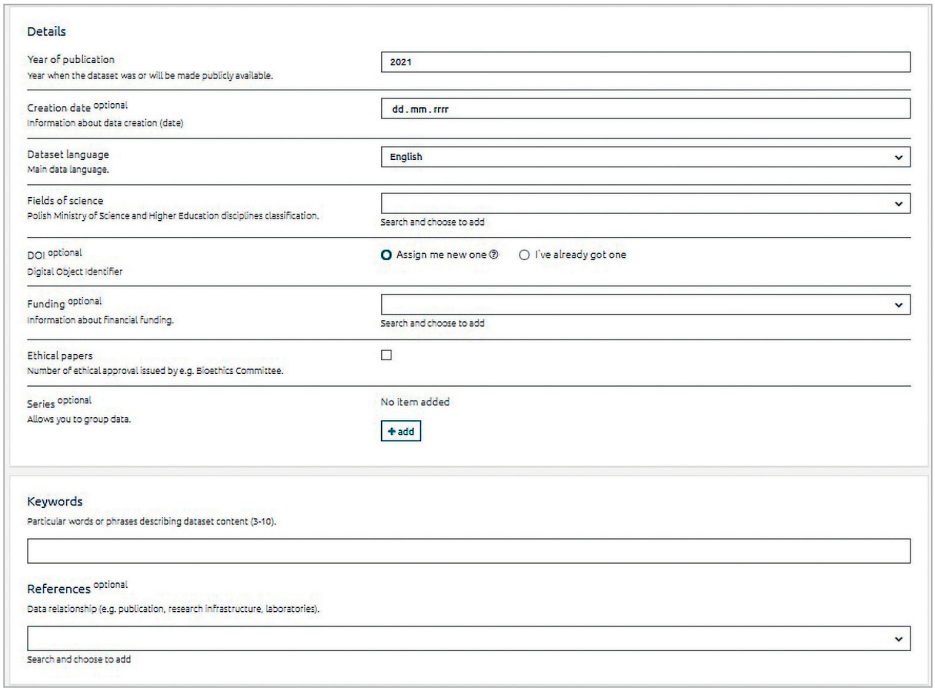


Fig. 1.12. A form for filling in metadata for the dataset

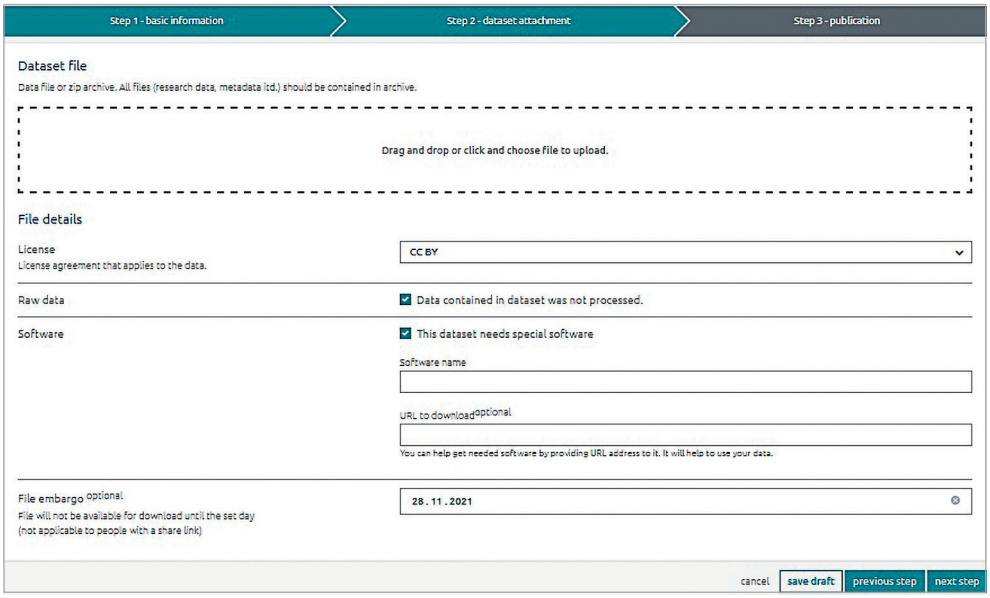


Fig. 1.13. A form for adding a data file and setting access rules

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.

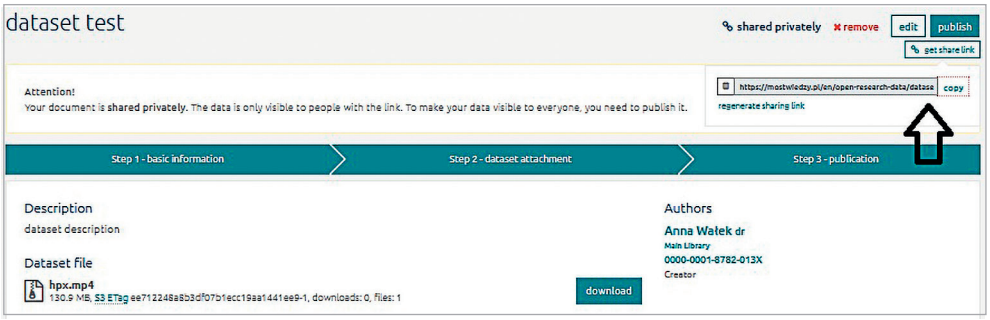


Fig. 1.14. A private link option for the dataset in draft

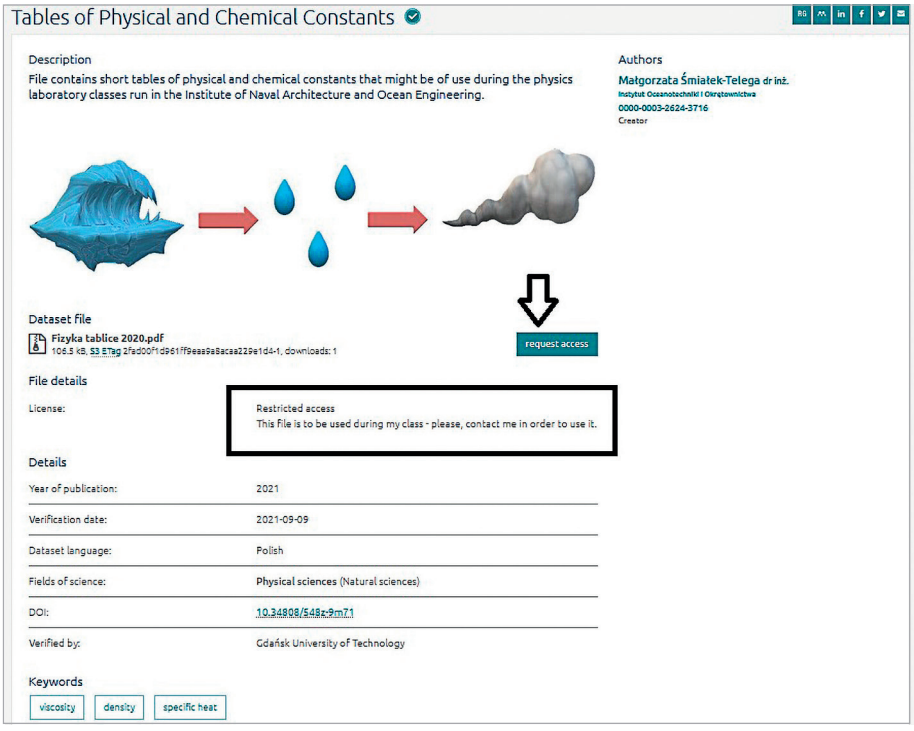


Fig. 1.15. An example of the dataset with a restricted access

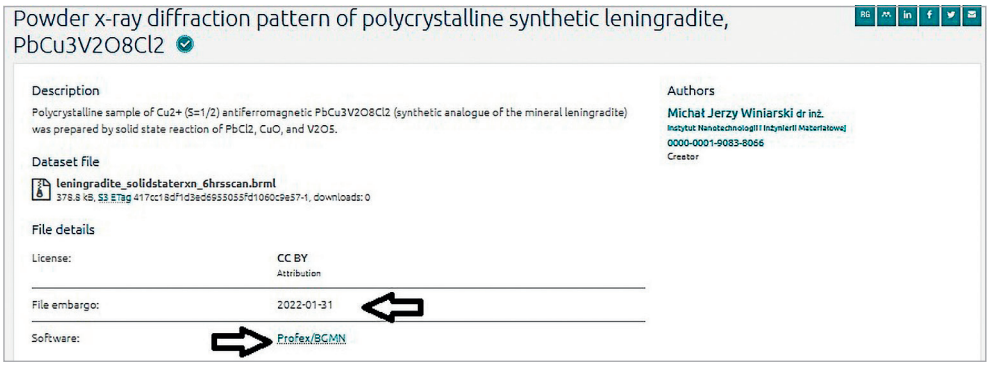


Fig. 1.16. An example of the dataset with the embargo and external software usage

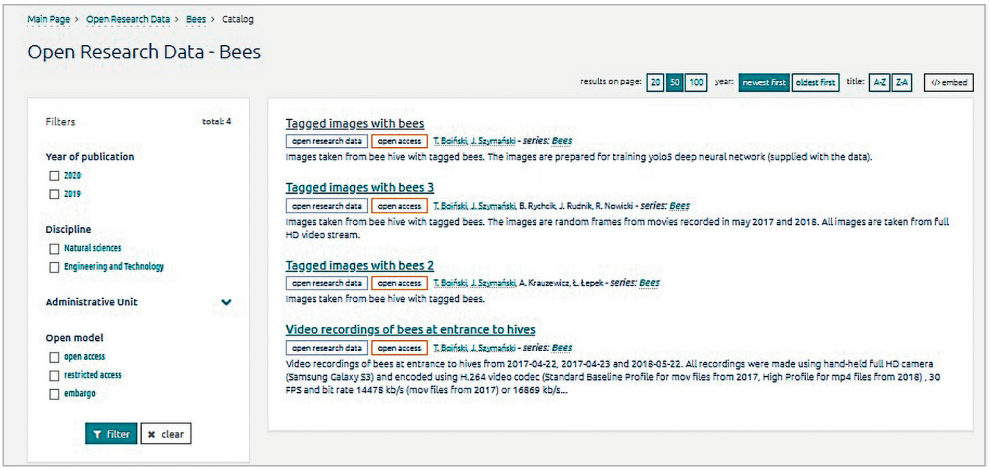


Fig. 1.17. An example of datasets grouped into series

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.

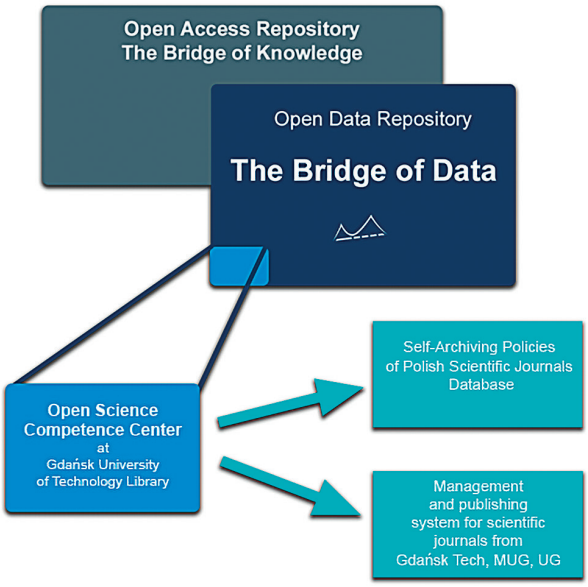


Fig. 1.18. The scheme of the Open Science Competence Center

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.

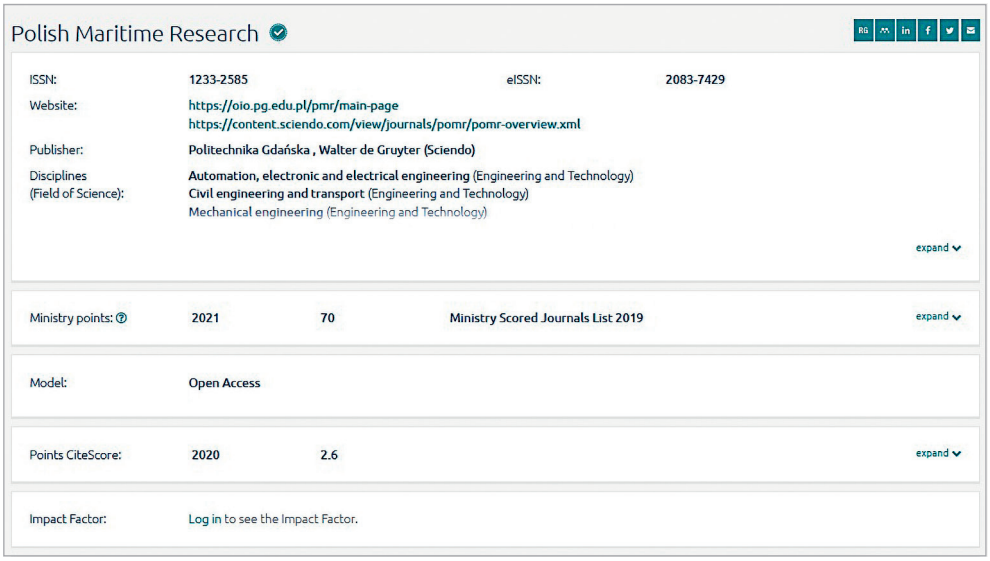


Fig. 1.19. An example of the journal's record in the policies database

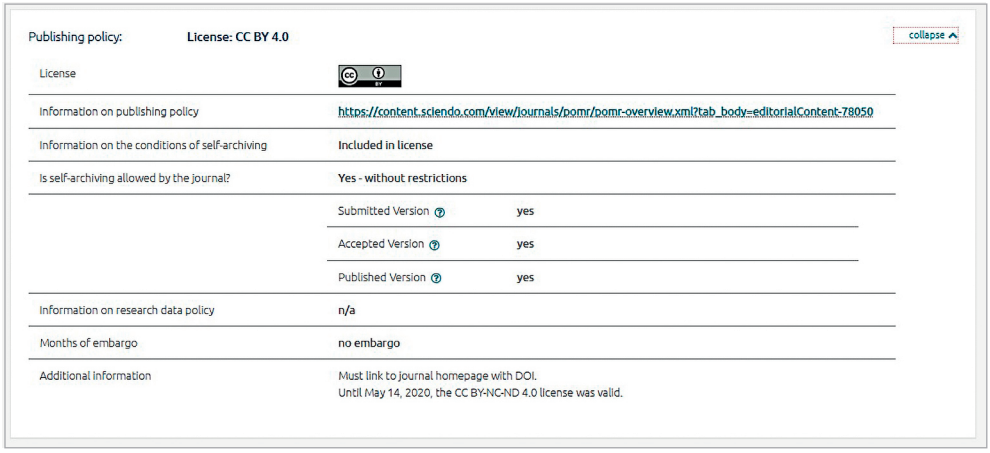


Fig. 1.20. An example of the journal publishing policy in details

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.

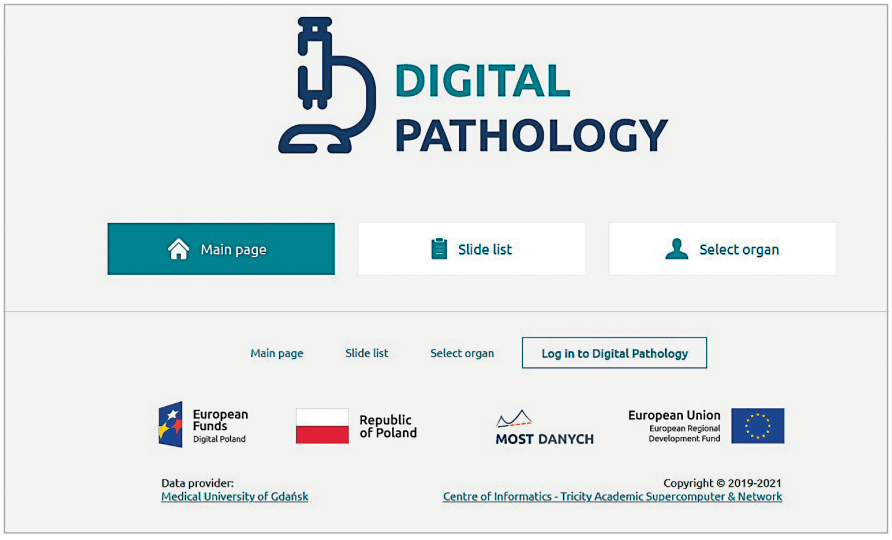


Fig. 1.21. The main page of the Virtual Microscope (digital pathology) tool

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.

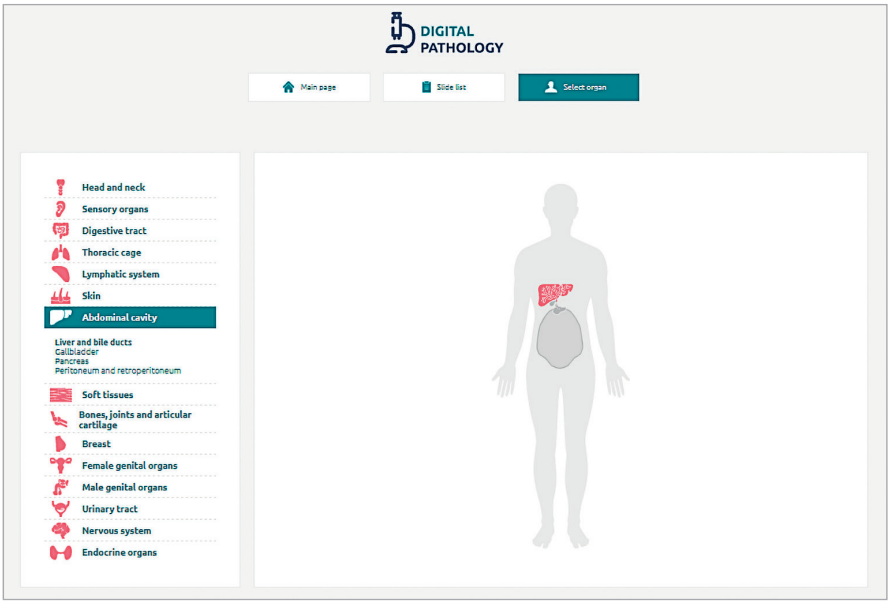


Fig. 1.22. The view of the selected organ

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.

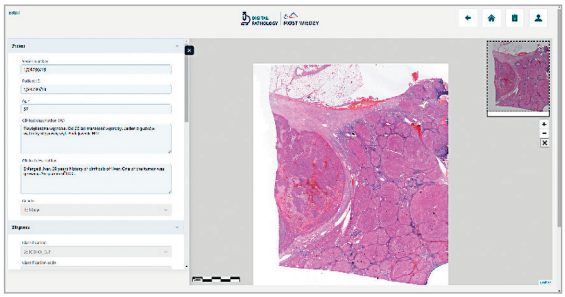


Fig. 1.23. The preview and navigation

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.

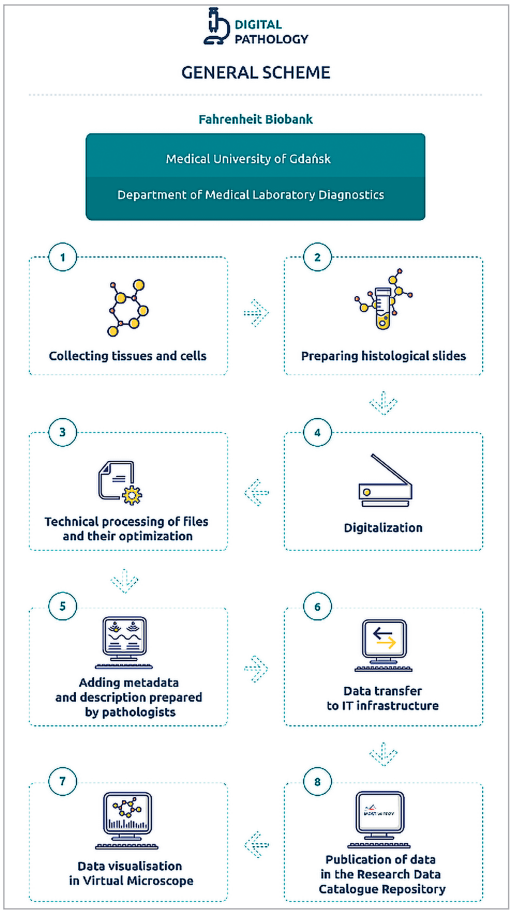


Fig. 1.24. General scheme of the Virtual Microscope – Digital Patology database workflow

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

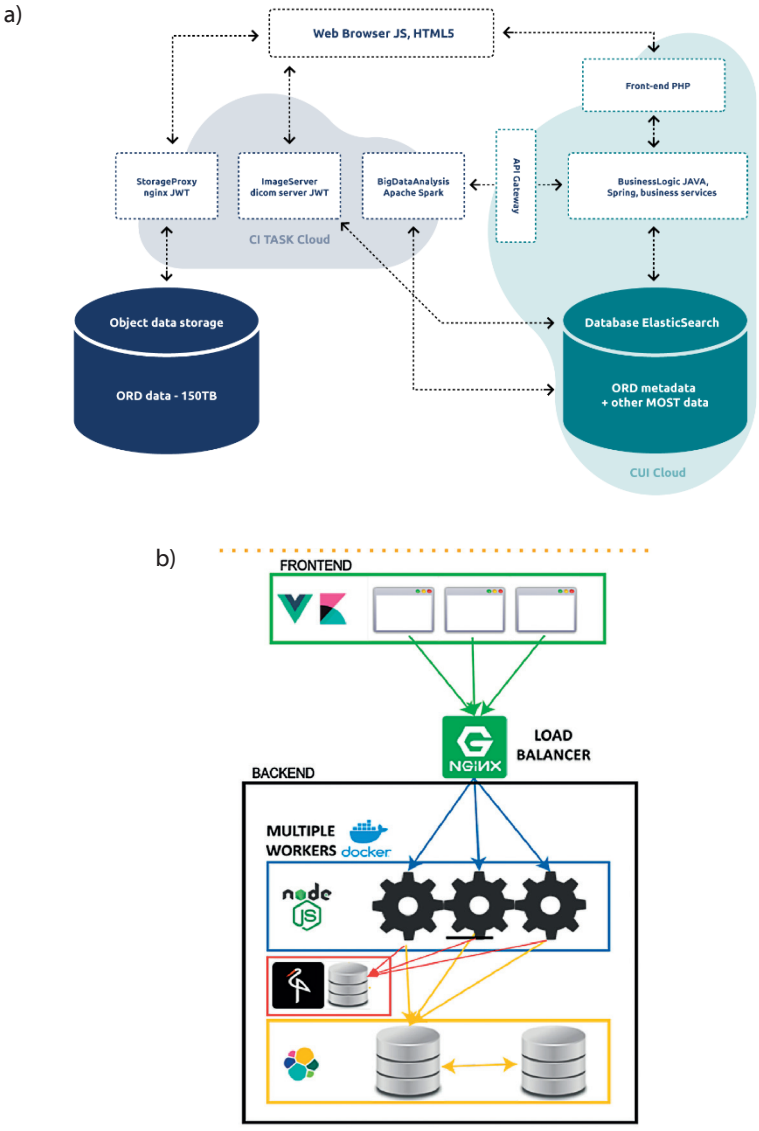


Fig. 1.25. High-level architecture of the platform

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).

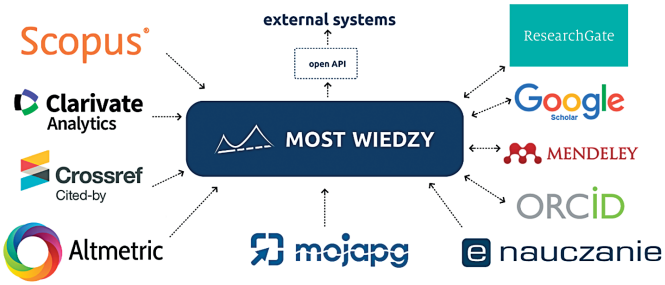


Fig. 1.26. Interoperability of Bridge of Knowledge platform

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.

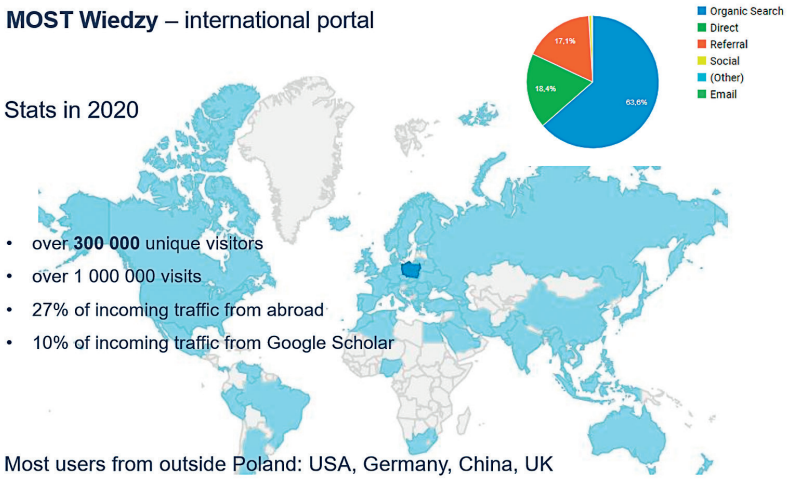


Fig. 1.27. Traffic on the Bridge of Knowledge platform

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).

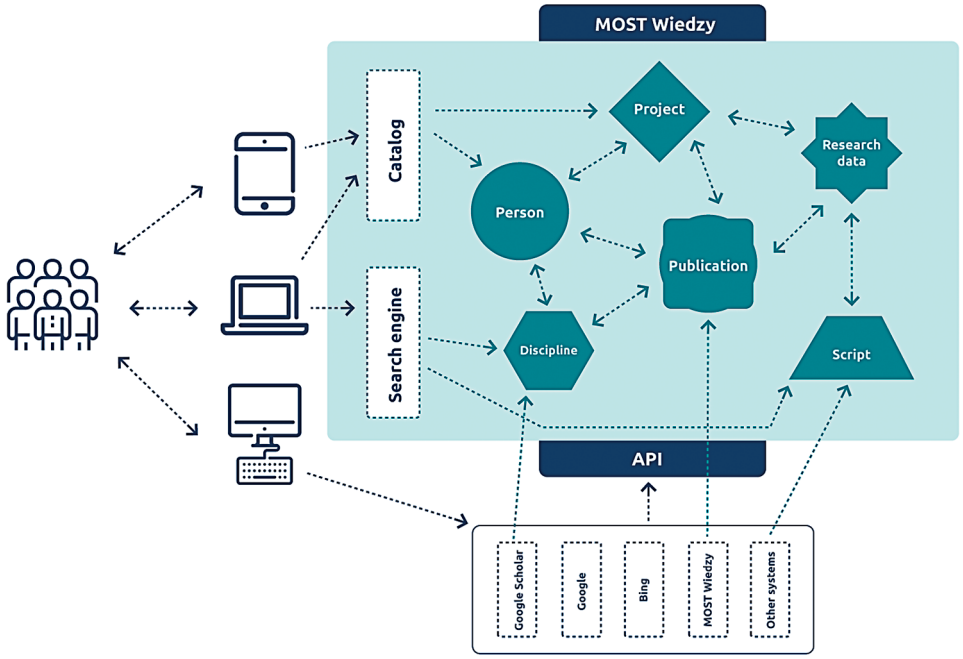


Fig. 1.28. The organisation of data on the Bridge of Knowledge platform

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.

Data Analysis in Bridge of Data

Wojciech Artichowicz[1*], Krzysztof Drypczewski[2], Jerzy Proficz[2]

[1] Faculty of Civil and Environmental Engineering, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)
[2] Center of Informatics Tri-Citry Academic Supercomputer and network (CI TASK), Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: wojciech.artichowicz@pg.edu.pl; ORCID: 0000-0002-7582-1052

Abstract

The chapter presents the data analysis aspects of the Bridge of Data project. The software framework used, Jupyter, and its configuration are presented. The solution's architecture, including the TRYTON supercomputer as the underlying infrastructure, is described. The use case templates provided by the Stat-reducer application are presented, including data analysis related to spatial points' cloud-, audio- and wind-related research.

Keywords: Jupyter, Kubernetes, Bridge of Data, Big Data, data analysis, interactive computing, Stat-reducer

https://doi.org/10.34808/x55q-sz53_dyr_roz2

Introduction

One of the main goals of the Bridge of Data project is to provide a repository filled with research data complying with an Open Access policy. The project is being implemented by three Gdańsk universities: Gdańsk University of Technology (leader), University of Gdańsk and Medical University of Gdańsk. These organisations perform a wide range of research and are involved in data gathering and dissemination.

The repository itself is being developed by two teams from the Gdańsk University of Technology: Centre of IT Services (CUI) and Centre of Informatics Tricity Academic Supercomputer & Network (CI TASK). The former focuses on the front-end components, such as the user interface and input data validation, and the latter focuses on the backend, including data storage and analysis.

CI TASK provides network and computes services for Pomeranian researchers, including all of the public universities in the region. These services are based on advanced ICT (Information and Communication Technology) infrastructure with the TRYTON

supercomputer (Krawczyk, Nykiel and Proficz, 2015). This supercomputer is based on the cluster architecture. It consists of about 1600 compute nodes, each containing two processors (Intel Xeon Processor E5 v3, 2.3GHz, Haswell), with 12~physical cores (24 logical, based on Hyperthreading technology) and 128 GB RAM. Its total compute power is 1.48 PFLOPS, and its total weight is over 20 metric tons – 40 racks.

The main contribution of the chapter is an overall description of data analysis tools deployed within the Bridge of Data project. It presents the proposed services and typical examples of their usage, including a description of a spatial points' cloud-, audio- and wind-related research, which shows the extended capabilities of the proposed tools. The examples are prepared and documented to be easily reused and applied for other research problems.

Moreover, technical details of the proposed solution are described, especially the Jupyter platform dedicated for easy-to-use graphical user interface (GUI) based on the web technologies. This open-source solution supports several programming languages, including Python, Scala and R and provides an easy access to several important data analysis frameworks such as Apache Spark.

The data analysis within the scope of the Bridge of Data project is performed using TRYTON, within its separate (sub-)system managed by the OpenStack (Rosado and Bernardino, 2014) and Kubernetes (Bernstein, 2014) cloud platforms. Access to the computing power is provided by the Jupyter computing platform (Perkel, 2018), which provides a frontend console for researchers based on Python and other scripting languages. Support for Big Data processing, including MapReduce and beyond, is provided by the Apache Spark (Zaharia et al., 2016) analytics engine. The data used in this solution can be fetched from the Bridge of Knowledge repository or provided from external sources. Within the project's scope, Jupyter was enriched with template use cases – the Stat-reducer application, available for the users as open-source.

The following section presents the Jupyter configuration used to provide services supporting the data analysis, including its architecture and deployment on the CI TASK TRYTON supercomputer. The following section presents the Stat-reducer application, including the template use cases. The document is concluded with the final remarks and possible future directions of the works.

Services for data analysis

Jupyter is an open-source computing platform that enables cooperation with various technologies (libraries, programming languages, etc.). From the user's point of view, Jupyter provides a web application that is an engine for interactive task submission, data processing, and presentation of results. The most crucial element of the Jupyter Project is Jupyter Notebook, i.e., documents that define the tasks to be performed. A notebook can consist of many elements: hypertext, source code in multiple languages (Python, R, Scala, etc.), and presentation modules with associated data (tables, pictures, charts, etc.). Jupyter provides a rich programming interface that allows Notebooks to be easily expanded with additional functionalities (Ragan-Kelley et al., 2014), (Perkel, 2018).

The individual elements of a Notebook are commands which run in an interactive REPL (read-eval-print loop) console. Then, commands are passed to the JupyterKernel, which processes the user requests (executes code, communicates with a database, sends commands to external systems, etc.) and returns the response.

The JupyterLab project is responsible for the presentation of Notebooks to the user. It provides a web application that serves as an integrated development environment (IDE), equipped with a text editor, file browser, and the ability to control the flow of task execution (Kluyver et al., 2016).

JupyterHub is a Notebook server designed to make the Jupyter environment available to users over the Internet. It allows for building a fully integrated system: JupyterLab, access to storage space (Notebooks, data scripts, etc.), the configuration of individual kernels, authorisation and authentication system, etc. so that users do not have to carry out the local installation process themselves. Due to the ability to integrate JupyterHub with container technologies, it is possible to separate the environment of individual users completely and easily scale the system.

As part of the Bridge of Data project, a Jupyter-based computing environment was launched on the TRYTON supercomputer (Krawczyk, Nykiel and Proficz, 2015) located in CI TASK. Fig. 2.1 presents the architecture of the solution.

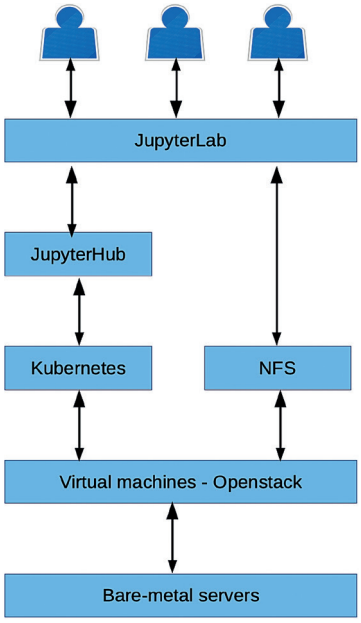


Fig. 2.1. The system architecture of the analysis system

The proposed system was launched in a private cloud managed by Openstack. In compliance with the DevOps methodology, the creation of virtual machines is fully automated with Ansible.

Containerisation is an operating system-level virtualisation technique that can isolate a user space within a single system kernel (kernel) (Boettiger, 2015). With container orchestration tools such as Docker, it is possible to deliver the application along with the entire environment (dependencies, libraries, configuration, etc.), as shown in Fig. 2.2. This approach increases application portability, enables more efficient use of resources (in comparison with virtual machines), and facilitates the use of continuous code integration and delivery (CI / CD) techniques.

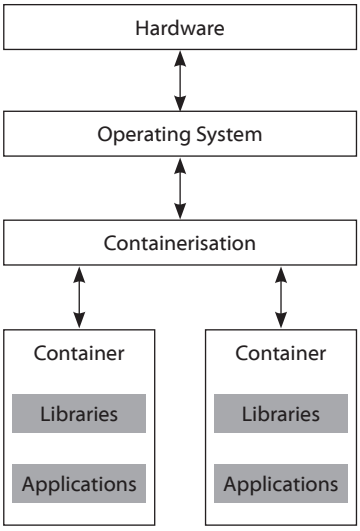


Fig. 2.2. Containerisation schema of the analysis system

Due to the high complexity of the system, its modules were separated using containers. Kubernetes (Bernstein, 2014) was selected as the container orchestration platform. It provides functionality to control the allocation of resources (per container/application/user), schedule the execution of tasks (allocation of pods to machines), adds support for network traffic between containers, application monitoring and automatic response to errors (Brewer, 2015). The computation run in Kubernetes is divided into 1) containers, 2) pods: groups of cooperating containers with a shared network, 3) services: groups of pods that create an application.

The following virtual machines were used for the Kubernetes deployment: 1) networking (i.e., ingress / egress controllers): 2 × 2 CPUs, 5 GB RAM, 20 GB SSD, 2) management: (Kubernetes services) 3 × 4 CPUs, 10 GB RAM, 40 GB SSD, 3) workers (services and pods runtime): 3 × 8 CPUs, 20 GB RAM, 80 GB SSD.

JupyterHub is deployed on the Kubernetes platform; the installation and configuration processes are performed with Ansible (automation) and Helm ("package manager" for Kubernetes). JupyterHub prepares a separate environment for each user based on a JupyterLab container. It is launched as a new pod available only for the current session

duration. To extend the functionality, a custom version of the JupyterLab installation has been prepared with the following characteristics:

JupyterKernels: Python, Scala, R, access to the standard Ubuntu Linux environment (bash, gcc, vi, etc.), possibility to install modules with pip and conda, up to 16 GB RAM per user/container, support for Apache Spark (Zaharia et al., 2016) analytics engine – deployed for Big Data processing, including map-reduce and other more complex programming models, preconfigured modules for data science and machine learning.

The data in each user's home directory is persistent, i.e. it is stored between logins. Access to the data is handled by an NFS server installed on a virtual machine, which has access to a 500 GB Ceph volume. Each user can use up to 10 GB of disk space.

User authorisation and authentication are handled by Keycloak, which provides a single sign-on (SSO) system for all CI TASK services. Thanks to the use of Keycloak, each JupyterHub instance can assign access and permissions to individual users and groups. Network traffic between users and the Notebook server is TLS encrypted and uses certificates obtained from Let's Encrypt.

The current version of the aforementioned software can be easily modified to include changes such as adding new modules and JupyterLab extensions, increasing the available computing resources (instance CPU, RAM and disk space), and scaling the Kubernetes cluster up or down.

Stat-reducer – use cases templates

The Stat-reducer functionality offered as part of the Bridge of Knowledge (MOST Wiedzy) platform is not a typical service delivered in the form of application functionality. The Stat-reducer service provides the Bridge of Knowledge platform users with a collection of Jupyter notebooks presenting complete examples of analyses of various types: from audio/video data, through environmental data, to the analysis of three-dimensional scans.

The essence of the functionality of Stat-reducer is the presentation of the capabilities of the Bridge of Knowledge platform in terms of processing and analysing data of various types:

loading and processing audio and video data, e.g. making periodograms and histograms of sound frequencies; point cloud analysis, e.g. convex hull determination and analysis of the principal components of the object; analysis and generalisation of tabular data, e.g., making histograms, determining the characteristic statistics of a dataset, etc.

These possibilities are presented in the form of Jupyter notebooks. Thanks to this, the user can easily use the code snippets from the notebooks. It can be said that Stat-reducer is documentation in the form of ready-to-use code. In the following subsections, some examples of Jupyter notebooks are provided as part of the Stat-reducer functionality.

Analysis of spatial data in the form of a point cloud

The data used in this example results from a laser scanning process of a fuel tank object. The coordinates of the points are saved in a text file in which the columns are separated by an empty character. The first step in the analysis is importing the data into memory. For this purpose, it is most convenient to use the Pandas module of the Python language, the basic object of which is a so-called data frame. This module offers the functionality of importing and exporting data from many different formats (SQL, CSV and text files, HDF5, JSON, etc.).

The second step of the analysis is the computation of the basic characteristics of the point cloud, such as the mean, standard deviation, minimum, maximum and quartiles. The easiest way to do the above calculations is to use the describe () method of the Pandas data frame.

The third step is to visualise the set and histograms of each x, y and z coordinates. For this purpose, the data frame method called hist() was used. The spatial plot of the point cloud presented in Fig. 2.3 was made with the Matplotlib library, whereas the coordinate histograms are depicted in Fig. 2.4.

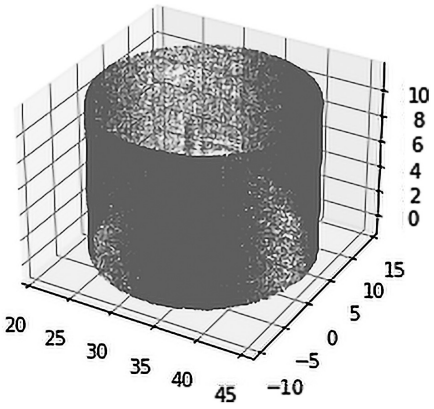


Fig. 2.3. Visualisation of the point cloud

Another issue raised in the analysis was the determination of the principal components of the set and their graphical presentation. To calculate the principal components, the data was centred, and then the PCA class of the sci-kit-learn library was used, which offers many basic machine learning algorithms. The result of the calculations is presented in Fig. 2.5.

The last element of the analysis was the creation of a convex hull of the set and then saving it in the STL format, which enables the file to be processed in software dedicated to preparing objects for 3D printing. The convex hull was determined using the ConvexHull function from the Scipy library (scipy.spatial). Convex hull export was performed using the NumPy-stl Python library. The analysis output is presented in Fig. 2.6, which shows the original set of points and its convex hull.

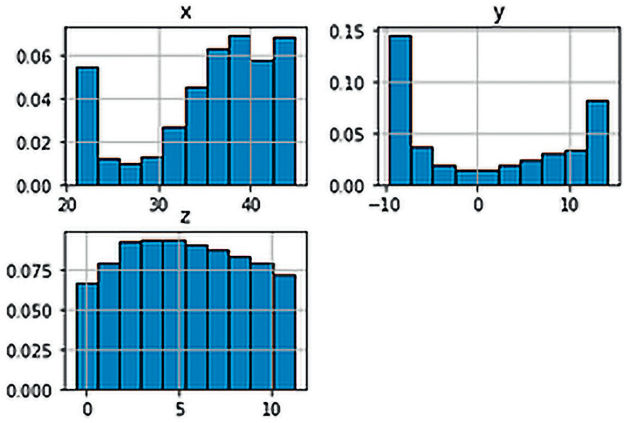


Fig. 2.4. Histograms of the point cloud coordinates

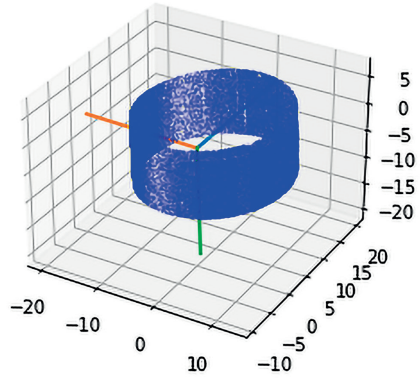


Fig. 2.5. Centred dataset (points) and its principal components (lines)

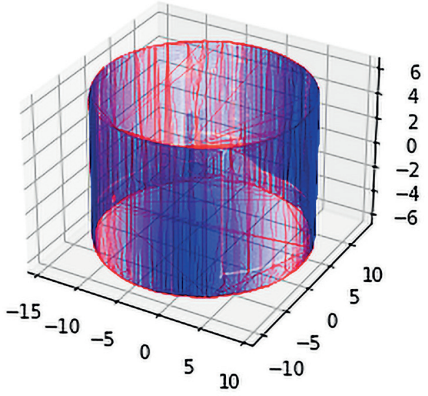


Fig. 2.6. The plot of the point cloud (blue) and its convex hull (red)

Audio analysis

The data used in this example consists of video files with sound, which contain recordings of bees flying into a hive. The focus of the analysis presented in this example is a visualisation of the audio present in the video files.

The first step in the analysis was to extract the sound from the video files and save them as .wav files. There are many possible approaches to this problem. However one of the simpler ones is to use the moviepy library, which offers the required functionality. Using the functionality of the os module of the standard Python library and the moviepy library, the sound was extracted from each of the video files and saved to disk as a WAV file. Then, the functionality of the Scipy library (scipy.io) was used to load the audio channels present in the files. Again, using the Scipy library (scipy.signal) and the Matplotlib library, a function visualising the sound in the waveform plots, periodograms and frequency histograms was developed. To make the exploratory analysis of many files convenient, an interactive widget in the form of a drop-down list was instantiated from the ipywidgets module. Each entry in the list refers to an audio file. The functionality was easily achieved by the interact function from the ipywidgets module, which was used as a decorator, to the visualising function. Finally, the created Jupyter Notebook document allows an audio file to be selected for which the abovementioned graphs are to be made. The effect of the abovementioned activities is the Jupyter Notebook document, including an interactive drop-down widget, which is presented in Fig. 2.7.

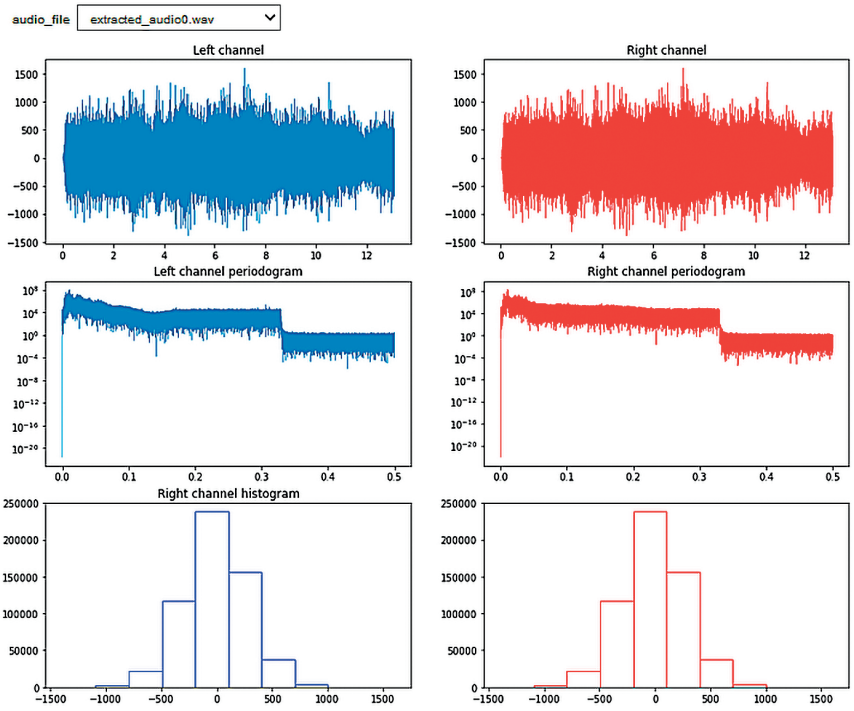


Fig. 2.7. Interactive Jupyter Notebook document for the graphical analysis of the audio files

Wind data analysis

The data used in this example relate to wind speed measurements at the Nowa Pasłęka location, made in 2008–2017 by the IMGW. The analysis presented here is a state-of-the-art analysis of the wind data.

Raw data containing information about the date and time of measurement, wind speed and wind direction is stored in an MS Excel© file. The most convenient way to handle the data is to read it into a Pandas data frame using the read_excel() function.

After reading the data into the data frame, it is necessary to check whether the data contains incorrect values. Such functionality is provided by the data frame object in the form of the isna() method, which identifies rows with incorrect values. Deleting rows containing incorrect data can be done using the Pandas data frame method called dropna().

After cleaning the data of incorrect values, a plot of the wind speed and direction is usually made. For this purpose, the Matplotlib library was used. The wind time-series is shown in Fig. 2.8.

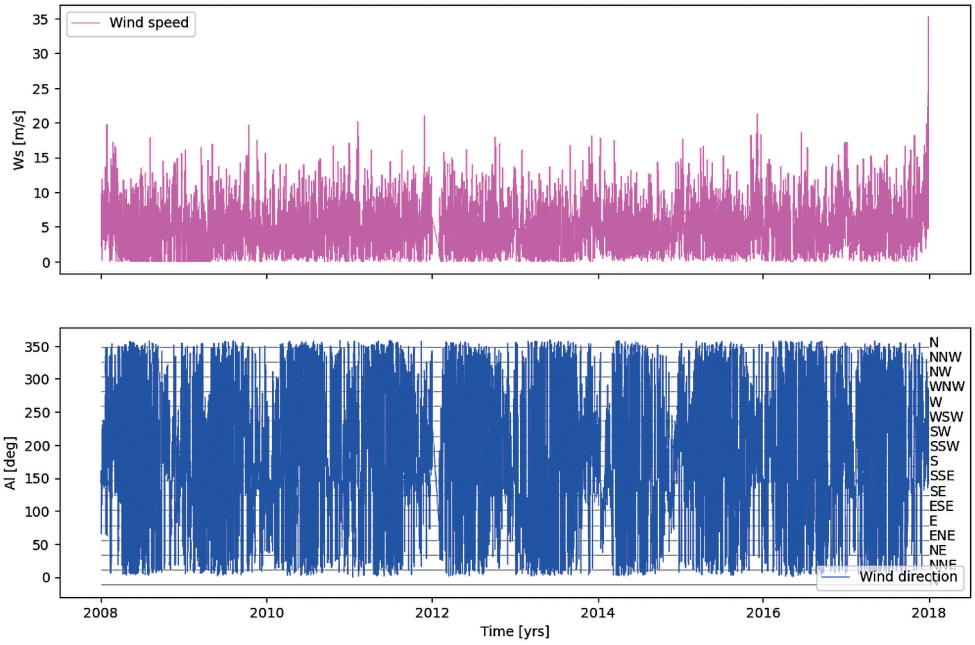


Fig. 2.8. The plot of the wind speed and direction

To generalise the data, the basic statistics of the dataset should be calculated: minimum, maximum, quartiles, mean and standard deviation. The most convenient way to do this is to use the describe() method of the Pandas data frame.

A complete wind data analysis requires a wind speed and direction histogram and an estimation of the parameters of the Weibull distribution applied to the wind speed data. The wind speed histogram can be made using the hist() function from the Matplotlib library (Fig. 2.9).

Wind analysis uses a form of the Weibull distribution other than that available in the Scipy.stats library. Therefore, the most convenient solution would be to use and adapt the rv_continuous object representing the continuous distribution offered by this library. For this purpose, a new distribution object inheriting from rv_continuous is created, containing the required parameter set. The created object represents the Weibull distribution given by the following probability density function formula:

The distribution parameters (,) were determined by the maximum likelihood method. Following this method, a nonlinear equation with respect to the parameter was obtained, which was solved using the newton function from the Scipy.optimize library. The density of the estimated Weibull distribution is shown in Fig. 2.9 with the blue line.

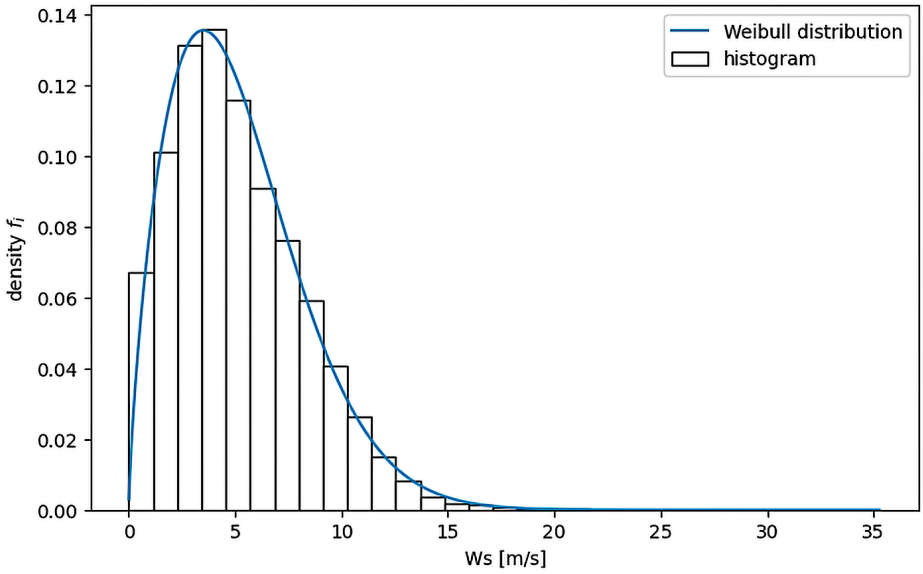


Fig. 2.9. Wind velocity histogram and estimated Weibull distribution density

The last element of wind data analysis is the plot of the so-called wind rose, i.e. a chart that is a histogram of wind speed and direction. Such a plot can be made using the windrose library for Python. The wind rose for the analysed data is presented in Fig. 2.10.

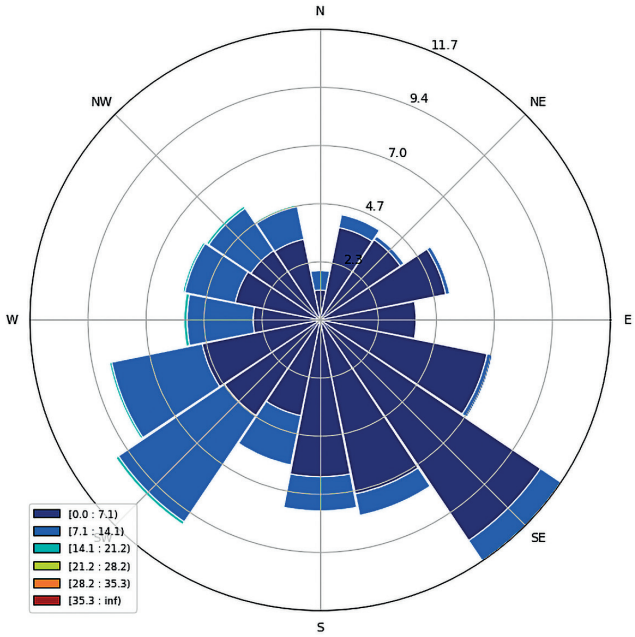


Fig. 2.10. The wind rose, of example data

Final remarks

Data analysis is an important part of the Bridge of Knowledge platform functionality. It is performed using CI TASK compute infrastructure: TRYTON supercomputer (Krawczyk, Nykiel and Proficz, 2015). The provided solution is based on the Jupyter framework deployed within OpenStack (Rosado and Bernardino, 2014) and Kubernetes (Bernstein, 2014) platforms. Jupyter supports researchers as a compute frontend, providing a set of services based on Python and other scripting languages. The framework was widely tested and enriched with a specially developed Stat-reducer package containing useful use case templates.

Future works within the scope of the project and during its maintenance period will cover the following activities: supporting the researchers with new packages dedicated to the framework, maintaining the high performance and scalability of the platforms, providing more use cases and templates.

We believe that Open Research Data analysis supported by the Bridge of Data project will help scientists perform better research work.

References

Bernstein, D. (2014) 'Containers and cloud: From lxc to docker to kubernetes', *IEEE Cloud Computing*, 1(3), pp. 81–84. DOI: 10.1109/MCC.2014.51.

Boettiger, C. (2015) 'An introduction to Docker for reproducible research'. *ACM SIGOPS Operating Systems Review*, 49(1), pp. 71–79. DOI: 10.1145/2723872.2723882.

Brewer, E. A. (2015) 'Kubernetes and the path to cloud native'. *Proceedings of the sixth ACM symposium on cloud computing*, pp. 167–167. DOI: 10.1145/2806777.2809955.

Krawczyk, H., Nykiel, M. and Proficz, J. (2015) 'TRYTON Supercomputer Capabilities for Analysis of Massive Data Streams', *Polish Maritime Research*, 22, 3 (87), pp. 99–104, DOI: 10.1515/pomr-2015-0062.

Kluyver, T. et al. (2016) 'Jupyter Notebooks-a publishing format for reproducible computational workflows', in Loizides, F., Schmidt, B. (ed.) *Positioning and Power in Academic Publishing: Players, Agents and Agendas. ELPUB*, pp. 87-90. DOI: 10.3233/978-1-61499-649-1-87.

Perkel, J.M. (2018) 'Why Jupyter is data scientists' computational notebook of choice', *Nature*, 563 (7732), pp. 145–146. DOI: 10.1038/d41586-018-07196-1.

Ragan-Kelley, M. et al. (2014) 'The Jupyter/IPython architecture: a unified view of computational research, from interactive exploration to communication and publication', *AGU Fall Meeting Abstracts*, pp. H44D-07. Available at: https://agu.confex.com/agu/fm14/meetingapp.cgi/Paper/27181 (Accessed: 8th June 2021).

Rosado, T. and Bernardino, J., (2014) 'An overview of openstack architecture', *Proceedings of the 18th International Database Engineering & Applications Symposium*, pp. 366–367. DOI: 10.1145/2628194.2628195.

Zaharia, M. et al. (2016). Apache spark: a unified engine for big data processing, *Communications of the ACM*, 59(11), pp. 56–65. DOI: 10.1145/2934664.

2

Part

Data Management Practices

The Digital Tissue and Cell Atlas and the Virtual Microscope

Jarosław Skokowski[1*], Marika Bolcewicz[1], Kamila Jendernalik[1], Thierry van de Wetering[1], Jacek Gulczyński[1], Anna Lewandowska[1], Leszek Kalinowski[1]

[1] Department of Medical Laboratory Diagnostics, Faculty of Pharmacy, Medical University of Gdańsk (3a Marii Skłodowskiej-Curie Street, Gdańsk, Poland)

* Correspondence author: jaroslaw.skokowski@gumed.edu.pl; ORCID: 0000-0002-3079-3502

Abstract

With the cooperation of the CI TASK (Center of lnformatics Tri-Citry Academic Supercomputer and network) and the Gdańsk University of Technology, the Medical University of Gdańsk undertook the creation of the Digital Tissue and Cell Atlas and the Virtual Microscope for the needs of the Bridge of Data project. In the beginning, an extensive collection of histological and cytological slides was carefully selected and prepared by pathomorphology experts. After processing and digitising, the specimens were sent to servers of the TRYTON Supercomputer, where storing, searching for and scrolling through the images in the Virtual Microscope was made available. The collection consists of twenty thousand high definition images of human tissues and cells accompanied by structurised clinical metadata. Creating a digital atlas and a virtual microscope is an answer to modern education challenges that shape digital competence and are open to modern technologies. The main idea behind the creation of the information tools and digital image data repositories is using them for the purpose of education and as a basis for the creation of new methods of long-distance education. Those resources are shared under the terms of the open Creative Commons license (CC BY-SA), making it possible for teachers, students, and entrepreneurs to use the images safely and process content included in the presented materials without intellectual property infringement.

Keywords: whole slide image (WSI), virtual microscopy, digital atlas of histology, digital microscopy, digital pathology, virtual pathology slides

https://doi.org/10.34808/x55q-sz53_dyr_roz3

Introduction

The modern world is changing rapidly, at a pace never observed before; the time of the pandemic accelerated the pace of those changes, especially in the field of services and telecommunication networks. Restrictions in mobility on the one hand and the process of facilitating telecommunication, on the other hand, caused humanity to swiftly move to the digital reality. On a daily basis, people make use of advanced information technology that they use for remote work or education. Broadband Internet makes it possible to quickly search for and scroll through huge and diverse databases, while cloud solutions allow for scrolling through them on a PC, laptop and a smartphone, even in places where the Internet connection is weak.

The necessity of distance learning, the value of which was verified by the COVID-19 pandemic, is slowly becoming the new paradigm of education. However, a dynamic development of new technologies and the development of science alone shows that knowledge presented to students in some fields has already become outdated when it is being taught. For that reason, it became necessary to support education through allowing for easy and free access to constantly updated and credible digital education resources. Moreover, sharing new technologies and interesting sources of information with the teachers increases their competence and serves as a chance to prepare further generations to function in a digitised world (Koczy, 2020).

The amount and diversity of open data also encourages individuals to search through the Internet on their own and creates the conditions for self-studying. The race in sharing data and encouraging potential users to access created information platforms motivates scientific communities to share their data sources in a way that is attractive and convenient for the user.

For the purposes of the project Bridge of Data – Multidisciplinary Open System Transferring Knowledge – stage II: Open Research Data, the team from Medical University of Gdańsk undertook the creation of the Digital Tissue and Cell Atlas and the Virtual Microscope with the cooperation of the CI TASK IT Centre and the Gdańsk University of Technology (Gdańsk Tech). Carefully selected histological and cytological specimens were gathered by experts in the field of pathomorphology (Greater Poland Cancer Centre, "Copernicus Sp. z o.o.", Medical University of Gdańsk). Specimens were processed, stained and scanned in the MUG's Department of Medical Laboratory Diagnostics (DMLD) and after digitisation, the images were sent to the servers of the TRYTON supercomputer where storing high definition images was made possible. The created repository of Open Research Data (ORD) consists of structurised metadata describing twenty thousand images that can be viewed in the Virtual Microscope. Creating the Digital Tissue and Cell Atlas, along with the Virtual Microscope is the answer to the challenges of modern education that shapes digital competence and is open to modern technologies.

Processing, staining, digitisation

Biological material is delivered to MUG's DMLD as tissue that is not preserved, paraffin tissue blocks or histological specimens. Depending on the type of the received bi-

ological material, particular procedures are implemented in order to obtain high quality digitised images. After being put in histology cassettes, tissue retrieved directly from the patient needs to undergo the process of tissue preservation which, according to the standards, takes anywhere from 24 to 48 hours. The preservation process alone serves to stop the autolysis, which is autodigestion by enzymes present within the cell, and the metabolic processes, thus allowing for the natural structure of the cell to be preserved. A 4% formaldehyde water solution (10% neutral buffered formalin) is a universal preservative that stabilises the structure of the cell by creating methyl bonds (Litwin and Gajda, 2011).

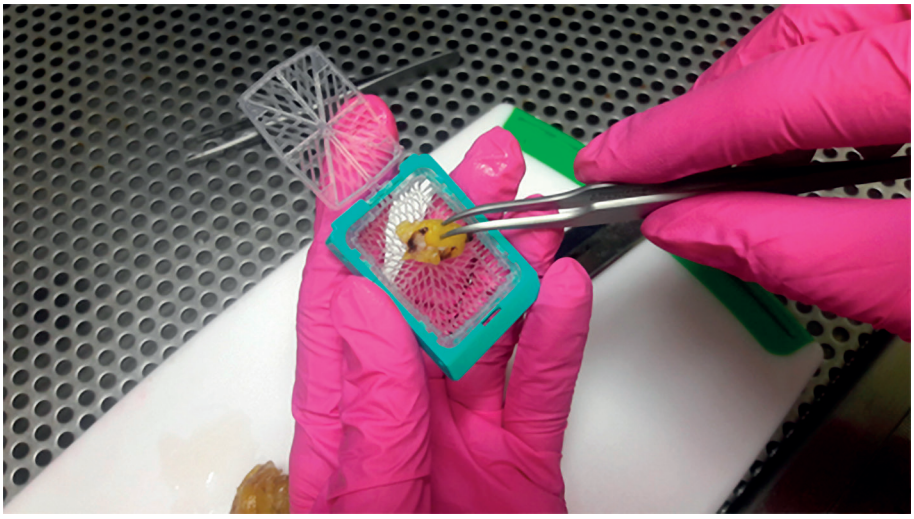


Fig. 3.1. Inserting the tissue into a histological cassette. Kreft, 2020

The tissue prepared in the process described in the previous paragraphs undergoes further steps which are shown in Tab. 3.1. Firstly, the material is processed with the help of an automatised system. The preservative and water are removed from the tissue under vacuum and then the tissue is saturated with liquid paraffin at high temperature. Thanks to modern technology, this process has been reduced from several to merely around 2 hours. Then the tissue is submerged in special cassettes in liquid paraffin which hardens at low temperature. That is how paraffin blocks are created; during this process, the material gains homogeneity and a sufficient degree of hardness which is especially important during further processing (Litwin and Gajda, 2011).

The obtained strips are put in a water bath where the tissue, as well as the paraffin surrounding the tissue, is decompressed and smoothed out, which allows the strip to be fixed on a slide that is later annealed in an incubator set to 60°C. After this step is completed, the slide is ready to be stained by implementing one of the histological staining techniques used in the laboratory (Litwin and Gajda, 2011).

Staining histological slides is a necessary process that makes it possible for the particular cellular structures to be contrasted so that the image achieved under the micro-

scope is legible and unambiguous. In order to achieve that, the slides are stained according to one or a few techniques that will make the specific diagnostic characteristics of the collected tissues visible. The principal technique of positive staining is the hematoxylin and eosin stain (H&E). In this method, the basic hematoxylin dye shows affinity with basophilic structures, such as the nuclei, and stains them blue, while the acidic eosin binds with acidic structures, e.g. cytoplasm, and stains them red. This type of staining provides a general overview of the structure of tissues and cells (Sawicki and Malejczyk, 2012).

After performing H&E staining, additional staining techniques are implemented in order to make specific cellular structures (those in the centre of the attention of a certain researcher) visible. Histochemical staining allows for the identification and localisation of certain chemical substances and biochemical processes taking place in cells and tissues. Substances detected by using this histological technique are for example carbohydrates, lipids, amino acids, proteins and nucleic acids (Sawicki and Malejczyk, 2012).

Methods of molecular biology, such as immunohistochemistry (IHC), immunofluorescence (IF) and a cytogenetic method of fluorescence in situ hybridisation (FISH), are also used in histology (Lewandowska-Ronnegreen, 2018).

The immunohistochemic method makes use of antibodies that bind with antigens present in the tissues. This type of binding is visualised thanks to a coloured detection system directed against the antigen-antibody complex that is created. The result of a positive reaction that confirms the presence of the antigen is brown color visible on the histological slide (Lewandowska-Ronnegreen, 2018).

Just like immunohistochemistry, immunofluorescence also relies on the reaction of the antibody with the antigen. The fundamental difference is a different detection system; antibodies are marked with fluorescent dye and the visual effect of the antigen-antibody complex is glowing in a particular range of the light spectrum (Lewandowska-Ronnegreen, 2018).

In contrast to the IHC method, the fluorescence in situ hybridisation measures the amount of a certain protein; it determines the gene amplification level, namely the number of copies of the gene present in the observed cells. This method relies on the use of gene-complementary probes and the detection system contains probes that fluorochromes are bound to. After a positive reaction takes place, the fluorescent microscope is used to count the signals whose colour depends on the type of fluorochrome used (Lewandowska-Ronnegreen, 2018).

Histological slides stained in the process of the previously presented techniques for the purposes of the Bridge of Data project are later digitised in order to be added to the repository of Open Research Data (Most Wiedzy – The Bridge of Knowledge Research Data Catalogue). During this process, scanners dedicated to this particular goal are used and this creates a digital version of the histological slides. DMDL is in possession of a highly advanced Pannoramic 250 scanner (3DHistech) which combines optical microscope and digital camera functions. The native format of the obtained images is MRXS which, for the purposes of the project, is converted to the universal DICOM format widely used in modern biomedicine while the high resolution of the images is maintained. The Pannoramic 250 camera captures the images of all histological slides with the use of

two available lenses with magnification factors of 20x and 40x and the autofocus function. The device has different types of lighting and diverse observation methods. It allows for scanning microscope positive stain slides in a bright field and additional filters make it possible to capture fluorescent images as well (www.3dhistech.com).

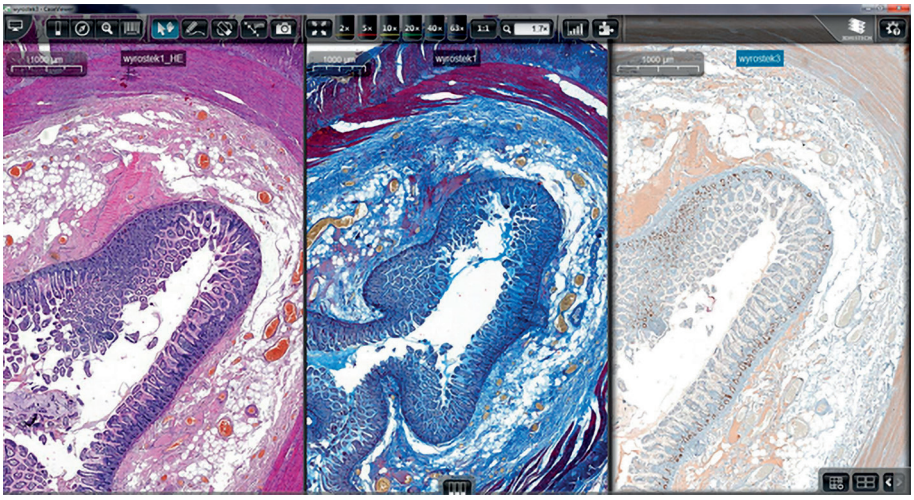


Fig. 3.2. Summary of histology staining techniques. From the left: H&E, HC (Masson's Trichrome), IHC (antigen KI-67). Bolcewicz, 2021

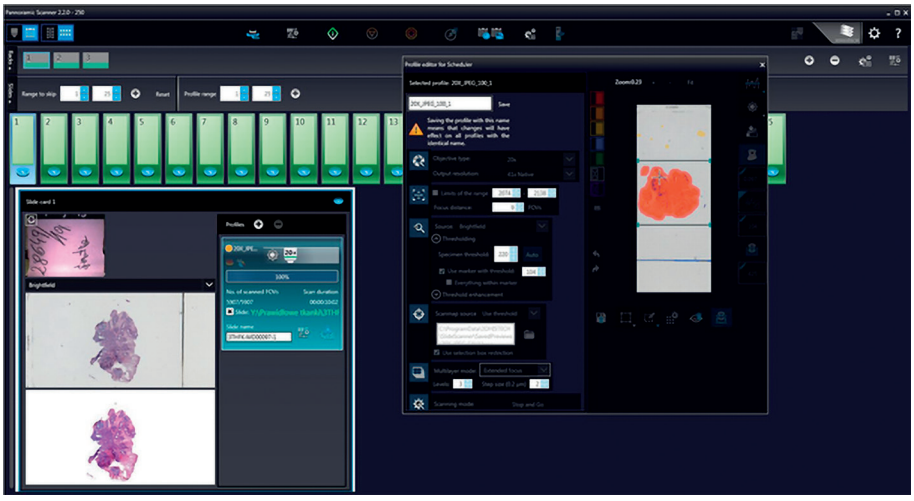


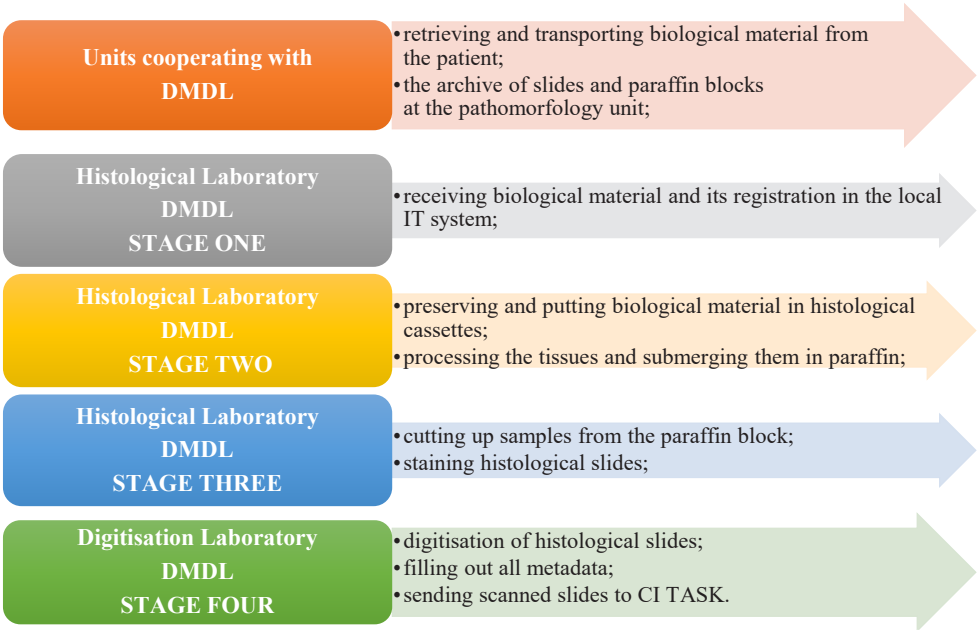
Fig. 3.3. Software supporting the Pannoramic 250 3DHistech histological scanner. Kreft, 2021

Software dedicated to support the scanner is equipped with multifunctional scanning modes that bring out the depth of the tissue, making observation of the image seem like looking at a slide under a light microscope. Additional tools to analyse and process obtained

scans allow for a preliminary assessment to be made according to the standards by the DMDL staff in order to ensure the highest possible quality of images (www.3dhistech.com).

Tab. 3.1

Flowchart of managing biological material in the Bridge of Data project at DMDL



Metadata, sending data to CI TASK

The main goal of the DMDL as a participant of the Bridge of Data project is to create a vast database of image data numbering twenty thousand digitised histological slides along with their detailed descriptions that are jointly called metadata. Metadata constitutes a vital element of every database, because it provides information and describes attributes of particular records. It applies both to technical and research data, allowing for its later identification and further use. However, using metadata requires the preparation of a special key that makes it easier to describe the images. Because of that, the team from the Medical University of Gdańsk created dictionaries with definitions of particular parameters of histological slides and their digital record. They contain attributes assigned to categories that additionally bear an alphanumerical code which is assigned in order to facilitate the procedure of sending data to CI TASK (a project partner).

Metadata used for the project goals was divided into five main categories and contains information such as:

- data concerning the patient (age, sex, the year of a study and a clinical description); information concerning the diagnosis according to the International Classification of Diseases (ICD) and a histological description of the slide, if there are abnormal

images present; histological data concerned with the digitisation process – the type of histological scanner, the lens magnification used and the format of the image; information about the histological slide – the type of the tissue and the material, the way it was collected and the type of staining; additional organisational data – names of cooperating units, the internal number given by the DMDL.

Storing and managing all of the information connected with the digitised resources is possible thanks to a local IT system. After registering the records, a special file with the "yml" extension is generated automatically and contains all metadata essential to describe the digital slides. Packets of data and high definition images are sent with the help of a modern IT infrastructure to CI TASK. A high channel capacity connection (with a speed of 10 Gigabytes per second) allows the images and elements of data to be sent directly to the data centre storage where the created database (after the specialists' approval) will be published in the system of Open Research Data of the Bridge of Data project.

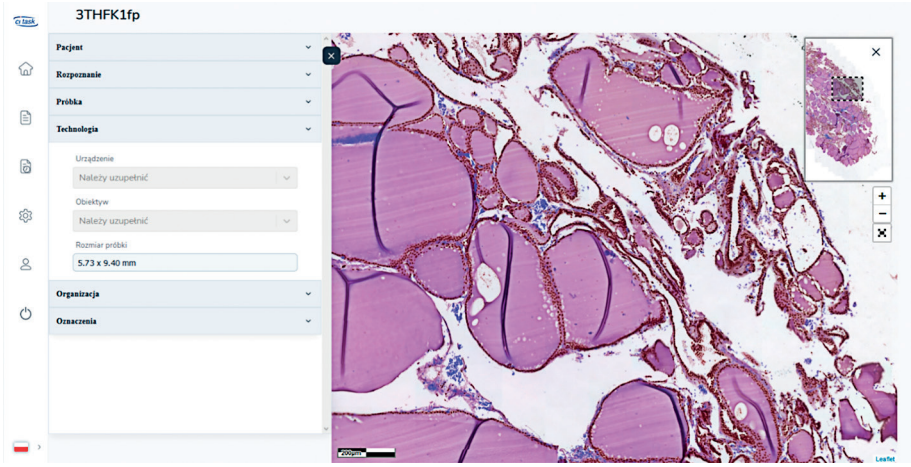


Fig. 3.4. View of thyroid tissue in the Virtual Microscope. Van de Wetering, 2020

Sharing research data – microscopic images

Just like at the beginning of the XIX century, when the development of the optic microscopy and the appearance of histochemical stains complementing the primary H&E stain (hematoxylin and eosin) took place, at the turn of the XXI Century, microscope images started to appear in digitised form. The term "digital pathology" has permanently entered the scientific language for good and the number of publications containing just those two words is steadily rising (around two thousand occurrences in the years 1995–2000 and over eight thousand occurrences during the last 5 years, according to the Pubmed database) (Dee and Meyerholz, 2007). According to the definition, this field of science is concerned with creating a digitised version (digitisation) of a real histological slide and saving it on a physical medium or in a cloud, as well as sharing it in order to analyse or interpret those slides. Whole slide imagining allows for an assessment of the

microscopic image at the place where it was created and consulting any other establishment at the same time which makes it possible to make the final diagnosis in a really short time (Farahani, Parwani and Pantanowitz, 2015). Quick and easy access to those types of images is also extremely important. Additionally they can be connected to LIS (Laboratory Information Systems) where they supply other clinical data of the patients.

The didactic aspect of databases containing digitised microscopic images (both histological and histopathological) is also vital. They can serve as a source of knowledge for the students, medical residents and specialist doctors. Especially those mentioned last can access high definition images of rare cases when paper sources cannot provide images with such resolution parameters. Creating databases in various establishments can also be a starting point for interestablishment research to take place when a pooled analysis of rare cases (occurring only as few or several cases in every establishment) can take place. It helps not only with diagnostics, but also didactics. Of course the important issue of setting threshold parameters arises, if it comes to sharing data, such as the definition of the image, the quality of the scanner and its lenses, as well as the digitised version of the output files, which can constitute a kind of barrier for certain establishments. On the other hand, it enforces actions whose goal is to enhance the quality of work of a digital pathology lab (definition, the number of layers depending on the thickness of the slide, colour scale, output file format, level, degree and image compression algorithm). Similarly not every establishment will be able to purchase a scanner that will make it possible to digitise specimens with non-standard measurements of the microscopic slide. Standardisation and the quality of the microscopic slides alone, as well as routine staining and special staining are also important, but the topic is beyond the scope of this work (Prochorec-Sobieszek, 2016).

Commercial companies concerned with the development of AI (artificial intelligence) will also be allowed to use the created image databases. On this basis, they can develop algorithms for the assessment of digital images which, in the future, will assist pathologists during diagnoses (Prochorec-Sobieszek, 2016).

After evaluating progress in the field of pathology, we can assume that digital pathology will keep making an appearance in other fields more frequently, in order to be used during the process of teaching students and doctors, as well as to assist specialists in the process of diagnosis, and also as a source of valuable scientific material for researchers.

References

Dee, F.R. and Meyerholz, D.K. (2007) 'Teaching medical pathology in the twenty-first century: virtual microscopy applications', *Journal of Veterinary Medical Education*, 34(4), pp. 431–436. DOI: 10.3138/jvme.34.4.431.

Farahani, N., Parwani, A.V. and Pantanowitz, L. (2015) 'Whole slide imaging in pathology: advantages, limitation, and emerging perspectives', *Pathology and Laboratory Medicine International*, 7, pp. 23–33. DOI: 10.2147/PLMI.S59826.

Feldman, M. D. (2015) 'Whole slide imaging in pathology: what is holding us back?', *Pathology and Laboratory Medicine International*, 7, pp. 35–38. DOI: 10.2147/PLMI.S81743

Koczy, S. (2020) '*Nowoczesne technologie w pracy nauczyciela jako szansa na edukację przyszłości*'. Availaible at: https://www.metis.pl/content/view/3456/105/ (Accessed: 8th April 2020)

Lewandowska-Ronnegreen, A. (2018) '*Techniki laboratoryjne w biologii molekularnej*'. Wrocław: MedPharm Polska, pp. 314, 346–351.

Litwin, J.A. and Gajda, M. (2011) '*Podstawy technik mikroskopowych*'. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, Wydanie VII, poprawione i rozszerzone, pp. 49–55, 60–68.

Prochorec-Sobieszek, M. (2016) 'Future perspectives of digital pathology', *Journal of Oncology*, 66(4), pp. 277–284. DOI: 10.5603/NJO.2016.0054.

Sawicki, W. and Malejczyk, J. (2012) '*Histologia*', Warszawa: Wydawnictwo Lekarskie PZWL, pp. 4–7.

Towards Open Research Data in the Economics Discipline

Piotr Kasprzak[1*], Magdalena Szuflita-Żurawska[2]

[1] Faculty of Management and Economics, Department of Finance, Gdańsk University of Technology (79 Traugutta Street, Gdańsk, Poland)
[2] Gdańsk University of Technology Library, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: pkasprzak@zie.pg.gda.pl; ORCID: 0000-0002-5570-4000

Abstract

Nowadays, Open Research Data, as one of the three pillars of Open Science (along with Open Access and Open Scholarly Communication), is gaining enormous attention from different academic and commercial environments. A wide range of scientific disciplines represent and produce different types of data and at the same time, gather different issues and problems in terms of sharing and dissemination research output. This chapter aims to briefly describe the current state of the scientific data for the business and economics disciplines. It presents the opportunities and concerns in those particular scientific fields where very often conducting research is supported not only by public funds but also by business and industry providers. Specific data types (such as financial data) need a particular approach to maintain them, especially in terms of the licensing, preservation and sharing openly. Their proper management can have a significant impact on the knowledge-based economy.

Keywords: Open Research Data, economics discipline, tax system, taxation, open data, tax information source, tax reporting

https://doi.org/10.34808/x55q-sz53_dyr_roz4

Introduction

Open Science has an undoubted impact on conducting research. It has changed the scientific methods and practices and scholarly communication and dissemination of research findings at national and international levels. According to the Organization for Economic Co-operation and Development (OECD), Open Science includes free access

to scientific articles, access to data from publicly funded research, and collaboration in research activities enabled by ICT tools (OECD, 2015). Nowadays, Open Access to scholarly papers is relatively well established in developed countries; however, making research data publicly available is during in transition period. Generally, Open Research Data (ORD) refers to all scientific data that is freely accessible and can be reused and redistributed without restrictions.

Standard practices include a Data Management Plan (DMP) to grant application process (e.g. for grants from Polish National Science Centre) to maximize knowledge about scientific projects from the initial phase or during the research project (such as those financed by Horizon 2020 and Horizon Europe, assume the sending of revised versions of plans during the project). The creation of DMP as a "living document" provides a description and evidence of produced data results in making research as time-efficient and safe as possible in terms of losing it due to, e.g. technical mishaps. Moreover, making accessible the research data is progressively common within Publishing Houses as part of their Open Science policies. To improve the quality of shared data and optimize their reusability, the community associated with research data management activities has developed the best practices called FAIR principles (Findable, Accessible, Interoperable, Re-usable). Those principles have human researchers in mind, but they focus on enhancing the availability of computers and machines to retrieve and use data automatically.

All scientific disciplines and sub-disciplines are characterised by their scholarly communication practices and methods, and their traditions in sharing knowledge and academic culture. Sharing research data varies between different fields and is determined by several factors such as geographical, technical and methodological features (Akers and Doty, 2013). Economics is considered to be a rather homogeneous field represented together by empirical and theoretical areas of conducting research. For instance, according to Dawson and Rascoff (2006), applied economics strongly depends on finding data (datasets) from commercial providers, governmental agencies, and academia, and collecting it through fieldwork. Additionally, supporting initiatives such as the Open Economics project (https://openeconomics.net/projects/) – a collaboration of Open Knowledge Foundation and Cambridge University – affect the popularisation of good practices of ORD available for preservation, transparency, analysis and replication.

It has always been known that information is a valuable asset. We strive to know more and more, we often incur high expenses to obtain information. More often, however, it is possible to obtain free access to knowledge, e.g. under the so-called Open Research Data. Although access to open data is rather identified mainly with science, the publication of research results and dissemination of knowledge, i.e. popularising activity, it is worth noting that access to knowledge is also becoming a significant competitive advantage, especially for enterprises. Companies that want to achieve more than their competitors are looking for data and information about their clients, contractors, changes in the market, and other business trends. Access to information becomes a valuable unit of assets (Schwab et al., 2011), although, for example, in accounting, this concept has long been known as an element of so-called intangible assets.

Currently, there is a growing trend in promoting universal and free access to up-to-date, reliable knowledge, including in the field of economy, finance and management (Sobczyk, 2016). The concept of open access to data is of general public interest, although one can find a number of examples of stakeholder groups who, due to the need to incur significant expenditures, especially financial ones, are reluctant to share their knowledge and experience (Skrzypek, 2018).

Economics is the science of making choices, allocating resources in the process of managing, producing and producing. Economics is also a science that collects and organises reliable knowledge about management. It is difficult to imagine making rational allocation decisions optimising a specific goal without having an appropriate amount of information (Czarny, 2011).

Enterprises are more often starting to use information and knowledge as factors of business success. The development of enterprises, and even their existence on the market, increasingly depends on the information and knowledge at their disposal (Olak, 2011). A new type of capital appears in enterprises, which consists mainly of knowledge, but also "raw" data and information. The contemporary success of a company is built on the basis of good knowledge of the market and the possibility of predicting what its environment will be like. The task of the information system is to recognise weak signals from the environment, analyse them and assess their suitability for building a competitive advantage for the enterprise (Górecka, 2018).

The aforementioned fields of science and researchers activity is based on so-called Research Data. More often, this data is a common good that is available in digital form for every Internet user (Strzelczyk, 2017).

According to a number of definitions in the literature, research data includes:

- Text documents,
- Numerical data,
- Surveys and survey research results,
- Video and sound recordings,
- Algorithms and schemes,
- Test and simulation results,
- Samples.

It should be noted that in the case of enterprises, the most important data are that about the market, customers, and their resources and preferences (Sobczyk, 2016). A problem for many companies, however, is the multitude of data (their quantity), and therefore skilful filtering of only those that are necessary for a given company is needed. Currently, we have access to many databases, information websites, financial statements and economic reports. In addition, companies obtain information on a daily basis by conducting their own research (Sosnowski, 2019). According to the literature, the methods of obtaining information are divided into basic research methods and detailed methods.

The first involves collecting information available to everyone from various types of publications, reports, etc. These are simpler and more informative methods. The detailed methods, on the other hand, require more effort and work to be performed. However, the

information is deeper in details and provides greater benefit to the company, and it must be conducted in order to find the solution for a specific problem. Usually one leading technique and two or more complementary ones are chosen and therefore used. It all depends on the needs of the enterprise and the budget (Rzemieniak, 2012).

The main test methods are:

- **Observation** – an imperceptible, intentional and systematic way of perceiving the tested objects in their natural conditions.
- **Participant observation** – in which the researcher tries to enter the environment she/he is interested in and collect information from the inside.
- **Open observation** – an observation in which research objects are aware that they are the subject of a study, which may lead to falsification of information.
- **Tests** – which belong to a wider group of projection techniques aimed at assessing the personality, attitudes and motives of research objects.
- **Survey research** – a set of questions prepared for a group of respondents, taking place without the participation of the researcher. It is strictly defined by her/him in the form of a questionnaire, and the manner of its delivery and receipt is also developed.
- **Panel method** – i.e. researching the same group of people several times in order to collect information on changes in the formation of various market or industry events at a specific time.
- **Interview** – a conversation between the researcher and the object of interest, the purpose of which is to obtain the necessary information.
- **Experiment** – examines the phenomena caused by the researcher in the environment and the conditions specified by her/him.

However, it should be noted that companies most often conduct research for their own needs and do not share the information obtained with others. Below (Tab. 4.1), the main conclusions in the form of opportunities and advantages, as well as threats and disadvantages, resulting from the widespread sharing of obtained and processed data, both by enterprises and specialised units (government agencies, economic intelligence agencies, research institutions) are presented.

In addition to in-house research, including marketing, enterprises obtain information from many other sources. In order to obtain information, entities often use the services of specialised agencies, consulting companies or public opinion polling centres (Sobczyk, 2016).

It is also common for companies to use databases which, due to their specific characteristics, can be divided into the following repositories:

- institutional/national: e.g. Central Statistical Office, European Statistical Office (EUROSTAT), Central Banks, Data Archiving and Networked Services (multi-domain science portal of the Netherlands created by a coalition of scientific governmental organisations of this country),
- thematic: e.g. UK Data Archive (data repository from social sciences and humanities), SeaDataNet (international oceanographic project), ADS Social Data Archive,

- general: e.g. FigShare, World Data System (open data from various fields, created in 2010 by the International Scientific Organization International Council for Science), OpenDOAR, RepOD,
- paid, with the possibility of ordering personalised data, e.g. Amadeus – European Business Information, EMIS Database.

Tab. 4.1

Open data in economics – opportunities and advantages as well as threats and disadvantages

Opportunities and Advantages	Threats and Disadvantages
– possibility of establishing cooperation, especially at an international level and between various industries, – access to market data from around the world, – inventing new and improving existing inventions, – reducing the cost of obtaining data that is particularly difficult to access, – disseminating good practices in business, – non-compliance with certain processes by several companies at the same time (e.g. analysis of the same data by several companies at the same time), – possibility of specialisation and establishing cooperation on the business-science line, – ORD increase the efficiency of economic research and its diffusion – ORD increase cross-sector collaboration – cost savings (reducing direct costs of scientific findings for R&B sector), – provides for long-term safe storage of data, – increases citation rates and enhancing scientific reputation, – sharing data creates a public good, – enhancing and popularisation of using research methods, – maximising accountability and knowledge generation.	– possibility of "usurpation" of data, their inappropriate citation or failure to provide information about the author at all, – concerns of some stakeholder groups about losing the status quo or fear of losing market share, – reluctance to share data for free, especially when obtaining the information was associated with incurring significant expenditure, – reducing the importance of companies specialising in the commercial collection and analysis of information, – fear of access to enterprise data by government agencies and the tax administration, – problems related to patenting inventions and scientific discoveries – making information available earlier may harm the patenting process, – contractual obligations may prohibit data sharing, – very complex data value cycle in economics – data may be used for the wrong or inappropriate purposes, – difficulty generating qualitative data and concerns about sharing data before the researcher(s) have had a chance to publish due to this.

*Source***:** own elaboration

In addition to the "information" (considered to be true), there are also terms that have a completely different meaning, i.e. a negative meaning for both the entrepreneur, and contractors and customers (Łęgowik-Małolepsza et al., 2017). These are:

- pseudo-information – information that comes from different sources, but relates to the same fact or event,
- metainformation – information about information,
- misinformation – the concept of false information based on lies, manipulation and willful misrepresentation,
- parainformation – subjective information, based on an interpretation made in an individual way, which may lead to different conclusions or opinions.

Without the doubt, companies may resort to manipulating and misinforming about a given product or service, but in the long run, this is not profitable (after all, a company that does not apply an exemplary policy itself may one day fall victim to another entity that will deliberately want to introduce them. in error).

An example of actions aimed at curbing manipulation or misleading practices may be the amendment to the Act on Trading in Financial Instruments introduced in 2017 (Kwieciński, 2019). Pursuant to Polish law, the legislator introduced eight types of manipulative behaviour. The manipulative behaviour listed in the Act includes placing orders or entering into transactions that are misleading or may mislead as to the actual demand, supply or price of a financial instrument; placing orders or entering into transactions which result in an unnatural or artificial pegging of financial instruments; placing orders or concluding transactions with the intention of causing other legal effects than those for which the legal act is actually performed; placing orders or concluding transactions while misleading market participants or taking advantage of their error as to the price of financial instruments. It should be noted that transparency and universal access to information in the case of trading in securities is particularly important because the capital market is, next to the banking sector, one of the most important elements of the financial system and is subject to special protection. An efficiently operating, stable and balanced financial system plays a key role in the economy of any country, enables its development and, consequently, improves the wealth of the society. Its destabilisation may disturb the balance of this economy. Ensuring the stability of the financial system belongs to the state authorities which have been entrusted with the supervisory functions.

Moreover, due to the universal access to the financial data of listed companies, these entities are required to report data on a quarterly basis and publish them on the Internet, and, like other commercial companies, they are required to approve and publish the annual financial statements to the competent registry court. Consequently, the possibility of manipulation and deliberate misleading is negligible, as investors have virtually unlimited access to data, which allows them to quickly verify individual messages. This, in turn, proves that the exchange of information and access to data is in the interest of various groups of capital market stakeholders.

Promoting open, free and universal access databases is in everyone's interest. In the era of globalisation and more and more universal access to the Internet, companies are

not making any sense in terms of access to information and sharing it with other entities. Creating open databases to which various stakeholders have access is in the interest of everyone, as evidenced by numerous clusters, e.g. the automotive, craft, and agri-food clusters. A free database of an economic and financial nature is another such potential cluster.

An example of database including datasets meeting the criteria of openness described above (including publishing according to CC-BY policy) is the platform introduced by the Gdańsk University of Technology where both academic staff as well as external users may publish their data. The platform is available at mostwiedzy.pl, where datasets are published in English and can be used for free. One of the co-authors of the following paper regularly publishes the datasets obtained during the academic and scientific research[1].

As Rosa (2008) notes, a feature of knowledge-based clusters is a high concentration on a limited area of entrepreneurs, investors and scientists, and the existence of frequent formal and informal contacts between them. Industry or regional associations, often supported by public funds, are also a platform for these contacts. Therefore, the basis for the exchange of ideas and experiences in the case of economic, financial, reporting and statistical data can be common, open, data repositories. It should be clearly emphasised that nowadays, in the era of widespread use of the Internet, a cluster may also become a jointly, though independently, created open database, to which everyone has access. Everyone can also be the author of some part of the information contained within it, and thus participate in the general development of the given industry.

What is more, access to broad data also means building a knowledge-based economy. This is a concept that emerged in the 1990s. An economy in which knowledge is created, absorbed, transferred and used more effectively by enterprises, organisations, individuals and communities, fosters it quickly and has a positive impact on societies. The development of such an economy is based on the significant use of knowledge and experience; such a country can develop faster than others, because it generates faster technical and organisational progress, has better educated people and uses human and productive capital more effectively (Skrzypek, 2011).

Conclusions

"Open Science" is an "umbrella" term where its different aspects such as ORD have to be singled out and investigated separately to better understand their opportunities and tangible problems. The above examples demonstrate the great challenge in opening research in the economics discipline, which has substantial implications for the business and industry sectors. In return, Open Science might greatly influence international visibility and recognition, and contribute to the knowledge diffusion and credibility that is one of the most critical issues for research worldwide. The ongoing debate on sharing data is far from settled, and scholars represent different opinions and practice it differently. Many of the concerns relate to the legal constraints and human participation; however, proper data management and sharing are at the core of valuable research. There is

[1] https://mostwiedzy.pl/pl/search/openResearchData?s=piotr%20kasprzak.

a severe concern that mostly commercially used open data leaves no trace and it is rather hard to assess its coverage and the actual state of affairs. However, the obvious benefits such as labour cost savings, productivity improvements, and reducing research duplications might decrease tensions related to commercialisation and openness. The right approach that will focus on developing policy positions on Open Science and commercialisation, promoting and supporting open collaboration between different stakeholders, and targeting business and industry in research findings can facilitate a transition and maximise the value of ORD in the economics discipline.

References

Akers, K.G., and Doty, J. (2013) 'Disciplinary differences in faculty research data management practices and perspectives', *International Journal of Digital Curation*, 8(2), pp. 5–26. DOI: 10.2218/ijdc.v8i2.263 (Accessed: 6th January 2021).

Dawson, M., and Rascoff, M. (2006) '*Scholarly communications in the Economics discipline. A report commissioned by JSTOR*'. DOI: 10.18665/sr.22340.

Czarny, B. (2011) '*Podstawy ekonomii*'. PWE, Warszawa.

Górecka, K. (2018) 'Źródła informacji wykorzystywane w systemie wczesnego reagowania w przedsiębiorstwach', *Studia i Prace Kolegium Zarządzania i Finansów*, 163, pp. 143–164, DOI: 10.33119/SIP.2018.163.8.

Kwieciński, J. (2019) 'Przestępstwo manipulacji na rynku kapitałowym w świetle ustawy o obrocie instrumentami finansowymi i prawa wspólnotowego', *Studia i Prace Kolegium Zarządzania i Finansów*, 174, pp. 9–31, DOI: 10.33119/SIP.2019.174.1.

OECD (2015) 'Making Open Science a Reality', *OECD Science, Technology and Industry Policy Papers*, 25, OECD Publishing, Paris, DOI: 10.1787/5jrs2f963zs1-en.

Olak, A. (2011) 'Kapitał intelektualny i wiedza jako czynniki nowoczesnego zarządzania finansami – studium przypadku', *Zeszyty Naukowe SGGW - Ekonomika i Organizacja Gospodarki Żywnościowej*, 88, pp. 151–164, Available at: http://sj.wne.sggw.pl/article-EIOGZ_2011_n88_s151/ (Accessed: 25th May 2022).

Rosa, K. (2008) 'Klaster jako forma powiązań przedsiębiorstw w gospodarce opartej na wiedzy', *Zeszyty Naukowe Wyższej Szkoły Humanitas. Zarządzanie*, 1, pp. 69–78, Available at: https://www.sbc.org.pl/publication/11535#description (Accessed: 25th May 2022).

Rzemieniak, M. (2012) '*Badania marketingowe w podejmowaniu decyzji menedżerskich*', Lublin: Wydawnictwo Politechniki Lubelskiej, Available at: http://bc.pollub.pl/Content/1376/PDF/badania.pdf (Accessed: 25th May 2022).

Schwab, K., et al. (2011) 'Personal Data: The Emergence of a New Asset Class,' World Economic Forum Report. Available at: https://www.weforum.org/reports/personal-data-emergence-new-asset-class/ (Accessed: 25th May 2022).

Skrzypek, E. (2011) 'Gospodarka oparta na wiedzy i jej wyznaczniki', Nierówności społeczne a wzrost gospodarczy, 23, pp. 270–285, Available at: https://www.ur.edu.pl/storage/file/core_files/2012/7/2/8756a245950c4fa64366f06281bd5bcc/022.pdf (Accessed: 25th May 2022).

Skrzypek, E. (2018) 'Bariery dzielenia się wiedzą w organizacji w warunkach społeczeństwa informacyjnego', Nierówności Społeczne a Wzrost Gospodarczy, 53, pp. 34–48, DOI: 10.15584/nsawg.2018.1.2.

Sobczyk, G. (2016) 'Rola i źródła informacji rynkowych w działalności przedsiębiorstw – wyniki badania', Annales Universitatis Mariae Curie-Skłodowska. Sectio H. Oeconomia, 50(3), pp. 173–183, DOI: 10.17951/h.2016.50.3.173.

Sosnowski, P. (2019) 'Systematyzacja pojęć związanych z metodami i źródłami pozyskiwania informacji w kontekście infobrokeringu', Studia Politologiczne, 54, pp. 45–67, DOI: 10.33896/SPolit.2019.54.2.

Strzelczyk, E. (2017) 'Otwarte dane badawcze – kolejny krok do otwierania nauki, Bibliograficzne bazy danych: perspektywy i problemy rozwoju', III Konferencja Naukowa Konsorcjum BazTech, Kraków. Available at: http://ebibojs.pl/index.php/Mat_konf/article/view/616 (Accessed: 25th May 2022).

A Text as a Set of Research Data. A Number of Aspects of Data Acquisition and Creation of Datasets in Neo-Latin Studies

Grzegorz Kotłowski[1*], Jacek Pokrzywnicki[1], Elżbieta Starek[1]

[1] Classical Philology Division, Institute of Classical and Slavonic Studies, Faculty of Languages, University of Gdańsk (51 Wita Stwosza Street, Gdańsk, Poland)

* Correspondence authors: grzegorz.kotlowski@ug.edu.pl; ORCID 0000-0003-3689-6462

Abstract

In this paper, the authors, who specialise in part in neo-Latin studies and the history of early modern education, share their experiences of collecting sources for Open Research Data sets under the Bridge of Data project. On the basis of inscription texts from St. Mary's Church in Gdańsk, they created 29 Open Research Data sets. In turn, the text of the lectures of the Gdańsk scholar Michael Christoph Hanow, Praecepta de arte disputandi (Recommendations on the art of discussion) written in 1754 and a series of 17th-century syllabuses from classes at the Gdańsk Academic Gymnasium became the core of two other Open Research Data collections. The article discusses: the authors' approach towards the text (this approach primarily takes into account the traditionally understood philological research paradigm), and then the method of obtaining data from church (church parish) or state institutions (Polish Academy of Sciences Gdańsk Library). In the authors' opinion, the issue to which they devote the most attention is also operations on the original text, i.e. transcribing the original record into a character that a modern recipient of Latin texts can expect – both an amateur of history or old literature, as well as a professional researcher of the history and writings of the early modern period.

Keywords: Neo-Latin studies, inscriptions, Academic Gymnasium in Gdańsk, research dataset, transcription, data acquisition, St. Mary's Church in Gdańsk; epigraphy

https://doi.org/10.34808/x55q-sz53_dyr_roz5

Under the Bridge of Data project, we have prepared[2] twenty-nine datasets based on the texts of inscriptions from St. Mary's Church in Gdańsk (e.g. Kotłowski and Starek, 2019.; Starek and Kotłowski, 2019). Under the same project, we have prepared two datasets containing texts from sources printed in the 17th century and from handwritten sources created a century later (Pokrzywnicki, 2019; Pokrzywnicki, 2020). The printed sources involved syllabuses for lessons in Gdańsk Academic Gymnasium, bound together with other prints in a volume with the following reference number: Ma3920 8o (Oelhaf et al., 1645–1653). The handwritten source text, on the other hand, is a transcript of lectures on the theoretical foundations of the discussion. These lectures series' title in Latin is: Praecepta de arte disputandi. In 1754, they were delivered at Gdańsk Academic Gymnasium by Michael Christoph Hanow (1695–1773) – a versatile scholar who held the position of a professor of philosophy at this school for many years (Hanow, 1754).

Since the title of our article quotes the text as the main element of the dataset, we deem it appropriate to provide a brief explanation of our approach to the texts and why we used this type of methodology. We will not, however, try to define the term "text". Anyone interested in the subject should refer to the definitions provided in dictionaries (e.g. Cluysenaar, 2006). Further, we would like to describe how we acquire old Latin texts. We also feel it is necessary to touch on the issue of text transcription, that is, transferring the text from the source to a more accessible medium (for on-line research datasets, it is obviously an electronic file). This is because the transcription procedure directly affects the form of the record.

Let us start with stating the following: in our daily research and academic work, we do not focus on theoretical reflection. However, and it must be clearly emphasised, we are far from belittling the importance of newer formal or functional text analysis (for an account of some contemporary textological research, see, e.g. Bartmiński and Niebrzegowska, 2012, pp. 7–10). As a general rule, we act as practising text analysts. Sometimes, we also move towards descriptive textology. We do so while interpreting a text and place it within a specific literary, historical and social context. We are, however, most interested in the coherence of lexis and syntax: our analysis primarily boils down to organising and verifying the Latin text in these two aspects. The text we are analysing should, as we assume, become understandable for the users of Latin. However, it is necessary to provide a translation (in this case, a translation into Polish) for people who, for various reasons, are not going to use the Latin original. The quality of the translation also depends on the ordering of the lexis and the syntax of the text. Only with a properly reconstructed text are we able to effectively recognise the meanings conveyed in the Latin-language source text. It follows from the above considerations that we act in accordance with the editorial tradition of classical philology, using the tools or procedures that are needed in order to reconstruct a multidimensional text, in line with the origin of the word 'text', which

[2] In the paper, the plural form is always used. However, we would like to clarify that the research datasets with inscription texts were prepared by Elżbieta Starek and Grzegorz Kotłowski. Two other datasets with the texts created by the teachers of Gdańsk Academic Gymnasium were prepared by Jacek Pokrzywnicki.

derives from 'texere' – "weave a fabric", but which, in ancient times, was also used to describe the literary texture, or style (Quintilianus, 1996, pp. 514–515).

Extracting texts for research datasets

Preparing research datasets always starts with the acquisition of the text. It should also be taken into account that the texts we work on while preparing a dataset or while doing any research are not physically available to us: they were created several hundred years ago and for that reason, they are quite unique. Moreover, the origin of these texts is also, as is the case with the inscriptions, related to religious activities.

Working with inscriptions not only makes in necessary to visit the interior of the church, but also to establish contact with those officially responsible for the building. In the Polish reality, the person who can grant permission to work inside a church is usually the pastor of the parish administering the church. In our practice, although not with published datasets, we also had to negotiate for consent to work in the church interior with competent representatives of religious communities. This happened when the church with the inscriptions of interest to us was administered by a community of this type (e.g. the Church of St. Trinity in Gdańsk).

The contract, usually a verbal one, provides for the time and scope of works that we anticipate to carry out in the church's interior. Since the in situ works mainly consist in making a preliminary version of the inscription reading and preparing appropriate photographic records of the inscription, it is important to schedule the works with the person administering the temple. The point is that the works that sometimes require a certain reorganisation of the church space (e.g. moving some movable elements of the interior) should not disrupt the religious activities within the church.

In the case of handwritten or printed texts dated before 1801 (the so-called old prints), the procedure is rather formalised. Trying to access the texts from these two groups, we follow the recommendations of the libraries. The rules governing access to unique collections are formulated in a generally similar way by various libraries – not only in the territory of the Republic of Poland; it is necessary to obtain a reader's card or to present an ID document (which involves providing our own personal data), to fill in a special declaration indicating the purpose for which you are applying for access to the collections (usually research) or to supply recommendations from a person with extensive research experience (Michalska, 2019, pp. 108–110). The institution with the largest collection of early printed books and manuscripts about old Gdańsk is the Polish Academy of Sciences, the Gdańsk Library. The terms and conditions of using this library's old collections provide for all of the above-mentioned methods of verifying the person attempting to access the collections (Polska Akademia Nauk Biblioteka Gdańska, 2019a).

Usually, however, the use of unique datasets is subject to further restrictions. The libraries with the collections of manuscripts and old prints recommend that readers should work with microfilms or scans of ordered prints or manuscripts if these institutions have such substitutes within their collections. However, working with the research datasets, we assumed that the data constituting the dataset would be totally generated in

an independent manner, i.e. without the use of substitutes offered by the Polish Academy of Sciences Gdańsk Library. We signaled our wish to create our own photographic documentation that would then be included in the planned datasets, in the official correspondence with the Special Collections Department of the Polish Academy of Sciences Gdańsk Library. The director of this unit agreed in a formal letter "for the use of […] digital copies of the resources originating from the Gdańsk Library, or fragments thereof, as illustrative material" (Polska Akademia Nauk Biblioteka Gdańska, 2019b). This library, acting in accordance with internal regulations, also granted a non-exclusive license for an indefinite term, for the following fields of exploitation:

- recording and reproduction, production of copies of reproductions using a specific technique, including printing, reprographic and digital techniques;
- uploading objects and parts thereof into computer memory;
- disseminating the objects on the Bridge of Data project website, specifying the place of storage for such objects;
- use of the objects in informational texts, as part of advertising or press publications (Polska Akademia Nauk Biblioteka Gdańska, 2019b).

The Polish Academy of Sciences Gdańsk Library, in all of the above-mentioned circumstances, recommended informing that original sources, i.e. photographed manuscripts and printed texts, are stored in the library's collection (Polska Akademia Nauk Biblioteka Gdańska, 2019b).

Certain operations on the texts located in research datasets – text retrieval and transcription

One of the important components of the research datasets prepared by us are files with transcribed texts. The transcription process implies several changes to the original record. It should now be emphasised that the assumptions made for the transcription of the inscriptions differ from the assumptions used when transcribing manuscripts and old prints. The transcription of inscriptions was fully formalised and settled. This is due to the fact that all of the transcripts we have included in the datasets were published beforehand (see: Kotłowski and Starek, 2014). On the other hand, the transcriptions of Michael Christoph Hanow's De arte disputandi lectures (Hanow, 1754) and the syllabuses of classes at the Gdańsk Academic Gymnasium (Oelhaf et al., 1645–1653) were prepared as part of the datasets and we do not treat them as definitive. Such a suggestion was proposed in the description of the dataset containing the text by Michael Christoph Hanow (see introductory remarks in: Pokrzywnicki, 2019).

What does the transcription we use in all of the inscriptions from St. Mary's Church in Gdańsk look like? It must be said right away that we did not fully apply the principles developed by the historians analysing the inscriptions (for the principles, see: Szymański and Trelińska, 2003). We assumed that the inscriptions we publish are to be readable not only by specialists who usually transcribe inscriptions without deciphering any abbreviations (see e.g.: Grześkowiak-Krwawicz and Szyszkiewicz, 1988). We tried to reproduce

the Latin text as accurately as possible. That is why we have deciphered all of the abbreviations. We have maintained uppercase and lowercase letters; we also tried to keep the same word order in individual lines. This was the case, for example, in the epitaph of the Choene-Jaski family (Kotłowski and Starek, 2020. DOI: 10.34808/03by-1j90). The photograph, which was also included in the research dataset (see Fig. 5.1), clearly shows that all lines start equally from the left side.

Fig. 5.1. Inscription from the epitaph of the Choene-Jaski family | (photographer: Lucyna Lewandowska)

Despite the abbreviations used, the inscription does not form a compact column. In our transcription, the deciphered abbreviations completely disturbed the original column composition. That is why we decided to propose a centered composition.

Paulo Chöne, dict(o) Iaskii, Avo et Patri,
Henr(ico), Paulo, Marco et Iacobo Fratrib(us)
Andreas, Ser(enissimorum) Elect(orum) Brand(erburgensium) Ioach(imi) Frid(erici),
Io(annis) Sigis(mundi) et Geor(gii) Gulielmi a cons(iliis) secret(iis)
Et Israel Chöne, d(ictus) Iaskii, Consul Reip(ublicae) Ged(anensis)
Hoc obser(vantiae) amor(is)q(ue) posuere monumentum
Anno D(OMI)NI MDCXXVIII.

[Paul Chöne a.k.a. Jaski, grandfather and father, brothers Henry, Paul, Mark and Jakob, Andrew, secret counselor of the brightest electors of Brandenburg, Joachim III Frederick, John III Sigismund and George Wilhelm I, and Izrael Chöne called Jaski, councilor of the city of Gdańsk, erected this monument of honor and love in 1628]

Note the abbreviations with a full stop. We were always deciphering them. For example, in the first line, the third word "dict(o)"(original: dict.), in the second line, the first "Henr(ico)" (originally: Henr.), and in the sixth – "Fratrib(us)" (originally: Fratrib.), etc. What may not always be the case, we also solved the popular abbreviation for the enclitic 'que' (this is the case in the sixth line, where in the original, along with an additional abbreviation for the word 'amor(is)', we read: amorq;). The same applies to another com-

mon abbreviation in the word "D(omi)NI" (in the original: "DNI" with the sign '–' above the letters 'NI'). However, we left the full stops if they appeared between words.

The transcription of the inscription on the sepulchral plate of Thomas Tympfius (see also Fig. 5.2–5.3 and Starek and Kotłowski, 2020l) indicates, in turn, how we tried to maintain the original size of the letters:

TYMPFIUS hic terra vitalis ubi aura recedet
Ossa gelata dari vult reditura sciens.
Euro flante plagis e quat(t)uor, ossa reviset
Et feret ad coelos aura calorq(ue) redux.
Ante DIeM CVrans, en! fVnera na(m) bene VIVIt
Na(m) parat, et Late et possIt obIre bene.

[Believing in the resurrection, Tympfius wants to lay his bones here when the soul leaves the earth. He will see the bones again when the wind blows from all four corners of the world, and the breath and warmth returned will lift them to heaven. Because caring for the funeral during his lifetime, he lives well, he prepares himself so that he can die well.]

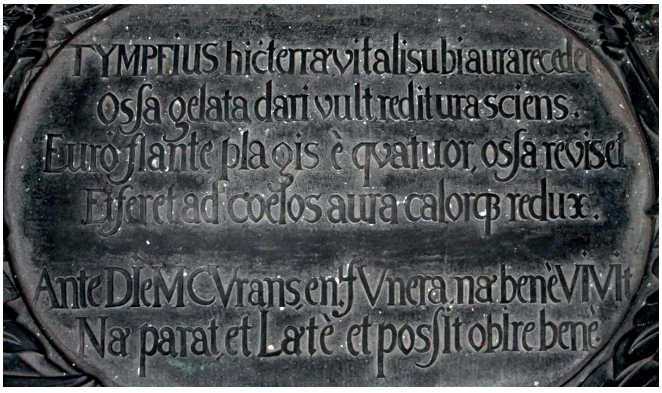


Fig. 5.2. Inscription on the sepulchral plate of Thomas Tympfius (photographer: Lucyna Lewandowska)

This poetic inscription is provided with a chronogram (fifth and sixth lines). In order to reflect the presence of a chronogram in the transcription indicating the date of the inscription, it was necessary to leave capital letters. We also found it necessary to keep the letter 'V' for the Roman numeral 'five'. This letter in other words, for example in the "quattuor" in the fourth line, has been transcribed as 'u'.

Finally, let us add that we deciphered conventional abbreviations using widely available dictionaries of abbreviations in ancient Latin inscriptions (e.g. Cappelli, 1995; Węcowski, 2010; Winiarczyk, 1995; Kloos, 1973, pp. 111–125).

In the transcription of the manuscript (Praecepta de arte disputandi by Michael Christoph Hanow) and of the old prints (syllabuses from the Academic Gymnasium in Gdańsk), we applied a number of principles that partially (e.g. in terms of spelling) comply with the instructions of the international Latinist society Academia Latinitati

Fovendae (see: Sallmann, 1992), for the purposes of research datasets. At the same time, they fail to follow the rules of the 17th or 18th centuries (II, Manuzio and I, 1566; Lipsius et al., 1660; Manuzio, Manuzio and Keller, 1738). Why such difference? Well, both of the above-mentioned datasets do not contain texts that are fully prepared for publication. Rather, they should be treated as a preliminary analysis of a text extracted from old sources. In these transcripts, we simply introduced a series of solutions, developed within our own research practice. Below are some examples of such solutions.

Fig. 5.3. Inscription dedicated to Thomas Tymphius – right in the centre of the sepulchral plate of the Zimmerman family (photographer: Lucyna Lewandowska)

Capital letters

First of all, we avoided capital letters at the beginning of words. The decision to eliminate the excess of capital letters appeared to be the right one, because in early modern Latin texts, there are no clearly defined rules on this matter. Therefore, whoever decides to transcribe a Latin text is forced to make their own decisions: when to start a word with a capital letter, and when to start with a lower case. It is irrelevant whether we are dealing with a manuscript or a printed text (the former is usually more flexible in terms of spelling).

In any case, we tried to use the commonly accepted recommendations for the use of a capital letter at the beginning of a word. Therefore, we left or used a capital letter:

- in a word that starts a new sentence
- in proper nouns
- in words derived from proper nouns (adjectives and adverbs)
- in titles (including modified and abbreviated) of literary works
- in common words that indicate a specific thing or person, e.g. Salvator – Savior, i.e. JCH; Christmas holidays – Natalitia Christi

Abbreviations

Wherever possible, we deciphered conventional abbreviations. Such a procedure allows for comparing the understanding of a shortened word proposed in the transcription and the original notation of this abbreviation,
for example:
"[Daniel Lagus] […] continuabit […] analysin Testam[enti] Novi […]"
[Daniel Lagus] […] will […] continue to comment on [Greek excerpts] from the New Testament] (See: 106, fol [buntur verso] and IMG_3569 in Pokrzywnicki, 2020: DOI: 10.34808/823g-1y16), and:
"P[relo] P[ublico] Dominica I Adventus Christi"
[From the official printing house [Gdańsk City Council] on the first Sunday of Advent] (see: 108, fol [A1 verso] / quam, IMG_3577 Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16), or:
"S.S. – sacrosancta theologia, Trinitas" [the most holy theology, the Trinity (see: 108, fol [A1 verso] / quam, IMG_3577 Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16).

Numerals

We left ordinal numbers in the Roman system when they meant:

- a number / part of a literary work,
  for example:
  "[…] impraesentiarum pollicetur certo se (si Deus robur corporis et artuum concesserit) librum III et IV Institutionum semestris spatio methodo solita absoluturum".
  [Now [Peter Oelhaf] promises that in this semester he is certainly going to finish explaining, as he has done so far, the third and fourth Institutiones, if God gives strength to his body and its [sc. body] members] (see: 106, fol. A2 [verso] / maiora, IMG_3567 in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16), or:
  "Absolvi tamen [1] de Deo Trinuno, [2] de peccato originis, cum quo propter materiam cognatam articulum XVIII de libero arbtitrio et XIX de causa peccati coniunxi […]".
  ["But I finished 1) On Triune God, 2) On the origin of sin, and with him – because of the thematic relationship – article 18, On free will, and added [article] nineteen On the cause of sin […]" (see: 109, fol. A2 [recto] / sterio, IMG_3585.JPG109, fol. A2 [recto] / sterio, IMG_3585.JPG in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16).
- numerals accompanying the names of sovereigns,
  for example:
  "Historias a Friderici I usque ad nostra tempora chronologice et politice pertexturus"
  ["[Peter Oelhaf] is going to start with a history lecture (from Frederick I to our times)"] (see: 106, fol. A2 [verso] / maiora, IMG_3567 in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16).

Spelling

We have changed the original spelling, using the currently used Latin dictionaries. They are popular reference dictionaries available on-line (see: Gaffiot, 1934; Georges, 1879–1880; Lewis, 1987) and in hard copy version (see: Glare, 2004; Korpanty, Bobrowski and Brodka, 2001–2003).

For example

- elimination of diphthongs: "ceptarum" instead of "caeptarum" (see: IMG_3567, page A2 [verso] / maiora in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16)
- elimination of the letter j: "iuris" instead of "juris" (see: IMG_3567, page A2 [verso] / maiora] in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16)
- elimination of letters not falling within the classic set of 24 Latin letters, e.g. "Oelchafius" instead of "Ölchafius" (see: 107, fol. nimi-, IMG_3573 in Pokrzywnicki, 2020. DOI: 10.34808/823g-1y16)

Punctuation

We resigned from the rhetorical punctuation, which takes into account such elements of an extended complex sentence, called period in rhetoric, such as colon (membrum in Latin) and komma (incisum in Latin). Instead, we introduced punctuation marks in a manner similar to the rules applicable in Polish. However, we also assumed that the potential user of the dataset would be able to change the punctuation in the most appropriate way in his or her opinion. In any case, the Societas Latinitati Provehendae instruction, cited earlier, recommends not to overuse punctuation marks (see: Sallmann, 1992, 447, 455).

Conclusions

All of the examples presented above show how extensive are the activities of the creator of Open Research Data sets. It is possible to freely shape the content of the dataset, which must first be acquired, following the rules imposed by the persons or by the institutions storing the original source material. It is obvious: when deciding what should be found in such a set, individual research preferences are important. The creator of datasets with a background in classical philology (or neo-Latin studies) will be primarily interested in reading and presenting a text that will seem understandable to all potential recipients. Rather, it is the text files that will form the backbone of such an open dataset. These files can appear within the set individually. However, the combined formula seems to be better, where next to the text file with the transcription of the source text there is a photograph of that source. In the case of little-known languages – Latin is certainly one of these today – you should also consider introducing text files with translations. This applies especially to those texts that, according to the creator or authors of the open dataset, could be useful for people who do not know the source language. This group of texts

includes inscriptions. After all, they are clearly present in the public space and, if only for this reason, they may interest people who do not have appropriate linguistic skills. More specialised texts – and we consider syllabuses or lecture notes as such – do not (rather) require translation. People interested in obtaining this type of textual data will probably have an appropriate linguistic background. These advanced users of datasets will probably also be willing to modify the acquired data according to their own research methodologies. Such an operation seems even desirable: the research dataset should not only serve its creator.

References

Bartmiński, J. and Niebrzegowska, S. (2012) '*Tekstologia*'. Warszawa: Wydawnictwo Naukowe PWN.

Cappelli, A. (1995) 'Lexicon abbreviaturarum: Usate nelle carte e codici specialmente del medio-evo; riprodotte con oltre 14000 segni incisi; con l'aggiunta di uno studio sulla brachigrafia medioevale, un prontuario di sigle epigrafiche, l'antica numerazione romana ed arabica ed i segni indicanti monete, pesi, misure, etc. ; corredata con 9 tavole fuori testo = Dizionario di abbreviature latine ed italiane'. 6th edn. (Manuali Hoepli). Milano: Hoepli.

Cluysenaar, A. (2006) *'Text'*, in Childs, P. and Fowler, R. (eds.) The Routledge dictionary of literary terms. Abingdon [England]: Routledge, pp. 237–238.

Gaffiot, F. (1934) '*Dictionnaire illustré latin-français*'. Paris: Libraire Hachette.

Georges, K. E. (ed.) (1879-1880) '*Ausführliches lateinisch-deutsches Handwörterbuch aus den Quellen zusammengetragen und mit besonderer Bezugnahme auf Synonymik und Antiquitäten unter Berücksichtigung der besten Hülfsmittel*' (2 vols). 8th edn. Leipzig: Hahnsche Buchhandlung. Available at: http://www.zeno.org/Georges-1913 (Accessed: 2nd January 2021).

Glare, P.G. (ed.) (2004) '*Oxford Latin Dictionary*'. Oxford: Clarendon Press.

Grześkowiak-Krwawicz, A. and Szyszkiewicz, K. (1988) 'Zabytki epigraficzne kościoła św. Mikołaja w Gdańsku', *Rocznik Gdański*, 48(1), pp. 152–178.

Hanow, M.C. (1754) '*Praecepta de arte disputandi*' (a manuscript from the collection of Polish Academy of Sciences Gdańsk Library, reference number: Ms. 512).

Kloos, R.M. (1973) '*Einführung in die Epigraphik des Mittelalters und der frühen Neuzeit*'. 2nd edn. (Die Kunstwissenschaft). Darmstadt: Wissenschaftliche Buchgesellschaft.

Korpanty, J., Bobrowski, A. and Brodka, D.e.a. (eds.) (2001-2003) '*Słownik łacińsko-polski*' (2 vols). Warszawa: Wydaw. Szkolne PWN.

Kotłowski, G. and Starek, E. (2014) 'Łacińskie inskrypcje w kościołach Gdańska: Bazylika Mariacka'. Pelplin: Bernardinum.

Kotłowski, G. and Starek, E. (2019) 'Epitaph of George Posse in St. Mary's Church in Gdańsk'. DOI: 10.34808/canj-3539.

Kotłowski, G. and Starek, E. (2020) '*Alms-panel of Anton Moeller in St. Mary's Church in Gdańsk*'. DOI: 10.34808/ka5s-kc18.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Bartholomew Wagner in St. Mary's Church in Gdańsk*'. DOI: 10.34808/tccm-wq24.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Choene-Jaski family in St. Mary's Church in Gdańsk*'. DOI: 10.34808/03by-1j90.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Dorothy and John Brandes in St. Mary's Church in Gdansk*'. DOI: 10.34808/knzz-yy15.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Edward Blemke in St. Mary's Church in Gdańsk*'. DOI: 10.34808/68p4-vh21.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Ferber family in St. Mary's Church in Gdańsk*'. DOI: 10.34808/sptw-bh75.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Gabriel Schumann and his son Gabriel in St. Mary's Church in Gdańsk*'. DOI: 10.34808/799p-y810.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of George Hojer in St. Mary's Church in Gdańsk*'. DOI: 10.34808/j9tq-z081.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Guldenstern family in St. Mary's Church in Gdańsk*'. DOI: 10.34808/5xw4-fj43.

Kotłowski, G. and Starek, E. (2020) '*Epitaph of Henry Giese in St. Mary's Church in Gdańsk*'. DOI: 10.34808/s2bg-w279.

Kotłowski, G. and Starek, E. (2020). '*Epitaph of James Schadius in St. Mary's Church in Gdańsk*'. DOI: 10.34808/hj1m-0764.

Kotłowski, G. and Starek, E. (2020). '*Epitaph of John Schroeder in St. Mary's Church in Gdańsk*'. DOI: 10.34808/ryg8-p650.

Kotłowski, G. and Starek, E. (2020). '*Epitaph of Oehm family in St. Mary's Church in Gdańsk*'. DOI: 10.34808/ngvf-ar66.

Kotłowski, G. and Starek, E. (2020). '*Plaque of Dorothy and John Brandes in St. Mary's Church in Gdańsk*.' DOI: 10.34808/f6fc-4e72.

Lewis, C.T. (1879) '*A Latin dictionary: Founded on Andrews' ed. of Freund's Latin dictionary*'. Rev., enl., and in great part rewritten. Oxford: Clarendon Press. Available at: http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3atext%3a1999.04.0059 (Accessed: 2nd January 2021).

Lipsius, I. et al. (1660) '*Apparatus philologiae, sive Iusti Lipsii Orthographia secundum iam edita, cum auctuario et opusculis aliis*'. Edited by Johann M. Dilherr. Nürmberg: Tauber, Johann; Felsecker, Wolfgang E.

Manuzio, A., Manuzio, P. and Keller, C. (1738) 'Epitome ortographiae Aldi Manutii Paulli filii, Aldi nepotis. Ex libris antiquis, grammaticis, etymologia, Graeca consuetudine, nummis veteribus, tabulis aureis, lapidibus: Cui in hac editione praefixa est tota prior pars Orthographiae Latinae Christophori Cellarii, regulas generales continens et quaestiones disceptans; inserta vero opportuna additamenta ex altera eiusdem parte, quae de vocibus singillatim agit, deprompta. Ad usum scholarum Veronensium'. Verona: Berno, Pier Antonio.

Manuzio, Aldo (II), Manuzio, P. and Aldo Manuzio (I) (1566) 'Orthographiae ratio ab Aldo Manutio Paulli filio collecta ex libris antiquis, grammaticis, etymologia, Graeca consuetudine, nummis veteribus, tabulis aereis, lapidibus amplius MD. Interpungendi ratio, notarum veterum explanatio, Kalendarium vetus Romanum e marmore descriptum cum Paulli Manutii, patris, commentariolo, de veterum dierum ratione et Kalendarii explanatione; Aldi Manutii avi De vitiata vocalium et diphthongorum prolatione parergon'. Vene-

zia: Manuzio, Aldo II. Available at: http://digital.onb.ac.at/OnbViewer/viewer.faces?doc=ABO_%2BZ220110408 (Accessed: 2nd January 2021).

Michalska, M.M. (2019) 'Udostępnianie rękopisów i starodruków w Bibliotece Narodowej', *Rocznik Biblioteki Narodowej*, 50, pp. 107–119. DOI: 10.36155/RBN.50.00006.

Oelhaf, P. et al. (1645-1653) '*Typus et Syllabi lectionum publicarum et operarum Athenaei Gedanensis*'. Danzig (a set of old prints from the collection of Polish Academy of Sciences Gdańsk Library, reference number: Ma 3920 8o, items 106–109).

Pokrzywnicki, J. (2019) '*Hanow – Praecepta de arte disputandi – transcription and photographs*'. DOI: 10.34808/01f6-n909.

Pokrzywnicki, J. (2020) '*Rhetoric at school – a selection of the syllabi from the Academic Gymnasium in Gdańsk – transcription and photographs*'. DOI: 10.34808/823g-1y16.

Polska Akademia Nauk Biblioteka Gdańska (2019a) '*PAN Biblioteka Gdańska. Regulamin korzystania z Czytelni Zbiorów Historycznych*'. Gdańsk. Available at: http://bgpan.gda.pl/wp-content/uploads/2019/03/Regulamin-Czytelni-Zb-Hist__popr.pdf (Accessed: 2nd January 2021).

Polska Akademia Nauk Biblioteka Gdańska (2019b) '*Zgoda na wykorzystanie kopii cyfrowych obiektów z zasobów PAN Biblioteki Gdańskiej*', 8th March.

Quintilianus, M.F. (1996) '*The institutio oratoria of Quintilian: In four volumes*'. (The Loeb classical library, 126). Cambridge, Mass.: Harvard Univ. Press. Available at: https://archive.org/details/institutioorator03quinuoft/page/n7/mode/2up (Accessed: 8th January 2021).

Sallmann, K. (1992) '*Normae orthographicae et orthotypicae Latinae. Zasady ortografii łacińskiej (do stosowania w piśmie i druku)*'. Przekład polski: Marian Plezia, Meander. Rocznik poświęcony kulturze świata starożytnego, 47(9–10), [441] – 457.

Starek, E. and Kotłowski, G. (2019) '*Epitaph of John Sinclair in St. Mary's Church in Gdańsk*'. DOI: 10.34808/shrp-gx28.

Starek, E. and Kotłowski, G. (2020a) '*Epitaph of Anna and Daniel Zierenberg in St. Mary's Church in Gdańsk*'. DOI: 10.34808/jv8e-tx54.

Starek, E. and Kotłowski, G. (2020b) '*Epitaph of Anna Loitz in St. Mary's Church in Gdansk*'. DOI: 10.34808/swz2-6712.

Starek, E. and Kotłowski, G. (2020c) '*Epitaph of Bartholomew Schachmann in St. Mary's Church in Gdańsk*'. DOI: 10.34808/x0p2-q577.

Starek, E. and Kotłowski, G. (2020d) '*Epitaph of Clemens and Gabriel Coelmer in St. Mary's Church in Gdańsk*'. DOI: 10.34808/egb7-dt65.

Starek, E. and Kotłowski, G. (2020e) '*Epitaph of Eggert von Kempen in St. Mary's Church in Gdańsk*'. DOI: 10.34808/p5z7-5q34.

Starek, E. and Kotłowski, G. (2020f) '*Epitaph of George Wildtberger in St. Mary's Church in Gdańsk*'. DOI: 10.34808/g76j-se78.

Starek, E. and Kotłowski, G. (2020g) '*Epitaph of John Botsack in St. Mary's Church in Gdańsk*'. DOI: 10.34808/xp84-rm12.

Starek, E. and Kotłowski, G. (2020h) '*Epitaph of Martin Ramus in St. Mary's Church in Gdańsk*'. DOI: 10.34808/jyp0-8e44.

Starek, E. and Kotłowski, G. (2020i) '*Foundation text of St. Mary's Church in Gdańsk*'. DOI: 10.34808/secj-9y72.

Starek, E. and Kotłowski, G. (2020j) '*Sepulchral plate of Dorothy and Matthias Zimmermann in St. Mary's Church in Gdańsk*'. DOI: 10.34808/sgwt-z104.

Starek, E. and Kotłowski, G. (2020k) '*Sepulchral plate of James von Werden in St. Mary's Church in Gdańsk*'. DOI: 10.34808/a9pm-gh43.

Starek, E. and Kotłowski, G. (2020l) '*Sepulchral plate of Thomas Tympfius in St. Mary's Church in Gdańsk*'. DOI: 10.34808/5x5z-zq54.

Starek, E. and Kotłowski, G. (2020m) '*Tombstone of Simon Bahr in St. Mary's Church in Gdańsk*'. DOI: 10.34808/k79m-mt87.

Szymański, J. and Trelińska, B. (2003) '*Instrukcja wydawnicza dla źródeł epigraficznych*'. Lublin: Wydaw. UMCS.

Węcowski, J. (2010) '*Słownik skrótów łacińskich*'. FAN MEDIA: Warszawa.

Winiarczyk, M. (1995) '*Sigla latina in libris impressis occurentia: Cum siglorum graecorum appendice = Skróty łacińskie w książkach drukowanych: z dodatkiem skrótów greckich*'. 2nd edn. Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Surf Zone Currents in the Coastal Zone of the Southern Baltic Sea – a Modelling Approach

Aleksandra Dudkowska[1*], Aleksandra Cupiał[1]

[1] Division of Physical Oceanography, University of Gdańsk (8 Jana Bażyńskiego Street, Gdańsk, Poland)

* Correspondence author: aleksandra.dudkowska@ug.edu.pl ; ORCID: 0000-0001-9781-8826

Abstract

Nearshore currents in a multi-bar non-tidal coastal zone environment located in the Southern Baltic Sea are studied. Spatiotemporal seaward-directed jets – so-called rip currents – are an important part of the nearshore current system. In previous research, Dudkowska et al. (2020) performed an extended modelling experiment to determine the wave conditions that are conducive to the emergence of rip currents. In this paper, the collected dataset is presented to facilitate the reuse of this data for other purposes.

Keywords: Southern Baltic Sea; coastal currents; rip currents

https://doi.org/10.34808/x55q-sz53_dyr_roz6

Introduction

The Baltic Sea is a non-tidal inner sea, thus surf zone currents are generated mostly by waves. The dominating part of the current system is the so-called longshore current, and the other important part is the undertow. In such hydrodynamic conditions, a multi-bar seabed forms. Apart from these two typical flows in some specific hydro- and lito-morphological conditions, so-called rip currents can occasionally form, which are seaward-directed quasi-steady jets. Flows of this type may initiate a recess in the bar and even form a channel in some specific conditions. This type of rip is called a channel rip current and is the most commonly documented in barred surf zones (Castelle et al., 2016)

The most widely used method of observation of such kind of flows in the environment is carried out using free-floating drifters placed on the surface of the water. Such measurements, which belong to the group of Lagrangian measurements, were carried out on diverse coasts, for example in Western Australia (Johnson and Pattiaratchi, 2004), and along the coast of the Gulf of Guinea (Castelle et al., 2016; Floc'h et al., 2018). In the

Southern Baltic coastal zone, the occurrence of rip currents was confirmed by Schönhofer in 2014, who designed and used the drifters shown in Fig. 6.1 (Schönhofer and Dudkowska, 2021). In the series of Lagrangian measurements performed during field surveys at Coastal Research Station (CRS) Lubiatowo, he registered several cases of a flow with features characteristic of rip currents. The same drifters were used in subsequent research in this region, described in (Dudkowska et al., 2020).

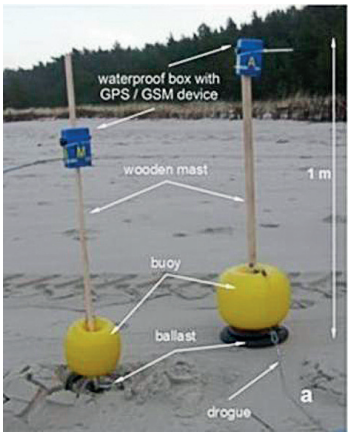


Fig. 6.1. Drifters designed for rip current measurement (Schönhofer and Dudkowska, 2021)

With the current technique, Lagrangian measurements in storm conditions are very difficult and perhaps impossible. Therefore, it is problematic to study near-shore flows at the site in a wide range of water and morphological conditions. Dudkowska et al. (2020) applied the modelling approach to comprehensively investigate this phenomenon. They carried out simulations for all wave conditions likely in the studied region. The range of the sea states was determined based on long-term buoy measurements. The tracks of virtual drifters were simulated by the XBeach numerical model (Roelvink et al., 2010). Another numerical model – SWAN (Booij, Holthuijsen and Ris, 1996) – was used for the purpose of calibration. The dataset published in the MOST Wiedzy repository (The Bridge of Knowledge) contains all of the data collected during this study. It was used to determine the incidence of rip currents on the Polish coast of the Baltic Sea. However, the modelled flows can be the basis or inspiration for further research into the characteristics of flows in the surf zone. The data from stored files can also serve as boundary conditions for detailed modelling of flows on nested grids.

Study area

The research was conducted for an area which is located in the southern Baltic Sea, in Polish Marine Areas adjacent to the coastline in the vicinity of the village of Lubiatowo. Hydrodynamic and morphodynamic phenomena in this area are well studied as The Coastal Research Station (CRS) – field laboratory of the Institute of Hydro-Engi-

neering of the Polish Academy of Sciences (IBW PAN) – is situated here. This part of the coast is considered to be a representative part of the southern Baltic coast. The nearshore bathymetry is characterised by: the gentle slope of the seabed of about 1.5%, locally near the coastline up to 4%; the occurrence of usually five offshore bars: four fixed and one temporarily, at a distance from the coastline approximately: 100–120 m, 200 m, 400–450 m, 650–850 m. The typical bathymetry for the Lubiatowo area was used for modelling (Fig. 6.2). The main depth of the water is about 8 m at 1 km from the shoreline and increases to about 15–17 m at 2 km and up to 25 m at 9 km.

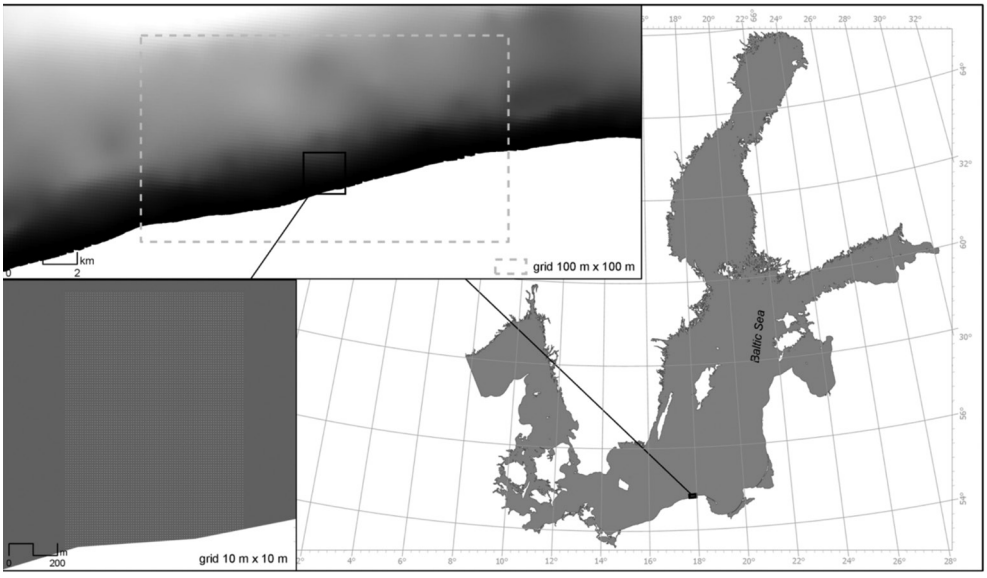


Fig. 6.2. Area of interest and the computational meshes used in numerical modelling: SWAN external grid 100 m ×100 m; SWAN (internal) and XBeach grid 10 m ×10 m (Dudkowska et al., 2020)

Methods

The numerical reconstruction of the coastal flow was performed with the use of the following spectral numerical wave models: SWAN (Simulating WAves Nearshore) cycle III (Booij, Holthuijsen and Ris, 1996) and XBeach (Roelvink et al., 2010). The XBeach model solves coupled 2D horizontal equations for wave propagation, flow, sediment transport, and bottom changes. The applied model setting are: (i) neglecting of sediment transport and bottom changes, (ii) wave-current interaction option, (iii) stationary wave boundary condition at the seaward boundary, which means that a uniform, constant wave energy distribution is set, based on the given values of wave height and period; Neumann lateral boundaries, (iv) the roller model which can provide a shoreward shift in a wave-induced

setup, return flow and longshore current. In order to simulate the rip channel, the gap in the bar was artificially introduced into the bathymetry. The presented dataset consists of the output of the XBeach model simulations (xboutput.nc) for the wave conditions described. The model parameters are included in text file (params.txt).

Fig. 6.2 shows the two numerical meshes used for the calculations. (1) SWAN external grid, with generalised bathymetric data over the area between the coastline and a certain location offshore; spatial resolution: 100 m x 100 m. (2) Internal grid, used both for SWAN and XBeach, distributed over the area up to 1 km offshore with detailed bathymetric data; spatial resolution: 10 m × 10 m. All SWAN parameters were chosen according to the previous extended studies, including the model configuration and results validation in the area of study (Gic-Grusza and Dudkowska, 2014).

Two integral wave parameters, calculated form the wind wave spectrum, are used in this work. The first one, significant wave height (Hs), is a statistical parameter that is used as a representation of the state of the sea at given place and time. Calculated from the energy variance of the spectrum (by means of a zero-order spectral moment), it corresponds to the average height of the highest one-third of all of the waves in the given time interval. In the most general terms, this parameter might be interpreted as the wave height as seen by a human observer. The second considered wave parameter, mean wave period (T), is calculated from the spectral moments of higher order. It represents the averaged time interval between the instances of two consecutive wave peaks (or other corresponding points on the wave, i.e. troughs) passing through a certain point in space. In other words, by corresponding to the average time of upward or downward zero crossing in an analysed time series, this parameter provides insight into the present state of the sea as well as the speed of wave propagation.

Data

Data stored in the Bridge of Knowledge, Research Data Catalog repository are in the NetCDF format. NetCDF is a commonly used data format that allows for easy storage and transfer of large sets of multidimensional data. Information is stored in binary format in the form of multidimensional arrays. Apart from the data itself, NetCDF files have a header with metadata, which makes them self-describing. In a NetCDF file, one can find the following components: dimensions, variables and attributes. Time and geographic coordinates as well as station numbers can be dimensions; variables store actual data: i.e. millimetres of rainfall, height of the significant wave in metres, wind speed in metres per second, etc.; attributes describe individual variables as well as the whole NetCDF file. Information about the dimensions, variables and attributes is stored in a header. Even though the files can contain large datasets, NetCDF allows for access of a selected subset of data without the need to read the entire file. The NetCDF format is independent of the operating system and program that is used to open and analyse the files.

This format is widely used on various platforms that share all kinds of scientific data, from satellite observations, wave data measured in situ with buoys, to meteorological forecasts, and so on. It is commonly used by spatial analysts working with geographic information systems (GIS). This data format, which is built from software libraries for data management and storage and an accompanying interface, was created at the University Corporation for Atmospheric Research (UCAR) which curates its development (Esri 2020, Physical Sciences Laboratory NOAA 2020, Unidata 2020).

Results

The drifters' movement, which depends on the direction of the wave propagation, can be seen in Fig. 6.3 and Fig. 6.4. These figures represent two exemplary wind wave conditions, described by pairs Hs = 4 m and T = 9 s, and Hs = 1 m and T = 7 s, respectively. The red arrows show the direction of the wave propagation. As expected, the velocity of the drifters increases faster and reaches higher values during more severe conditions (Hs = 4 m). When waves propagate parallel to the shore, the modelled drifters follow the direction of wave propagation independently of the modelled conditions (left columns of Fig 6.3 and Fig. 6.4). During more severe conditions, the maximum drifter velocity is 1.5 m/s. Such drifter velocity is simulated by model runs with wave propagation directions between NNW and NWW (panels 2–4 in the left column, and panels 1–2 in the right column of Fig. 6.3). When the direction of propagation is set up as perpendicular to the shore, the drifter's velocities, especially for the drifter that is released in the vicinity of the rip channel, are lower or reach higher values with a delay. In the 4th and 5th panel in the right column, the drifter released in the vicinity of the rip channel moves off shore before turning and changing direction towards the shore, although it still moves at a certain angle to the shore, unlike the two remaining drifters that follow the direction of wave propagation.

During calmer conditions, the highest velocity drifters reach around 0.5 m/s (Fig. 6.4). The lowest velocities, below 0.5 m/s, occur in the instances of wave propagating perpendicular to the shore (directions: NNE to N, or panels 2–4 in the right column of Fig. 6.4). In panels 2–5 in the right column, the drifter released in the vicinity of the rip channel moves off shore before turning and moving alongshore. In one instance (2nd panel) the middle drifter is only visible moving away from the shore. In panels 3–5, all drifters are driven by another/additional force than just wave propagation, since they are not moving with the direction of wave propagation. In the 5th and, in part, in the 4th panel, the blue and red drifter move alongshore.

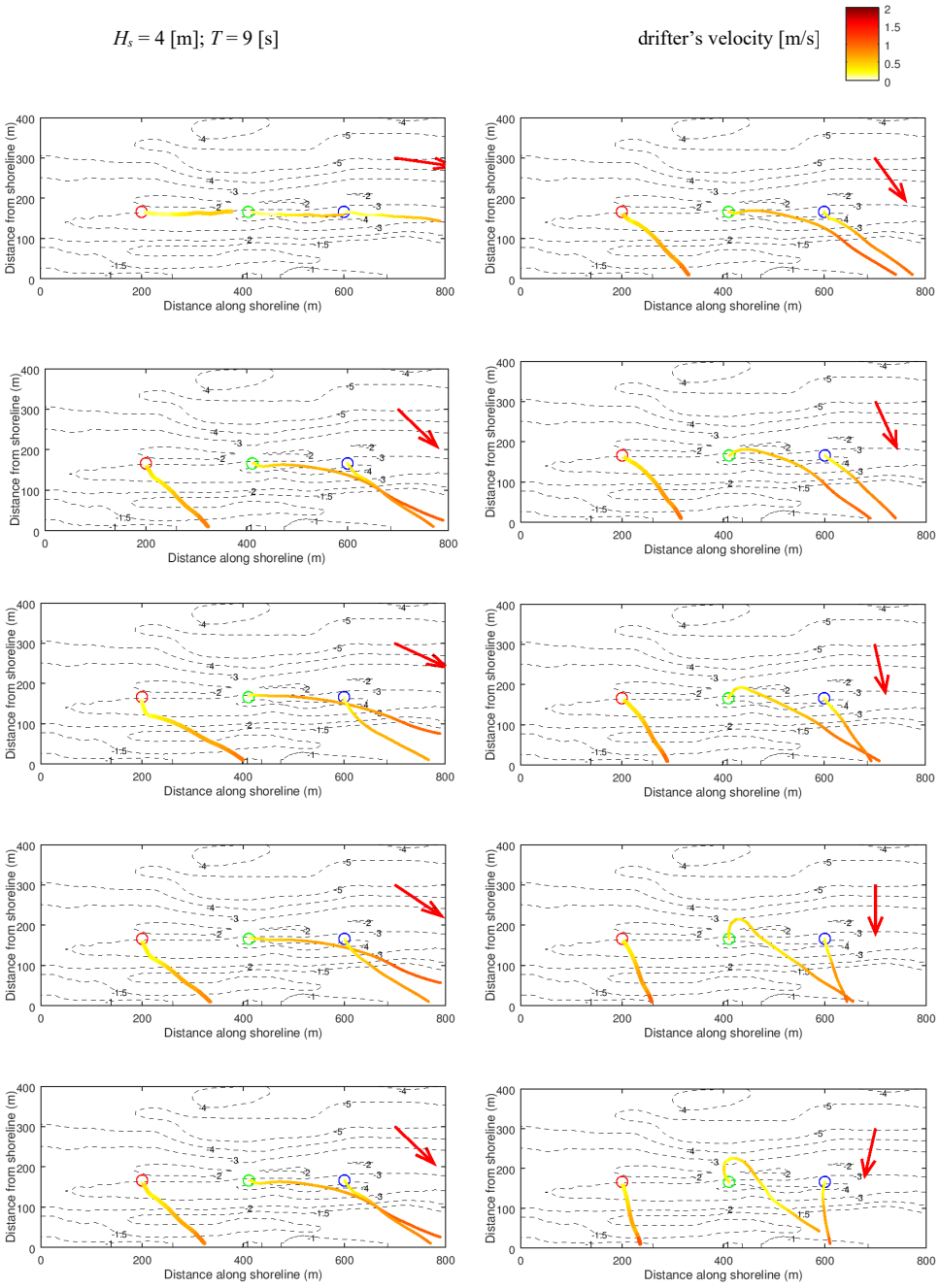


Fig. 6.3. Tracks of virtual drifters obtained by numerical simulation, example 1

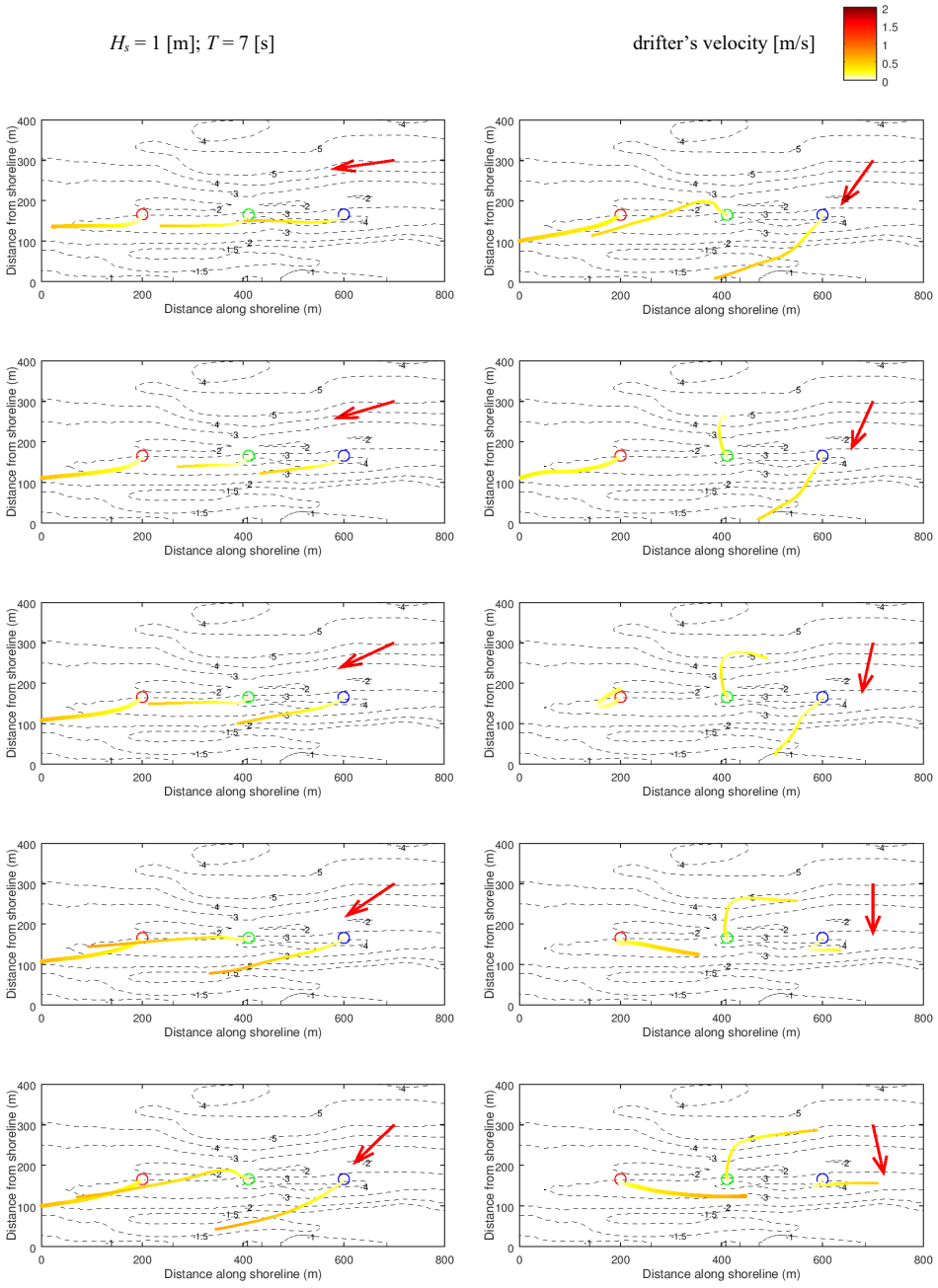


Fig. 6.4. Tracks of virtual drifters obtained by numerical simulation, example 2

References

Booij, N., Holthuijsen, L.H. and Ris, R.C. (1996) 'The SWAN wave model for shallow water', In: *Proceedings of the 25th International Conference on Coastal Engineering*, Orlando, USA, 1, pp. 668–676. DOI: 10.1061/9780784402429.053.

Castelle, B. et al. (2016) 'Rip current types, circulation and hazard', *Earth-Science Reviews*, 163, pp. 1–21. DOI: 10.1016/j.earscirev.2016.09.008.

Cieślikiewicz, W. et al. (2017) 'Extreme bottom velocities induced by wind wave and currents in the Gulf of Gdańsk', *Ocean Dynam*ics, 67(11), pp. 1461–1480. DOI: 10.1007/s10236-017-1098-4.

Dudkowska, A. et al. (2020) 'Rip currents in the non-tidal surf zone with sandbars: numerical analysis versus field measurements', *Oceanologia*, 62(3), pp. 291–308. DOI: 10.1016/j.oceano.2020.02.001.

ESRI (2020) '*What is netCDF data?*', viewed 30th December 2020, Available at: https://pro.arcgis.com/en/pro-app/latest/help/data/multidimensional/what-is-netcdf-data.htm (Accessed: 25th May 2022).

Floc'h, F. et al. (2018) 'Flash rip statistics from video images', *Journal of Coastal Research*, 81, pp. 100–106. DOI: 10. 2112/SI81-013.1.

Gic-Grusza, G. and Dudkowska, A. (2014) 'Modeling of wind wave induced sediment transport in the coastal zone of polish marine areas (southern Baltic), Baltic International Symposium (BALTIC)', *2014 IEEE/OES: Measuring and modeling of multi- scale interactions in the marine environment*, 26–29th May 2014, Tallinn, Estonia IEEE/OES, pp. 1–5. DOI: 10.1109/ BALTIC.2014.6887860.

Johnson, D. and Pattiaratchi, C. (2004) 'Transient rip currents and nearshore circulation on a swell-dominated beach', *Journal of Geophysical Research: Oceans*, 109, C02026. DOI: 10.1029/2003jc001798.

Physical Sciences Laboratory NOAA (2020) '*Data Resources: What is netCDF?*', Available at: https://psl.noaa.gov/data/gridded/whatsnetCDF.html (Accessed: 25th May 2022).

Roelvink, D. et al. (2010) '*XBeach Model Description and Manual, Delft University of Technology*', User Manual, Delft, The Netherlands, Available at: https://xbeach.readthedocs.io/en/latest/user_manual.html (Accessed: 25th May 2022).

Schönhofer, J. and Dudkowska, A. (2021) 'Rip currents in the southern Baltic Sea multi-bar nearshore zone', *Continental Shelf Res*earch, 212, 104324. DOI: 10.1016/j.csr.2020.104324.

Unidata (2020) '*Network Common Data Form (NetCDF)*', Available at: https://www.unidata.ucar.edu/software/netcdf/ (Accessed: 25th May 2022).

Using Synchronously Registered Biosignals Dataset for Teaching Basics of Medical Data Analysis – Case Study

Tomasz Kocejko[1*]

[1] Department of Biomedical Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: tomkocej@pg.edu.pl; ORCID: 0000-0002-2304-8942

Abstract

Medical data analysis and processing strongly relies on the data quality itself. The correct data registration allows many unnecessary steps in data processing to be avoided. Moreover, it takes a certain amount of experience to acquire data that can produce replicable results. Because consistency is crucial in the teaching process, students have access to pre-recorded real data without the necessity of using additional equipment for data acquisition. The analyzed sample dataset consists of raw signals of ECG, Body Impedance and Body temperature recorded synchronously in laboratory conditions. The data are sampled with 250Hz sampling frequency and are framed in blocks. Students gain a chance to acquire, exchange and process the medical data in simulated conditions. Pre-recording data provides the opportunity to teach certain techniques that can be used in real life scenarios but in a control replicable environment.

Keywords: biosignals processing; ECG, bioimedance; synchronous measurements; education

https://doi.org/10.34808/x55q-sz53_dyr_roz7

Background

This case study shows how pre-recorded data can be used in the course of teaching basic algorithms of biosignal processing. The data were used in a course of a medical data processing and analysis in its laboratory part. The purpose of the laboratory exercises is to prepare students for a different tasks and problems which they may encounter on the job. In the case of biomedical signal processing, these tasks can by roughly divided into

two categories: signal measurements and signal acquisition and processing. Biomedical engineers with a specialty in medical informatics should focus on acquisition and processing the most. The biomedical measurement part should be covered by medical professionals. Therefore, the laboratory exercises were built on top of the data pre-recorded in a controlled environment, maintaining focus on the precision of the measurements and reliability of the data.

The Data

The data are a complementary part of the experiment designed to demonstrate how to use network protocols to transmit, receive and process medical data. The dataset contains biomedical signals of ECG, impedance and temperature acquired simultaneously at a 250Hz sampling rate.

Scope of the laboratory

The whole laboratory course is divided into four stages: Data Transmission, Data Acquisition, Data Visualisation and Data Processing. After completion of this four stages, participants can take part in a challenge relying on solving a real life problem related to the biomedical signals (like detecting the heart rate based on the ECG signal). In every stage, students become familiar with code and techniques necessary to write the software in C++. Although there are more suitable programming languages for signal processing (such as Octave/Matlab or Python), the goal of this course is to prepare students to write an end-to-end application for biosignal acquisition and processing that can be deployed on embedded devices. Therefore the recommended programming language is C++ with the QT framework ([https://www.qt.io/).

Data Transmission and Acquisition

During the class, the data (ECG, impedance and temperature) are broadcast using the UDP protocol. This simulates obtaining the data directly from the measuring devices over popular protocols like TCP/IP, UDP (Zubairi, Misbahuddin and Tasadduq, 2010) (Lamberti and Sanna, 2005), bluetooth (Mulyadi et. al., 2009) or ZigBee (Jung and Lee, 2008). Most often, such devices send data bit by bit or as information containing more than one character. The students are obliged to become familiar with the technology, protocols and algorithms used for: both data transmission and data acquisition. They also have to create software for sending and receiving the data.

The essential parts of programming are demonstrated on diagrams and followed by exemplary code like is presented in Fig. 7.1.

Fig. 7.1. Exemplary code for communication using UDP protocol (QT C++)

The broadcast information is wrapped in a certain data frame:
<Impedance, ECG, Temp>
where:

"<" is the beginning of the data frame,
">" is the end of the data frame
" , " is a separator
"ECG" is the ECG signal amplitude
"Impedance" is the impedance signal amplitude
"Temp" is the amplitude of the temperature

In binary form, a single batch of data is 7 bytes (56 bits). The first 24 bits are impedance samples, the next 24 bits are ECG, the last 8 bits are samples acquired by means of a thermistor.

Data Visualisation and Processing

Knowing the data frame, students have to write software to collect the data belonging to a certain biosignal and display them in real time. They also have to create a buffer for each signal.

This task requires writing a software program that will record data into a buffer of a given length. Students are presented with a simple example, like the one presented in Fig.7.2, of how to continuously display a sin wave transmitted from a function generator. On top of that software, they have to write their own solution for biosignal visualisation.

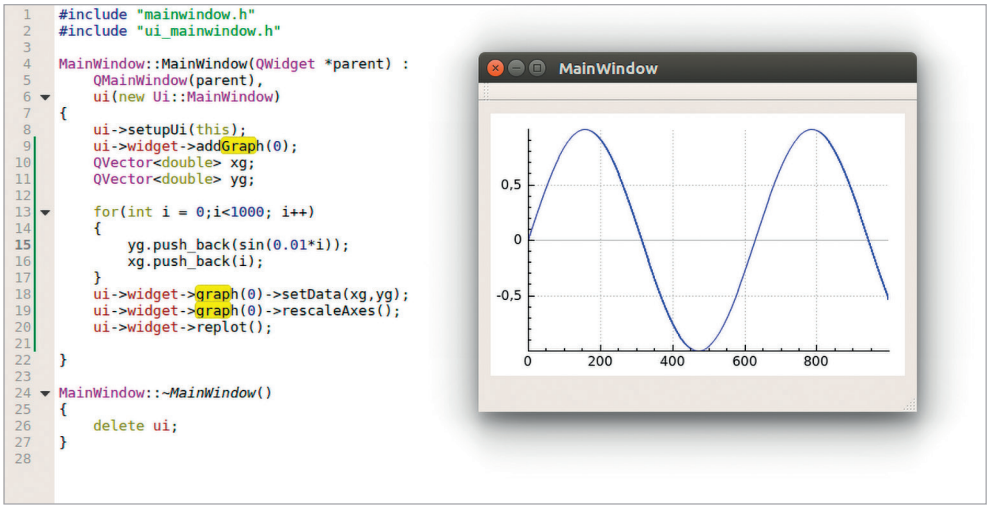


Fig. 7.2. Example of data visualisation using a simple buffer

Assuming that at this point students are familiar with the data structure, with the data frame and know how to split data and collect them using buffers, it is time for signal processing. This sometimes requires additional libraries for signal processing. In the described case study, students learn how to use the Armadillo C++ library for linear algebra and scientific computing [6]. Using additional libraries often requires changing the data format for the one suitable for signal processing. In this case, it is also desirable to perform signal filtration in the frequency domain. Students are provided with exemplary software for Fast Fourier Transform – FFT (Fig. 7.3) and precise comments on how to modify it for use with the available dataset.

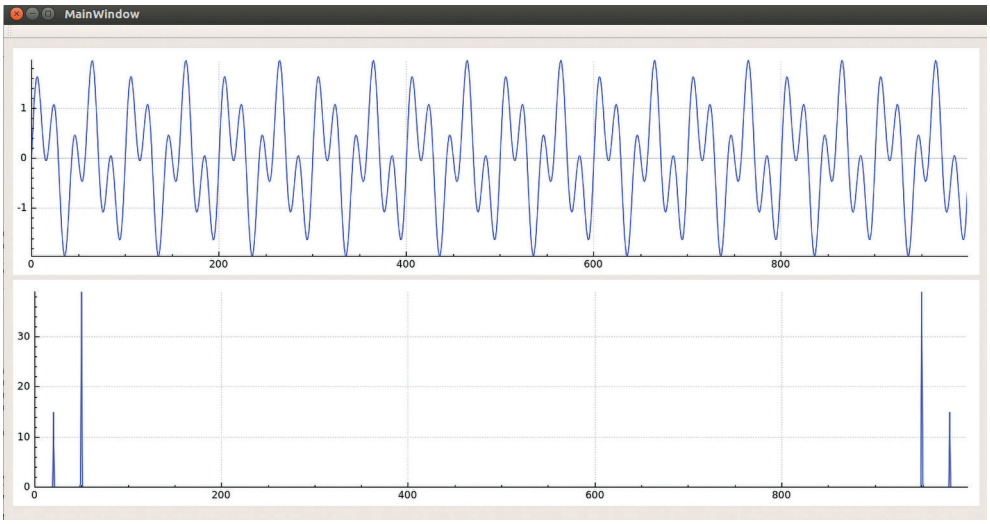


Fig. 7.3. Example of signal processing

Conclusions

Dealing with biosignals is difficult as they are prone to all sorts of noise and distortions. Using the data pre-recorded in a control environment gives students a chance to focus on the algorithms. In the described case, data were broadcast to create the illusion that they came directly from measuring equipment (such as an ECG). The teaching process is more consistent due to the fact that all examples refer to the same case and to the same data. It also helps to compare results between students and different groups.

Data quality and availability

Dataset DOI:

10.34808/x3f9-fh19

Dataset License

CC-BY-NC

References

[https://www.qt.io/] (Accessed: February 2021)

Jung, J.Y., and Lee, J.W. (2008) 'ZigBee device access control and reliable data transmission in ZigBee based health monitoring system'. In 2008 *10th International Conference on Advanced Communication Technology*, pp. 795–797, DOI: 10.1109/ICACT.2008.4493875.

Lamberti, F., and Sanna, A. (2005) 'A solution for displaying medical data models on mobile devices', 4th WSEAS International Conference on Software Engineering, Parallel & Distributed Systems, 16, pp. 1–7. Available at: https://dl.acm.org/doi/abs/10.5555/1365774.1365790 (Accessed: 25th May 2022).

Mulyadi, I.H. et al. (2009) 'Wireless medical interface using ZigBee and Bluetooth technology'. In *2009 Third Asia International Conference on Modelling & Simulation*, pp. 276–281, DOI: 10.1109/AMS.2009.134.

Zubairi, J.A., Misbahuddin, S., and Tasadduq, I. (2010) 'Emergency medical data transmission systems and techniques'. In *Handbook of Research on Advances in Health Informatics and Electronic Healthcare Applications: Global Adoption and Impact of Information Communication Technologies*, pp. 169–188, DOI: 10.4018/978-1-60566-030-1.ch011.

3

Part

Data Descriptors

Data Descriptor

Educational Dataset of Handheld Doppler Blood Flow Recordings

Tomasz Kocejko[1*]

[1] Department of Biomedical Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: tomkocej@pg.edu.pl; ORCID: 0000-0002-2304-8942

Abstract

Vital signals registration plays a significant role in biomedical engineering and education process. Well acquired data allow future engineers to observe certain physical phenomena as well learn how to correctly process and interpret the data. This dataset was designed for students to learn about Doppler phenomena and to demonstrate correctly and incorrectly acquired signals as well as the basic methods of signal processing. This paper presents three corresponding datasets consisting of 21 recording of signals acquired from the neck area with USG gel applied and 21 distorted recordings acquired without gel.

Keywords: Doppler; biosignals; education

https://doi.org/10.34808/x55q-sz53_dyr_roz8

Specification table (data records)

Subject area	Biosiognals processing
More specific subject area	Education on biosignals processing based on audio recordings
Type of data	raw recordings time series time-frequency domain
How the data was acquired	Three data sets were created, a dataset containing RAW recordings from a portable Doppler blood flow detector, a dataset containing the time series representation of the recorded signals, and finally a dataset containing the time-frequency representation of the recorded signals.

Data format	".wav" (raw recordings) ".csv" (time series) ".jpg" (time-frequency domain)
Data source location	MOST Wiedzy Open Research Data Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	CC BY-NC

Background

Handheld Doppler allows blood flow to be measured and the heart rate to be assessed. In comparison to other medical devices (such as electronic fetal monitors) these devices are easy to operate. It was observed that a handheld Doppler monitor causes less maternal discomfort than the Pinard foetal stethoscope and therefore its reliability is being evaluated for foetal heart rate monitoring (Kamala et al., 2018), (Dyson, Jeffrey and Kluckow, 2017), (Byaruhanga et al., 2015). The simplicity and availability of handheld Doppler monitors makes it a good source of vital sign measurements.

Vital signals registration plays a significant role in the biomedical engineering and education process. Medical professionals rely on valid and reproducible information for their clinical assessments. The validity of the measurement reflects the ability to provide information on the assessed value of real phenomena, while the reproducibility refers to the ability to perform repeated measurements of the phenomenon (Fletcher R., Fletcher S. and Fletcher G., 2012), (Worster et al., 2003). Well acquired data allow future engineers to observe certain physical phenomena as well learn how to correctly process and interpret the data. However, data processing and data acquisition often require different sets of skills. Biomedical engineering professionals should be well trained in data processing. Therefore it is crucial for their education to provide reliable data recorded in a control environment. It reduces potential distortion within the data that can appear during data recordings.

Methods

The data were acquired by means of a portable Doppler blood flow detector. Utilised device allows a fast assessment of blood flow in the peripheral vessels to be performed, including in a patient with a weak pulse. The used device was equipped with a universal 5 MHz ultrasound probe. The data were recorded from the neck area with and without gel. The raw signals were recorded through line-in input/output with a 22050Hz sampling rate. These data were recorded for educational purposes and are intended for biomedical engineering students and students interested in data science. The sole purpose of these data is to demonstrate the difference between correctly and incorrectly recorded data and to show basic signal processing methods.

Data records

This data set was designed for students to learn about Doppler phenomena and to demonstrate correctly and incorrectly acquired signals. Three data sets were created,

a dataset containing RAW recordings from a portable Doppler blood flow detector, a dataset containing the time series representation of the recorded signals, and finally a dataset containing the time-frequency representation of the recorded signals. This dataset contains a visual representation of the performed analysis. The files are stored in '.jpg' file format. The length of each window segment was set to 200; overlapping was introduced. The overlap between windows was set to 120.

The first dataset contains the raw recording. Each recording contains 20 seconds of well established signal. The whole set consists of 21 recording of signals acquired from the neck area with USG gel applied and 21 distorted recordings acquired without gel. The raw data are stored in .wav file format.

The time series are stored in .csv file format. Samples are separated with commas (the ',' character). The time-frequency analysis has a graphic representation stored in .jpg file format.

The exemplary wave form representation is presented in Fig. 8.1 while Fig. 8.2 shows its time-frequency analysis.

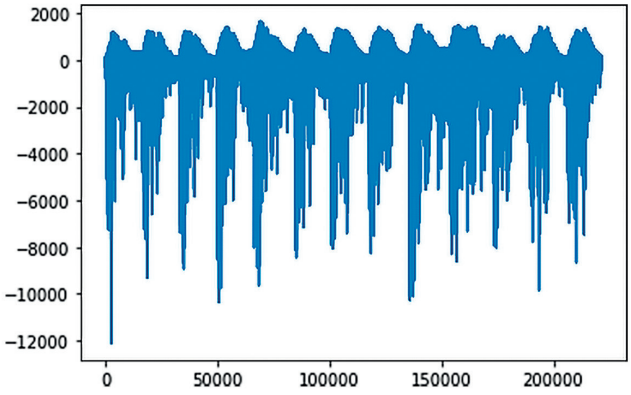


Fig. 8.1. Waveform of exemplary recording

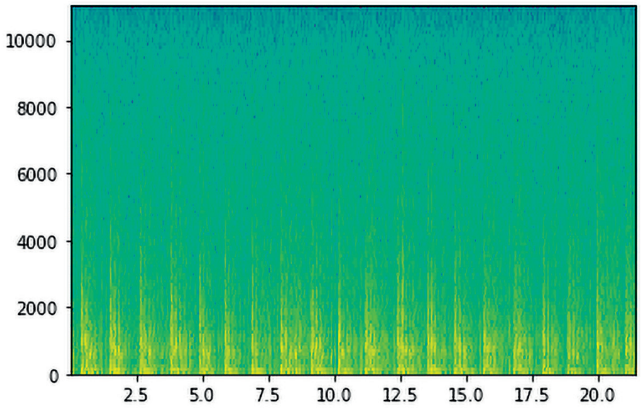


Fig. 8.2. Time-frequency analysis of exemplary recording

Data quality and availability

Dataset DOI
10.34808/nmgg-xm45
10.34808/z8vj-sg17
10.34808/hx3e-nh64

Dataset License
CC-BY-NC

References

Kamala, B.A. et al. (2018) 'Intrapartum fetal heart rate monitoring using a handheld Doppler versus Pinard stethoscope: a randomized controlled study in Dar es Salaam', *International Journal of Women's Health*, 10, pp. 341–348, DOI: 10.2147/IJWH.S160675.

Dyson, A., Jeffrey, M., and Kluckow, M. (2017) 'Measurement of neonatal heart rate using handheld Doppler ultrasound', *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 102(2), pp. F116–F119. DOI: 10.1136/archdischild-2016-310669.

Byaruhanga, R. et al. (2015) 'Use of wind-up fetal Doppler versus Pinard for fetal heart rate intermittent monitoring in labour: a randomised clinical trial', *BMJ Open*, 5(1), e006867, DOI: 10.1136/bmjopen-2014-006867.

Fletcher, R.H., Fletcher, S.W., and Fletcher, G.S. (2012) '*Clinical epidemiology: the essentials*' Lippincott Williams & Wilkins.

Worster, A. et al. (2003) 'Reliability of vital signs measured at triage', *European Journal of Emergency Medicine*, 10(2), pp. 108–110. DOI: 10.1097/00063110-200306000-00008.

Data Descriptor

Testing the Diagnostic Utility of Recombinant *Toxoplasma Gondii* Chimeric Antigens – Generated Datasets

Bartłomiej Ferra[1*]

[1] Department of Molecular Biotechnology and Microbiology; Faculty of Chemistry, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: barferra@pg.edu.pl; ORCID: 0000-0003-2176-3966

Abstract

The datasets titled *Toxoplasma gondii* recombinant chimeric antigens – IgM and IgG ELISAs – mouse serum samples and *Toxoplasma gondii* recombinant chimeric antigens – IgG and IgM ELISAs – human serum samples contain absorbance measurements obtained during serological tests using mouse and human sera in enzyme-linked immunosorbent assay (ELISA) tests based on recombinant chimeric antigens. The datasets allows a comparison of absorbance values obtained for individual recombinant chimeric antigens in relation to the whole *Toxoplasma* lysate antigens (TLA) used in commercial tests.

Keywords: *Toxoplasma gondii*, recombinant chimeric antigen, diagnostic utility, toxoplasmosis

https://doi.org/10.34808/x55q-sz53_dyr_roz9

Specification table (data records)

Subject area	Immunology; Parasitology; Diagnostics
More specific subject area	Serological diagnostics
Type of data	Text
How the data was acquired	The data was collected using a commercially available Multiscan FC (Thermo Scientific) microplate photometer
Data format	Mixed (raw and processed)

Experimental factors	The obtained measurement data were obtained for: healthy animals and within a specified period of time after infestation with the parasite patients in various stages of the disease and healthy people
Experimental features	The data was obtained directly in the laboratory and the averaged measurement values were used to perform the data analysis
Data source location	MOST Wiedzy Open Research Data Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Toxoplasmosis is a zoonotic infection with the protozoan parasite *Toxoplasma gondii*; which can affect most warm-blooded animals from mammals to birds all over the world. Human infection of *T. gondii* is generally asymptomatic, however, diagnosis of *T. gondii* infection is of great medical importance, especially for pregnant women and immunocompromised patients (e.g. AIDS and cancer patients) (Tenter, Heckeroth and Weiss, 2000). Primary infection of a pregnant woman is often associated with foetal infection, which can lead to miscarriage, stillbirth or severe neonatal malformation (Dunn et al., 1999; Fatoohi et al., 2002).

Routine diagnosis of T. gondii infection in humans is based on the detection of specific anti-*T. gondii* antibodies in serum samples. Serological techniques are suitable for the analysis of a large number of samples and play a major role in the diagnosis of toxoplasmosis. Among the available serological tests, ELISA has been adapted for the detection of IgM, IgG and IgA class antibodies. The serological methods for detection of specific *T. gondii* antibodies have usually involved the preparation of TLA, the production of which is expensive and laborious, moreover, a quantity of the antigen mixture is difficult to standardise. In recent years, the utility of recombinant antigens has been demonstrated in the diagnosis of toxoplasmosis in humans (Holec-Gąsior, 2013). The use of recombinant antigens for the diagnosis of *T. gondii* infection has proven highly beneficial for improving the standardisation of the method since the antigen composition of the test is precisely known. Furthermore, the advantages of these proteins are that the production cost of antigens is reduced and more than one defined antigen can be used for the detection of specific antibodies. A relatively new approach is the construction of a new generation of recombinant products so-called chimeric antigens which can replace native antigens from a lysed whole parasite.

The datasets titled *Toxoplasma gondii* recombinant chimeric antigens – IgG and IgM ELISAs – mouse serum samples and *Toxoplasma gondii* recombinant chimeric antigens – IgM and IgG ELISAs – human serum samples have been used to assess the diagnostic usefulness of recombinant chimeric antigens composed of immunodominant fragments of parasite antigens such as apical membrane antigen 1 (AMA1), surface antigen 2 (SAG2),

dense granule antigen 1 (GRA1), and rhoptry antigen 1 (ROP1). The first dataset was used to determine the humoral immune response dynamics during murine experimental toxoplasmosis using tetravalent recombinant chimeric proteins. The second dataset was used to determine the diagnostic usefulness of recombinant chimeric antigens in indirect IgM and IgG ELISAs with human serum samples. These datasets were used to develop the figures and tables included in the publication under the title The first study on the usefulness of recombinant tetravalent chimeric proteins containing fragments of SAG2, GRA1, ROP1 and AMA1 antigens in the detection of specific anti-*Toxoplasma gondii* antibodies in mouse and human sera (Ferra et al., 2019).

Methods

Ethical statements

The data contained in the datasets required obtaining appropriate approvals from ethical committees. Mouse serum samples were obtained according to experimental procedure guidelines provided by the Polish Local Ethics Commission for Experiments on Animals No. 9 in Lodz (Agreement 75/ŁB639/2012). The human serum samples were collected as part of the project entitled "Toxoplasmosis—facts and myths. Educational initiative raising social awareness about the infection with protozoan *Toxoplasma gondii*", the "Our Children" foundation, funded from the Civil Initiatives Fund of the Ministry of Labour and Social Policy (FIO 2008, contract No. 813). To ensure the anonymity of the subjects, only the date of serum collection and the individual's immune status in regard to anti-T. gondii specific antibodies and IgG avidity were disclosed.

ELISAs tests with mouse and human serum samples

To assess the reactivity of anti-T. gondii IgM and IgG antibodies from mouse sera with the native (TLA) and recombinant antigens, an immunoenzymatic assay was used, as described previously (Ferra et al., 2019). The diagnostic utility of recombinant antigens in detecting T. gondii infection in human sera was determined according to the IgM and IgG ELISA procedures described previously (Ferra et al., 2019).

Data records

All of the obtained data contained in the datasets are the values of the measured absorbance. The data was collected using a commercially available Multiscan FC (ThermoFischer Scientific) microplate photometer. Data obtained from the microplate reader were saved to external memory in the form of text files which were then imported into Microsoft Excel spreadsheets. The obtained measurement values for individual serum samples were then segregated into appropriate groups depending on the phase of the disease, as well as depending on the antibody titre. Finally, the obtained datasets, containing measurement results for large pools of serum samples, ranged in size from several dozen to several hundred kB. The datasets were subjected to statistical analysis using the SigmaPlot 14.0 software (Systat Software), as described previously (Ferra et al., 2019). The

results of the performed statistical analyses will be added to the datasets soon. A sample graph (Fig. 9.1) and a table (Tab. 9.1) of the performed statistical analysis are presented below.

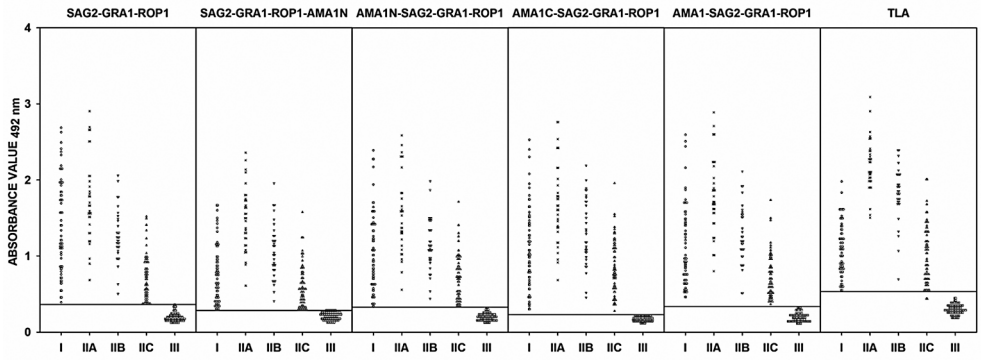


Fig. 9.1. Comparison of immunoreactivity, in IgG ELISA using SAG2-GRA1-ROP1, SAG2-GRA1-ROP1-AMA1N, AMA1N-SAG2-GRA1-ROP1, AMA1C-SAG2-GRA1-ROP1, and AMA1-SAG2-GRA1-ROP1 chimeric proteins, as well as TLA, with sera from patients with suspected acute (I) and chronic T. gondii infection (IIA–IgG titre>300 IU/ml, IIB–IgG titre between 101–300 IU/ml, and IIC IgG titre≤100 IU/ml), and from seronegative individuals (III). The horizontal lines represent the cut-off values (Ferra et al., 2019)

Tab. 9.1

IgG ELISA tests of recombinant antigens and TLA performances to discriminate among samples from suspected acute (I) and chronic (II) phase of T. gondii infection patient groups vs. control (III) group (Ferra et al., 2019)

ANTIGEN	calculated cut-off	ROC cut-off	AUC	Sensitivity, %	Specificity, %
SAG2-GRA1-ROP1	0.3641	0.3560	0.9980	100	97.8
SAG2-GRA1-ROP1-AMA1N	0.2882	0.2875	1.0000	100	100
AMA1N-SAG2-GRA1-ROP1	0.3286	0.3240	1.0000	100	100
AMA1C-SAG2-GRA1-ROP1	0.2294	0.2508	1.0000	100	100
AMA1-SAG2-GRA1-ROP1	0.3389	0.3478	1.0000	100	100
TLA	0.5344	0.4523	0.9999	99.5	100

Fig. 9.1 shows the plots obtained for each of the recombinant chimeric antigens. For comparative purposes, the previously tested trivalent recombinant antigen SAG2-GRA1-

ROP1 and the commercially used TLA antigen preparation were used (Ferra, Holec-Gąsior and Kur, 2015). The figure shows the values of the measured absorbance for each of the tested serum groups. Each single dot corresponds to the mean absorbance value obtained for a single serum sample. The size of a single plot obtained for one antigen preparation is several MB.

Serum groups:

I – suspected acute T. gondii infection (IgM +; IgG +; low avidity), n=64
II – chronic T. gondii infection (IgM –; IgG +; high avidity), n=128
III – control group (IgM –; IgG –), n=92

Sensitivity and specificity were determined using the cut-off value obtained by ROC analysis for the best discrimination.

Tab. 9.1 contains one of the most important statistical parameters calculated for antigen preparations tested in diagnostic tests. The table compares the cut-off values calculated by the standard method (based on a separate pool of seronegative serum samples) with the cut-off value obtained using the Receiver Operating Curve (ROC) analysis, which is based on all sera in relation to the results of the commercial test. The ROC analysis was performed to obtain the Area Under the Curve (AUC), the sensitivity, and the specificity percentages for the different groups of the compared sera. ROC analysis is currently one of the best statistical tools for determining and comparing the diagnostic utility of antigen preparations.

Data quality and availability

All serum samples were previously stored at –20°C. The serum samples were thawed immediately before testing, thoroughly mixed, and then their appropriate dilution was prepared. All of the tested serum samples were clear and did not contain any sediments. Quality-certified microplates (Nunc) with an appropriate sorption surface were used for the tests. The tests were performed on the same antigen preparations, different fractions of purified proteins were not used. All buffers used at various stages of the tests were prepared on an ongoing basis from certified reagents and then autoclaved. The components were weighed on an analytical balance, the liquid volumes were measured using serological pipettes or automatic pipettes with the appropriate volume range. All tests were performed with the use of multichannel pipettes (Eppendorf) and a Wellwash microplate washer (ThermoFischer Scientific). Test microplates were incubated at each stage at a constant temperature of 37°C in a StabiliTherm incubator (ThermoFischer Scientific). In order for the data to be of adequate quality and reliable, each sample of sera was tested twice. If the measured absorbance values were close to each other, the measurement was considered correct. The mean was calculated for the obtained duplicates. Moreover, for the test pool of serum samples used on a given day, tests were performed for all antigen preparations on the same day. In addition, an internal negative and positive control was used for the same serum samples on each test plate.

Datasets DOI
10.34808/w7qp-2j94
10.34808/bw9j-yq04

Acknowledgements
This research was funded by The National Science Centre, Poland, under the research project no. UMO-2015/17/B/NZ6/03480.

References

Dunn, D. et al. (1999) 'Mother-to-child transmission of toxoplasmosis: risk estimates for clinical counselling', *Lancet* 353 (9167), pp. 1829–1833. DOI: 10.1016/S0140-6736(98)08220-8.

Fatoohi, A. F. et al. (2002) 'Cellular immune responses to recombinant antigens in pregnant women chronically infected with Toxoplasma gondii', *Clinical and Diagnostic Laboratory Immunology*, 9(3), pp. 704–707. DOI: 10.1128/CDLI.9.3.704-707.2002

Ferra, B., Holec-Gąsior, L. and Kur, J. (2015) 'A new *Toxoplasma gondii* chimeric antigen containing fragments of SAG2, GRA1, and ROP1 proteins – impact of immunodominant sequences size on its diagnostic usefulness', *Parasitology Research*, 114, pp. 3291–3299. DOI: 10.1007/s00436-015-4552-6.

Ferra, B.T. et al. (2019) 'The first study on the usefulness of recombinant tetravalent chimeric proteins containing fragments of SAG2, GRA1, ROP1 and AMA1 antigens in the detection of specific anti-Toxoplasma gondii antibodies in mouse and human sera', *PLoS ONE*, 14 (6), e0217866. DOI: 10.1371/journal.pone.0217866.

Holec-Gasior, L. (2013) 'Toxoplasma gondii recombinant antigens as tools for serodiagnosis of human toxoplasmosis: current status of studies', *Clinical and Vaccine Immunology*, 20(9), pp. 1343–1351. DOI: 10.1128/CVI.00117-13.

Tenter, A.M., Heckeroth, A.R. and Weiss, L.M. (2000) 'Toxoplasma gondii: from animals to humans', *International Journal of Parasitology*, 30 (12–13), pp. 1217–1258. DOI: 10.1016/S0020-7519(00)00124-7.

Data Descriptor

Long-Term Measurement of Physiological Parameters – Child Dataset

Małgorzata Szczerska[1*]

[1] Department of Metrology and Optoelectronics, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: malszcze@pg.edu.pl; ORCID: 0000-0003-4628-6158

Abstract

The dataset titled "Long-term measurement of physiological parameters – child is one dataset" of the bigger series named Long-term measurement of physiological parameters. The dataset contains physiological parameter measurements such as skin temperature and resistance, blood pulse, as well as the stress detection marker, which can have a value of 0 when there is no stress detected or 1 when stress appeared. Additionally, the dataset contains information about the value of the device battery charge. The measurements were conducted for a long period of 120 minutes.

Keywords: temperature measurement; blood pressure measurement; skin resistance

https://doi.org/10.34808/x55q-sz53_dyr_roz10

Specification table (data records)

Subject area	Metrology, Electronics, Sensors, Biomedical engineering
More specific subject area	Physiological parameters measurements
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology by the use of a dedicated measurement setup consisting of a heart rate sensor, a skin resistance sensor, an accelerometer, and an infrared thermometer.
Data format	Text document in .log format.
Experimental factors	The data contained in the dataset were not processed.

Experimental features	The measurement setup consisted of: a Plethysmographic heart rate sensor BH1790GLC from ROHM Semiconductor, a skin resistance sensor, accelerometer LSM6DSL from STMicroelectronics, a skin thermometer (infrared radiation sensor) MLX90615 from Melexis
Data source location	MOST Wiedzy Open Research Data Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

In the case of neuro-diverse children (such as autism spectrum disorder (ASD), one of the most important things which has an influence on the efficiency of therapy is a low level of the child's stress (Jędrzejewska-Szczerska et al., 2015). However, the main difficulty in such a process is the inability of the teacher or child to recognise such stressful situations. This is because, on the one hand, ASD children find it very difficult to recognise their own emotions, and on the other hand, they do not use body language, which makes it impossible for the teacher to recognise their pupil's emotions. Therefore, many research groups focus on emotion recognition research to improve interaction with ASD children (Landowska A. et al, 2014; Kołakowska et al., 2017; Tomaczak et al., 2018). The main challenge in such an activity is to correctly measure the physiological parameters of the child and to correlate the changes in the measurand value with emotion changes. The measurements have to be carried out as long-term and real-time measurements. Furthermore, the measurement devices must be designed in such a way that they can be accepted by the person with high sensory sensitivity. As a consequence, many devices are dedicated and personalised equipment.

The dataset, Long-term measurement of physiological parameters - child, provides the possibility for other researchers and engineers to compare and validate their measurement devices and measurement procedures.

Methods

The measurements were carried out by the use of a dedicated measurement setup, which consists of the wearable bracelet with sensors, charger and dedicated software (Fig. 10.1).

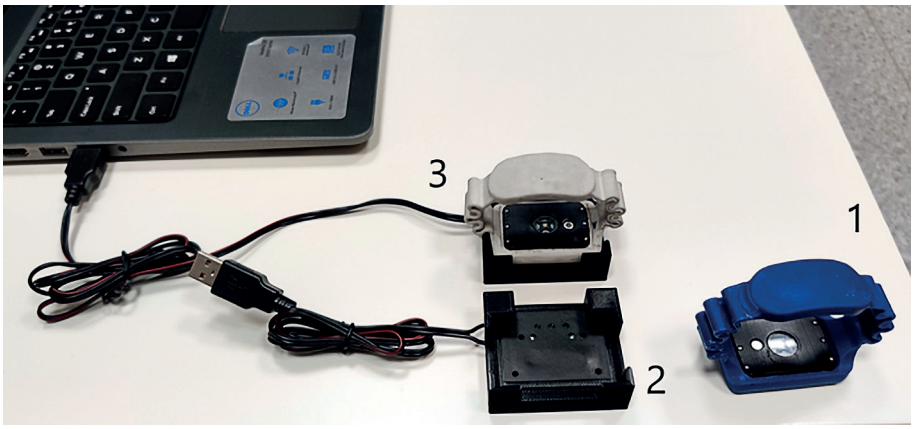


Fig. 10.1. Measurement setup: 1 – wearable bracelet with sensors, 2 – charger, 3 – data transfer

The measurement setup consists of a BH1790GLC heart rate sensor (ROHM Semiconductor), a skin resistance sensor, an LSM6DSL accelerometer (STMicroelectronics), and an MLX90615 skin thermometer (infrared radiation sensor) (Melexis). The main part of the device is an embedded system built around a PSoC6 system-on-a-chip integrated circuit from Cypress. The system and the algorithms used to detect the stressful situation were described in detail by Tomczak (Tomczak et al., 2020).

Data quality and availability

This dataset can be used by other research groups to validate their measurement process and to compare their algorithms which are used to recognise stressful situations based on a change of the values of physiological parameters.

Dataset DOI
10.34808/x75s-2h03

Dataset License
CC-BY-NC

Acknowledgments
This work was supported in part by the National Centre for Research and Development, Poland (NCBiR), under Project 17/02/2018/UD, in part by the Gdańsk University of Technology, and in part by the DS Programs of Faculty of Electronics, Telecommunications and Informatics.

References

Jędrzejewska-Szczerska, M. et al. (2015) 'System supporting behavioral therapy for children with autism', *Journal of Innovative Optical Health Sciences*, 08(03), 1541008, DOI: 10.1142/S1793545815410084.

Kołakowska, A. et al. (2017) 'Automatic recognition of therapy progress among children with autism', *Scientific Reports,* 7, 13863, DOI: 10.1038/s41598-017-14209-y.

Lanowska, A. et al., (2014) 'Selection of physiological parameters for optoelectronic system supporting behavioral therapy of autistic children', *Proceedings SPIE 9290,* Photonics Applications in Astronomy, Communications, Industry and High Energy Physics Experiments, 92901Q, DOI: 10.1117/12.2075020.

Tomczak, M.T. et al. (2018) 'Support for Employees with ASD in the Workplace Using a Bluetooth Skin Resistance Sensor–A Preliminary Study', *Sensors*, 18(10), 3530, DOI: 10.3390/s18103530.

Tomczak, M.T. et al. (2020) 'Stress Monitoring System for Individuals With Autism Spectrum Disorders', *IEEE Access*, 8, pp. 228236-228244, DOI: 10.1109/ACCESS.2020.3045633.

Data Descriptor

Porous Phantoms Mimicking Tissues – Investigation of Optical Parameter Stability Over Time

Monika Ronowska[1*]

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

[*] Correspondence author: mronowska@gmail.com

Abstract

In terms of optical parameters, optical phantoms can now replace live tissues and be used to validate optical measurement methods. Therefore, whether these parameters would be maintained after storage for 6 months was examined. The absorption and scattering coefficients were obtained from the measured transmittance and reflectance measurements taken 6 months apart and then compared. All of the measurements were conducted using the same experimental setup consisting of an integrating sphere, a light source with a wavelength of 635 nm and a detector. The optical phantoms on which the research was performed were prepared from silicone and glycerol in various proportions.

Keywords: absorption; coefficient stability; optical parameters; optical phantoms

https://doi.org/10.34808/x55q-sz53_dyr_roz11

Specification table (data records)

Subject area	Metrology, Electronics, Biomedical engineering
More specific subject area	Optical parameters measurements
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology by the use of a dedicated measurement setup consisting of integrating sphere, light source and detector
Data format	xls file

Experimental factors	The data contained in the dataset were not processed
Experimental features	The measurement setup consisted of: 4P-GPS-053-SL integrating sphere (Labsphere Inc., North Sutton, NH, USA) with a Spektralon® coating, LDM635-03-08X25 laser diode module with a wavelength of 635 nm (red) and an optical power of 3 mW, L-100 lux meter with a dedicated measuring head (Sonopan, Białystok, Poland)
Data source location	MOST Wiedzy Open Research Data Catalogue, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Phantoms, due to their optical properties, mimic the way light is scattered inside their structure, therefore they can effectively replace biological tissues. As a result, they are used in many fields of physics, engineering and medicine. They are used for the registration of reference measurements using optical devices and measuring techniques (Wróbel et al., 2016) (Karpienko et al., 2016)we highlight how the use of blood phantoms enables to investigate the phenomena that otherwise are almost impossible to be noticed.”,”container-title”:”Journal of Innovative Optical Health Sciences”,”DOI”:”10.1142/S1793545816500127”,”journalAbbreviation”:”Journal of Innovative Optical Health Sciences”,”source”:”ResearchGate”,”title”:”Blood equivalent phantom vs. whole human blood, a comparative study”,”volume”:”9”,”author”:[{„family”:”Karpienko”,”given”:”Katarzyna”},{„family”:”Gnyba”,”given”:”Marcin”},{„family”:”Majchrowicz”,”given”:”D.”},{„family”:”Wróbel”,”given”:”Maciej”},{„family”:”Szczerska”,”given”:”Malgorzata”}],”issued”:{„date-parts”:[[„2015”,11,1]]}}],”schema”:”https://github.com/citation-style-language/schema/raw/master/csl-citation.json”} , calibration of optical devices (Listewnik et al., 2021), and light distribution planning using the physical geometry of the tissue (Wróbel et al., 2016). They replace both porous tissues such as the brain or lungs (Listewnik et al., 2021) and those with an internal vascular system (Feder et al., 2016). Moreover, optical phantoms have a number of additional advantages such as relatively low manufacturing cost, easier storage, and durability.

Therefore, the study was performed to determine the stability of the optical parameters of phantoms after storage for 6 months in order to verify their reusability (Listewnik et al., 2021).

Methods

The absorption and scattering coefficient were obtained on the basis of transmittance and reflectance measurements using the experimental setup presented in Fig. 11.1 com-

posed of a 4P-GPS-053-SL integrating sphere (Labsphere Inc., North Sutton, NH, USA) with a Spektralon® coating, a laser diode module with a wavelength of 635 nm (red) and an L-100 lux meter with a dedicated measuring head (Sonopan, Białystok, Poland).

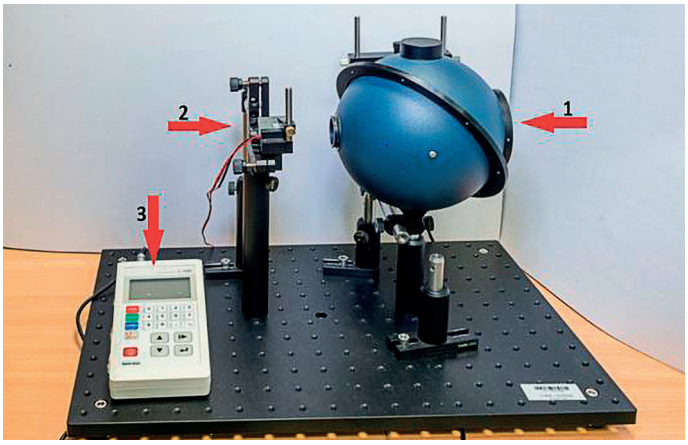


Fig. 11.1. Measurement setup: 1 – integrating sphere 2 – light source, 3 – lux meter

The integrating sphere has a rotation function which was enabled to investigate both optical parameters in one measuring system. The transmittance and reflectance were examined in a series of measurements performed with an interval of 6 months. In order to obtain the most accurate measurements results, the same conditions were recreated and the same experimental setup was used.

The optical phantoms, on which the research was performed, were prepared from polydimethyl-siloxane (PDMS, Sylgard®184, Dow Corning, Midland, MI, USA) and glycerol in various proportions. One of the samples contained 2 mL of glycerol while the second included 5 mL.

Data quality and availability

This dataset can be used by other research groups to validate their measurement process as well as compare their algorithms which are used to assess the stability of absorption and scattering coefficients of optical phantoms.

Dataset DOI
10.34808/85r2-yb55

Dataset License
CC-BY-NC

Acknowledgments
This work was supported in part by the DS Programs of the Faculty of Electronics, Telecommunications and Informatics of the Gdańsk University of Technology.

References

Feder, I. et al. (2016) 'Experimental results of full scattering profile from finger tissue-like phantom', *Biomedical Optics Express*, 7(11), pp. 4695–4701. DOI: 10.1364/BOE.7.004695.

Karpienko, K. et al. (2016) 'Blood equivalent phantom vs. whole human blood, a comparative study', *Journal of Innovative Optical Health Sciences*, 09 (02), 1650012. DOI: 10.1142/S1793545816500127.

Listewnik, P. et al. (2021) 'Porous Phantoms Mimicking Tissues—Investigation of Optical Parameters Stability Over Time', *Materials*, 14 (2), 423. DOI: 10.3390/ma14020423.

Wróbel, M.S. et al. (2016) 'Nanoparticle-free tissue-mimicking phantoms with intrinsic scattering', *Biomedical Optics Express,* 7(6), pp. 2088–2094. DOI: 10.1364/BOE.7.002088.

Data Descriptor

Non-Contact Temperature Measurements Dataset

Aleksander Mroziński[1*]

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

[*] Correspondence author: aleksander.mrozinski@pg.edu.pl; ORCID 0000-0003-1235-2735

Abstract

The dataset titled The influence of the distance of the pyrometer from the surface of the radiating object on the accuracy of measurements contains temperature measurements using a selection of four commercially available pyrometers (CHY 314P, TM-F03B, TFA 31.1125 and AB-8855) as a function of the measuring distance. The dataset allows a comparison of the accuracy and measuring precision of the devices, which are very important features in the reliable non-contact prediction of COVID-19 symptoms without interference from external disturbances during fast patient recognition.

Keywords: temperature measurement; pyrometer; measuring distance; COVID-19

https://doi.org/10.34808/x55q-sz53_dyr_roz12

Specification table (data records)

Subject area	Metrology, Electronics, Sensors, Biomedical engineering
More specific subject area	Non-contact temperature measurements
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology using commercially available pyrometers (AB-8855, CHY 314P, TM-F03B and TFA 31.1125) on a special optical stand with an Omega BB703 temperature blackbody calibrator
Data format	The tables are in .xlsx format
Experimental factors	The data contained in the dataset were not processed

Experimental features	The selected emissivity of the pyrometer was identical to the target plate emissivity (ε = 0.95)
Data source location	MOST Wiedzy Open Research Data Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

During the COVID-19 pandemic, non-contact temperature measurement has become one of the ways to monitor the spread of the disease (Guan et al., 2020; Zhang et al., 2020). There are more and more devices on the market that allow for the quick assessment of people's health. One such device is the pyrometer, which can measure temperature without being in contact with the measuring object, such as the human skin (Costanzo and Flores, 2020; Ebeid, Selem and Abd El-kader, 2020). Unfortunately, there are some limitations that could affect the results: the ambient radiation, the absorption of radiation by the atmosphere, scattering, and the optical resolution (Fig.12.1).

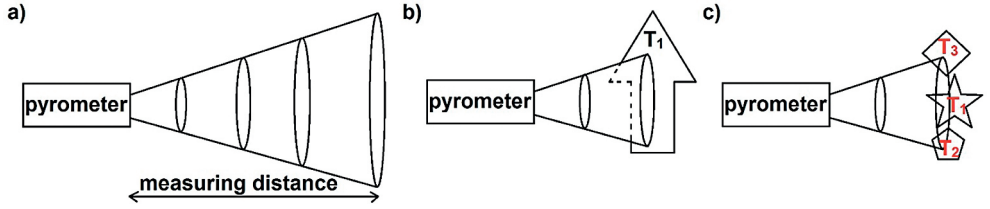


Fig. 12.1. (a) Schematic of the relation between the measuring distance and the field of view; (b) a measurement carried out correctly; and (c) ambient radiation negatively affecting the results

The dataset, The influence of the distance of the pyrometer from the surface of the radiating object on the accuracy of measurements, has been designed to help understand how important it is to choose the right equipment due to differences in optical construction. The dataset contains comparative results for four commercially available devices with different optical resolutions and consists of two parts: temperature measurements as a function of the measuring distance and the calculated field of view for each pyrometer based on the known optical resolutions.

Methods

An Omega BB703 blackbody temperature calibrator with a target aperture of 2.9×10^{-2} m was used to generate thermal radiation. The temperature was measured using the following pyrometer models: CHY 314P, TM-F03B, TFA 31.1125 and AB-8855. The selected emissivity of the pyrometer (ε) was identical to the target plate emissivity and equal to 0.95. The blackbody calibrator and pyrometers were mount-

ed on stands on an optical bench so that the pyrometers were on the same axis as the blackbody target. There were no obstacles in the measuring path. The target set temperature was 81°C and the room temperature was 23°C during the entire experiment. The target temperature was selected so that it was within the measuring range of all devices and, at the same time, was easy to distinguish when the results contained radiation not only from the target (Fig. 12.1c). The only parameter that was changed was the distance between the pyrometer and the radiating target. The exact distance was measured with a ruler attached to the optical bench; the smallest division on the ruler was 10^{-3} m. The distance was varied between 5×10^{-2} m and 1 m.

Data quality and availability

All measurements were collected at a stable blackbody target and laboratory temperature using pyrometers with the same emissivity factor. Catalogue data were used to calculate the field of view of three of the pyrometers (CHY 314P, TFA 31.1125 and AB-8855). For the TM-F03B, unfortunately, the seller provides only information on the scope in which measurements are allowed. Nevertheless, the data contained in the dataset allow for a fair comparison of all pyrometers.

Dataset DOI
10.34808/wtyv-n266

Acknowledgements
The generation of this dataset was supported by a statutory research grant from Gdańsk University of Technology.

References

Costanzo, S. and Flores, A. (2020) 'A Non-Contact Integrated Body-Ambient Temperature Sensors Platform to Contrast COVID-19', *Electronics*, 9(10), 1658. DOI: 10.3390/electronics9101658.

Ebeid, A. G., Selem, E. and Abd El-kader, S. M. (2020) 'Early Detection of COVID-19 Using a Non-contact Forehead Thermometer', *6th International Conference on Advances Intelligent Systems and Computing*, AISI 2020. DOI: 10.1007/978-3-030-58669-0_29.

Guan, W. et al. (2020) 'Clinical Characteristics of Coronavirus Disease 2019 in China', *New England Journal of Medicine*, 382(18), pp. 1708–1720. DOI: 10.1056/nejmoa2002032.

Zhang, J. et al. (2020) 'Therapeutic and Triage Strategies for 2019 Novel Coronavirus Disease in Fever Clinics', *The Lancet Respiratory Medicine*, 8(3), pp. e11–e12. DOI: 10.1016/S2213-2600(20)30071-0.

Data Descriptor

Determination of Changes in Viscosity of Hydrogel Depending on Shear Rates

Szymon Mania[1*], Robert Tylingo[1], Katarzyna Kozłowska-Tylingo[2]

[1] Department of Chemistry, Technology and Biotechnology of Food, Faculty of Chemistry, Gdańsk University of Technology (11/12 Narutowicza Street, Gdańsk, Poland)
[2] Department of Pharmaceutical Technology and Biochemistry, Faculty of Chemistry, Gdańsk University of Technology (11/12 Narutowicza Street, Gdańsk, Poland)

* Correspondence author: szymon.mania@pg.edu.pl; ORCID: 0000-0003-2748-0393

Abstract

The datasets entitled Determination of changes in viscosity of hydrogel depending on shear rate contain the results of viscosity measurements using a Brookfield viscometer, with different kinds of spindles and shear rates. The data allowed the used hydrogel preparations to be characterised and their functional parameters, as substances modifying the rheology of thickeners and determining the effect of shear rate on the viscosity of hydrogels, to be assessed.

Keywords: viscosity, hydrogel, shear rate, Brookfield viscometer

https://doi.org/10.34808/x55q-sz53_dyr_roz13

Specification table (data records)

Subject area	Materials engineering, Biomedical engineering
More specific subject area	Measurements of rheological properties of hydrogel
Type of data	Text
How the data was acquired	The data were collected at the Gdańsk University of Technology using a Brookfield viscosity test machine with the Rheocalc V2.7 software
Data format	The tables are in .xls and .csv formats
Experimental factors	The data contained in the datasets were not processed

Experimental features	The dynamic viscosity data were collected under different kinds of spindles and shear rates
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The datasets are accessible and are publicly and freely available for any research or educational purposes

Background

Rheology is an appropriate method for characterising hydrogel mechanical properties since it is quick, sensitive, requires small sample sizes, and is revealing of differences in architecture such as degree of crosslinking, proximity of the glass transition, structural homogeneity/heterogeneity, and molecular weight. There are a number of reviews of the rheological properties of hydrogels composed of proteins, polysaccharides, or both (Tylingo, Mania and Szwacki, 2016). Compared to other synthetic biomaterials, hydrogels have physical properties similar to living tissue due to their relatively high water content, softness and plasticity. The polymer chains that make up the hydrogel network can be linked by chemical bonds, or their structure can be maintained by molecular bonds, additional ionic forces, hydrogen bonds or hydrophobic interactions. In the first case, hydrogels, depending on the size of the polymer-water interaction and the density of connections between the polymer chains, reach an equilibrium swelling state and are called solid or chemical gels. In the second described case, the hydrogels are called reversible or physical gels, respectively. Physical and chemical hydrogels can adopt a wide variety of macromolecular structures consisting of cross-linked or entangled linear homopolymers, linear copolymers, or block and graft copolymers. The network can be stabilised by the reaction of a polyion and polyvalent ions, or two polyions, resulting in complexes containing hydrogen bonds and others (Swarbcick and Boylan, 2000). Hence, the use of viscosity measurements to assess the quality of created hydrogels is a good analytical tool.

Our dataset series entitled "Determination of changes in viscosity of hydrogel depending on shear rate", has been created to enable the determination of the rheological properties of hydrogels composed of natural polymer solutions. The datasets contain: dynamic viscosity, speed and torque of the spindle, shear stress and the information of the temperature and spindle type used for measurement. These data allow the determination of the flow curve of the tested fluid.

Methods

Most rotational viscometers / rheometers work according to the Searle principle: A motor drives a spindle inside a fixed cup. The rotational speed of the spindle is preset and produces the motor torque that is needed to rotate the measuring bob. This torque has to overcome the viscous forces of the tested substance and is therefore a measure for its viscosity. A rotational viscometer measures the dynamic viscosity [η] of a sample.

Newton's Law defines the dynamic viscosity η as the shear stress divided by the shear rate (Mezger, 2011). When we measure the viscosity on a rotational viscometer, we apply to the sample a certain shear stress or a certain shear rate, respectively. The physical properties speed and torque can be translated into the rheological properties shear rate and shear stress if the measurement is performed using a standard measuring geometry according to ISO 3219. The viscometer's speed is converted into shear rate using a conversion factor (the factor depends on the measuring geometry used) and the torque is also converted into shear stress using a conversion factor. Usually, this is automatically done by the instrument.

$$\text{speed} \times \text{conversion factor} = \text{shear rate}$$
$$\text{torque} \times \text{conversion factor} = \text{shear stress}$$

In a typical rotational test, the shear rate is preset. This means that the viscometer translates the chosen shear rate into speed and measures the resulting torque, which it then translates into shear stress. Such a measurement, where the shear rate is increased and the resulting shear stress is measured, is called a flow curve.

During the test, the effect of adding a thickener in a wide range of concentrations on the change of the viscosity of a hydrogel prepared on the basis of a natural polymer was checked. All samples were prepared for measurement as follows. Before starting the measurement, the liquid sample was thermostated for at least 5 hours in a laboratory chamber at the measurement temperature.

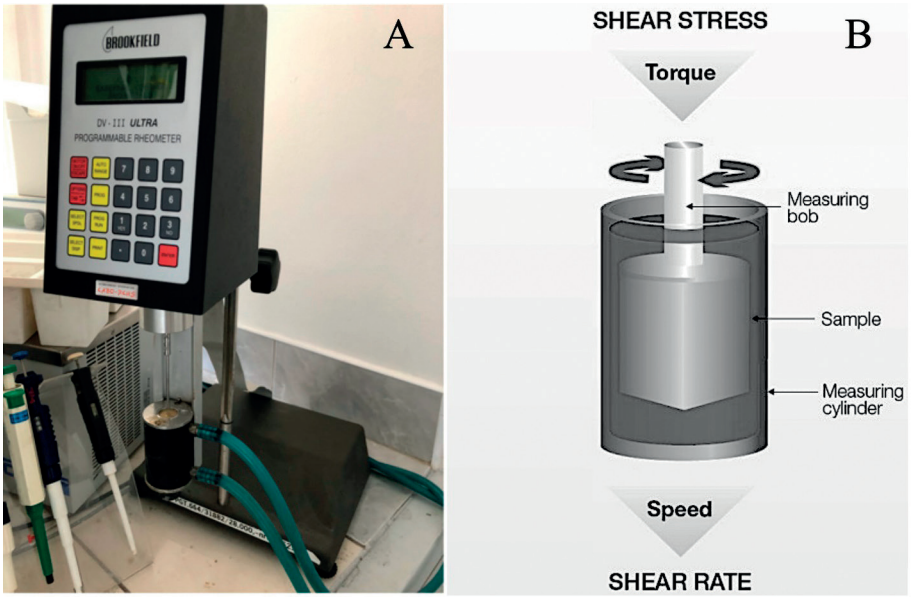


Fig. 13.1. Presentation of the main measuring device of the Brookfield DV-III+ viscometer: (A) main measurement unit, (B) principle of operation

Based on a preliminary visual viscosity assessment, the appropriate sample volume was transferred to the measuring cell and mounted in the thermostatic jacket of the apparatus. Next, after resetting the viscometer indications, the appropriate spindle (LV SC4 – 27 spindle and shear rates from 1.7 to 34 s-1, LV SC4 – 25 spindle and shear rates from 1.1 to 55 s-1, LV SC4 – 18 spindle and shear rates from 6.6 to 330 s^{-1}) was installed and the measurement was started. Measurements were performed in time intervals with a stabilisation time of the viscometer indications of 10 minutes for each shear rate.

Data quality and availability

All measurements were collected using a Brookfield DV-III+ viscometer (Labo Plus), using the Rheocalc V2.7 software and typical operating parameters, described within each experimental file. The datasets may be opened with any calculation processing software capable of recognising text or XLS files.

The flow curves were measured in triplicate. Before starting each measurement, the instrument readings were zeroed so as not to generate spindle stress before starting the measurement (measurement error below 3%). Additionally, before each series of measurements for the same tests, the viscometer was calibrated on the basis of selected certified viscosity standards in the form of silicone oil in the range of 100–100,000 mPas at 25°C.

Datasets DOI
10.34808/0d90-0v75
10.34808/63qr-f068
10.34808/vmm8-4d37

Datasets License
CC-BY-NC-SA

References

Mezger, T. (2011) 'The Rheology'. Handbook. *European Coatings Tech Files*, 3rd ed., Hanover: Vincentz Network.

Tylingo, R., Mania, S. and Szwacki, J. (2016) 'A novel method for drop in drop edible oil encapsulation with chitosan using a coaxial technique', *Reactive and Functional Polymers*, 100, pp. 64–72. DOI: 10.1016/j.reactfunctpolym.2016.01.001.

Swarbcick, J. (2000) '*Encyclopedia of Pharmaceutical Technology*', Informa Healthcare, USA, New York, Available at: https://gmpua.com/Process/EncyclopediaPT.pdf (Accessed: 25th May 2022).

Data Descriptor

Mechanical Properties of Human Stomach Tissue

Szymon Grymek[1*]

[1] Faculty of Mechanical Engineering and Ship Technology, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: sgrymek@pg.edu.pl; ORCID: 0000-0003-4393-8982

Abstract

The dataset entitled Determination of mechanical properties of human stomach tissues subjected to uniaxial stretching contains: the length of the sample as a function of the corresponding load (tensile force) and the initial values of the average width and average thickness of the sample. All tests were conducted in a self-developed tensile test machine: PGTissueTester. The dataset allows the coefficients of various models of incompressible and nearly incompressible hyperelastic materials used to model human tissues to be determined.

Keywords: human stomach; biomechanical properties; hyperelastic materials

https://doi.org/10.34808/x55q-sz53_dyr_roz14

Specification table (data records)

Subject area	Mechanical engineering, Biomedical engineering, Materials engineering
More specific subject area	Measurements of mechanical properties of living tissues
Type of data	Text
How the data was acquired	The data were collected at the Gdańsk University of Technology using the PGTissueTester tensile test machine equipped with a commercially available KMM20 force transducer and a video-extensometric system
Data format	The tables are in .xls format
Experimental factors	The data contained in the dataset were not processed
Experimental features	The samples were obtained from fragments of dissected human stomachs

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The mechanical properties of human tissues are important variables used in the computer modelling of surgical procedures, optimisation of operative techniques, designing of surgical equipment and virtual reality surgical simulations (Brouwer et al., 2001; Delingette and Ayache, 2004; Rosen et al., 2008). One of the methods to estimate the range of tissue's mechanical properties is an uniaxial stretching (Jia et al., 2015).

Our dataset, Determination of mechanical properties of human stomach tissues subjected to uniaxial stretching, has been created to enable the determination of the mechanical properties of human stomach tissue. The dataset contains: the length of the sample as a function of the corresponding load (tensile force) and the initial values of the average width and average thickness of the sample. This suffices to determine the coefficients of various models of incompressible and nearly incompressible hyperelastic isentropic materials (Martins et al., 2006).

Methods

The tests were conducted in a self-developed tensile test machine: PGTissueTester 1.0 (Rotta and Grymek, 2020).

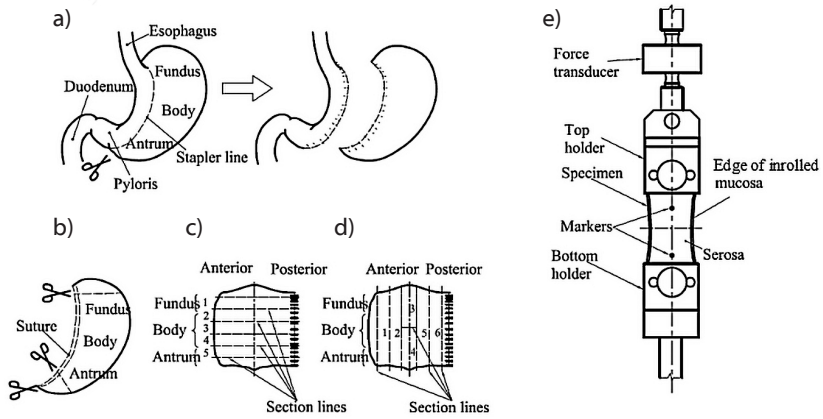


Fig. 14.1. Preparation of the test samples: (a) extent of resection during sleeve gastrectomy; (b) removal of unnecessary parts of resected stomach; (c) preparation of circumferential samples; (d) preparation of longitudinal samples; (e) fixing in holders

All samples were prepared as shown in Fig.14.1. The initial values of the average width and average thickness were calculated from 5 measurements (for each dimension) made with a calliper. Every sample was stretched at a constant speed of 0.4 mm/s. The measurement of deformations (distance between markers) was carried out by the non-contacting video extensometer method. The measurement of load (tensile force) was carried out by the KMM20 force transducer (range 0–50 N).

The measurement procedure utilised was described in detail by Rotta et al. (2019).

Data quality and availability

All of the stomach specimens were transported from the operating room to the laboratory within 5 min after resection and tested within an hour of the process. The test samples did not go through the preconditioning process because they were fresh; instead, the samples were pre-stressed to a force of 0.5 N. The tests were carried out in an air-conditioned room, at a temperature of 20°C and 60% relative humidity of the ambient air.

The relative error in the measurement of load (tensile force), sample width and deformation (distance between markers) was less than 2%. But, there is an uncertainty issue of calculating the relative error in the measurement of the sample thickness. Nevertheless we estimated that it should not be greater than 15%.

Dataset DOI
10.34808/kp2c-yb14

Acknowledgements
The generation of this dataset was supported by statutory research grants from Gdańsk University of Technology and from the Medical University of Gdańsk.

References

Brouwer, I. et al. (2001) 'Measuring in vivo animal soft tissue properties for haptic modeling in surgical simulation', *Studies in Health Technology and Informatics*, 81, pp. 69–74. DOI: 10.3233/978-1-60750-925-7-69.

Delingette, H. and Ayache, N. (2004) 'Soft tissue modeling for surgery simulation', *Handbook of Numerical Analysis*, 12, pp. 453–550., DOI: 10.1016/S1570-8659(03)12005-4.

Jia, Z.G., Li, W. and Zhou, Z. R. (2015) 'Mechanical characterization of stomach tissue under uniaxial tensile action', *Journal of Biomech*anics, 48 (4), pp. 651–658. DOI: 10.1016/j.jbiomech. 2014.12.048.

Martins, P.A.L.S., Natal Jorge, R.M. and Ferreira, A.J.M. (2006) 'A Comparative Study of Several Material Models for Prediction of Hyperelastic Properties: Application to Silicone-Rubber and Soft Tissues', *Strain*, 42(3), pp. 135–147. DOI: 10.1111/j.1475-1305.2006.00257.x.

Rosen, J. et al. (2008) 'Biomechanical properties of abdominal organs in vivo and post-mortem under compression loads', *Journal of Biomechanical Eng*ineering, 130(2), 021020. DOI:10.1115/1.2898712.

Rotta, G. and Grymek, S. (2020) 'Test Stand for Multi-option Stretching of Soft Tissues', in Śnieżek, L. (ed.), *Fundamentals of Machine Design. Selected problems*. Warszawa: Wydawnictwo WAT, pp. 287–296.

Rotta, G. et al. (2019) 'Mechanical properties of the human stomach under uniaxial stress action', *Current Science*, 116 (11), pp. 1886–1893. DOI: 10.18520/cs/v116/i11/1886-1893.

Data Descriptor

Measurement of the Temporal and Spatial Temperature Distribution on the Surface of PVCP Tissue Phantom Illuminated by Laser Dataset

Anna Sękowska-Namiotko[1*]

[1] Department of Metrology and Optoelectronics, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: annsekow@student.pg.edu.pl; ORCID: 0000-0003-4628-6158

Abstract

The dataset entitled Measurement of the temporal and spatial temperature distribution on the surface of PVCP tissue phantom illuminated by laser was obtained with a laboratory set-up for characterisation of the thermal properties of optical tissue phantoms during laser irradiation. The dataset contains a single image file representing the spatial temperature distribution on the surface of a PVCP tissue phantom. This thermal image was captured at the moment when the temperature reached its maximum value as a result of irradiation with a dermatological laser.

Keywords: IR thermography, lasers in medicine, thermal properties of tissues, optical skin phantoms, tissue-mimicking phantoms

https://doi.org/10.34808/x55q-sz53_dyr_roz15

Specification table (data records)

Subject area	Metrology, Electronics, Sensors, Biomedical engineering
More specific subject area	Thermal imaging
Type of data	Image
How the data was acquired	The data was collected at the Gdańsk University of Technology by the use of a dedicated measurement setup consisting of a thermographic camera and dermatological laser system
Data format	Image file in .bmp format

Experimental factors	The data contained in the dataset were not processed
Experimental features	The measurement setup consisted of: VIGOcam v50 thermographic camera from VIGO System S.A., pulsed dermatologic 20W diode-laser emitting at 975 nm
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The use of lasers in medicine is constantly increasing. They have proven useful in many fields of medicine, both in diagnostics and therapy. In dermatology, they are used in the removal of superficial skin lesions such as discoloration, pathological and vascular lesions (Szymańczyk et al., 2017). Nevertheless, before a new laser is introduced into clinical practice, its ability to interact with tissues has to be carefully examined and it has to be properly calibrated. In order to do this, optical skin phantoms can be used (Jędrzejewska-Szczerska et al., 2015; Wróbel et al., 2015). Such phantoms closely imitate the thermal and optical properties of real human skin (Bashkatov et al.,2018). To be able to create such materials, the correlation between the proportions of ingredients and properties of the obtained phantoms has to be examined. Therefore, to perform this task, appropriate laboratory set-ups, such as the one used in this study, are produced.

The dataset, Measurement of the temporal and spatial temperature distribution on the surface of PVCP tissue phantom illuminated by laser, provides the possibility for other researchers and engineers to compare and validate the results of measurements obtained for other tissue phantoms as well as tissue samples and for different laser systems.

Methods

The measurements were carried out by the use of a dedicated measurement setup, which consists of a thermographic camera and dermatological laser system with a 975 nm diode laser module (Fig. 15.1).

The measurement setup consists of a VIGOcam v50 thermographic camera from VIGO System S.A. and a pulsed dermatologic 20W diode-laser emitting at 975 nm. The pulsed diode laser was described in detail by Piechowski (Piechowski et al., 2012). A detailed description of the system and experimental procedure were provided by Wróbel (Wróbel et al., 2016).

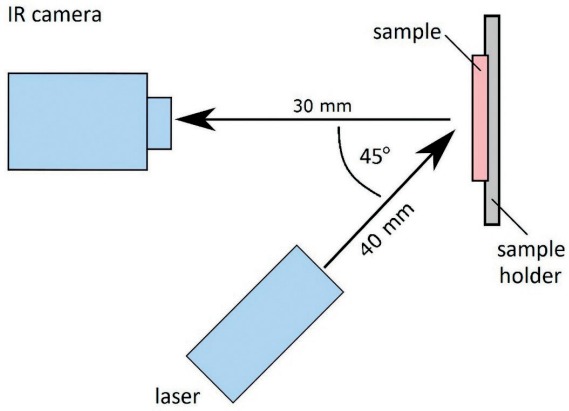


Fig. 15.1. Measurement setup

Data quality and availability

This dataset can be used by other research groups to compare the results of measurements obtained for another tissue phantoms as well as tissue samples and to validate their measurement process with the use of different laser systems.

Dataset DOI
10.34808/ga0d-gj04

Dataset License
CC-BY-NC

Acknowledgments
This work was supported in part by the DS Programs of the Faculty of Electronics, Telecommunications and Informatics of Gdańsk University of Technology.

References

Bashkatov A. et al. (2018) 'Model of optical phantoms thermal response upon irradiation with 975 nm dermatological laser', *Proceedings of SPIE*, 10716, Saratov Fall Meeting 2017: Optical Technologies in Biophysics and Medicine XIX, 1071604, DOI: 10.1117/12.2317566.

Jędrzejewska-Szczerska M. et al. (2015) 'Investigation of photothermolysis therapy of human skin diseases using optical phantoms', *Proceedings of SPIE*, 9447, 18th International School on Quantum Electronics: Laser Physics and Applications, 944715, DOI: 10.1117/12.2177790.

Piechowski L. et al. (2013) 'Pulsed dermatologic 20W diode-laser emitting at 975 nm', *Proceedings of SPIE*, 8703, Laser Technology 2012: Applications of Lasers, 870304, DOI: 10.1117/12.2014998.

Szymańczyk J. et al. (2017) 'Application of the laser diode with central wavelength 975 nm for the therapy of neurofibroma and hemangiomas', *Journal of Biomedical Optics,* 22(1), 010502, DOI: 10.1117/1.JBO.22.1.010502.

Wróbel M. et al. (2015) 'Use of optical skin phantoms for preclinical evaluation of laser efficiency for skin lesion therapy', *Journal of Biomedical Optics*, 20(8), 085003, DOI: 10.1117/1.JBO.20.8.085003.

Wróbel M. et al. (2016) 'Use of optical skin phantoms for calibration of dermatological lasers', *Proceedings of SPIE*, 9961, Reflection, Scattering, and Diffraction from Surfaces V, 996113. DOI: 10.1117/12.2240064.

Data Descriptor

Herbarium of Division of Marine Biology and Ecology as the Primary Basis for Conservation Status Assessments in the Gulf of Gdańsk

Ilona Złoch[1*]

[1] Department of Marine Biology and Ecology, Faculty of Oceanography and Geography, University of Gdańsk (8 Jana Bażyńskiego Street, Gdańsk, Poland)

[*] Correspondence author: ilona.zloch@ug.edu.pl; ORCID: 0000-0002-3233-2021

Abstract

The dataset titled Herbarium of Division of Marine Biology and Ecology University of Gdańsk (DMBE) is a research herbarium encompassing specimens of vascular plants and algae hosted by the Laboratory of Marine Plant Ecology at the University of Gdańsk, Poland. The aim of Herbarium is to preserve marine plant and algae collections mostly from the Gulf of Gdańsk, but the herbarium also holds specimens from other parts of the world.

Keywords: herbarium, algae, plants, the Gulf of Gdańsk

https://doi.org/10.34808/x55q-sz53_dyr_roz16

Specification table (data records)

Subject area	Herbarium, Vascular plants, Algae
More specific subject area	marine plant and algae collections from Gulf of Gdańsk
Type of data	Text
How the data was acquired	The data was collected during annual Gulf of Gdańsk monitoring
Data format	The tables are in .xlsx format
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible for any research or educational purposes under CC BY licence

Measurement Type	taxonomic inventory
Design Type(s)	data integration objective, database creation objective

Background

Herbarium of Division of Marine Biology and Ecology University of Gdańsk (DMBE) is a research herbarium encompassing specimens of vascular plants and algae hosted by the Laboratory of Marine Plant Ecology at the University of Gdańsk, Poland. The main aim of Herbarium is to preserve marine plant and algae collections mostly from the Gulf of Gdańsk. Storing the specimens allows users to return to them, after even long time, to check the identification of a specimen, to describe new species or to classify them based on their morphological and genetic traits. Recently, new purposes for herbarium specimens have arisen, so they are being utilised for purposes for which they were not initially intended, for example medical research and phenology. The application of herbarium to taxonomy, ecology, biodiversity conservation, evolution, and other disciplines, makes it a valuable research tool.

Our collection of preserved plant and algae specimens is used for scientific study to understand plant diversity in time and space, conservation and threats, and to provide an educational resource for learning.

To conclude, the DMBE herbarium is particularly rich in the genera Chara and individuals under protection, what makes this collection an extensive reference collection of Gulf of Gdańsk macrophytes.

Methods

Specimens in good condition were chosen and collected directly from environment; the snorkeling or scuba-diving method were also implemented. After collection of each specimen, a unique collection number was assigned. Most specimens in the herbarium are pressed, dried plants. These pressed specimens are mounted on sheets of card and stored in flat folders.

Data records

The dataset currently comprises records for 572 specimens representing 50% of the estimated total number of specimens. The collection specimens are mounted on herbarium sheets.

The file 'Herbarium_of_Division_of_Marine_Biology_and_Ecology' follows the Darwin Core Archive (DwCA) specifications for the field, holding one record per specimen with all core specimen fields (herbarium info, current identification, collector info, dates, location, notes, etc.). Each occurrence record is uniquely identified within the DMBE Herbarium database by a field called id (DwCA).

The structure of the dataset is presented in Fig. 16.1.

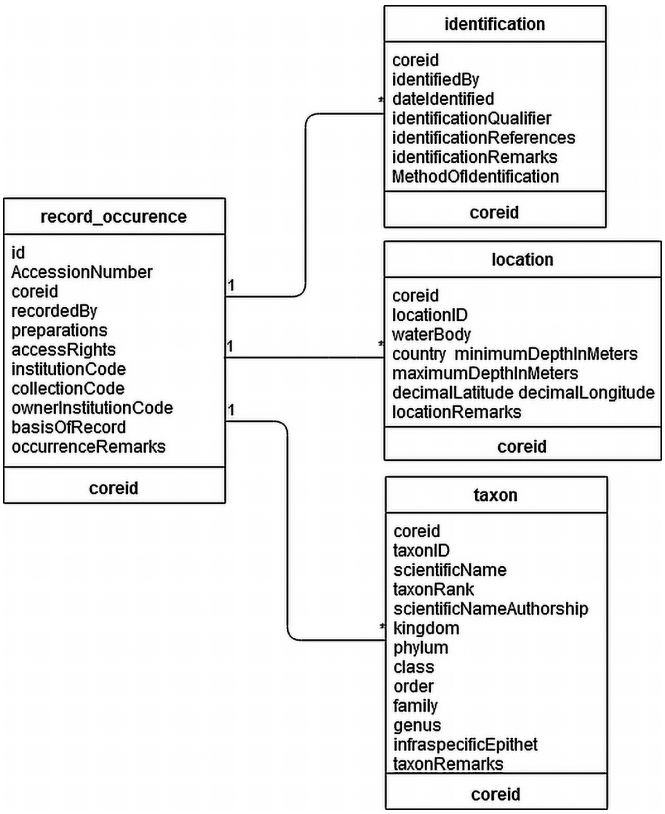


Fig. 16.1. Diagram of the main tables of the dataset presenting the column headers and the relations between the tables

The Herbarium holds a wide variety of specimens; however, the main body of the collection comprises the Charales species (Fig. 16.2).

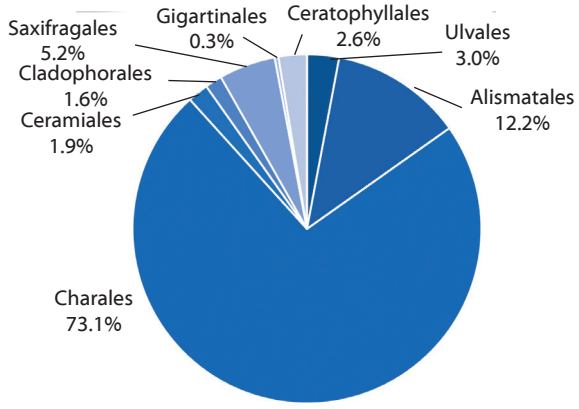


Fig. 16.2. Taxonomic coverage of the dataset

The temporal coverage of the database runs from 2013 until 2019 (Fig. 16.3).

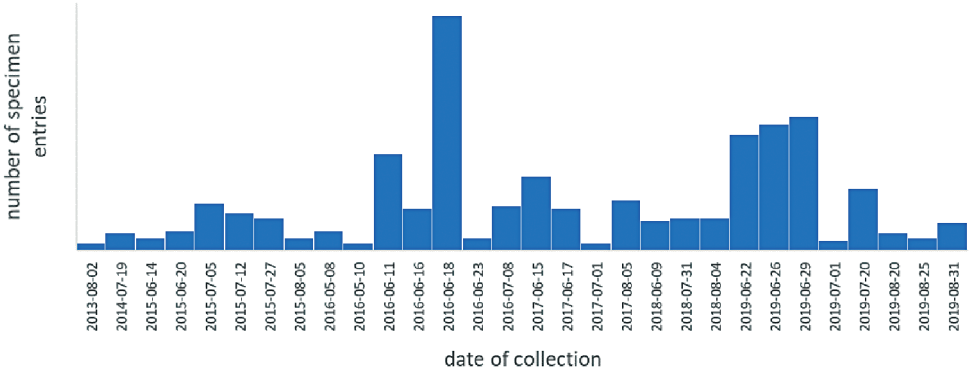


Fig. 16.3. Chronological composition of dataset by date of collection

Data quality and availability

The data included in the dataset has not been processed in any way (raw data). The data quality of the dataset has been checked with several taxonomic authorities.

Dataset DOI
10.34808/jh0b-ne37

Dataset License
CC-BY

Data Descriptor

Macrophytobenthos in the Puck Bay in 2010–2018 Dataset

Aleksandra Zgrundo[1*]

[1] Faculty of Oceanography and Geography, University of Gdańsk (8 Jana Bażyńskiego Street, Gdańsk, Poland)

* Correspondence author: aleksandra.zgrundo@ug.edu.pl; ORCID: 0000-0001-7700-5510

Abstract

The dataset titled Biomass of macrophytobenthos in the Puck Bay in 2010-2018 contains data on the qualitative composition and biomass of macrophytobenthos (flower plants and macroalgae) in samples collected in the Puck Bay area (Gulf of Gdańsk, southern Baltic Sea) at 20 stations between 2010–2018. The data was supplemented with additional information: values of measured parameters of water and sediment, e.g. temperature and salinity of water, sediment granulometry, and selected photos documenting the identified taxa. The data show the changes that have occurred in macrophytobenthic communities over the years.

Keywords: macrophytobenthos; plant communities; biomass; algae; angiosperms; Puck Bay; Baltic Sea

https://doi.org/10.34808/x55q-sz53_dyr_roz17

Specification table (data records)

Subject area	Botany, Ecology, Life Sciences
More specific subject area	Benthic plant communities
Type of data	Tables, Images, Map

How the data was acquired	The data was collected during projects conducted at the University of Gdańsk. Samples of macrophytobenthos were taken with Van Veen or Ekman grab between 2010 and 2013 and by a scuba diver using DAK in 2018. The plant material was kept at -20°C until analysis. All plants were identified to the lowest taxonomic level possible and sorted. The identified taxa were dried at 60°C to a stable weight in a Binder FED53 (Tuttlingen, Germany) drier and weighed to the nearest 0.001 g using an MS204S Mettler Toledo (Greifensee, Switzerland) scale. Biomass values were converted to g·m-2. Depending on the project, the physicochemical parameters of the water (i.e. temperature, salinity, dissolved oxygen, chlorophyll and total particulate matter content) and sediment (i.e. granulometry, organic carbon and organic nitrogen content) were also measured
Data format	Raw. The tables are in .xlsx format, Images are in .jpg format, Map is in .pdf format
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes. Share alike
Related research articles	Sokołowski, A., Ziółkowska, M. and Zgrundo A. (2015) 'Habitat-related patterns of soft-bottom macrofaunal assemblages in a brackish, low-diversity system (southern Baltic Sea)', Journal of Sea Research (103), pp. 93–102 Bełdowska, M., Jędruch, A., Zgrundo, A., Ziółkowska, M., Graca, B. and Gębka K. (2016) 'The influence of cold season warming on the mercury pool in coastal benthic organisms', Estuarine, Coastal and Shelf Science (171), pp. 99-105

Background

The rich plant meadows of the Puck Bay were severely degraded from 1960 to 1980 due to industrial plant exploitation (e.g. Furcellaria lumbricalis (Hudson) J.V.Lamouroux, Fucus vesiculosus L., Zostera marina L.), eutrophication and industrial pollution (Pliński 1982, Kruk-Dowgiałło 1991). However, the plant communities in the study area are still considered exceptionally rich and diverse compared to other marine ecosystems of the Polish coast (Gic-Grusza et al., 2009).

Access to data on the qualitative and quantitative composition of the plant communities in the Puck Bay is limited as the data produced within commercial projects are not made available and monitoring within the framework of the National Environmental Monitoring (Państwowy Monitoring Środowiska), although carried out since 2006, includes only 2 sites investigated twice a year. Therefore, the data collected in the Biomass of macrophytobenthos in the Puck Bay in 2010-2018 database, obtained during various

projects carried out at the Institute of Oceanography, University of Gdańsk, are a valuable complement to the monitoring data. The database makes it possible to trace changes in benthic plant communities with respect to composition and structure, and to supplement or verify information on the extent of occurrence of particular algal and vascular plant taxa.

Methods

The field studies were conducted in the Puck Bay (western shallow water part of the Gulf of Gdańsk, Baltic Sea) on benthic flora (macroalgae and flowering plants) along with environmental parameters measurements. The material for qualitative and quantitative analysis of macrophytobenthos was collected at 21 sites (Fig. 17.1) between 7th July 2010 and 14th October 2018. During 2010–2013, the samples of macrophytobenthos were taken with Van Veen or Ekman grab and in 2018 by a scuba diver using DAK. At all stations, 3–4 subsamples were collected. All subsamples were put in plastic bags and transported to the laboratory in a cool box. The plant material was kept at –20°C until analysis. All plants were identified to the lowest taxonomic level possible and sorted. The identified taxa were dried at 60°C to a stable weight in a Binder FED53 (Tuttlingen, Germany) drier and weighed to the nearest 0.001 g using an MS204S Mettler Toledo (Greifensee, Switzerland) scale. The biomass values were converted to g·m-2. The fieldwork and laboratory works were carried out according to recommendations of Helcom Combine (Part C, Annex C9) and the Polish National Monitoring Programme (Kruk-Dowgiałło et al., 2010).

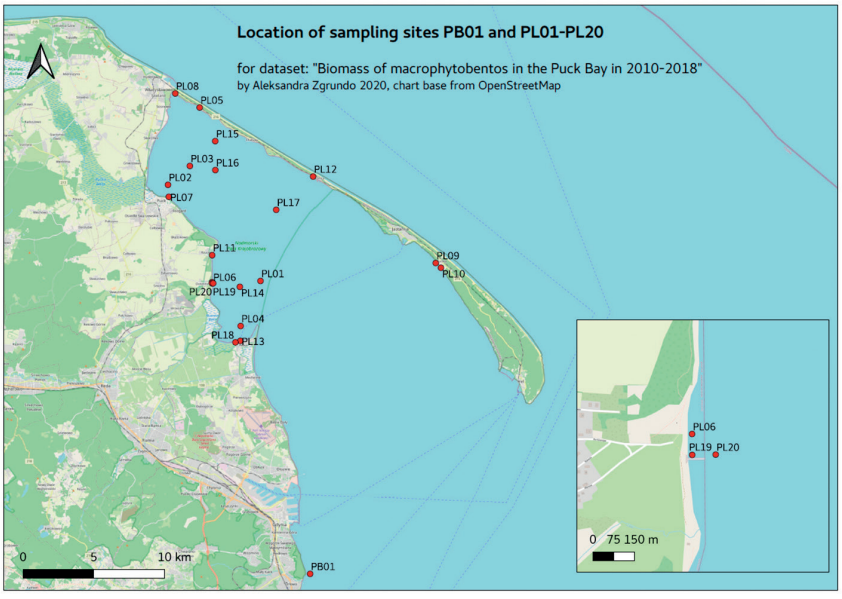


Fig. 17.1. Location of sampling sites

The morphological identification of macroalgae and flowering plants was based mostly on local key floras, e.g. Pliński and Hindak (2012), Pliński and Surosz (2013), Pliński and Szmeja (2013) and widely recognised literature, e.g. Braune and Guiry (2011), Bunker et al. (2017).

Simultaneously with field works aimed at macrophytobenthos collection, depending on the project, the physicochemical parameters of the water (i.e. temperature, salinity, dissolved oxygen, chlorophyll and total particulate matter content) and sediment (i.e. granulometry, organic carbon and organic nitrogen content) were measured. The detailed methodologies of collection of the samples and analyses of the environmental parameters are included in the publications of Sokołowski, Ziółkowska and Zgrundo (2015) and Bełdowska et al. (2016).

Data quality and availability

In order to ensure the quality of the data obtained, sampling and analysis of the samples were performed using recommended methods and protocols (Helcom Combine, Part C, Annex C9; Kruk-Dowgiałło et al., 2010). The correctness of the data recording in the database was verified by a group of 31 students using the database during online classes.

The database contains gaps in relation to the physicochemical parameters of the water and sediment.

Dataset DOI
10.34808/tnq7-nd63

Dataset License
CC-BY-SA

Acknowledgements
The generation of this dataset was supported by the Ministry of Science and Higher Education (project NN304 162237), the National Science Centre (project UMO-2011/01/B/ST10/07697) and a statutory research grant from the University of Gdańsk.

References

Bełdowska, M. et al. (2016) 'The influence of cold season warming on the mercury pool in coastal benthic organisms', *Estuarine, Coastal and Shelf Science* (171), pp. 99–105. DOI: 10.1016/j.ecss.2016.01.033.

Braune, W. and Guiry, M.D. (2011) '*Seaweeds. A colour guide to common benthic green, brown and red algae of the world's oceans*', A.r.g. Gantner Verlag, pp. 601.

Bunker, F. et al. (2017) '*Seaweeds of Britain and Ireland*', Marine Conservation Society, pp. 312.

Gic-Grusza, G. et al. (2009) 'Atlas of Polish marine area bottom habitats', *Environmental valorization of marine habitats*, Broker-Innowacji, Gdynia, p. 180.

Helcom Combine, Part C, Annex C9, https://helcom.fi/media/documents/Manual-for-Marine-Monitoring-in-the-COMBINE-Programme-of-HELCOM_PartC_AnnexC9.pdf (Accessed: 9th February 2020).

Kruk-Dowgiałło, L. (1991) 'Long-term changes in the structures of underwater meadows of the Puck Lagoon', *Acta Ichthyologica et Piscatoria, XXI Supplement*, pp. 77–84. DOI: 10.3750/AIP1991.21.S.09.

Kruk-Dowgiałło, L. et al. (2010) 'Macroalgae and angiosperms' [in Polish], Kruk-Dowigiałło et al. (eds.), *Przewodniki metodyczne do badań terenowych i analiz laboratoryjnych elementów biologicznych wód przejściowych i przybrzeżnych*. Biblioteka Monitoringu Środowiska, ISBN 978-83-61227-36-6, pp. 33–64.

Pliński, M. (1982) 'The distribution and biomass of phytobenthos in the internal part of Puck Bay', *Studia i Materiały Oceanologiczne* (39), Biologia Morza (6), pp. 195–217.

Pliński, M. and Hindak, F. (2012) 'Chlorophyta (Green Algae). Filamentous Green Algae (Ulotrichophycae, Conjugatophycae and Charophyceae)' [in Polish], Pliński M. (eds.). *Flora Zatoki Gdańskiej i wód przyległych (Bałtyk Południowy)*, 7/2. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, pp. 150.

Pliński, M. and Surosz, W. (2013) 'Red Algae ad Brown Algae' [in Polish], Pliński M. (eds.). *Flora Zatoki Gdańskiej i wód przyległych (Bałtyk Południowy)*, 6. Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, pp. 158.

Pliński, M. and Szmeja, J. (2013) 'Seed plants – Spermatophyta (Magnoliopsida & Liliopsida)' [in Polish], Pliński M. (eds.). *Flora Zatoki Gdańskiej i wód przyległych (Bałtyk Południowy),* 8, Gdańsk: Wydawnictwo Uniwersytetu Gdańskiego, pp. 104.

Sokołowski, A., Ziółkowska, M. and Zgrundo A. (2015) 'Habitat-related patterns of soft-bottom macrofaunal assemblages in a brackish, low-diversity system (southern Baltic Sea)', *Journal of Sea Research*, 103, pp. 93–102. DOI: 10.1016/j.seares.2015.06.017.

Data Descriptor

Cyanobacterial and Algal Strains in the Culture Collection of Baltic Algae (CCBA)

Filip Pniewski[1*]

[1] Department of Marine Ecosystems Functioning, Faculty of Oceanography and Geography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

* Correspondence author: filip.pniewski@ug.edu.pl; ORCID: 0000-0003-0313-6627

Abstract

The dataset titled Microalgal strains from "Culture Collection of Baltic Algae (CCBA)" is a representation of cyanobacterial and microalgal cultures isolated from the Baltic Sea. It is a unique catalogue of strains of the dominant and rare species found in the Baltic phytoplankton and microphytobenthos assemblages. The main purpose of the collection is to extend the knowledge on the Baltic microbial communities by providing high quality material for research, education and commercial ventures. The increasing interest of the biotechnological industry in marine biological resources makes microalgal strain collection an important provider of resource material for future technological developments.

Keywords: Baltic, microalgae, culture collections

https://doi.org/10.34808/x55q-sz53_dyr_roz18

Specification table (data records)

Subject area	culture collections, microalgae, cyanobacteria
More specific subject area	Cyanobacterial and microalgal strains isolated from the Baltic Sea
Type of data	Text and micrographs
How the data was acquired	The data were collected during routine activities in the collection
Data format	The table is in .xlsx format; micrographs are in .jpg and .tif formats
Measurement Type	Taxonomic inventory
Design Type(s)	Data integration objective, database creation objective

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible for any research, educational and commercial purposes under CC BY licence

Background

The dataset contains information on the cyanobacterial and microalgal strains maintained at the Culture Collection of Baltic Algae (CCBA) at the Institute of Oceanography UG. The collection maintains cyanobacterial and algal strains isolated from the Baltic Sea and additionally several strains collected from a wide range of habitats. The culture collection specialises in the Polish region and conducts continuous enlargement of the resources by the addition of new isolates. The cultures are unialgal, but mostly non-axenic. The strains are available for research, educational and commercial purposes. The main research objectives of the collection include: (1) isolation of new strains from Baltic and Polish inland waters with special attention paid to algal species threatened by extinction, rare, harmful as well as typical of the Polish region, (2) extension of the knowledge on algal biodiversity, and (3) investigation of different environmental factors to determine the optimal conditions for algal growth. With the increasing interest in microalgal biomass as a high-quality resource material for biotechnology, the role of culture collections is increasing. Therefore, sharing information on the microalgal resources has become an important task allowing for the development of new eco-friendly technologies in various commercial sectors.

Methods

Water and sediment samples were collected using a plankton net and sediment corer. Subsequently, cyanobacteria and microalgal strains were isolated with standard microbiological isolation techniques, including: serial dilution, plating and by using a micromanipulator. The strains were grown in f/2 and BG-11 media which were specifically designated for marine microalgae and cyanobacteria, respectively. Stable unialgal cultures were then included into collection resources and subjected to microscopic analysis. The strains were identified based on their morphological features and micrographs were taken. Strains of special interest were also investigated using molecular biology tools and data on their genetic diversity was collected (i.e., by generating their 18S rDNA and ITS sequences).

Data records

The collection contains 86 original strains, both planktonic and benthic. They represent the diversity of both community types. They are maintained as unialgal, non-axenic strains in liquid media prepared using natural Baltic water. They are kept under constant culturing conditions, i.e. low light of 50 mmol of photons m-2 s-1 and a temperature of

18˚C, ensuring their stable growth. The strains maintained in the CCBA collection cover various taxonomical groups, including: cyanobacteria, diatoms and green algae.

The dataset provides: (1) basic information on the year and location of hte environmental samples from which the strains were isolated, and their culturing conditions, (2) strain taxonomic affiliations, and (3) information on available genetic data. The dataset was prepared in agreement with the Darwin Core standard. The taxonomic coverage presented in the dataset is shown in Fig. 18.1.

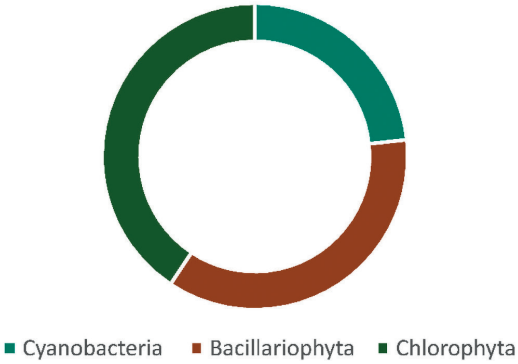


Fig. 18.1. Taxonomic content of the dataset

Data quality and availability

The data included in the dataset was not processed in any way (raw data). The data quality of the dataset was consulted with taxonomic authorities.

Dataset DOI
10.34808/6myr-f916

Dataset License
CC-BY-NC-ND

References

Anderson, R.A. et al. (2005) 'Appendix A – Recipes for freshwater and seawater media', in: Anderson, R. A. (ed.) *Algal culturing techniques*. Burlington: Elsevier/Academic Press, pp. 429–538.

Hoshaw, R. W. and Rosowski, J. R. (1979) 'Methods of microscopic algae', in: Stein, J. R. (ed.) *Handbook of phycological methods. Cultures methods & growth measurements*. Cambridge: University Press, Cambridge, pp. 53–67.

Pniewski, F., Friedl, T. and Latała, A. (2010) 'Identification of diatom isolates from the Gulf of Gdańsk: Testing of species identifications using morphology, 18S rDNA sequencing and DNA barcodes of strains from the Culture Collection of Baltic Algae (CCBA)', *Oceanological and Hydrobiological Studies*, 39(3), pp. 3–20. DOI: 10.2478/v10009-010-0031-7.

Data Descriptor

Occurrence of Cyanobacteria in the Gulf of Gdańsk (2008–2009)

Justyna Kobos[1*], Anna Krakowiak[1], Agata Błaszczyk[1], Anna Toruńska-Sitarz[1]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

* Correspondence author: justyna.kobos@ug.edu.pl; ORCID: 0000-0002-9360-2187

Abstract

Blooms of cyanobacteria develop each summer in the Baltic Sea. Collecting complete data on this phenomenon is helpful in understanding the changes taking place in the Baltic Sea and forecasting the occurrence of these phenomena in the future. This dataset includes unpublished information about the occurrence of cyanobacteria in the Gulf of Gdańsk (Southern Baltic) in 2008 and 2009. The presented data combines basic physicochemical and biological parameters of the sampled waters, therein: water temperature, salinity, chlorophyll a, occurrence of diazotrophic cyanobacterial species from Nodularia, Aphanizomenon and Dolichospermum genera, their biomass and abundance, concentration of detected cyanotoxins. The obtained results indicate the presence of toxic cyanobacteria species in the Gulf of Gdansk. The abundance and biomass of cyanobacteria and the concentration of toxins changed during the season. The highest concentration of nodularin was recorded on 29th of July 2009 when Nodularia spumigena dominated in the phytoplankton.

Keywords: cyanobacteria; bloom; cyanotoxins; nodularin; Nodularia spumigena

https://doi.org/10.34808/x55q-sz53_dyr_roz19

Specification table (data records)

Subject area	Earth Science/Environmental Sciences/Ecology
More specific subject area	Cyanobacterial Harmful Algal Blooms
Design Type(s)	Database creation objective/ Compilation of data from various sources

Measurement Types(s)	Salinity, Temperature, Cyanobacterial diversity, Cyanobacterial biomass, Chlorophyll a, Diversity of cyanotoxins, Concentrations of cyanotoxins in cyanobacteria cells and water
Technology Type(s)	pH/conductivity multifunction meter, Spectrophotometer, Microscope, High Performance Liquid Chromatography
Factor Type(s)	Geographic location, Temporal interval
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes.

Background

Toxic phytoplankton blooms that appear as result of natural water eutrophication are often intensified by anthropogenic influence. Blooms change the water colour and turbidity, which looks as if it has been covered with thick paint. Their negative impact on the environment is also associated with the reduction in phytoplankton biodiversity and the presence of toxic species. In the Baltic Sea, blooms of filamentous cyanobacteria are reported regularly every year in late summer (Mazur-Marzec et al., 2006). Currently, blooms of Aphanizomenon flosaquae and Nodularia spumigena are typical phenomena for waters of the Baltic proper; whereas, Dolichosspermum spp. occurs primarily in coastal waters. It is believed that salinity is one of the most important abiotic factors that determines the structure of phytoplankton communities and the development of cyanobacterial blooms. Water temperature, is another important determinant of the mass growth of cyanobacteria from the Nodularia genus. The blooms can be particularly harmful when the nodularin (NOD)-producing species, N. spumigena, dominates.

With EU Directive (2006/7/CE), monitoring of cyanobacterial proliferation became mandatory but no precise framework was provided for monitoring. Current WHO alert levels for bathing waters are based on parameters such as chlorophyll a (Chl a) and cyanobacterial density.

Our dataset, Cyanobacterial occurrence in the Gulf of Gdansk (https://doi.org/10.34808/14tr-n964), contains the results of environmental studies on cyanobacterial blooms in the waters of the Gulf of Gdansk during the summer seasons of 2008 and 2009. The data concerns more than 90 samples collected from the coastal zone and the open waters of the Gulf. The mean water temperature values measured at the Gdynia boulevard and during cruises in the central part of the Gulf of Gdańsk were similar and were about 18°C. During the summer season, average Chl a concentrations were comparable in different parts of the gulf (~ 5 μg dm^{-3}, boulevard; ~7 μg dm^{-3}, central part). Only in the coastal zone, high chlorophyll concentrations were recorded during phytoplankton blooms. The results indicate that when N. spumigena formed blooms, they were always

toxic. The mass presence of this NOD-producing species was observed sporadically at the Gdynia boulevard, with only 3 samples from 2008 and 3 samples from 2009. In the central part of the Gulf of Gdańsk, no toxins were found in the 26 samples collected in 2008. While in 2009, nodularin was identified in 3 of 26 phytoplankton samples.

The accumulation of harmful biomass of cyanobacteria in places used for bathing and recreation may pose a threat to persons who bathe in the water. Therefore, it is necessary to monitor the possibility of toxic blooms every year.

Methods

The r/v 'Oceanograf 2' of the Institute of Oceanography, University of Gdańsk, was used for sample collection in the Gulf of Gdańsk. In addition, samples were collected from one onshore station situated at the end of the pier in Gdynia. Water and phytoplankton samples were collected, extracted, and Chl a, NOD, and microscopic analysed according to the methods described by Mazur-Marzec et al. (2006).

Data quality and availability

Sampling and analysis of the samples were performed using recommended methods and protocols (HELCOM, Meriluoto and Spoof, 2005).

Dataset DOI
10.34808/14tr-n964

Dataset License
CC-BY-NC

Acknowledgements
The generation of this dataset was supported by a statutory research grant from University of Gdańsk.

References

HELCOM '*Guidelines for monitoring of phytoplankton species composition, abundance and biomass*', www.helcom.fi (Accessed: 15th January 2021).

Macário, I.P.E. et al. (2017) 'Stepwise strategy for monitoring toxic cyanobacterial blooms in lentic water bodies', *Environmental Monitoring and Assessment*, 189 (12), 620, DOI: 10.1007/s10661-017-6292-9.

Mazur-Marzec, H. et al. (2006) 'Toxic Nodularia spumigena blooms in the coastal waters of the Gulf of Gdańsk: a ten-year survey', *Oceanologia*, 48 (2), pp. 255–273. http://www.iopan.gda.pl/oceanologia/482mazur.pdf.

Meriluoto, J. and Spoof, L. (2005) 'SOP: Analysis of microcystins by high-performance liquid chromatography with photodiode array', in Meriluoto J., G. A. Codd (ed.) *TOXIC: Cyanobacterial Monitoring and Cyanotoxin Analysis*, Åbo: Åbo Akademi University Press, pp. 77–85.

Persson, K-J., Stenroth, P. and Legrand, C. (2011) 'Effects of the filamentous cyanobacterium Nodularia of fitness and feeding behaviour of young-of-the-year (YOY) Eurasian perch (Perca fluviatlils)'. *Toxicon*, 57 (7–8), pp. 1033–1040. DOI: 10.1016/j.toxicon.2011.04.007.

Data Descriptor

Mercury in various components of the environment (Gulf of Gdańsk, southern Baltic Sea)

Magdalena Bełdowska[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk, (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: magdalena.beldowska@ug.edu.pl; ORCID: 0000-0002-3585-9283

Abstract

Total mercury concentration was determined monthly form January 2012 to May 2013 in the water (microlayer and the surface), plankton, epilithon, epiphyton, benthos and sediment. The study was conducted in the coastal zone of the Gulf of Gdańsk. The abundance and biomass of plankton, benthos; the concentration of suspended particulate matter, particulate organic carbon and nitrogen; δ13C and δ15N in the suspended matter; sediment and water parameters were also measured. The Hg concentration in the biotic and abiotic components of the marine coastal zone was higher during the warm cold season.

Keywords: Mercury, coastal zone, estuarium, microlayer, surface water, suspended particulate matter, phytoplankton, zooplankton, phytobentos, zoobenthos, epilithon, epifithon, sediment

https://doi.org/10.34808/x55q-sz53_dyr_roz20

Specification table (data records)

Subject area	Marine chemistry
More specific subject area	Biogeochemistry of mercury
Type of data	text
How the data was acquired	The analysis were conducted in the Institute of Oceanography University of Gdańsk and in the Institute of Oceanology of the Polish Academy of Sciences

Data format	xlsx
Experimental factors	no
Experimental features	Chemical analysis; biological analysis
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	Creative Commons Attribution Non Commercial

Background

Climate change in the southern Baltic region manifests itself inter alia by warming of the cold season. Warming of the cold season leads to an increase of sea temperature, enabling the growth of phytoplankton. This contributes to the accumulation of mercury entering the Gulf via dry deposition by the plankton and inclusion of this metal into the food chain, instead of deposition to the sediments. Additionally, the lack of icing causes the biomass of plankton in early spring to be larger than in years where icing occurs. In a similar way, icing affects the biomass of fauna and flora of the sediments of the Bay. In this case, in years with no icing, the average annual mercury pool in the phytobenthos is higher than in years where icing would last for 90 days (Bełdowska and Kobos, 2016; Bełdowska et al., 2016).

The extension of the vegetative season (and accompanied by a lack of icing) and improvement of the water quality, contribute to extended coverage of sediments by flora. Underwater meadows formed in the Puck Bay are a habitat where animals are abundant – many of them herbivores, which are directly accumulating mercury by ingestion of phytobenthos. In this way, intense development of phytobenthos contributes to faster inclusion of mercury into food webs (Bełdowska et al., 2015).

Climate changes contributes on one hand to the depuration of water and sediments by flora, but on the other hand they transfer mercury relatively biounavailable for animals from sediments to higher trophic levels. It was also observed that in coastal sediments, in the absence of icing, mercury persisted in bioavailable forms for a larger proportion of the year (Bełdowski et al., 2018).

Methods

The collected samples were freeze-dried and analysed by means of a thermo-desorption advanced mercury analyser (AMA 254). Qualitative analysis of plankton was carried out in accordance with the procedure set out by Helcom Combine (2014). Macrobenthic organisms were identified to the highest possible taxonomic separation using the taxonomic keys of Bernatowicz and Wolny (1969), Braune and Guiry (2011), Barnes (1995) and Żmudzinsk, (1990).

Data records

Samples and environmental data were collected at three coastal stations: Chałupy (54 44 20.605 N, 018 33 54.912 E), Osłonino (54 40 06.516 N, 018 27 58.377 E) and Gdynia (54 29 18.740 N, 018 34 06.401 E). Analyses were conducted at the University of Gdańsk, Gdynia, Poland. The tables are in .xlsx format.

Data quality

All chemical measurements were controlled by analysis of reference materials within the certified ranges. The dataset is accessible and is publicly and freely available for any research or educational purposes.

Data availability

Dataset DOI
10.34808/5pe0-g310

Dataset License
CC-BY-NC

Acknowledgements
This study was performed within the framework of the National Science Centre project No. 2011/01/B/ST10/07697.

References

Barnes, R. S., (1995) '*The Brackish-water Fauna of Northwestern Europe*', Cambridge University Press, p. 304.

Bełdowska, M. et al. (2015) 'Macrophyta as a vector of contemporary and historical mercury from the marine environment to the trophic web', *Environmental Science and Pollution Research,* 22, pp. 5228–5240. DOI: 10.1007/s11356-014-4003-4.

Bełdowska, M. et al. (2016) 'The influence of cold season warming on the mercury pool in coastal benthic organisms', *Estuarine, Coastal and Shelf Science*, 171, pp. 99–105. DOI:10.1016/j.ecss.2016.01.033.

Bełdowska, M. and Kobos, J. (2016) 'Mercury concentration in phytoplankton in response to warming of an autumn – winter season', *Environmental Pollution,* 215, pp. 38–47; DOI: 10.1016/j.envpol.2016.05.002.

Bełdowski, J. et al. (2018) 'Seasonal changes of mercury speciation in the coastal sediments', *Journal of Soils and Sediments,* 18, pp. 3424–3436. DOI: 10.1007/s11368-018-1993-4.

Bernatowicz, S. and Wolny, P. (1969) '*Fishing Botany*', National Agricultural and Forestry Publisher, Warszawa (in Polish).

Braune, W. and Guiry, M. D. (2011) '*Seaweeds. A Colour Guide to Common Benthic Green*','Brown and Red Algae of the World's Oceans' Koeltz Scientific Books, Koenigstein.

HELCOM COMBINE (2014) Part C. 'Programme for monitoring of eutrophication and its effects. Annex C-6: Guidelines concerning phytoplankton species composition, abundance and biomass'.

Żmudzinski, L. (1990) '*The Fauna of the Baltic Sea*', Atlas Macrofauna, Wydawnictwo Szkolne i Pedagogiczne, Warszawa, p. 196 (in Polish).

Data Descriptor

Methodology of the Morphometry Study on Baltic Herring

Jakub Idczak[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: jakub.idczak@ug.edu.pl; ORCID: 0000-0001-5459-1393

Abstract

Acoustic techniques are used in the estimation of the abundance of Baltic herring. Investigations of the relationship between the Baltic herring individual target strength TS and the total fish length L, significant in the acoustic assessment, demonstrated its dependence on the study area location. It motivated the detailed analysis of the relationship between herring from the southern Baltic ICES Subdivisions 24, 25, and 26, in which Poland is responsible for the acoustic herring abundance assessment. The modelling approach, based on the Modal Based Deformed Cylinder Model approximation, was used to study the backscattering. The detailed shape of the body and swim bladder of herring individuals was considered. The article describes the methodology of acquiring data for the model study containing information on the detailed shapes of the bodies and swim bladders of 74 herring individuals obtained in the area of the Polish coastal zone (ICES Subdivision 26).

Keywords: Baltic herring, fish morphometry, X-ray images, numerical modelling of backscattering by fish, target strength, abundance estimation

https://doi.org/10.34808/x55q-sz53_dyr_roz21

Specification table (data records)

Subject area	Fish biomass estimation
More specific subject area	Modelling of acoustic backscattering by southern Baltic herring
Type of data	X-ray images, morphometric measurements and swimbladder countours

How the data was acquired	Fish for research - during the cruise on the fishing vessel X-rays images – taken at the hospital in Hel
Data format	.jpg, .dwg, .txt, .cdr, .xlsx
Experimental factors	Baltic herring
Experimental features	Real-time X-ray living fish analysis
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	CC-BY-NC

Introduction

The published datasets (X-ray images of Baltic herring and X-ray images of Baltic herring. Data analysis) contain detailed information on the shapes of the swim bladders and fish bodies of herring from ICES Subdivision number 26. These data were collected during the author's doctoral dissertation (Idczak, 2015) and were used to model the backscattering characteristics of herring individuals obtained from the Polish coastal zone of the Baltic Sea.

Considering that there is variation in the ontogenesis of herring in various regions of the Baltic Sea (ICES Reports 2004; 2005a, b; 2006), which may have an influence on the geometrical shape of their swim bladders, we can also expect geographic variation in the shape of this organ. An indirect confirmation of this may be the target strength study of herring individuals carried out for various areas of the Baltic Sea (Rudstam, Lindem and Hansson, 1988; Rudstam, 1999; Didrikas and Hansson, 2004; Peltonen and Balk, 2005; Kasatkina, 2009), the value of which largely depends on the morphometric parameters of the fish's swim bladders (Gorska, 2007; Fässler et al., 2008; Fässler and Gorska, 2009; Gorska and Idczak, 2010). These studies revealed a geographic differentiation in the backscattering characteristics of the Baltic herring individuals.

Earlier studies of the shape of the swim bladder of Baltic herring based on X-ray images was performed only for fish individuals obtained from the Baltic Sea area along the Swedish coast (ICES subdivisions 25, 27, 29S) (Håkansson and Arrhenius, 2002; Swedish Museum of Natural History and Danish Institute for Fisheries Research). In order to study the shapes of the swim bladders of fish from other areas of the Baltic Sea, a collection of X-ray images for herring obtained from other areas of the Baltic Sea had to be developed.

Methodology of research on the geometric shape of the swim bladder and the body of herring individuals.

In order to precisely determine the shape of the swim bladder of the South Baltic herring individuals, a research methodology was developed that does not cause significant changes in the shape of this organ in relation to its natural shape (while the fish is in the marine environment) (Idczak, Gorska and Arciszewski, 2011; Idczak and Kniaź-Kubacka, 2012). The biggest challenge in studying the natural shape of the herring's swim blad-

der using X-rays was the necessity to obtain live fish and keep them alive during the research. To achieve this goal, the following actions were undertaken:fish for research were obtained from shallow depths, minimising the risk of distortion of the swim bladder as a result of a sudden change in hydrostatic pressure;a method of storing and transporting the fish (from the moment they are caught until carrying out the X-ray examinations) was developed, assuring the good condition of the fish individuals in order to maintain the natural shape of their swim bladders; the places for collecting, storing fish and taking X-ray images were located in close proximity to each other (to shorten the time of fish transport as much as possible); maximum attention was given when restraining the fish immediately before X-ray examinations (in order not to cause any damage).

Collection of research material

The collection of biological material for X-ray examinations took place in November 2011. The herring was collected from the HEL-125 fishing boat in the Gulf of Gdańsk, south-east of the Hel Peninsula. During the fishing, a pelagic trawl with a mesh side of 11 mm with a horizontal span of about 18 m and a vertical span of about 20 m was used. The trawl was towed at an average speed of approximately 3.5 knots. During transport to the place of X-ray examinations, the collected herring individuals were properly secured using plastic bags with sea water and oxygen.

X-ray studies of herring individuals

Before placing the fish on the X-ray table, it had to be immobilised without being killed. For this purpose, a solution of clove oil and 40% ethanol was used (Hazen and Horne, 2003, 2004) for which the proportions of the individual ingredients are given in Tab.21.1 (Horne and Jech, 2001). The fish were placed in the solution for several seconds. It resulted in their anesthesia, but did not cause a lethal effect. The use of the solution prevented each fish from moving while it was being X-rayed. This allowed for quick X-ray examinations for the desired arrangement of fish on the X-ray table.

Tab. 21.1

Composition and concentration of the solution used for anesthesia of herring individuals (Horne and Jech, 2001)

Concentration (ppm)	Clove oil (ml)	$C_2 H_5 0H$ (40%) (ml)	Water (l)
40	0.4	3.6	10

X-ray studies of 220 specimens of herring were carried out on 03–04.11.2011. In order to study the morphometry of the swim bladder and the body of the fish, two images were taken for each individual: one in the lateral position and one in the dorsal position.

When the image was taken in the dorsal position, the fish was placed on its stomach on the X-ray table and supported on both sides with paper wedges. The X-ray machine parameters were set to low values (applied voltage = 42kV and the product of the current through the X-ray tube (mA) and the X-ray exposure time (s) = 1.2 mAs for fish placed in a lateral position, and 42 kV and 1.6 mAs, respectively, for fish placed in the dorsal position) so that the images of the fish tissue were recorded on the X-ray film.

Before taking the X-rays, in order to identify the fish, an individual metal number was placed next to each fish individual, visible on the X-ray image. The individual fish in the photo had the same number in the lateral and dorsal positions. After the photo was taken, the number was permanently attached to the tail fins of the fish. Then the fish were placed in a container for subsequent geometric measurements of their bodies.

While waiting for the X-ray examinations, the fish were kept in specialised aquariums filled with seawater. For the period of transport for the X-ray examinations, they were re-packed in plastic bags with sea water and oxygen. Only those fish whose form was assessed as good, similar to the form during the stay of the fish in its natural environment, were selected for X-ray tests. All selected fish fled while attempting to be caught with a net in the aquarium and swam naturally in the entire depth of the water. On the first day of the research, X-rays were taken of 136 herring individuals, the next day – the remaining 84 individuals.

Scanning of X-ray films

Next, images were selected for morphometric studies of the fish, and for this purpose a method of converting photos from analog to digital was developed. The task was rather challenging because the photos were available only in the form of X-ray films, which made it impossible to simply scan them. For this purpose, a scanner with a transparency adapter was used, enabling X-ray film scanning.

Measurements of fish morphometric parameters

Shortly after the end of the X-ray examinations, the body lengths of the fish were measured (total length of the fish L – measured with the caudal fin and the standard length of the fish Lb – without the caudal fin, width wb and height hb of the fish bodies). Then, with the help of AutoCAD, the morphometric parameters of the swim bladders of the fish from the digitised X-ray images were measured (length Lsb, width wsb and height hsb).

In this way, the morphometric parameters of 135 herring individuals were measured. It was observed that due to the good condition of these fish individuals, from the moment they were caught from the sea until they were placed on the X-ray table, the swim bladders visible in the X-ray images retained their shapes similar to their natural shape, i.e. the shape of the swim bladders of herring found in their natural environment.

Digitisation of fish body and swim bladder contours

Next, in order to obtain input data for the modelling of acoustic wave backscattering on herring individuals, taking into account the detailed shapes of the fish bodies and swim bladders, the contours of the bodies and swim bladders had to be digitised from the X-ray images. For this purpose, a methodology for reading contours and a method for saving them as digital data that could be used in MATLAB as input data was developed.

For the digitisation of contours, only those X-ray images in which the contours of the fish bodies and swim bladders were clearly visible were selected. Digitisation was carried out using two programs: CorelDRAW and AutoCAD. Using the first program, the contours of the bodies and swim bladders of the fish were read. Using the second program, the contours were saved as digital data. Coordinates (x, y) were read for each contour point (in steps of 1 mm). In this way, from the X-ray images of 220 fish individuals, the contours of the bodies and swim bladders of 74 fish individuals were digitised.

The data prepared in this way can be used as input data in model studies of the backscattering characteristics of herring individuals taking into account the detailed shapes of fish bodies and swim bladders. In addition, the existing collection of X-ray images of herring obtained off the coast of Sweden has been expanded to include X-ray images of fish from the Polish coastal zone.

Data availability

Dataset DOI
10.34808/s4sj-b755
10.34808/80xc-f462

Dataset License
CC-BY-NC

References

Didrikas, T. and Hansson, S. (2004) 'In situ target strength of the Baltic Sea herring and sprat', *ICES Journal of Marine Science*, 61 (3), pp. 378-382. DOI: 10.1016/j.icesjms.2003.08.003.

Fässler, S.M.M. et al. (2008) 'Differences in swimbladder volume between Baltic and Norwegian spring-spawning herring: Consequences for mean target strength', *Fisheries Research*, 92(2–3), pp. 314–321. DOI: 10.1016/j.fishres.2008.01.013.

Fässler, S.M.M. and Gorska, N. (2009) 'On the target strength of Baltic clupeids', *ICES Journal of Marine Science*, 66(6), pp. 1184–1190. DOI: 10.1093/icesjms/fsp005.

Gorska, N. (2007) 'On target strength of Baltic herring', *Proceedings of 2007 ICES Annual Science Conference*, Helsinki, Finland, 17–21th September, 2007, ICES Council Meeting 2007/H:7 (CD-version). Available at: https://www.ices.dk/sites/pub/CM%20Doccuments/CM-2007/H/H0707.pdf (Accessed: 25th May 2022).

Gorska, N. and Idczak, J. (2010) 'On the acoustic backscattering by Baltic herring and sprat', *Hydroacoustics*, 13, pp. 89-100. Available at: http://pta.eti.pg.gda.pl/journal/paper.py?id=469 (Accessed: 25th May 2022).

Hazen, E.L. and Horne, J. K. (2003) 'A method for evaluating the effects of biological factors of fish target strength', *ICES Journal of Marine Science*, 60 (3), pp. 555-562. DOI: 10.1016/S1054-3139(03)00053-5.

Hazen, E.L. and Horne, J. K. (2004) 'Comparing the modelled and measured target-strength variability of walleye pollock, Theragra chalcogramma', *ICES Journal of Marine Science*, 61(3), pp. 363–377. DOI: 10.1016/j.icesjms.2004.01.005.

Horne, J. and Jech, M. (2001) 'Baltic Sea manual', *ICES Working Group on Fisheries Acoustics, Science and Technology*.

ICES (2004) 'Report of the Study Group on Baltic Fish and Fisheries Issues in the BSRP (SGBFFI)', 3–5th February 2004, Riga, Latvia, *ICES Council Meeting 2004/H:04*. Available at: https://www.ices.dk/sites/pub/CM%20Doccuments/2004/H/SGBFFI04.pdf.

ICES (2005a) 'Report of the Study Group on Baltic Fish and Fisheries Issues in the BSRP (SGBFFI)', 9–13th June 2005, Riga, Latvia, *ICES Council Meeting 2005/H:05*. Available at: https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/bcc/2005/SGBFFI05.pdf.

ICES (2005b) 'Report of the Study Group on Multispecies Assessment in the Baltic (SGMAB)', 13–17th June 2005, Riga, Latvia, *ICES Council Meeting 2005/H:06*. Available at: https://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/bcc/2005/sgmab05.pdf.

ICES (2006) 'Report of the Study Group on Baltic Fish and Fisheries Issues in the BSRP (SGBFFI)', 9–13th June 2006, Vilnus, Lithuania, *ICES Council Meeting 2006/BCC:03*.

Idczak, J., Gorska, N. and Arciszewski, B. (2011) 'Study of swimbladder morphometryof Baltic herring and sprat (development of measurement methodology)', *Hydroacoustics,* 14, pp. 61–68. Available at: http://pta.eti.pg.gda.pl/journal/paper.py?id=499.

Idczak, J. and Kniaź-Kubacka, N. (2012) 'Backscattering properties of southern Baltic herring', *Hydroacoustics*, 15, pp. 57-64. Available at: http://pta.eti.pg.gda.pl/journal/paper.py?id=531.

Idczak, J. (2015) '*Mechanism of acoustic wave backscattering by Baltic herring (Clupea harengus), doctoral dissertation*', Uniwersity of Gdańsk, Faculty of Oceanography and Geography, Institute of Oceanography, Gdynia.

Kasatkina, S.M. (2009) 'The influence of uncertainty in target strength on abundance indices based on acoustic surveys: examples of the Baltic Sea herring and sprat', *ICES Journal of Marine Science*, 66(6), pp. 1404–1409. DOI: 10.1093/icesjms/fsp086.

Peltonen, H. and Balk, H. (2005) 'The acoustic target strength of Herring (Clupea harengus L.) in the northern Baltic Sea', *ICES Journal of Marine Science*, 62(4), pp. 803–808. DOI: 10.1016/j.icesjms.2005.02.001.

Rudstam, L., Lindem, T. and Hansson, S. (1988) 'Density and in situ target strength of herring and sprat: a comparison between two methods of analyzing single beam sonar data', *Fisheries Research*, 6(4), pp. 305–315. DOI: 10.1016/0165-7836(88)90001-X.

Rudstam, L.G. et al. (1999) 'Comparison of target-strength distributions and fish densities obtained with split- and single-beam echo sounders', *Fisheries Research*, 42(3), pp. 207–214. DOI: 10.1016/S0165-7836(99)00047-8.

Data Descriptor

Early Oceanographical Data Collected by the Institute of Oceanography, University of Gdańsk

Maciej Matciak[1*], Adam Krężel[1]

[1] Department of Physical Oceanography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: maciej.matciak@ug.edu.pl; ORCID: 0000-0001-9707-9349

Abstract

Three data sets entitled Water currents in Głębinka Passage in late spring of 1975, Hydrometeorological and hydrochemical conditions in the Gulf of Gdańsk in the vicinity of Vistula river mouth in July of 1977, and Gulf of Gdańsk monitoring conducted by the Institute of Oceanography, University of Gdańsk, in 1981–1994 contain archival field measurement results from the Gulf of Gdańsk (the southern Baltic). The data can be used for the assessment of the long-term changes in the marine coastal environment and for validation in retrospective modelling.

Keywords: archival data, field measurements; seawater properties; meteorological conditions; Gulf of Gdańsk

https://doi.org/10.34808/x55q-sz53_dyr_roz22

Specification table (data records)

Subject area	Physical and chemical oceanography, Meteorology
More specific subject area	Field measurements of seawater properties and meteorological parameters
Type of data	Text
How the data was acquired	The data was collected at the Institute of Oceanography, University of Gdańsk, with standard oceanographical and meteorological measurement devices used in fieldwork
Data format	The tables are in .xlsx format

Experimental factors	The data contained in the dataset were not processed
Experimental features	Water sampling, direct measurement in the field, laboratory analysis
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The datasets are accessible and are publicly and freely available for any research or educational purposes

Background

In 1974, when the Institute of Oceanography at University of Gdańsk (IO UG) acquired its first hydrographic research vessel, fields works in the coastal waters of the Gulf of Gdańsk (the southern Baltic) started. The non-biological research was mainly focussed on the water temperature and salinity variability, local sea current patterns, biogenic substances as well as the oxygen content in the water and meteorological observations. The results obtained in the 1970s concerned the interaction between waters of different origin, e.g. in the vicinity of the Vistula river mouth or in the Głębinka Passage connecting shallow and deep parts of Puck Bay (sub-area of the Gulf of Gdańsk). After the first experiences were gained, the Institute of Oceanography initiated multiannual monitoring of the coastal waters of the Gulf of Gdańsk, which was conducted in 1981–1994.

Methods

Oceanographical and meteorological parameters were measured at the sampling stations whose locations depended on the aim and area of study. Standard methods of measurements in the marine sciences were used. Water samples taken in the sea were transported to the laboratory and then analysed. Before 1985, in situ water salinity was determined with the use of Knudsen's chemical definition adapted for the Baltic Sea waters (Trzosińska, 1977) and after this year, by means of water conductivity measurements according to the recommendation of UNESCO (1981). Similarly, about 1985, the oceanographic reversible thermometers were replaced by electronic ones. During field campaigns, the following measurement devices were used: water samplers, including a Nansen water sampler equipped with a reversible thermometer; BPW-2 current meters; a white disc of 0.3 m in diameter (Secchi disc) for determining water transparency; CTD (conductivity-temperature-depth) automatic probes (a probe constructed in the Laboratory of Oceanographic Equipment at IO UG went into use in 1985 and a probe bought from Meerestechnik GmbH went into use in 1990); transmissometer constructed in the Laboratory of Oceanographic Equipment (IO UG) for determining the light-beam attenuation coefficient of green light (535 nm), anemometer, and psychrometer. A detailed description of the measurement methods can be found in the manuscripts of Pęcherzewski et al. (1977), Nowacki (1980) and e.g. Nowacki et al. (1986).

Data quality and availability

Oceanographic data were digitalised based on original measurement protocols. During this process, they were verified and they were excluded in the case of any doubt with respect to their quality.

Datasets DOI
10.34808/fadj-np77
10.34808/pg10-8e59
10.34808/ggag-hv21

Datasets License
CC-BY-NC

References

Nowacki, J. (1980) '*The role of submerged strait in water exchange in Puck Bay*' (in Polish) Ph.D. thesis, Institute of Oceanography, University of Gdańsk, pp. 115.

Nowacki, J. et al. (1986) '*Research on hydrology and hydrochemistry of the Gulf of Gdańsk in 1986*' (in Polish) Dept. Phys. Ocean., Institute of Oceanography University of Gdańsk, manuscript, pp. 45.

Pęcherzewski, K. et al. (1977) '*Hydrometeorological and hydrochemical conditions in the marine area near Vistula river mouth*' (in Polish) Dept. Phys. Ocean., Institute of Oceanography University of Gdańsk, pp. 57.

Trzosińska, J. (1977) 'Basic ionic composition of Baltic water. Baltic water, its composition and properties' (in Polish), *Studia i Materiały Oceanol*ogiczne, 17, pp. 165–180.

UNESCO (1981) 'Background papers and supporting data on the practical salinity scale 1978', UNESCO *Technical Paper in Marine Science,* 37, pp. 145.

Data Descriptor

High-Resolution Wind Wave Parameters in the Area of the Gulf of Gdańsk During 21 Extreme Storms

Gabriela Gic-Grusza[1], Aleksandra Dudkowska[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: aleksandra.dudkowska@ug.edu.pl; ORCID: 0000-0001-9781-8826

Abstract

This dataset contains the results of wind-wave parameter modelling in the area of the Gulf of Gdańsk (Southern Baltic). For the simulations, a high resolution SWAN model was used. The dataset consists of the significant wave height, the direction of the wave approaching the shore and the wave period during 21 historical, extreme storms. The storms were selected by an automatic search over the 44-year-long significant wave height time series.

Keywords: coastal zone; Gulf of Gdańsk hydrodynamics; SWAN model; wind-waves

https://doi.org/10.34808/x55q-sz53_dyr_roz23

Specification table (data records)

Subject area	Physical Oceanography, Coastal Zone, Wind-waves
More specific subject area	Extreme storm conditions
How the data was acquired	Numerical modelling using SWAN
Data format	Text files
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The principal goal of the calculation performed is to gain some preliminary insights into the intensity of water movement generated by wind waves in the Gulf of Gdańsk (located in the southern Baltic Sea) during severe storms. Baltic storms appear to be important phenomena from several points of view: first, as threatening to the safety of human activity, and secondly, they may have beneficial effects on the natural environment causing oxygenating inflows from the North Sea and refreshing waters in deeper layers, and in particular, the bottom waters. The third point is that Baltic storms may set in motion sediment transport in the coastal zone and may induce sedimentation processes at the bottom (Cieślikiewicz et al., 2017).

In the area of the Gulf, the generation and transformation of waves is limited by an open water fetch distance and complex bathymetry. A sheltering effect in the Gulf reduces wave energy compared to open coast areas and protects the Gulf from storms (Cieślikiewicz et al., 2016).

This dataset contains the results of SWAN (Booij et al, 1999) modelling in the area of the Gulf during 21 extreme storms selected based on hindcast wind wave data in the period 1958–2001. The wind wave fields' characteristics over the Baltic Sea were taken from the 44-year hindcast generated within the framework of the HIPOCAS project (Cieślikiewicz and Paplińska-Swerpel, 2008) using the WAM model (WAMDI Group, 1988).

Methods

Wave parameters during 21 extreme storms (output from the WAM model running in coarse resolution), as well as wind parameters, were used as the boundary conditions for the SWAN model operating in a high resolution grid covering the area of the Gulf of Gdańsk. The SWAN numerical model allows the estimation of surface gravity wave parameters for a given seabed topography, wind field, sea state, and the current field. SWAN determines the generation of waves by wind and wave propagation in time and space, taking into account a number of physical phenomena determining the wave field. These are mainly shoaling, refraction, nonlinear interaction between waves, and wave energy dissipation caused by whitecapping. Additionally, it includes energy dissipation by friction of the bottom and refraction due to ocean currents (Urbański et al., 2008). The SWAN model simulation was performed in this study on a grid with a spatial resolution of 200m × 200 m, covering the entire Gulf of Gdańsk. Boundary conditions in the form of monochromatic wave parameters were set on the Gulf of Gdańsk's open border. This northern border of the computational grid of the SWAN model includes 11 WAM grid points. All of the analysed extreme storms occurred during similar wind conditions, namely with a similar wind speed of several metres per second from the northern direction. All of the SWAN simulations were performed in stationary mode, and the results are given for the peak of each storm. For the modelling, high-resolution bathymetric data provided by the Naval Hydrographic Office in digital form prepared by the Maritime Institute in Gdańsk was used.

Data records

The dataset contains SWAN model output in the form of text files: significant wave height, peak period and wave direction for 21 extreme storms (63 files in total). The significant wave height (Hs) and mean wave period (T) are integral wave parameters calculated from the wave spectrum. The physical meaning of Hs is the average of wave height taken over the given time interval, calculated from the highest one-third of all waves. The T parameter can be interpreted as the averaged time interval between two corresponding points on the wave, i.e. troughs.

Data quality and availability

The SWAN model has been successfully applied in extreme wave conditions but not in the Baltic Sea. Although there have been several studies focussed on the modelling of the wave field in this region, SWAN model was only applied in a few cases (e.g. Reda and Paplińska, 2002). These data are accessible in the MOST Wiedzy Open Research Catalog.

Dataset DOI
10.34808/b8ar-2651

Dataset License
CC-BY

References

Booij, N., Ris, R.C. and Holthuijsen, L. H. (1999) 'A third-generation wave model for coastal regions. Model description and validation', *Journal of Geophysical Research*, 104 (C4), pp. 7649–7666. DOI: 10.1029/98JC02622.

Cieslikiewicz, W. et al. (2017) 'Extreme bottom velocities induced by wind wave and currents in the Gulf of Gdansk', *Ocean Dynamics*, 67, pp. 1461–1480. DOI: 10.1007/s10236-017-1098-4.

Cieślikiewicz, W., Dudkowska, A. and Gic-Grusza, G. (2016) 'Port of Gdansk and Port of Gdynia's exposure to threats resulting from storm extremes', *Journal of Polish Safety and Reliability Association*, 7(1), pp. 29–36. Available at: http://jpsra.am.gdynia.pl/upload/SSARS2016PDF/Vol1/SSARS2016-CieslikiewiczDudkowskaGicGrusza.pdf (Accessed: 25th May 2022).

Cieślikiewicz, W. and Paplińska-Swerpel, B. (2008) 'A 44-year hindcast of wind wave fields over the Baltic Sea', *Coastal Engineering*, 55 (11), pp. 894–905. DOI: 10.1016/j.coastaleng.2008.02.017.

Reda, A. and Paplińska, B. (2002) 'Application of numerical wave model SWAN for wave spectrum transformation', *Oceanological Studies*, 31(1–2), pp. 5–21.

Urbański, J., Grusza, G. and Chlebus, N. (2008) 'A GIS-based WFD oriented typology of shallow micro-tidal soft bottom using wave exposure and turbidity mapping', *Estuarine, Coastal and Shelf Science*, 78(1), pp. 27–37. DOI: 10.1016/j.ecss.2007.11.025.

WAMDI Group (1988) 'The WAM model — A Third Generation Ocean Wave Prediction Model', *Journal of Physical Oceanography*, 18(12), pp. 1775–1810. DOI: 10.1175/1520-0485(1988)018<1775:TWMTGO>2.0.CO;2.

Data Descriptor

Long-term Hindcast Simulation of Currents, Sea Level, Water Temperature and Salinity in the Baltic Sea

Marek Kowalewski[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: marek.kowalewski@ug.edu.pl; ORCID: 0000-0002-8317-6989

Abstract

This dataset contains the results of numerical modelling of currents, sea level, water temperature and salinity over a period of 50 years (1958–2007) in the Baltic Sea. A long-term hindcast simulation was performed using a three-dimensional hydrodynamic model (PM3D) based on the Princeton Ocean Model (POM). The spatial resolution was 3 nautical miles, i.e. about 5.5 km. Currents, water temperature, and salinity were recorded at depths of 0 m, 5 m, 10 m, 15 m, 20 m, 30 m, 40 m, 50 m and 60 m.

Keywords: hydrodynamic model, currents, sea level, water temperature, salinity, Baltic Sea

https://doi.org/10.34808/x55q-sz53_dyr_roz24

Specification table (data records)

Subject area	physical oceanography, hydrodynamic model simulations
Type of data	multidimensional grid data obtained from the numerical model
Spatial resolution	3' latitude and 6' longitude, i.e., approximately 3 nautical miles (5.5 km)
Time resolution	currents: 3 hrs. sea level: 1 hr. water temperature and salinity: 12 hrs.
Data format	NetCDF (Network Common Data Form)
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Conventions	CF-1.8, CF Standard Name Table v72

Data accessibility	These data are available free of charge under the Creative Commons license (CC BY-NC-SA 4.0, https://creativecommons.org/licenses/by-nc/4.0/)

Background

The reconstruction of long-term changes in circulation, sea level, temperature and salinity of the Baltic Sea is possible thanks to simulations using hydrodynamic models. A long-term hindcast simulation was performed using a three-dimensional hydrodynamic model PM3D (Kowalewski and Kowalewska-Kalkowska, 2017), a new version of the M3D model (Kowalewski, 1997). Previously, the M3D model was already used for a 44-years (1958–2001) simulation hindcast at a resolution of 5 nautical miles in the HIPOCAS (Hindcast of dynamic processes of the ocean and coastal areas of Europe) project (Jędrasik, et. al., 2008). The new, improved version of the model (PM3D) was used for a 50-year (1958–2007) hindcast simulation at a resolution of 3 nautical miles (Jędrasik and Kowalewski, 2019).

Methods

The model is based on the Princeton Ocean Model (POM) developed at Princeton University (Blumberg and Mellor, 1987). Like POM, the M3D/PM3D model uses the Mellor-Yamada turbulence scheme and sigma vertical coordinates. The simulation was carried out for the Baltic Sea area (Fig. 24.1) in a rectangular grid of 3' (latitude) and 6' (longitude) resolution. The spatial resolution was 3 nautical miles, i.e. about 5.5 km. An open boundary was located between the Skagerrak and the Kattegat along the parallel connecting Skagen and Göteborg, where the water exchange with the North Sea takes place. A radiation condition based on Somerfield's concept for velocities vertically averaged and normal to the boarder plane was applied. Atmospheric forcing for the period 1958–2007 was performed using the regional atmospheric climate model: REMO (REgional MOdel) which was based on the numerical weather prediction model: EM (Europa-Model) of the German Weather Forecast Service (DWD).

Data records

The eastward and northward components of surface currents and at a depth of 5 m, 10 m, 15 m, 20 m, 30 m, 40 m, 50 m and 60 m were recorded every three hours in the computational grid of the model in unprojected form: LatLong, WGS 84 (EPSG: 4326). Analysis of the Baltic Sea current data spanning five decades showed a higher stability of subsurface current eddies relative to the stability of ephemeral surface currents (Jędrasik and Kowalewski, 2019). The spatial distribution of water temperature and salinity was recorded twice a day, at the same depths and in the same projection as the currents. The sea level spatial distributions were recorded in the model's computational grid with an hourly time interval. Sea level values are expressed in metres relative to the mean level of

the Baltic Sea in the PM3D model. Example results of a long-term simulation are shown in Fig. 24.2.

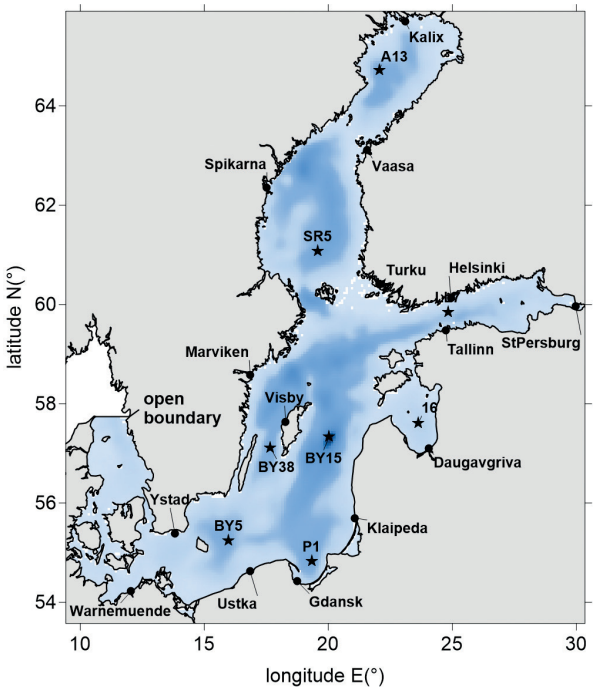


Fig. 24.1. Modelled area of the Baltic Sea and the location of open boundary, monitoring stations (stars) and tide gauges (circles) used for validation of the model results

Data quality

The modelled temperature and salinity distributions were validated for the period 1958–2007 (Jędrasik and Kowalewski, 2019) at selected depths for 8 monitoring stations (Fig. 24.1). The accuracy of the model results with respect to the level was evaluated at 15 sea level gauges, distributed evenly around the Baltic Sea (Fig. 24.1). The validation of the modelling results showed a root mean square error (RMSE) of the water temperature from 0.9 to 1.4°C, and in the case of salinity, the RMSE was from 0.16 to 0.5 psu for various stations in the Baltic Sea. The coefficients of correlation between the simulated and measured temperature ranged from 0.97 to 0.99. The validation of the sea level modelling results showed the RMSE of 7.2 to 18.7 cm and a correlation coefficient of 0.84 to 0.96 for different stations for various stations on the Baltic Sea.

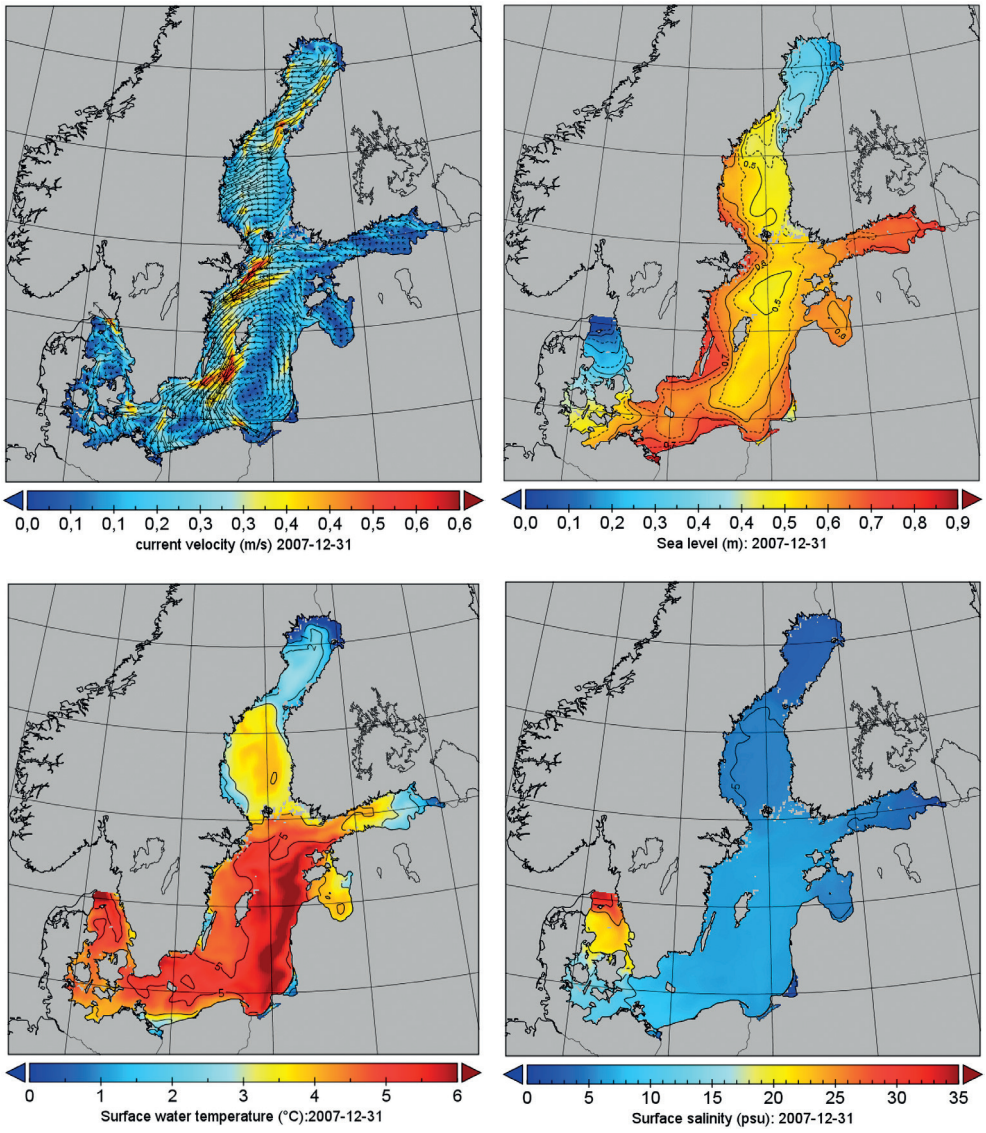


Fig. 24.2. Distribution of currents, sea level, water temperature and salinity in the Baltic Sea for the last day of the long-term simulation

Data availability

The dataset is available and is publicly and freely accessible for any research, educational and other non-commercial purposes under the Creative Commons license (CC BY-NC-SA 4.0, https://creativecommons.org/licenses/by-nc/4.0).

Datasets DOI
10.34808/4tx1-0476
10.34808/j1v9-ak44
10.34808/g7jm-8912

References

Blumberg, A.F. and Mellor, G. L. (1987) 'A description of the three-dimensional coastal ocean circulation model', [in] *Three-Dimensional Coastal Ocean Models*, N. Heaps, American Geophysical Union, pp. 1–16.

Jędrasik, J. et al. (2008) '44 years hindcast of the sea level and circulation in the Baltic Sea', *Coastal Engineering,* 55(11), pp. 849–860. DOI: 10.1016/j.coastaleng.2008.02.026.

Jędrasik, J., and Kowalewski, M. (2019) 'Mean annual and seasonal circulation patterns and long-term variability of currents in the Baltic Sea', *Journal of Marine Syst*ems, 193, pp. 1–26, DOI:10.1016/j.jmarsys.2018.12.011.

Kowalewski, M. (1997) 'A three-dimensional, hydrodynamic model of the Gulf of Gdańsk', *Oceanological Studies*, 26(4), pp. 77–98.

Kowalewski, M., and Kowalewska-Kalkowska, H. (2017) 'Sensitivity of the Baltic Sea level prediction to spatial model resolution', *Journal of Marine Systems,* 173, pp. 101–113, DOI: 10.1016/j.jmarsys.2017.05.001.

Data Descriptor

Areas of Updraft Air Motion in an Idealised Weather Research and Forecasting Model Simulation of Atmospheric Boundary Layer Response to Different Floe Size Distributions

Marta Wenta[1*]

[1] Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: marta.wenta@phdstud.ug.edu.pl; ORCID: 0000-0001-7264-820X

Abstract

Presented dataset is part of a numerical modelling study focusing on the analysis of the influence of sea ice floe size distribution (FSD) on the horizontal and vertical structure of convection in the atmosphere. The total area and spatial arrangement of the updrafts indicates that the FSD affects the total moisture content and the values of area averaged turbulent fluxes in the model domain. In fact, while convective updrafts occur in every model simulation, their intensity differs with varying FSD due to the changing extent and strength of breeze-like circulation. When the floes are tightly packed in the model domain (simulations with 1000 and 5000 floes) the updrafts are numerous but weaker due to weaker breeze-like circulation. In the simulations with smaller floe numbers (Nf =50 and Nf=100), the opposite situation takes place and the updrafts, while covering a smaller area, are stronger and thus are the values of total moisture and area averaged heat content for the model domain, as described in Wenta and Herman (2019).

Keywords: sea ice; floe size distribution; atmospheric boundary layer

https://doi.org/10.34808/x55q-sz53_dyr_roz25

Specification table (data records)

Subject area	Numerical modeling of ocean/sea ice-atmosphere interactions
More specific subject area	Floe size distribution influence on the atmospheric boundary layer

Type of data	Images
How the data was acquired	Generated with MatLab software from the WRF model output
Data format	tiff
Experimental features	
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	CC BY

Background

Sea ice cover in the Arctic is getting thinner and weaker. In consequence, it is more susceptible to breaking by storms, which accelerate the demise of sea ice (Graham et. al., 2019). Despite increased efforts to increase our knowledge about the changes taking place in the polar regions, many aspects of the response of the atmospheric boundary layer (ABL) to sea ice fragmentation are still poorly understood and require new parameterizations to properly reproduce the conditions in the Arctic (Vihma et.al., 2014). So far, most of the studies have focussed on the ABL modelling and observations over leads (e.g. Alam and Curry, 1995; Andreas and Cash, 1999; Marcq and Weiss, 2012), rarely considering multiple, differently oriented cracks between sea ice floes – typical features of marginal ice zone (MIZ). In a study by Wenta and Herman (2018), the problem of the ABL's response to heterogenous sea ice cover is approached differently. In a set of idealized WRF model simulations, they study the atmospheric response to various floes and leads' spatial and size distributions and demonstrate that the domain-averaged values are sensitive not only to sea ice concentration, but also to subgrid-scale spatial arrangement of ice floes and leads. The results clearly show that the spatial distribution and strength of updraft and downdraft regions associated with convective motion within the ABL are related to the underlying features of the ice cover (distribution of floes and leads). In the follow up study of Wenta and Herman (2019), the differences in area-averaged values of turbulent fluxes, total water vapour and liquid water content are explained based on the fact that floe size distribution (FSD) determines the spatial arrangement and intensity of convective cells, which in turn controls the exchange of heat and moisture.

The presented dataset includes the images and calculated total coverage of updraft areas for a number of WRF model simulations analysed in Wenta and Herman (2019). Analysis of this data allowed us to conclude that the convective updrafts occur in every one of our simulations, but their intensity differs with varying FSD due to the changing extent and strength of breeze-like circulation (Fig. 25.1a).

Methods

The Advanced Research Weather Research and Forecasting (ARW WRF) model is used in an "idealized mode". In this approach, the model is initialized with a single

sounding file consisting of the vertical profiles of potential temperature, water vapour mixing ratio, and the two components of wind velocity. The full air pressure and density, as well as other atmospheric variables, are initialized from that input sounding. The model domain is rectangular, with periodic boundaries in both horizontal directions, dimensions 200 × 200 grid points and a horizontal resolution of 100 m, thus covering an area of 20,000 × 20,000 m. The top boundary is set at 2000 m above the surface, close to the top of the inversion layer separating the ABL from the free atmosphere in the initial sounding. The air column is divided into 61 vertical levels with an exponential thickness distribution, from ~2 m at the surface to ~200 m at the top. Each simulation used in this analysis was performed for 14 h, with a time step of 1 s. The results were stored every 10 min.

In the presented dataset, three different sea ice concentration values are considered: c=50%, c=70% and c=90% as in Wenta & Herman (2019). For each ice concentration and two wind speed profiles, the model is launched for a series of simulations with different spatial arrangements of ice floes. Sea ice is represented as round floes with a power-law probability distribution of their radii $P(r) \sim r^{-\beta}$, with an exponent $\beta=1.8$ (values between 1.5 and 2.0 are typically observed in the marginal ice zone). Floe radii range from several tens of metres to over 4 km. The sea ice masks for WRF are obtained by first generating an ensemble of floes with a desired total ice surface area, corresponding to a prescribed c and number of floes Nf; subsequently, these floes are randomly placed within a temporary, very large rectangular domain at a very low ice concentration and without floe overlap, and a DEM model (Herman, 2016) is used to converge the initially loosely packed floe assemblage to obtain a desired ice concentration. A summary of all floe configurations considered is given in Wenta and Herman (2018).

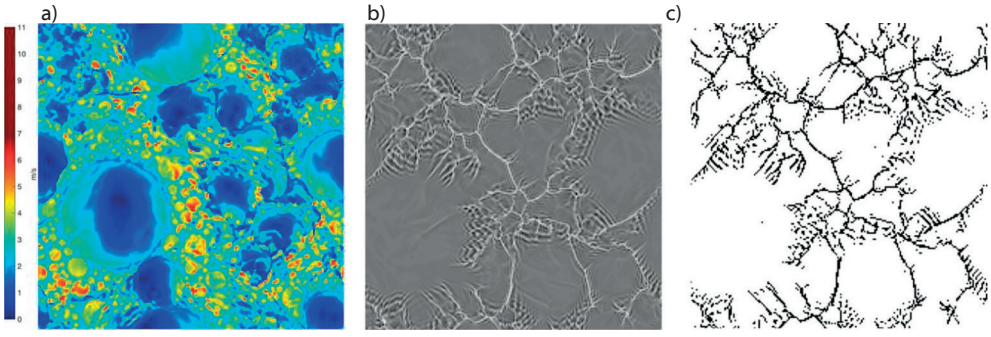


Fig. 25.1. Example of (a) Wind speed (m/s) within model domain for one of the applied FSDs and c=70%. High wind speed is associated with convergence zones related to upward air motions. (b) Grayscale image of upward air motion within the model area. (c) Pixie program output, upward areas marked in black

The study of convective structures within the model domain is based on the analysis of the vertical component of the air motion. In the first step of the analysis, gray-coloured maps of upward air motion areas are generated for every output time step of a given sim-

ulation (Fig.25.1b). In the next step, the program called Pixie (https://gitlab.com/seada-ta-software/pixie) is used to determine the size of updraft areas in every produced image. It is a Python script that applies simple methods of image recognition to distinguish the areas of upward air motion (Fig. 25.1c). It processes every image and counts the pixels for a given colour intensity threshold from 0 (black) to 255 (white). For the presented study, a threshold of 140 has been determined as the most suitable and applied for the whole analysis. Based on the computed number of classified pixels and the number of pixels present in the whole model domain, the total updraft area in km2 and their fractional coverage is computed.

Data Availability

Dataset DOI
10.34808/fwt2-bs21

Dataset License
CC-BY

Acknowledgments
The generation of this dataset was supported by the Polish National Science Center project No. 2015/19/B/ST10/01568 "Discrete-element sea ice modeling – development of theoretical and numerical methods".

References

Alam, A. and Curry, J. (1995) 'Lead–induced atmospheric circulations', *Journal of Geophysical Research*, 100(C3), pp. 4643–4651. DOI: 10.1029/94JC02562.

Andreas, E. and Cash, A. (1999) 'Convective heat transfer over wintertime leads and polynyas', *Journal of Geophysical Research.,* 104(C11), pp. 25721–25734, DOI: 10.1029/1999JC900241.

Graham, R. M. et.al. (2019) 'Winter storms accelerate the demise of sea ice in the Atlantic sector of the Arctic Ocean*, Scientific Reports,* 9, 9222, DOI:10.1038/s41598-019-45574-5.

Herman, A. (2016) 'Discrete-Element bonded-particle Sea Ice model DESIgn, version 1.3a – model description and implementation', *Geoscientific Model Development*, 9(3), pp. 1219–1241, DOI: 10.5194/gmd-9-1219-2016.

Marcq, S. and Weiss, J. (2012) 'Influence of leads widths distribution on turbulent heat transfer', *The Cryosphere*, 6 (1), pp. 143–156, DOI: 10.5194/tc-6-143-2012.

Vihma, T. et al. (2014) 'Advances in understanding and parameterization of small-scale physical processes in the marine Arctic climate system: a review', *Atmospheric Chemistry and Physics*, 14 (17), pp. 9403–9450, DOI: 10.5194/acp-14-9403-2014.

Wenta, M. and Herman, A. (2018) 'The influence of spatial distribution of leads and ice floes on the atmospheric boundary layer over fragmented sea ice', *Annals of Glaciology*, 59(76), pp. 213–230, DOI: 10.1017/aog.2018.15.

Wenta, M. and Herman, A. (2019) 'Area-Averaged Surface Moisture Flux over Fragmented Sea Ice: Floe Size Distribution Effects and the Associated Convection Structure within the Atmospheric Boundary Layer', *Atmosphere*, 10(11), 654, DOI: 10.3390/atmos10110654.

Data Descriptor

High Resolution Sea Ice Floe Size and Shape Data from Knox Coast, East Antarctica

Agnieszka Herman[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego Street, Gdynia, Poland)

[*] Correspondence author: agnieszka.herman@ug.edu.pl; ORCID: 0000-0001-5112-7165

Abstract

This dataset contains floe size distribution data from a very high resolution (pixel size: 0.3 m) optical satellite image of sea ice, acquired on 16 Feb. 2019 off the Knox Coast (East Antarctica). The image shows relatively small ice floes produced by wave-induced breakup of landfast ice between Mill Island and Bowman Island. The ice floes are characterised by a narrow size distribution and angular, polygonal shapes, typical for sea ice broken up by waves.

Keywords: sea ice; sea ice–wave interactions; satellite data; floe size distribution

https://doi.org/10.34808/x55q-sz53_dyr_roz26

Specification table (data records)

Subject area	Physical Oceanography, Polar science, Sea ice, Remote sensing
More specific subject area	Sea ice–wave interactions
How the data was acquired	Processing (image segmentation + object identification) of a very high resolution optical satellite image of sea ice
Data format	binary Matlab (.mat) files
Geographic location	65.55398°S, 101.9066°E
Date of image acquisition	16 Feb. 2019
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The floe size distribution (FSD) is an important characteristic of sea ice, influencing several bulk sea ice properties (e.g., its mechanical strength, lateral melting/freezing rates, etc.), as well as physical processes in the atmospheric/oceanic boundary layers over/under the ice cover. Therefore, an increasing number of research studies is dedicated to the analysis of the observed FSD data in different sea ice, weather and wave conditions, as well as to developing theoretical and numerical methods allowing FSD information to be taken into account in sea ice and climate models.

This dataset contains FSD data extracted from a very high resolution (pixel size 0.3 m) optical satellite image of sea ice (Fig. 26.1; original image acquired by Maxar Technologies, https://www.maxar.com/, and purchased from Overview, https://www.over-view.com/). Contrary to most existing similar datasets, the present one contains information on small ice floes (mean size 18, 13, and 51 m in sectors A, B, C, respectively) which have a narrow size distribution and polygonal, angular shapes, indicating that they were formed by fracturing of continuous ice cover by wave action in the time period shortly preceding the time of acquisition of the analysed image.

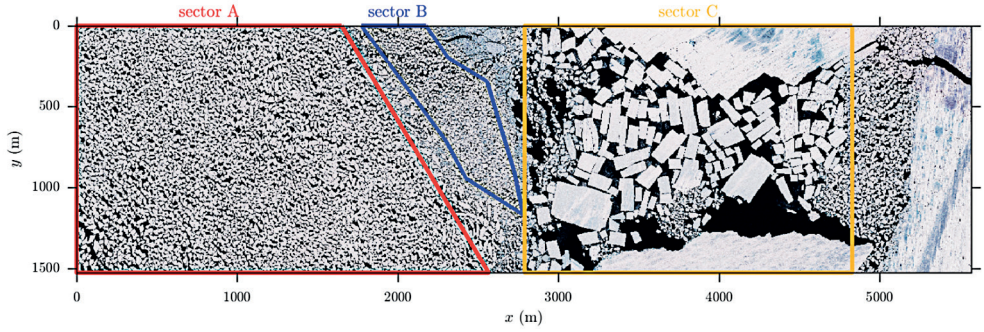


Fig. 26.1. Sea ice image used as the source of floe size and shape data created in this study, with the boundaries of analysis sectors A, B and C

The dataset contains data on the location, size, and selected geometrical properties of ice floes identified in three different parts of the image (sectors A, B, C in Fig. 26.1).

Methods

The fragments of the image from each sector (Fig.26.1) were processed as follows: k-means clustering was applied to the original image in order to identify 3 classes of pixels, corresponding to the open water, dark ice and light ice. Dark ice pixels fully surrounded by light ice were reclassified to light ice, the remaining dark ice to water. Then, a binary image was created (1=light ice, 0=water), and the tools from the Matlab Image Processing Toolbox, combined with ImageJ, were used for object identification (with manual corrections of floes' boundaries where necessary). Finally, selected geometric properties of each floe were computed, including surface area, minimum and maximum

caliper diameter, elongation, rectangularity and minor/major axes of a fitted ellipse (see Fig. 26.2 for an example). Details of the image processing are described in Herman, Wenta and Cheng (2021). The analysis of FSD data in Herman, Wenta and Cheng (2021) is based on methods described in Herman, Evers and Reimer (2018).

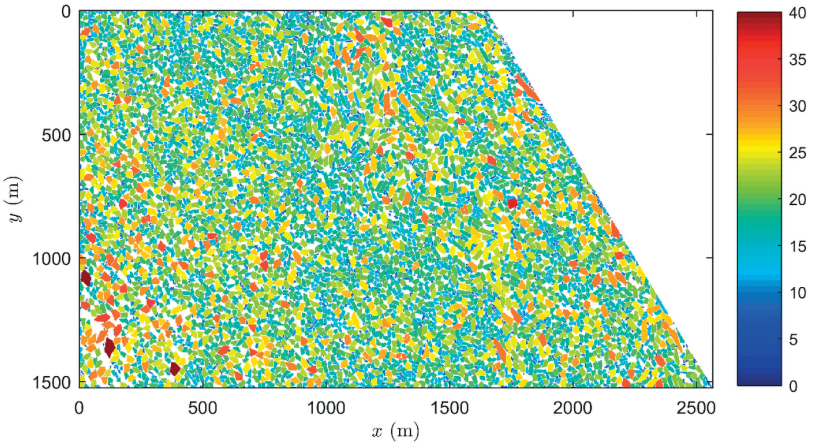


Fig. 26.2. Map of floe width (in m) from sector A, based on floe size data from the present dataset

Data availability

Dataset DOI
10.34808/h3ye-4d45

Dataset License
CC-BY-NC

Acknowledgements
The generation of this dataset was supported by Polish National Science Centre project No. 2018/31/B/ST10/00195 "Observations and modeling of sea ice interactions with the atmospheric and oceanic boundary layers".

References

Herman, A., Evers, K.-U., and Reimer, N. (2018) 'Floe-size distributions in laboratory ice broken by waves', *The Cryosphere*, 12(2), pp. 685–699, DOI: 10.5194/tc-12-685-2018.

Herman, A., Wenta, M., and Cheng, S. (2021) 'Sizes and shapes of sea ice floes broken by waves – a case study from the East Antarctic coast', *Frontiers in Earth Science*, 9, DOI: 10.3389/feart.2021.655977.

Data Descriptor

Long-Term GNSS Tropospheric Parameters for the Tropics (2001-2018) Derived from Selected IGS Stations

Zofia Baldysz[1*], Grzegorz Nykiel[1,2], Mariusz Figurski[1,2]

[1] Institute of Meteorology and Water Management – National Research Institute (61 Podleśna Street, Warsaw, Poland)

[2] Faculty of Civil and Environmental Engineering, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: zofia.baldysz@pg.edu.pl; ORCID: 0000-0002-8398-6703

Abstract

This paper describes dataset "Tropospheric parameters derived from selected IGS stations in the tropics for the years 2001-2018" contains GNSS-derived zenith tropospheric delay (ZTD), a posteriori corrected zenith wet delay (ZWD), and precipitable water vapour (PWV) time series. These troposphere-related data were estimated for the Jan 2001 – Dec 2018 period for 43 International GNSS Service (IGS) stations located across the global tropics. As one coherent strategy for the processing of the GNSS observations was adopted, the dataset is a robust source of long-term, homogeneous tropospheric time series, which can be used in meteorological and climate-related studies. It enables the examination of moisture patterns on numerous time scales, including seasonal and interannual variability.

Keywords: GNSS, PWV, water vapour, tropics, climate monitoring

https://doi.org/10.34808/x55q-sz53_dyr_roz27

Specification table (data records)

Subject area	Meteorology, Climate, Remote Sensing
More specific subject area	Atmospheric water vapour variability
Type of data	Text

How the data was acquired	The data were created based on GNSS observations (Jan 2001 – Dec 2018) from 43 IGS stations located across the global tropics and the ERA5 model as a source of meteorological parameters
Data format	Formatted ASCII files
Experimental factors	GNSS observations were processed using Bernese GNSS Software ver. 5.2
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The global navigation satellite systems (GNSS) signal passing through a neutral atmosphere is delayed due to atmospheric refraction. The size of this delay significantly affects coordinate estimation and thus has to be estimated during advanced processing of the GNSS observations. It was agreed to express its value towards the zenith direction (zenith tropospheric/total delay, ZTD) although it is estimated based on the slant direction between a receiver and each of the observed satellites. In general, it consists of delays caused by the hydrostatic (zenith hydrostatic delay, ZHD) and wet (zenith wet delay, ZWD) components of the atmosphere.

Due to the fact that determination of precise and accurate coordinates requires accurately estimated ZTDs, GNSS-derived tropospheric parameters have become a valuable source of meteorological and climate data, especially considering the fact that the GNSS permanent network has been operating since 1996. Although the ZTD alone can be used for climate and meteorological analysis (Baldysz et al., 2015), it can also be used for ZWD extraction, which is strongly related to the water vapour content in the atmosphere. In addition, using selected meteorological parameters, it can be converted into the precipitable water vapour (PWV) parameter (Bevis et al., 1992), which directly refers to the total water vapour content in the atmosphere. This is, in turn, the most important natural greenhouse gas. It is characterised by the single highest positive feedback on the surface temperature (Hansen et al., 1984) and dominates the effect of the Earth's surface temperature increase (Kiehl and Trenberth, 1997). It is also a key factor in the formulation of weather conditions, including severe weather events. Thus, monitoring of the water vapour content in the atmosphere as it changes over time, on numerous time scales from seasonal to interannual, is an important task in understanding its role in a changing climate.

Despite the fact, that GNSS is not a direct method of measuring PWV, application of an appropriate processing strategy enables utilisation of GNSS PWV time series as

a valuable source for long-term analysis. Its accuracy is similar to the PWV derived from radio sounding measurements (Baldysz et al., 2018).

The special role which the tropics plays in formulating global climate results from the fact that they are a major centre of atmospheric convection. Through atmospheric teleconnections, tropical weather patterns also have a significant impact on mid- and high-latitude weather conditions. As a consequence, analysis of the variability of GNSS-derived tropospheric parameters over the tropics, such as ZTD, ZWD and PWV, is an important task in climate monitoring.

Methods

ZTD is usually estimated during advanced processing of GNSS observations according to the following formula:

$$ZTD = ZHD + ZWD = mf_h(el) \cdot SHD + mf_w(el) \cdot SWD + mf_g(G_N \cos\alpha + G_E \sin\alpha) \quad (1)$$

where: and are the hydrostatic and wet slant delays in the direction of a satellite, and are mapping functions, dependent on the elevation angle () used to project them on the zenith direction, and denote azimuth () dependent gradients defined to the north and east direction, while is the gradient mapping function. Although during GNSS observation processing, ZTD estimation relies on the same steps, there are number of models, numerical approaches and calculation assumptions that affect its final reliability. Thus, they can limit the credibility of conclusions drawn based on them. In this dataset, the adopted GNSS processing strategy was verified through long-term comparison of the GNSS PWV and radio-sounding PWV, thus ensuring its high accuracy (Baldysz and Nykiel, 2019). Detailed information about the adopted processing strategy is given in Tab. 27.1.

Tab. 27.1

GNSS processing strategy used in this study.

GNSS processing parameter	value/name
software	Bernese 5.2 GNSS Software (Dach et al., 2015)
method	Precise point positioning (PPP)
data	30-second GPS
a priori ZHD	Vienna Mapping Function 1 (Boehm et al., 2006)
a posteriori ZHD	Saastamoinen model (1972)
troposphere mapping function	Vienna Mapping Function 1

gradients mapping function	Chen and Herring (1997)
cut-off angle	5°
ephemerides (clocks and orbits)	CODE R2 (Steigenberger et al., 2014)
antenna models	IGS08
reference frame	IGb08
ZWD estimation	Saastamoinen model using meteorological parameters from the ERA5 model
ZWD to PWV conversion	Water vapour weighted mean temperature based on the Bevis formula (Bevis et al., 1992) and the temperature from the ERA5 model

Generally, in the advanced processing of GNSS observations, the ZHD value is a priori and adopted from the numerical weather (or empirical) model, while the ZWD is estimated as an additional, unknown parameter. The sum of these two values is expressed as the ZTD. Nevertheless, the ZWD obtained in such a way contains both the wet delay and correction to the inaccurately modelled ZHD. Thus, in GNSS meteorology, the ZHD is also estimated a posteriori (based on observational or re-analysed meteorological data) and subtracted from the ZTD. For this purpose, the Saastamoinen (1972) model dependent on the total air pressure at the antenna height (, in hPa), station ellipsoidal latitude () and altitude (, in metres) is commonly used:

$$ZHD = \frac{0.0022767 \cdot P}{1 - 0.00266 \cdot \cos(2\varphi) - 2.8 \cdot 10^{-7} \cdot h} \tag{2}$$

Obtained in such a way, the ZWD is strictly related to the water vapour content in the atmosphere and consequently can be converted into the PWV, using the dependency:

$$PWV = \Pi(T_m) \cdot ZWD \tag{3}$$

where: is a dimensionless quantity dependent on the water vapour weighted mean temperature which can be retrieved via the following formula:

$$\Pi^{-1} = 10^{-8} \cdot \rho \cdot R_v \cdot (K_3/T_m + K_2') \tag{4}$$

where: is the specific gas constant of water vapour and equals 461.5 J/kg, is the density of water, while and are constant parameters related to the air refractivity (used values proposed by Ruger (2002)). In this dataset, was calculated using the Bevis formula (Bevis et al., 1992) and the surface temperature was derived from the ERA5 model (Hersbach et al., 2020).

The approach presented above was used to deliver hourly ZTD, ZWD (using a posteriori ZHD) and PWV for the 43 stations belonging to the International GNSS Service

(IGS) network (Johnston et al., 2017). The calculations covered the period from January 2001 to the end of December 2018. Fig. 27.1 shows the locations of the used stations.

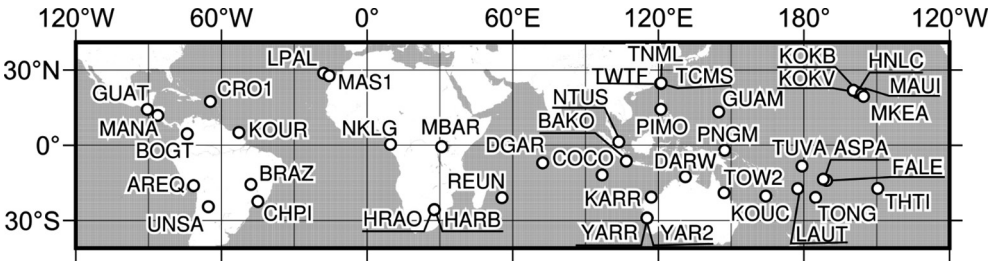


Fig. 27.1. Location of the IGS stations for which GNSS tropospheric parameters were estimated

Data quality and availability

GNSS ZTD, ZWD and PWV daily solutions were calculated as an average from hourly data. Daily solutions were adopted as reliable only when at least 16 hourly values were available (if less, the solution was removed). During the screening process, daily solutions exceeding the 3σ criterion were also removed. Note that the GNSS ZTD, ZWD and PWV daily solutions obtained in such way are not continuous. They may contains gaps related to the lack of observational data, its poor quality, etc. Also, we did not remove shifts that may have occurred as a result of a change in equipment at the station.

Dataset DOI
10.34808/9s0h-k459

Dataset License
CC-BY-NC-SA

Acknowledgements
Calculations were carried out at the Academic Computer Centre in Gdańsk

References

Baldysz, Z. et al. (2015) 'Investigation of the 16-year and 18-year ZTD Time Series Derived from GPS Data Processing', *Acta Geophysica*, 63, pp. 1103–1125, DOI: 10.1515/acgeo-2015-0033.

Baldysz, Z. et al. (2018) 'Assessment of the Impact of GNSS Processing Strategies on the Long-Term Parameters of 20 Years IWV Time Series', *Remote Sensing*, 10(4), 496, DOI: 10.3390/rs10040496.

Baldysz, Z. and Nykiel, G. (2019) 'Improved Empirical Coefficients for Estimating Water Vapor Weighted Mean Temperature over Europe for GNSS Application's, *Remote Sensing*, 11(17), 1995. DOI: 10.3390/rs11171995.

Bevis, M. et al. (1994) 'GPS meteorology: Mapping zenith wet delays onto precipitable water', *Journal of Applied Meteorology and Climatology,* 33(3), pp. 379–386, DOI: 10.1175/1520-0450(1994)033<0379:GMMZWD>2.0.CO;2.

Boehm, J., Werl, B. and Schuh, H. (2006) 'Troposphere mapping functions for GPS and very long baseline interferometry from European centre for medium-range weather forecasts operational analysis data', *Journal of Geophysical Research,* 111 (B2), B02406, DOI: 10.1029/2005JB003629.

Chen, G. and Herring, A. (1997) 'Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data', *Journal of Geophysical Research,* 102 (B9), pp. 20489–20502. DOI:10.1029/97JB01739

Dach, R. et al. (2015) 'Bernese GNSS software version 5.2'. In: *User Manual. Astronomical Institute*, University of Bern.

Hansen, J. et al. (1984) 'Climate sensitivity: Analysis of feedback mechanisms', *Climate Processes and Climate Sensitivity*, 29, pp. 130–163. DOI: 10.1029/GM029p0130.

Hersbach, H. et al. (2020) 'The ERA5 global reanalysis', *Quaterly Journal of the Royal Meteorological Society*, 146 (730), pp. 1999– 2049. DOI: 10.1002/qj.3803.

Johnston, G., Riddell, A. and Hausler, G. (2017) 'The International GNSS Service'. Teunissen, Peter J.G., & Montenbruck, O. (Eds.), *Springer Handbook of Global Navigation Satellite Systems* (1st ed., pp. 967–982). Cham, Switzerland: Springer International Publishing. DOI: 10.1007/978-3-319-42928-1.

Kiehl, J.T. and Trenberth, K.E. (1997) 'Earth's Annual Global Mean Energy Budget', *Bulletin of the American Meteorological Society*, 78(2), pp. 197-208, DOI:10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2.

Rueger, J. (2002) 'Refractive Index Formulae for Radio Waves'. In: *Proceedings of the FIG XXII International Congress.* Available at: https://www.fig.net/resources/proceedings/fig_proceedings/fig_2002/Js28/JS28_rueger.pdf (Accessed: 25th May 2022).

Saastamoinen, J. (1972) 'Atmospheric correction for the troposphere and stratosphere in ranging satellites'. In: *The Use of Artificial Satellites for Geodesy*, Geophysical Monography, 15. American Geophysical Union, pp. 247–251.

Steigenberger, P. et al. (2014) '*CODE repro2 product series for the IGS*', Astronomical Institute, University of Bern, DOI:10.7892/boris.75680.

Data Descriptor

Simulations of the Derecho Event in Poland of 11th August 2017 Using WRF Model

Mariusz Figurski[1, 2*], Grzegorz Nykiel[1, 2]

[1] Faculty of Civil and Environmental Engineering, Gdańsk University of Technology, Poland (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[2] Institute of Meteorology and Water Management - National Research Institute (61 Podleśna, Warsaw, Poland)

[*] Correspondence author: mariusz.figurski@pg.edu.pl; ORCID: 0000-0001-9602-5007

Abstract

This series contains datasets related to the forecasting of a severe weather event, a derecho, in Poland on 11 August 2017. The simulations were conducted using the Weather Research and Forecasting (WRF) model version 4.2.1 with different initial and boundary conditions of the pressure and model levels derived from 5 global models: Global Forecast System (GFS), Global Data Assimilation System (GDAS), European Centre for Medium-Range Weather Forecasts (ECMWF), and ERA5. Each simulation, shared as a separate dataset, was performed for two starting hours: at 00:00 and 12:00 UTC. The datasets contain about 280 meteorological parameters stored as 2D or 3D fields with high-spatial (2.5 km and 0.5 km domains) and temporal (10 minutes) resolutions. The three-dimensional fields are calculated at 50 levels up to 50 hPa. All data are stored in easily accessible NetCDF files.

Keywords: WRF; derecho; bow echo; severe weather forecast; GFS; GDAS; ECMWF, ERA5

https://doi.org/10.34808/x55q-sz53_dyr_roz28

Specification table (data records)

Subject area	Meteorology
More specific subject area	Severe weather event forecast
Type of data	Meteorological parameters from WRF model

How the data was acquired	Using WRF 4.2.1 model and five different global models as initial and boundary conditions (GFS, GDAS, ECMWF, ERA5 on pressure levels, ERA5 on model levels)
Data format	NetCDF files
Experimental factors	All factors and parametrisations were the same for all simulations and described in the paper in detail
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Severe weather events are occurring increasingly often, which is connected with climate change. They pose a direct threat to human health and life and cause numerous and costly material losses. Therefore, it is extremely important to forecast them accurately, both in terms of location and time of occurrence. Numerical weather prediction models (NWP) have been used for this purpose for several decades. With the development of algorithms and the increase in the processing power of computers, it is possible to obtain fast and reliable forecasts for up to tens of hours. Additionally, it is possible to obtain forecasts with time steps of even several minutes and a spatial resolution of less than 1 kilometre. It concerns especially mesoscale models designed for the forecasting of local phenomena. The accuracy of NWP forecasts depends on many factors: the parameterisation of the physical phenomena and model dynamics used, the accuracy of the geographical data, and the initial and boundary conditions, etc. These initial and boundary conditions are usually derived from global models and can vary significantly especially during local severe weather events. Thus, their utilisation may lead to different NWP results using the same model parameterisations.

These series of datasets are concerned with forecasting a severe weather event which is the derecho which occurred in Poland on 11th August 2017, and caused very significant tree stand and material damage. The datasets include simulations using a mesoscale model and various boundary and initial conditions.

Methods

The simulations were conducted using the Weather Research and Forecast (WRF) model which is designed for atmospheric research and weather forecasting (Skamarock et al., 2005). This implementation of the model has been adapted to work on the Tryton high performance computer located at Academic Computer Centre in Gdańsk. The

simulations were performed independently for five different sets of boundary and initial conditions, which were the following global models: Global Forecast System (GFS), Global Data Assimilation System (GDAS), European Centre for Medium-Range Weather Forecasts (ECMWF), and two versions of the ERA5 model – on pressure and model levels (here and after called ERA5P and ERA5M respectively). Three, one-way nested domains were used. This first domain covers most of Europe and has a spatial resolution equalling 12.5 km for the GFS, GDAS, and ERA5 models, and 7.5 km for forecasts using the ECMWF model. This difference is due to the resolution of the global models: 0.25° for GFS, GDAS, ERA5, and 0.125° for ECMWF. The spatial resolution of the second and third domains are equal for all simulations and are 2.5 km and 0.5 km, respectively. The first of this domain covers the area of Poland while the second one, the area where most of the damage occurred. Details of the locations of domains 2 and 3 are shown in Fig. 28.1. Also, the maximum reflectivity derived from WRF+GDAS at 20:00 UTC is presented for illustrative purposes. In the vertical, the WRF model was configured with an upper limit of 50 hPa and 50 levels using a vertical coordinate dependent on terrain height and hydrostatic pressure. All simulations were started at two epochs: 00:00 and 12:00 UTC, 11th August 2017. The length of the forecast was 24- and 12-hours, respectively. The results were saved with two different intervals: 1-hour for domain 1, and 10-minutes for domains 2 and 3. All of the basic parameters of the simulations are presented in Tab.28.1.

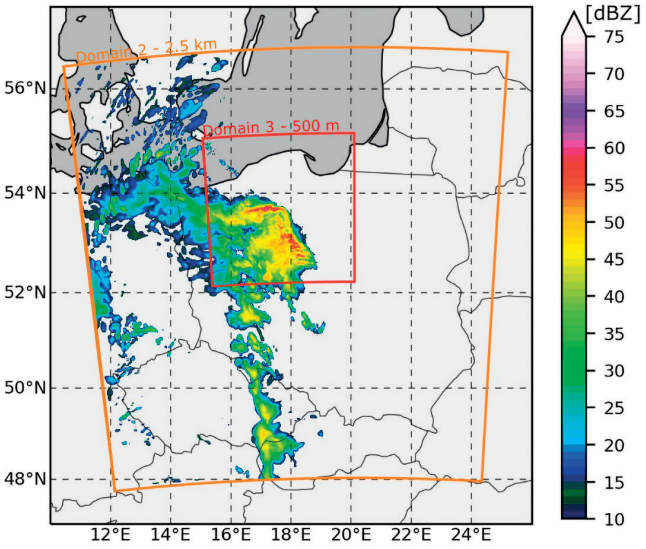


Fig. 28.1. Geographical localisation of two of the three WRF domains on the background on maximum reflectivity derived from the WRF+GDAS simulation at 20:00 UTC, 11th August 2017

Tab. 28.1

Details of the simulations performed using the WRF model.

Parameter	Name/Value				
Model name	WRF model version 4.2.1				
Global model (initial and boundary conditions)	GFS	GDAS	ERA5P	ERA5M	ECMWF
Global model resolution	0.25°	0.25°	0.25°	0.25°	0.125°
Number of global model levels	32	32	38	137	26
Global model vertical level types	pressure	pressure	pressure	model level	pressure
Number of domains	3				
Domain size (in nodes)	#1: 268 × 229 #2: 366 × 391 #3: 646 × 656				#1: 178 × 183 #2: 352 × 367 #3: 646 × 656
Domain resolutions	#1: 12.5 × 12.5 km #2: 2.5 × 2.5 km #3: 0.5 × 0.5 km				#1: 7.5 × 7.5 km #2: 2.5 × 2.5 km #3: 0.5 × 0.5km
Time resolution	#1: 1 hour #2 & #3: 10 min				
Number of levels	50 levels up to 50 hPa				
Time of forecast start	Two forecasts: at 00 and 12 UTC, 11th August 2017				
Time range of forecast	24 hours for forecast started at 00:00 UTC 12 hours for forecast started at 12:00 UTC				

It is worth noting that the parametrisation of all of the physical processes and model dynamics was the same for all simulations. A single-moment microphysics scheme with 6 hydrometeor classes (WSM6) (Zaidi and Gisen, 2018) was used, which is most suitable for high-resolution simulations (Hong and Lim, 2006). In the first domain, the convective processes were modelled using Grell-Freitas parameterisation (Grell and Freitas, 2014), while for second and third domains, explicit wet process physics were used. Moreover, we applied parameterisations of shortwave and longwave radiation according to the RRTMG radiation propagation scheme, which is a new version of RRTM (Iacono et al., 2008). To model the boundary layer processes, the Mellor Yamada Nakanishi Niino (MYNN) turbulence scheme with closure 2.5 was used (Nakanishi and Niino, 2009). The near-surface layer was parameterised according to the MYNN scheme (Nakanishi and Niino, 2006). Land topography, land use and soil type datasets were applied to the model at the WRF preprocessing stage. For the first and second domains, the standard

data contained in the WRF model geographic database (LULC) of IGBP MODIS and USGS GMTED2010 (30 arc seconds resolution) were used. These data are insufficient for high-resolution simulations (De Meij and Vinuesa, 2014, Jiménez-Esteve et al., 2017). Therefore, in our datasets, for the third domain (0.5 km), we prepared new geographic data with a higher resolution. Land use data with a 100-metre spatial resolution were prepared based on the CORINE Land Cover (CLC) project 2018. The terrain topography, with a 30-metre resolution, was prepared on the basis of the Shuttle Radar Topography Mission (SRTM).

Data quality and availability

All simulations are stored in separate, hourly NetCDF files. They can be easily accessed by any GIS software or script language. A NetCDF file contains data and metadata in the same file. This metadata contains information about the parameter name, dimensions and units. The datasets were verified in terms of the correctness of the data contained. No gross errors were found.

Datasets DOI:
10.34808/ceh3-7z70
10.34808/3s00-kn10
10.34808/ava3-yd08
10.34808/fkmj-dh09
10.34808/kxnc-rc32

Datasets License:
CC-BY-NC-SA

Acknowledgements
Calculations were carried out at the Academic Computer Centre in Gdańsk

References

De Meij, A. and Vinuesa, J.F. (2014) 'Impact of SRTM and Corine Land Cover data on meteorological parameters using WRF', *Atmospheric Research,* 143, pp. 351–370. DOI: 10.1016/j.atmosres.2014.03.004.

Grell, G. A. and Freitas, S. R. (2014) 'A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling', *Atmospheric Chemistry and Physics*, 14(10), pp. 5233–5250, DOI: 10.5194/acp-14-5233-2014.

Hong, S.Y. and Lim, J.–O.J. (2006) 'The WRF single–moment 6–class microphysics scheme (WSM6)', *Journal of the Korean Meteorological Society*, 42 (2), pp. 129–151. Available at: https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00937254&language=ko_KR (Accessed: 25th May 2022).

Iacono, M.J. et al. (2008) 'Radiative forcing by long–lived greenhouse gases: Calculations with the AER radiative transfer models', *Journal of Geophysical Research*, 113, D13103. DOI: 10.1029/2008JD009944.

Jiménez-Esteve, E.B. et al. (2017) 'Land use and topography influence in complex terrain area: A high resolution mesoscale modelling study over the Eastern Pyrenees using the WRF model'. *Atmospheric Research,* 202, pp. 49–62, DOI: 10.1016/j.atmosres.2017.11.012.

Nakanishi, M. and Niino, H. (2006) 'An improved Mellor–Yamada level 3 model: its numerical stability and application to a regional prediction of advecting fog', *Boundary- Layer Meteorology,* 119, pp. 397–407. DOI: 10.1007/s10546-005-9030-8.

Nakanishi, M. and Niino, H. (2009) 'Development of an improved turbulence closure model for the atmospheric boundary layer', *Journal of the Meteorological Society of Japan*, 87(5), pp. 895–912, DOI: 10.2151/jmsj.87.895.

Skamarock, W.C. et al. (2005) 'A description of the advanced research WRF version 2', *National Center For Atmospheric Research*, Boulder, Colorado, USA, DOI: 10.5065/D6DZ069T.

Zaidi, S.M. and Gisen, J.I.A. (2018) 'Evaluation of Weather Research and Forecasting (WRF) Microphysics single moment class-3 and class-6 in Precipitation Forecast', in: *MATEC Web of Conferences*, 03007, DOI: 10.1051/matecconf/201815003007.

Data Descriptor

Split-Beam Echosounder Data from Puck Bay

Patryk Damian Pezacki[1*]

[1] Faculty of Oceanography and Geography, Institute of Oceanography, University of Gdańsk (46 Marszałka Józefa Piłsudskiego, Gdynia, Poland)

[*] Correspondence author: patryk.pezacki@ug.edu.pl; ORCID: 0000-0003-3904-0176

Abstract

The acoustic data was collected in 2018–2019 in the Bay of Puck in the seasons: autumn, winter, spring. The data was collected during the day and night using three split-beam echosounders with frequencies of 38 kHz, 120 kHz and 333 kHz at a designated study area not far from the city of Hel, while the ship was sailing. To ensure data quality, the echosounders were calibrated and passive noise was measured.

Keywords: echosonda; puck bay; southern Baltic

https://doi.org/10.34808/x55q-sz53_dyr_roz29

Specification table (data records)

Subject area	Hydroacoustic study
More specific subject area	Hydroacoustic biological research in the Bay of Puck. Study of pelagic water
Type of data	Raw acoustic data
How the data was acquired	Using echosounder mounted on ship, while the ship was sailing and also was anchored
Data format	.RAW
Experimental factors	-
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	CC-BY-NC

Background

The data were collected as part of a doctoral thesis with the goal of sonar imaging pelagic concentrations in the Gulf of Puck .

The implementation of this goal has a significant practical dimension, allowing for more precise monitoring of the ecosystem in the Bay of Puck (its external part). This will also help to understand the behaviour of organisms residing in the most isolated part of the Gulf of Gdansk in their interaction with the environment. In order to understand the functioning of such a specific area of sea as the Bay of Puck, a representative amount of data is needed covering as many days as possible for each season of the year, and taking into account all times of the day.

Methods

The research work was carried out from the R/V Oceanograf. It is a steel construction catamaran designed for interdisciplinary maritime research. The main advantage of the ship, having a significant impact on the quality of the collected data, is the DP dynamic positioning system. Thanks to this system, the ship – in combination with the precise GPS RTK positioning system – can move with high accuracy along the given measurement profiles, and also maintain a fixed position at a research site. These conditions can be met with the sea state even up to 4 on the Douglas scale, with a favourable wind direction.

Description of hydroacoustic equipment

The basic tools used in the hydroacoustic research were 3 split-beam SIMRAD EK80 echosounders by Kongsberg Maritime. These are broadband sonars with central frequencies of 38kHz, 120kHz and 333kHz. The transducers for these sonars are divided into independent working parts: three for the 38 kHz sonar and four for the other two transducers. Such a structure allows for the possibility of determining the position of scattering targets in the sonar beam and the direction of their movement (e.g. for marine organisms). This is possible because, knowing the differences in the phases of the wave scattered from the object for the individual parts of the transducer, it is possible to determine the position of the scattering target in relation to the transducer beam.

The sonars were used in "continuous wave" (CW) mode, working in one non-modulated central frequency.

Research conducted with the use of three sonar sounders with acoustic frequencies of 38 kHz, 120 kHz and 333 kHz if effective for detecting organisms, respectively greater than 6 mm, 1.9 mm and 0.7 mm. These dimensions were calculated based on the scattering target detection condition ka > 1 (Simmonds and MacLenennan, 2005). Using sonars with different frequencies also allows the frequency response of scattering objects, such as fish assemblies, to be determined. This characteristic may be useful in the classification of these objects (Simmonds and MacLenennan, 2005).

Data records

Study area

The external Puck Bay, which is part of the Gdańsk Bay in the southern part of the Baltic Sea, was selected as the research area. The research area, with dimensions of 3 km by 2 km, the central part of which is in the position described by the coordinates: 54.60° N 18.74 ° E, was located in the Gulf of Puta not far from the city of Hel. The depth in the entire area varies from 52.0 m in the eastern part to 47.0 m in the western part.

Studies of the spatial distribution of organisms during day and night hours.

Data collected during the day and night were used to determine the distribution of organisms clusters in the water. Data was collected on the go. The measuring speed varied between 4.0 and 6.6 knots.

Echosounder calibration

The main purpose of sonar calibration is to adapt it to new hydrological conditions, changing e.g. due to the change of the season or the change of the research reservoir. The basic principles of calibration described by Foote et al. (1987), are still valid today, although some practical aspects have changed. Calibrations should be performed with standard calibration balls with known target strength. The choice of the calibration ball depends on the frequency at which the sonar is working. When calibrating the sonar, the ball should be placed directly under the transducer at the appropriate depth (in the far zone of the transducer). It is important that this depth remains as constant as possible throughout the calibration process. The ball should be suspended with the thinnest possible lines so that they do not affect the echo strength the ball target. The ball should be suspended from three points located on both sides of the ship in such a way that it can be freely moved in a plane parallel to the bottom. Thanks to this, during the calibration process, it will be possible to collect data from the entire horizontal cross-section of the transducer beam by properly moving the calibration ball under it. (Demer et al., 2015)

The software that we use to control echo sounders also has a calibration module. Using the data collected during the calibration, the software updates the system of gain settings and calculates the settings with a given directional characteristic of the transducer, which affects the sonar directionality functions and the precision of measuring the backscatter volumetric force of the returning acoustic signal.

Due to the fact that the weather was not conducive to calibrating the echosounders in the study area during the cruises and due to the limited cruise time, we calibrated them in the port at the beginning of the cruise, before starting the research. We decided to calibrate in the port because it was located relatively close to the study area and the hydrological conditions at the calibration site did not differ significantly from the conditions in the study area.

Noise measurements

In addition to collecting data in active mode, i.e. the standard operating mode of the sonar (transmitting an acoustic impulse and recording a returning echo), in accordance with ICES recommendations (Simmonds and MacLennan, 2005), data was also collected during the measurement sessions in the so-called passive mode. In this mode, the transducers only recorded the signals that reached them on their own, without transmitting acoustic pulses. The surrounding noise was recorded. In passive mode, data was collected over a period of approximately 10 minutes. Then, during the subsequent data analysis, this noise was eliminated from the data collected with the sonar in active mode.

Data quality and availability

In order to ensure the highest quality of data, the sonars were calibrated, and passive noise was recorded, which can be subtracted from the raw data.

Datasets DOI:
10.34808/b7sh-rt35
10.34808/ffm6-zk97

References

Demer, D.A. et al., (2015) 'Calibration of acoustic instruments', *ICES Cooperative Research Report*, 326, pp. 133. DOI: 10.25607/OBP-185.

Foote, K.G. (1987) 'Fish target strengths for use in echo integrator surveys', *The Journal of the Acoustical Society of America*, 82 (3), pp. 981-987. DOI: 10.1121/1.395298.

Simmonds, E.J. and MacLennan, D.N., (2005) '*Fisheries Acoustics: Theory and Practice*', Blackwell Publishing, Oxford, pp. 437.

Data Descriptor

Vehicle Detection and Speed Estimation Using Millimetre Wave Radar

Piotr Odya[1*]

[1] Department of Multimedia Systems, Faculty of Electronics, Telecommunications and Informatics; Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: pioodya@pg.edu.pl; ORCID: 0000-0003-0288-6178

Abstract

The dataset titled Data from 76- to 81-GHz mmWave Sensor located at S7 road contains data recorded employing an IWR1642 mmWave sensor from Texas Instruments. The data comes from two sessions lasting 24h each. The dataset provides the possibility to perform analyses related to car traffic intensity on one of the carriageways of the motorway heading to the Gdańsk metropolitan area. Based on the gathered data, it is possible to calculate the number of vehicles in particular time intervals and to estimate their speed.

Keywords: radar sensor; road; vehicle detection; vehicle speed

https://doi.org/10.34808/x55q-sz53_dyr_roz30

Specification table (data records)

Subject area	Traffic Study, Electronics, Sensors
More specific subject area	Vehicle Speed Estimation, Vehicle Detection
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology using an IWR1642 mmWave sensor from Texas Instruments installed on the S7 motorway.
Data format	The tables are in .csv format
Experimental factors	The data contained in the dataset were decoded from binary format recordings
Experimental features	Detection of vehicles and estimation of their speed in real-life conditions

Data source location	MOST Wiedzy Open Research Catalog Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes.

Background

The millimetre wave radar was one of the components of the intelligent road sign prototype developed as a part of the project "INZNAK: Intelligent Road Signs with V2X Interface for Adaptive Traffic Controlling". The radar sensor was installed in a measuring station, which also contained the following sensors:

- temperature, pressure and humidity,
- light intensity,
- precipitation,
- interior temperature,
- Bluetooth device scanner,
- 22GHz Doppler radar,
- video camera,
- sound intensity probe,
- accelerometer,
- lidar.

A Raspberry Pi microcomputer was used to collect information from the sensors, a Kontron microcomputer supervised the operation of the components, and connectivity was provided by using a router with an LTE modem.

Tests on the S7 road were aimed at verifying the operation of the sensors in real measurement conditions. The measuring station was mounted on a gantry pillar (Fig. 30.1) at a height of approx. 250 cm from the ground (the sensor was installed 40 cm from the bottom edge of the station). The distance from the edge of the right lane was about 420 cm.

Fig. 30.1. Measuring station (marked with a red rectangle) mounted on the gantry pillar

The tests proved that the Data from 76- to 81-GHz mmWave Sensor located at S7 road dataset can be applied for traffic analysis. Using relatively simple algorithms (based e.g. on distance analysis) should enable the process of automatic detection of vehicles and their speed.

Example data overlaid on an image from a camera mounted under the measurement station shows the functionality of the dataset: https://youtu.be/bKmTF2NT5Oc.

Methods

The evaluation board from Texas Instruments, based on an IWR1642 (rev. B) single-chip 76-GHz to 81-GHz mmWave sensor (Fig. 30.2), was used to collect the data (IWR1642 EVM (IWR1642BOOST) Single-Chip mmWave Sensing Solution User's Guide, 2018). The board is equipped with antennas and communicates with a computer using a UART<->USB interface.

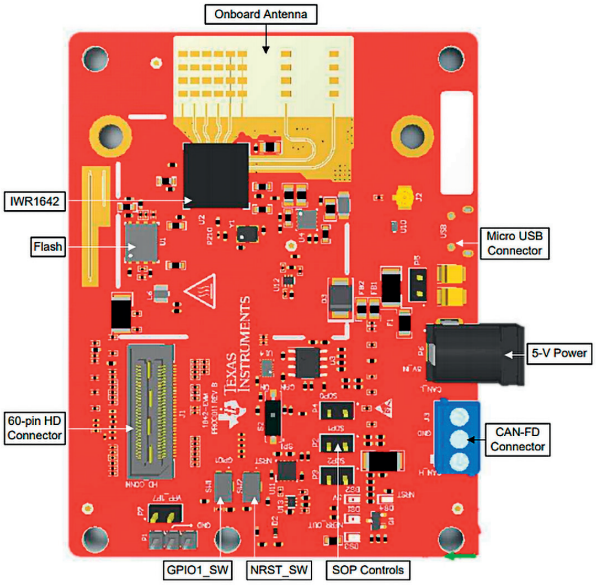


Fig. 30.2. IWR1642 evaluation board (IWR1642 EVM (IWR1642BOOST) Single-Chip mmWave Sensing Solution User's Guide, 2018)

Based on the supplied software, it is possible to adjust the parameters of the device (e.g. detection range, maximum object speed) to specific needs with the use of configuration files (MMWAVE SDK User Guide, 2019). The data in the dataset was gathered employing the following parameters:

1. November 26/27, 2019 – the maximum velocity was set to 152 km/h, the maximum distance was set to 147 meters, the accuracy of speed reading was about 1.2 km/h, and the accuracy of distance reading – about 2.2 m;

2. December 16/17, 2019 – the maximum velocity was set to 158 km/h, the maximum distance was set to 147 meters, the accuracy of speed reading was about 1.2 km/h, and the accuracy of distance reading – about 2.2 m.

The detection of vehicles driving on the adjacent roadway was disabled – only objects approaching the sensor were subject to detection.

The sensor data was stored in binary .dat format. Based on the documentation installed with mmWave SDK package (Millimeter Wave (mmw) Demo for XWR16XX, 2019), a python script was developed to decode the data into CSV format.

Data quality and availability

The data comes from two sessions lasting 24h each. Unfortunately, there are some gaps in the recorded data. These are caused by: errors in capturing data, applying corrections to radar settings (modifying the configuration files).

The gaps occur mainly between 9 and 11 pm.

Dataset DOI
https://doi.org/10.34808/2hra-xr19

Acknowledgements
The project entitled: "INZNAK: Intelligent Road Signs with V2X Interface for Adaptive Traffic Controlling (No. POIR.04.01.04-00-0089/16) is subsidised from the European Regional Development Fund by the Polish National Centre for Research and Development (NCBR).

References

IWR1642 EVM (IWR1642BOOST) 'Single-Chip mmWave Sensing Solution User's Guide', May 2017, revised July 2018.

'Millimeter Wave (mmw) Demo for XWR16XX, mmWave SDK' (2019).

'MMWAVE SDK User Guide' (2019).

Data Descriptor

AC Motor Voltage and Audible Noise Dataset

Beata Pałczyńska[1*]

[1] Department of Metrology and Information Systems, Faculty of Electrical and Control Engineering; Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: beata.palczynska@pg.edu.pl; ORCID: 0000-0002-2404-6500

Abstract

The dataset titled AC motor voltage and audible noise waveforms in ship's electrical drive systems with frequency converters contains the voltage and sound measurement results recorded in a marine frequency controlled AC drive system. The dataset is part of research focussing on the impact of the ship's electrical drive systems with frequency converters on vibrations and the level of audible noise on ships. The dataset allows the correlation between electric disturbances and vibration, as well as audible noise in marine frequency controlled AC drive systems to be investigated.

Keywords: ship's electrical drive system; frequency converter; sinusoidal LC filter; AC motor voltage; audible noise; electric disturbances

https://doi.org/10.34808/x55q-sz53_dyr_roz31

Specification table (data records)

Subject area	Metrology, Power Electronics, Acoustical engineering
More specific subject area	Electric disturbances measurements, audible noise measurements
Type of data	Text and audio files

How the data was acquired	The data acquisition was carried out at the laboratory of the Department of Ship Automatic Control at Gdynia Maritime University using a model of a marine drive system with a frequency converter. The AC motor voltage waveform was preconditioned by an LV 25-P voltage isolating sensor. A ½ inch condenser microphone and sound level meter (Sonopan IM 10) were used to measure the acoustic noise. Both types of analogue waveforms were processed by the digital measurement system equipped with SCXI-13-10 Terminal Blocks, SCXI-1141 Antialiasing Filters and a PCI-MIO-16XE-50 DAQ Board. The digital data was saved on the PC hard disc
Data format	Text files are in .txt format, audio files are in .wav format
Experimental factors	The data contained in the dataset were not processed
Experimental features	A correlation between the harmonics content in the motor voltage waveforms and the acoustic noise of the motor was found
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

In recent years, it is possible to observe a large number of frequency converters being applied to different drive systems with induction motors on ships, like in cranes, mooring winches, pumps, fans and in ship main drives (propellers). The power of the small fan motor can be less than 1 kW, but in the case of a ship's main engine, can reach even 6 MW. The efficiency of the frequency controlled AC drive system is very high, the motor speed control is very precise, overhaul is easy and not very frequent, and the frequency converters are not so expensive now. However, frequency converters have some disadvantages, because they generate electromagnetic disturbances in the wide frequency range. This happens due to the rectifier diodes and inverter transistors' commutation, which can be switched with the frequency from 2 kHz to 20 kHz. These disturbances, that are conducted to the system supply net and to the supply line of a motor, increase the vibrations of the whole drive system. Vibration components laying in the acoustic frequency range cause a significant increase of audible noise emitted by the drive system (Palczynska, Spiralski and Wyszkowski, 2005, pp. 73–76).

The dataset, AC motor voltage and audible noise waveforms in ship's electrical drive systems with frequency converters, has been designed to investigate the effect of electric disturbances on audible noise in frequency controlled AC drive systems. The conclusion of the study was to indicate that applying a filter connected at the output of the frequency converter not only reduces the level of electrical disturbances in the AC motor voltage

but also significantly decreases the level of an acoustic noise of the motor. The dataset contains the digital form of AC motor voltage (Fig. 31.1 – measuring point 9) and audible noise waveforms (Fig. 31.1 – measuring points 7 and 8). The dataset also includes measurement results for the repeated procedure with the output LC passive filter (Fig.31.1 – item no 3).

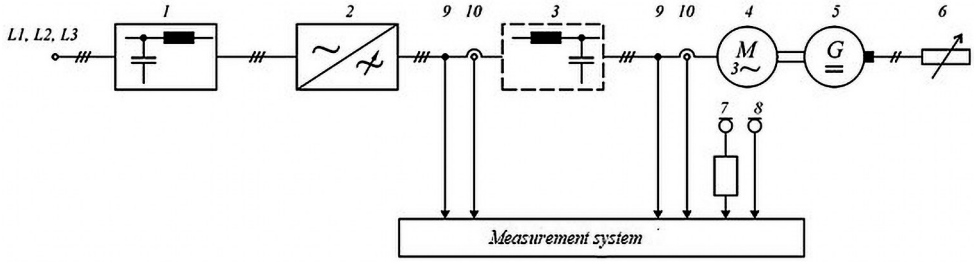


Fig. 31.1. Block diagram of the laboratory drive system. 1 – RFI filter, 2 – frequency converter, 3 – optimal sinusoidal output filter, 4 – induction motor, 5 – DC generator, 6 – adjustable resistor, 7 – Sonopan IM 10 sound level meter, 8 – microphone, 9 – LV 25-P voltage transducer, 10 – LA 55-P/SPI current transducer. (Wyszkowski, Spiralski and Winter, 2003)

Methods

The experiments based on a model of the marine drive system with a frequency converter were performed at the laboratory of the Department of Ship Automatic Control at Gdynia Maritime University. The measurements were carried out for two cases, first, the squirrel cage asynchronous motor was connected directly to the output of the frequency converter, then to confirm the thesis of the correlation between harmonics contents in motor voltage waveforms and acoustic noise of the motor, the sinusoidal LC passive filter was connected at the output of the frequency converter. During tests, the motor was nominally loaded with the use of the DC generator connected to an adjustable resistor. The sound level meter was pointed directly to the source, positioned at a distance of 1m from the motor lifting eye, in the motor axis of symmetry, from the drive end, 1 m above the floor.

The data acquisition of the audible noise level (A-weighed) and voltage waveforms were carried out with the digital measurement system. The AC motor voltage and audible noise waveforms were recorded in digital form for a few different output voltage frequencies fout (20 Hz, 23 Hz, 25Hz, 40 Hz, 46 Hz, 50 Hz) and a few different inverter's transistors switching frequencies fsw (3 kHz, 6 kHz, 12 kHz). Then the measurement procedure was repeated with the output LC passive filter, for all output voltage frequencies fout and two inverter's transistors switching frequencies fsw (6 kHz, 12 kHz). The sampling frequency of the PCI-MIO-16XE-50 DAQ Board was equal to 20,000 Hz. The values of the voltage samples in volts are smaller than the motor supply line voltage due to the attenuation brought by the voltage isolating transducer. The measured sound level has been A-weighed during experiments.

Data quality and availability

The standard uncertainties of voltage and acoustic measurement results were estimated using the B-type method. The relative uncertainty of the acoustic power measurements generated by an AC motor supplied from the frequency converter is less than 2.7% and depends, but not too much, on the inverter output frequency. However, when an output sinusoidal filter is used, the acoustic power generated by the AC motor decreases slightly, and is assessed with a relative measurement uncertainty of less than 2.6%. The main sources of uncertainty in sound level measurements are the sound meter errors related to the microphone directivity characteristics and the distance of the measuring points from the source. The relative uncertainty of the voltage measurements in both cases does not exceed 1.5%.

Dataset DOI:
10.34808/8cq0-2r35

Acknowledgements
The generation of this dataset was supported by a statutory research grant from Gdynia Maritime University.

References

Palczynska, B., Spiralski, L. and Wyszkowski, J. (2005) 'The influence of frequency controlled AC drives on vibration and audible noise level on ships', in: *Proceedings 10th IMEKO TC10 International Conference on Technical Diagnostics*, Hungarian Academy of Sciences, pp. 73–76.

Wyszkowski, J., Spiralski, L. and Winter, R. (2003) 'Digital Measurement system of electromagnetic disturbances and audible noise of marine frequency controlled AC drives', in: Book of summaries: *XVII IMEKO World Congress: Metrology in the 3rd millennium*: 22–27th June, 2003 Dubrovnik, Croatia, Croatian Metrology Society, pp. 2108–2111.

Data Descriptor

Impedance Spectra of RC Model as a Result of Testing Pulse Excitation Measurement Method Dataset

Grzegorz Lentka[1*]

[1] Department of Metrology and Optoelectronics; Faculty of Electronics, Telecommunications and Informatics; Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: grzegorz.lentka@pg.edu.pl; ORCID: 0000-0003-1221-7357

Abstract

The dataset titled Impedance spectra of RC model as a result of testing pulse excitation measurement method contains the impedance spectrum of an exemplary test RC model obtained using pulse excitation. The dataset allows presentation of the accuracy of the impedance spectroscopy measuring instrument, which uses the pulse excitation method to shorten the time of the whole spectrum acquisition.

Keywords: impedance measurement; impedance spectroscopy; pulse excitation

https://doi.org/10.34808/x55q-sz53_dyr_roz32

Specification table (data records)

Subject area	Metrology, Electronics, Non-destructive testing
More specific subject area	Impedance spectroscopy measurement
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology using a custom-built impedance spectroscopy analyser using pulse excitation and when measuring the RC test object model
Data format	The tables are in text format (Solartron)
Experimental factors	The data contained in the dataset were processed using linear approximation and after that, integrals were calculated to obtain the Fourier transform-based spectrum

Experimental features	The pulse time was set to 0.1 s, total acquisition time was set to 1000 s and excitation pulse amplitude was set to 1 V
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

One of the important applications of impedance spectroscopy is the diagnostics of protective anticorrosion coatings on the basis of their impedance spectra (O'Donoghue M. et al, 2003). Modern high-thickness anticorrosion coatings testing requires determination of the impedance spectrum in a wide frequency range, including very low frequencies (e.g., from 1 mHz). Due to the long measurement time at very low frequencies, traditional impedance spectroscopy cannot be used in field-worthy instruments (the total measurement time is longer than 1h). For this reason, the authors used square-pulse excitation to shorten spectrum determination down to a few minutes. (Hoja and Lentka, 2011). Unfortunately, for higher frequencies of the spectrum, the pulse signal energy decreases quickly (Fig. 32.1a).

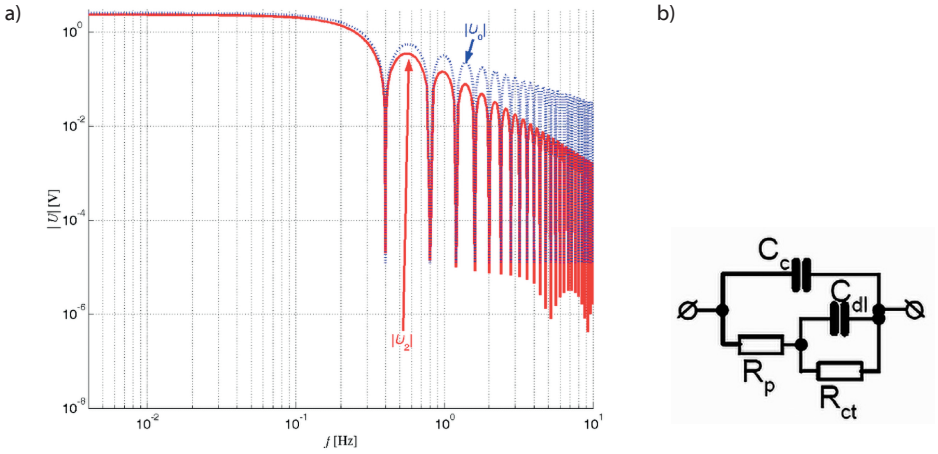


Fig. 32.1. a) Spectra of modulus of two signals: (Blue) Ideal square pulse |Uo| and (red) voltage across Zx |U2|, b) Baunier's RC model used as an object under test

The dataset, Impedance spectra of RC model as a result of testing pulse excitation measurement method, has been prepared to show the results of the impedance measurement technique employing pulse excitation and custom-made measuring instruments. The dataset contains the results for the test object (Fig. 32.1b) representing Bauniers model built using reference resistors and capacitors with known values thus allowing

the accuracy of the measurement method to be estimated. The used components are Cc = 314.6 pF, Rp = 9.935 GΩ, Cdl = 2.22 nF, and Rct = 4.969 GΩ. The tolerance of the components is 0.01% for the resistors and 0.1% for the capacitors, respectively. Highly stable components were used to assure the long-term repeatability and stability of the model.

Methods

A square pulse with an amplitude of 1 V and a duration equal to 0.1 s was generated using an LTC1668 16-bit digital-to-analog converter and analog output buffer using BUF634. The response signals: voltage across and current through the measured object were extracted using analog input circuitry using OPA627 amplifiers and then sampled using two LTC1420 12-bit analog-to-digital converters working simultaneously and storing the acquired signals into two separate blocks of RAM. The total duration of the acquisition is 1000 s. The samples are then taken from the memory buffer by the ARM processor to perform the calculations and to produce the final results – the impedance spectrum at the required frequency points (determined during the calculation phase). The calculation assumes three operations: in the first one, the linear approximation of the both acquired signals is performed, then the integral is calculated to obtain the Fourier transform of the voltage and current signal, e.g. the spectra of the voltage and the current, finally the impedance modulus spectrum is calculated by dividing point-by-point the modulus of the voltage by the modulus of the current as well as the calculation of the difference between the phase spectrum of the voltage and the phase spectrum of the current leading to the impedance phase spectrum. To estimate the resulting spectrum errors, the reference spectrum was calculated using the known values of the components.

Data quality and availability

All measurements were collected at laboratory temperature using a custom-built impedance analyser employing the pulse excitation impedance measurement method. The acquired data (voltage across and current through the test object) was used to calculate the impedance spectrum of the object under test. For calculation, the linearly interpolated data was integrated representing the Fourier transform. The obtained results were compared with the theoretically calculated impedance spectrum of the test object using the known reference values of the RC components.

Dataset DOI
10.34808/4gns-yv68

Acknowledgements
The generation of this dataset was supported by a statutory research grant from Gdańsk University of Technology.

References

O'Donoghue, M. et al. (2003) 'Electrochemical impedance spectroscopy: Testing coatings for rapid immersion service', *Materials Performance,* 42(9), pp. 36–41.

Hoja, J. and Lentka, G. (2011) 'Method using square-pulse excitation for high-impedance spectroscopy of anticorrosion coatings', *IEEE Transactions on Instrumentation and Measurement*, 60(3), pp. 957–964, DOI: 10.1109/TIM.2010.2060219.

Data Descriptor

X-ray Photoelectron Spectroscopy of Carboxylic Acids as Corrosion Inhibitors of Aluminium Alloys

Jacek Ryl[1*], Joanna Wysocka[2]

[1] Advanced Materials Center, Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)
[2] Institute of Biotechnology and Molecular Medicine (25 Kampinoska, Gdańsk, Poland)

[*] Correspondence author: jacek.ryl@pg.edu.pl

Abstract

The datasets, titled X-ray Photoelectron Spectroscopy studies of citric acid adsorption on aluminium alloy 5754 in alkaline media and X-ray Photoelectron Spectroscopy studies of various carboxylic acids adsorption on aluminium alloys in alkaline media, contain XPS studies of the corrosion inhibitory action of selected dicarboxylic acids towards commercially available aluminium alloy 5754 in alkaline media at pH=11. These datasets are part of a studies undertaken to determine the efficiency of carboxylic acids as green corrosion inhibitors. The datasets allow the analysis of raw XPS data and evaluation of the accuracy of the deconvolution models selected for this study, which is of key importance for any analytical technique based on the construction of a model to estimate experimental data.

Keywords: X-ray photoelectron spectroscopy; XPS; corrosion inhibitor; aluminium alloy; carboxylic acid

https://doi.org/10.34808/x55q-sz53_dyr_roz33

Specification table (data records)

Subject area	Electrochemistry, Corrosion Science, Materials Engineering
More specific subject area	X-ray Photoelectron Spectroscopy, Adsorption Studies
Type of data	Text

How the data was acquired	The data were measured with an Escalab 250Xi (ThermoFisher Scientific) using the Avantage 5.973 software at the Gdańsk University of Technology
Data format	The tables are in .avg and .vgd formats
Experimental factors	The data contained in the dataset were not processed
Experimental features	The XPS spectra were collected under identical conditions for each study on carboxylic acid
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Along with steel, aluminum alloys are the most commonly used construction materials with versatile applications, while the corrosion resistance of aluminium is determined by the presence of a so-called passive layer, a thin aluminum oxide film, which is unstable in acidic and alkaline environments. Its corrosion resistance may be enhanced by using corrosion inhibitors, but these are often expensive or toxic chemical compounds. There is impetus to investigate biodegradable and non-toxic alternatives; green corrosion inhibitors, often made of plant extracts or post-production wastes. An aluminium corrosion inhibitor's inhibitory action depends on the successful formation of the adsorbed layer, a barrier preventing the direct contact of the protected metal with the corrosive media. The adsorption mechanism and corrosion inhibition efficiency depend on the molecular structure of the chemical compound. The undertaken study aims to reveal the alteration of influence of the corrosion inhibitor based on the amount of carboxyl and hydroxyl functional groups within the molecule and the presence of an amine functional group. The studied carboxylic acids are citric, succinic, malic, maleic, tartaric, tricarballylic, and serine.

The datasets discussed within this data descriptor reveal the surface chemistry analysis with XPS of the adsorption of selected carboxylic acids at the surface of aluminium alloy in alkaline pH, on which the inhibitory action depends. X-ray photoelectron spectroscopy (XPS) is a modern analytical method used in many leading research centres worldwide. The signal resulting from the photoelectric effect differs depending on the electron binding energy resulting from the presence of various chemical bonds or other molecular interactions, moreover, it is generated at a depth of only a few nanometers below the surface of the sample, which makes XPS ultra-surface sensitive for the determination of the chemistry of thin adsorbed layers of corrosion inhibitors. However, for a more advanced analysis, a fitting procedure, deconvolution, should be performed.

The analysis requires the construction of fitting models, often quite complex, to properly consider each type of chemical bonds present at the investigated material's surface.

The superimposition of peaks from various chemical states, surface contamination due to air exposure, or the appearance of surface ionisation spectral shift in the case of insulating materials, are only a few things that need to be considered appropriately.

Our datasets allow for free access to the raw XPS data obtained during studies on aluminium alloys' green corrosion inhibitors. It has been shared to allow us to study and evaluate the fitting model applied by us in our studies. The results of the dataset titled X-ray Photoelectron Spectroscopy studies of citric acid adsorption on aluminium alloy 5754 in alkaline media are part of the manuscript (Wysocka et al., 2017), while X-ray Photoelectron Spectroscopy studies of various carboxylic acids adsorption on aluminium alloys in alkaline media contains the XPS results presented as a part of our second manuscript in this series (Wysocka et al., 2018).

Methods

The XPS results were obtained using an Escalab 250Xi multi-spectroscope from ThermoFisher Scientific, with the Avantage 5.973 software provided by the manufacturer. The spectroscope is equipped with a monochromatic AlKα X-ray source. The surveys were carried out at a binding energy of 150 eV, while for high-resolution measurements, the binding energy did not exceed 20 eV. The charge compensation was performed through a low-energy electron, and low-energy argon ion flow using a flood gun. The detailed experimental conditions are contained within each file in the dataset.

The datasets contain the results of the survey and high-resolution XPS measurements for the aluminium alloy 5754 exposed in bicarbonate buffer (pH = 11) without and with the addition of studied carboxylic acid in different concentrations. The exposure to the corrosive media was 100 minutes for each sample.

The data in the dataset titled X-ray Photoelectron Spectroscopy studies of citric acid adsorption on aluminium alloy 5754 in alkaline media contain citric acid adsorption studies at different concentrations, characteristic to various stages of interaction identified for this group of compounds. Stage 1 – concentration insufficient for effective 3D adsorption layer formation (2 and 4.5 mM), Stage 2 was the interstage, Stage 3 – full surface coverage by the corrosion inhibitor (9 and 20 mM), Stage 4 – an excess of corrosion inhibitor (100 mM). These results are compared to aluminium alloy surface chemistry in the absence of citric acid within the electrolyte (folder file: bufor).

The data in the dataset titled X-ray Photoelectron Spectroscopy studies of various carboxylic acids adsorption on aluminium alloys in alkaline media show the follow-up studies and reveal the variable interactions of aluminium alloy with different dicarboxylic acids (succinic acid, folder file bursz; malic acid, folder file jabl; maleic acid, folder file malein; tartaric acid, folder file winowy), tricarboxylic acids (citric acid, folder file cytr; tricarballylic acid, folder file trikarb) and one amino acid (serine, folder file seryna) when dissolved in the alkaline media. These studies were carried out at two concentrations of each compound, namely 4.5 and 20 mM, which correspond to the inhibitory action of Stage 1 and Stage 3 for citric acid. Different interaction derives from the altered amount of carboxylic and hydroxylic functional groups in each studied compound.

Data quality and availability

All measurements were collected using an Escalab 250Xi spectroscope (ThermoFisher Scientific) using the Avantage 5.973 software and typical operating parameters, as described within each experimental file. The dataset may be opened with Avantage or any other processing software developed for XPS analysis, such as CasaXPS, available at http://www.casaxps.com/.

Datasets DOI:
10.34808/s317-v790
10.34808/675p-aq50

Acknowledgements
The generation of this dataset was possible thanks to the financial support of the Polish Ministry of Science and Higher Education from the budget funds in the period 2016–2019 under Iuventus Plus project IP2015 067574.

References

Wysocka, J. et al. (2018) 'Carboxylic acids as efficient corrosion inhibitors of aluminium alloys in alkaline media', *Electrochimica Acta,* 289, 175–192. DOI: 10.1016/j.electacta.2018.08.070.

Wysocka, J., Krakowiak, S. and Ryl, J. (2017) 'Evaluation of citric acid corrosion inhibition efficiency and passivation kinetics for aluminium alloys in alkaline media by means of dynamic impedance monitoring', *Electrochimica Acta,* 258, 1463–1475. DOI: 10.1016/j.electacta.2017.12.017.

Data Descriptor

Crack Mouth Opening Displacement for EH36 Shipbuilding Steel Measurements Dataset

Jakub Kowalski[1*], Janusz Kozak[1]

[1] Faculty of Mechanical Engineering and Ship Technology, Institute of Naval Architecture and Ocean Engineering, Gdańsk University of Technology; (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: jakub.kowalski@pg.edu.pl; ORCID: 0000-0001-9603-1220

Abstract

The dataset titled EH36 steel for shipbuilding (plate thickness 50 mm) – CMOD – force record, a0/W=0.6 contains a CMOD (Crack Mouth Opening Displacement) – Force record which is the base for evaluation of the fracture toughness of structural steel. Bend specimens with a Bx2B section (B = 50 mm), and relative initial crack length a0/W=0.60 were used. The test was carried out at ambient temperature in accordance with the ISO 12135 standard. The dataset can be useful in the calibration of numerical material models which includes ductile failure. It can also be use as reference data in the comparison of the toughness designations of carbon steels.

Keywords: fracture toughness; CMOD; ductility

https://doi.org/10.34808/x55q-sz53_dyr_roz34

Specification table (data records)

Subject area	Mechanical engineering, mechanical properties of material, fracture toughness
More specific subject area	Ductility of structural steel
Type of data	Text
How the data was acquired	The data was collected at the Gdańsk University of Technology on a dedicated test stand. All calibrated voltage signals (from a force transducer and extensometer) were measured and recorded by a HBM Quantum MX840A device
Data format	The tables are in .ods format

Experimental factors	The data contained in the dataset were zero balanced and filtered
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The fracture toughness of material is one of the critical issues for marine structures built with thick plates. The Charpy impact test, which is typically used in case of thin and moderately thick plates, may not be sufficient in the case of thick structures. In such cases, some additional requirements are defined (DNV GL, 2019). On of these is a Fracture Mechanics (FM) based test – CTOD (Crack Tip Opening Displacement) test. The FM test is to be performed with specimens prepared from a full thickness plate. In this particular case (plate thickness 50 mm, specimen for three point bend test), the dimensions of the specimen are as follows: 50 × 100 × 460 mm and weight about 18 kg. The presented dataset is part of research on the scale effect and relative notch depth in fracture mechanics testing of high tensile steels for shipbuilding (Kowalski and Kozak, 2018b, 2018a)there is a risk of appearance of brittle fracture. This risk is reduced through the use of certified materials having guaranteed strength at a given temperature. A method which is most frequently used to determine brittle fracture toughness is the Charpy impact test, preformed for a given temperature. For offshore structures intended to work in the arctic climate, the certifying institutions more and more often require Crack Tip Opening Displacement (CTOD).

Methods

Fracture toughness test methods are standardised. The presented dataset fulfils the requirements of (British Standards Institute, 1991) and (ISO, 2016). A three point bend specimen with the breadth-to-width ratio B/W = 0.5 and relative crack length a0/W = 0.6 was used – see Fig. 34.1. The test was performed at ambient temperature (23 ±5°C). Crack extension during stable tearing was not included.

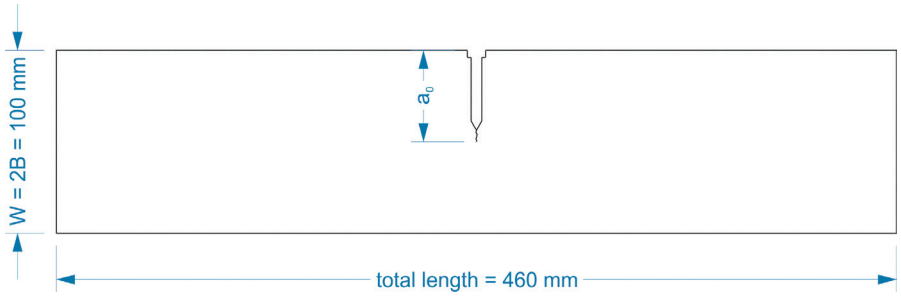


Fig. 34.1. Dimensions of applied specimen

The data was acquired on a ±250 kN dedicated stand in displacement controlled mode. CMOD was measured by an Epsilon 3541-010M-120M-LT extensometer, with a gauge length of 10 mm and range –2, +12 mm.

Data quality and availability

All of the applied measuring equipment was calibrated by an external, independent certification body. The laboratory estimated the measuring uncertainties. The collected data were manually checked, zero-balanced and filtered.

Dataset DOI
10.34808/y12d-ph76

Dataset License

Acknowledgements

The generation of this dataset was supported by a statutory research grant from Gdańsk University of Technology.

References

British Standards Institute (1991) 'BS 7448-1 Fracture mechanics toughness tests. Method for determination of KIc, critical CTOD and critical J values of metallic materials'.

DNV GL (2019) 'DNVGL Offshore standards OS-B101 Metallic Materials'. Available at: https://rules.dnvgl.com/docs/pdf/DNVGL/OS/2019-07/DNVGL-OS-B101.pdf (Accessed: 26th May 2022)

ISO (2016) 'ISO 12135:2016 Metallic materials — Unified method of test for the determination of quasistatic fracture toughness'.

Kowalski, J. and Kozak, J. (2018a) 'Numerical Model of Plastic Destruction of Thick Steel Structural Elements', *Polish Maritime Research*, 25(2), pp. 78–84. DOI: 10.2478/pomr-2018-0057.

Kowalski, J. and Kozak, J. (2018b) 'The Effect of Notch Depth on CTOD Values in Fracture Tests of Structural Steel Elements', *Polish Maritime Research*, 25(2), pp. 85–91. DOI: 10.2478/pomr-2018-0058.

Data Descriptor

Tribological Properties of Thermoplastic Materials Formed by 3D Printing by FDM Process

Jacek Łubiński[1*]

[1] Faculty of Mechanical Engineering and Ship Technology, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: jacek.lubinski@pg.edu.pl; ORCID: 0000-0003-0852-8400

Abstract

The dataset entitled 3D printed ABS thermoplastic vs. steel. Dry sliding wear test in constant load & velocity ring on flat configuration. Test parameters: print layer thickness and orientation. Test symbol: 019_h_4 contains: the time base (expressed in seconds and minutes), the friction torque for sliding friction, rotational velocity of the counter – specimen (velocity of sliding), friction coefficient, load in the friction contact interface, specimen temperature. Tests were conducted in an in-house developed tribological tester, the Tribometr PT-1. The test belongs to a testing program on the influence of technological parameters in 3D printing from thermoplastic materials by FDM process on friction in sliding with a steel counter – specimen without added lubrication.

Keywords: tribology, 3D printing, sliding friction, tribological wear, thermoplastic polymers

https://doi.org/10.34808/x55q-sz53_dyr_roz35

Specification table (data records)

Subject area	Mechanical engineering, tribology, sliding friction, experimental testing
More specific subject area	Measurements of tribological performance of a flat – on – flat unidirectional sliding contact of C45 annealed steel and ABS polymeric specimen (3D print by FDM process)
Type of data	Text
How the data was acquired	The data were collected at the Gdańsk University of Technology using the universal tribological test rig Tribometr PT-1

Data format	The tables are in .xls format
Experimental factors	Most of the data contained in the dataset were not processed. The friction coefficient was calculated in post-processing
Experimental features	The samples were manufactured in a 3D printing FDM process (Zortrax Inventure) from ABS filament
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

There is well-established interest in furthering knowledge on the tribological performance of the polymer-metal sliding combination under dry friction, which was presented, for example, in Mithun V. et al. (2016). The onset of rapid prototyping technologies sparked works on the impact of the FDM technology parameters on the tribological parameters. In the work by Kozior T. and Kundera C. (2018), the influence of the printing positions of ABS specimens on wear in sliding friction is observed. A similar study by Mithulin V. et al. (2016) investigates the tribological performance of ABS and PA6 polymer-metal sliding contact in dry friction with the focus on the influence of heat on the process. Despite the emerging new works on the problems of friction and wear in the sliding of components manufactured in a 3D printing process, the area of knowledge still remains in great need of further research effort.

Our dataset, 3D printed ABS thermoplastic vs. steel. Dry sliding wear test in constant load & velocity ring on flat configuration. Test parameters: print layer thickness and orientation. Test symbol: 019_h_4, was created to evaluate the friction coefficient in sliding friction in a unique combination of materials. The particular test is one in a series performed to look into the influence of additive manufacturing process parameters on the tribological interaction between a steel specimen and a structured ABS plastic specimen deposited in layers of hot filament setting incrementally. The dataset contains: the input parameters of the test (load, velocity) and output (friction torque/coefficient, temperature of the polymeric specimen). The observed torque and input load were used to calculate the friction coefficient.

Methods

The tests were conducted in a custom-developed tribological test rig, the Tribometr PT-1. The geometry of the test set-up is presented in Fig. 35.1. The top specimen (cylindrical) is made of C45 steel, annealed to approx. 45 HRC and ground to a roughness of Ra < 0.32 μm. The bottom specimen is printed in the FDM 3D printing process and is used without any machining of the usable faces.

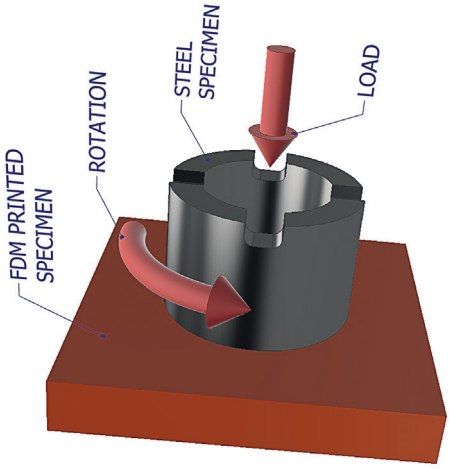


Fig. 35.1. Specimen setup geometry as used in the Tribometr PT-1 test rig

The ABS specimens are aged for a minimum of 48h to allow for relaxation of the residual stress and normal saturation with moisture and atmospheric gases. The test load is equal to FN = 410 N (mean contact pressure of pav = 0.5 MPa). The mean radius of the ring-shaped contact is equal to Rav = 0.0093 m = 9.3 mm. The rotational velocity of the spindle was set at 110 RPM (average sliding velocity of approx. 0.1 m/s). The test run lasted 30 min.

Data quality and availability

All of the polymeric (ABS) specimens were printed in a Zortrax Inventure printer (FDM process) from the same batch of filament. The specimens' technological variation comprised the thickness of a single layer of polymer deposited in one pass and the orientation of the specimen with respect to the printing table. The specimens were fabricated with the working face either parallel or perpendicular to the printing table. The tests were carried out in an air-conditioned room, at a temperature of 20°C and 60% relative humidity of ambient air.

The relative error in the measurement of load (force normal to the friction interface), friction torque, rotational velocity and temperature, was no greater than 1.5%. The linear dimensions of the contact zone were measured with the accuracy of 0.01 mm.

Dataset DOI
10.34808/tnbp-4r39

Acknowledgements
The generation of this dataset was supported by statutory research grants from Gdańsk University of Technology.

References

Hugo, I., Medellín-Castillo, J.E. and Pedraza, T. (2009) 'Rapid Prototyping and Manufacturing: A Review of Current Technologies', *Proceedings of the ASME 2009 International Mechanical Engineering Congress and Exposition, 4: Design and Manufacturing*. Lake Buena Vista, Florida, USA, pp. 609–621. ASME., DOI: 10.1115/IMECE2009-11750.

Kozior, T. and Kundera, C. (2018) 'An analysis of the impact of the FDM technology parameters on tribological properties', *Tribologia,* 5, pp. 33–40. DOI: 10.5604/01.3001.0012.7651.

Krasmik, V. and Schlattmann, J. (2017) 'Study of the friction and wear behavior of metal/polymer-metal multi-material configurations', *Key Engineering Materials*, 739, pp. 211–219, DOI: 10.4028/www.scientific.net/KEM.739.211.

Kar, M.K. and Bahadur, S. (1978) 'Micromechanism of wear at polymer-metal sliding interface', *Wear*, 46(1), pp. 189–202, DOI: 10.1016/0043-1648(78)90120-5.

Kulkarni, M.V. et al. (2016) 'Tribological behaviours of ABS and PA6 polymer-metal sliding combinations under dry friction, water absorbed and electroplated conditions', *Journal of Engineering Science and Technology*, 11(1), pp. 068–084. Available at: https://jestec.taylors.edu.my/Vol%2011%20issue%201%20January%202016/Volume%20(11)%20Issue%20(1)%20068-%20084.pdf (Accessed: 26th May 2022).

Boparai, K.S., Singh, R. and Singh, H. (2016) 'Wear behavior of FDM parts fabricated by composite material feed stock filament', *Rapid Prototyping Journal*, 22(2), pp. 350–357, DOI: 10.1108/RPJ-06-2014-0076.

Sudeepan, J. et al. (2014) 'Study of Friction and Wear of ABS/ZnO Polymer Composite Using Taguchi Technique', *Procedia Materials Science*, 6, pp. 391–400, DOI: 10.1016/j.mspro.2014.07.050.

Data Descriptor

Potential Energy Curves of Diatomic Alkali Molecules Datasets

Patryk Jasik[1*], Tymon Kilich[1]

[1] Department of Theoretical Physics and Quantum Information, Faculty of Applied Physics and Mathematics, Institute of Physics and Computer Science, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: patryk.jasik@pg.edu.pl; ORCID: 0000-0002-6601-0506

Abstract

The datasets described in this article contain potential energy curves for several diatomic systems. The data was obtained via high-performance computing using MOLPRO, a system of ab initio programs for advanced molecular electronic structure calculations. The datasets allow to model bond lengths, energy levels, spectra and time-evolution of molecular dimers for which the data are presented.

Keywords: diatomic molecules, dimers, potential energy curves

https://doi.org/10.34808/x55q-sz53_dyr_roz36

Specification table (Data records)

Subject area	Physics, Chemistry
More specific subject area	Atomic and molecular physics
Type of data	Numerical data
How the data was acquired	The data was obtained in calculations using the MOLPRO program package version 2012.1.43
Data format	The tables are in .csv format
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The datasets are available free of charge under Creative Commons Licenses

Background

Potential energy (hyper)surfaces describe energy of the electronic states of molecular systems and their changes with respect to the changes of the positions of the nuclei, creating the "energy landscape" of molecular geometry. It is an instrumental tool in the analysis of molecular conformations, transition states and chemical reaction dynamics (Thrular et al, 1987). In a diatomic molecule, in which there are only two atoms, the position of the nuclei can be expressed by only one coordinate, hence the potential energy surface reduces to a potential energy curve (PEC). Each curve corresponds to one electronic state of group representation and angular momentum. The ranges of the internuclear distances in the datasets depend on the described system. Our datasets consists of several chosen systems of diatomic molecules, created from pairs of alkali metal atoms. Such dimers are of particular interest in applications concerning ultracold (with an internal temperature in the range of mK) molecular systems, Bose-Einstein condensation, and coherent control of chemical reactions. Possible applications of strongly polar ultracold molecules include using long-range electric dipole-dipole interactions between polar molecules for the design of optical quantum systems. The internal degrees of freedom of polar molecules could be used as a medium for quantum information. The creation, storage as well as control of such molecules in optical lattices, created by a strong laser field, could be used to construct quantum computers (Pazyuk, 2015).

Methods

The potential energy curves were obtained using methods of quantum chemistry using Molpro, a software package developed for accurate ab initio quantum chemistry calculations (Werner et al., 2012; Werner et al., 2020). The potential energy curves were obtained within the Born-Oppenheimer approximation. The core electrons and relativistic effects were treated using the energy-consistent effective core potentials (ECP) from the Stuttgart/Cologne Group with the customised core polarisation potentials (CPP). The valence electrons were described using customised basis sets; the CASSCF and MRCI methods. A detailed discussion of this methodology is described in our previous works (Jasik et al., 2006; Wiatr et al., 2015; Jasik et al., 2018).

Data quality and availability

The quality of the datasets is realised by carefully optimised computational basis sets, and parameters of ECPs, as well as CPPs. The quality analysis of the potential energy curves available in the presented datasets is always based on a comparison with the results of other theoretical and experimental studies. As an example, Fig. 36.1. shows the comparison of the ground state of the NaRb molecule presented in the dataset with other available results. Overall, very good agreement is visible. All of the presented PECs are characterised by common spectroscopic parameters, such as bond length (Re), dissociation energy (De), or vibrational constant (we). Comparison of these features with data

provided by other authors is also used for verification of the quality of the dataset. A detailed analysis of the spectroscopic parameters is available in our previous works (Jasik et al., 2006; Wiatr et al., 2015; Jasik et al., 2018).

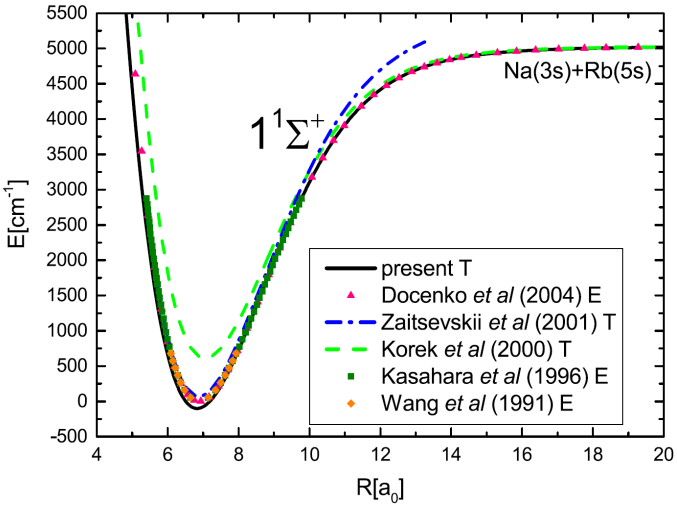


Fig. 36.1. Comparison of the ground 11Σ+ state of the NaRb molecule with experimental data obtained by (Docenko et al., 2004), (Kasahara et al., 1996), and (Wang et al., 1991), as well as with other theoretical results presented by (Zaitsevskii et al., 2001) and (Korek et al., 2000). The capital letter T refers to theoretical results and E denotes experimental data. (Wiatr et al., 2015)

Datasets DOI:

10.34808/t0yk-w861
10.34808/3kb8-st86
10.34808/bvx6-xs32
10.34808/1200-rr20
10.34808/1td6-pr92
10.34808/mjms-th03
10.34808/fj1r-d796
10.34808/75xd-bk46
10.34808/vpkc-xp73
10.34808/0vde-kw18

Datasets licenses

Creative Commons License (various variants)

Acknowledgements

Generation of these datasets would not have been possible without the help of our collaborators Józef E. Sienkiewicz, Marcin Wiatr and Hermann Stoll whom we would like to thank. The datasets described in these studies were calculated using computational

resources provided by Academic Computer Centre in Gdańsk (http://task.gda.pl) and using resources provided by the Wrocław Centre for Networking and Supercomputing (http://wcss.pl).

References

Docenko, O. et al. (2004) 'Potential of the ground state of NaRb', *Physical Review A*, 69(4), 042503. DOI: 10.1103/PhysRevA.69.042503.

Fuentealba, P. et al. (1982) 'A proper account of core-polarization with pseudopotentials: single valence-electron alkali compounds', *Chemical Physics Letters,* 89 (5), pp. 418–422, DOI: 10.1016/0009-2614(82)80012-2.

Jasik, P. and Sienkiewicz, J.E. (2006) 'Calculation of adiabatic potentials of Li2', *Chemical Physics*, 323(2–3), pp. 563–573, DOI: 10.1016/j.chemphys.2005.10.025.

Jasik, P. et al. (2018) 'Electronic structure and rovibrational predissociation of the $2^1\Pi$ state in KLi', *Physical Chemistry Chemical Physics*, 20 (27), pp. 18663–18670, DOI: 10.1039/C8CP02551G.

Kasahara, S. et al. (1996) 'High-resolution laser spectroscopy of the $X^1\Sigma^+$ and $(1)^3\Sigma^+$ states of $^{23}Na^{85}Rb$ molecule', *Journal of Chemical Physics,* 105 (4), 1341. DOI: 10.1063/1.472000.

Korek, M. et al. (2000) 'Theoretical study of the electronic structure of the LiRb and NaRb molecules', *Chemical Physics*, 256 (1), pp. 1–6, DOI: 10.1016/S0301-0104(00)00061-6.

Pazyuk, E.A. et al. (2015) 'Laser synthesis of ultracold alkali metal dimers: optimization and control', *Russian Chemical Reviews,* 84 (10), 1001. DOI: 10.1070/RCR4534.

Truhlar, D.G., Steckler, R. and Gordon, M.S. (1987) 'Potential energy surfaces for polyatomic reaction dynamics', *Chemical Reviews*, 87(1), pp. 217–236. DOI: 10.1021/cr00077a011.

Wang, Y.-C. et al. (1991) 'High resolution laser spectroscopy of the B 1Π–X 1Σ+ transition of $^{23}Na^{85}Rb$', *Journal of Chemical Physics*, 95 (9), 6229. DOI: 10.1063/1.461569.

Werner, H.J. et al. (2012) '*Molpro: a general-purpose quantum chemistry program package*', Wiley Interdisciplinary Reviews: Computational Molecular Science, 2 (2), pp. 242–253. DOI: 10.1002/wcms.82.

Werner, H J. et al. (2020) 'The Molpro quantum chemistry package', *The Journal of Chemical Physics*, 152(14), 144107. DOI: 10.1063/5.0005081.

Wiatr, M., Jasik, P. and Sienkiewicz, J.E. (2015) 'The adiabatic potentials of low-lying electronic states of the NaRb molecule', *Physica Scripta*, 90 (5), 054012, DOI: 10.1088/0031-8949/90/5/054012.

Zaitsevskii, A., et al. (2001) 'Energy and radiative properties of the low-lying NaRb states', *Physical Review A*, 63(5), 052504, DOI: 10.1103/PhysRevA.63.052504.

Data Descriptor

Minimal Sets of Lefschetz Periods for Morse-Smale Diffeomorphisms of a Connected Sum of g Real Projective Planes

Grzegorz Graff[1*], Adrian Myszkowski[1*]

[1] Faculty of Applied Physics and Mathematics, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence authors: grzegorz.graff@pg.edu.pl; ORCID: 0000-0001-5670-5729
adrian.myszkowski@pg.edu.pl; ORCID: 0000-0002-1011-4357

Abstract

The dataset titled Database of the minimal sets of Lefschetz periods for Morse-Smale diffeomorphisms of a connected sum of g real projective planes contains all of the values of the topological invariant called the minimal set of Lefschetz periods, computed for Morse-Smale diffeomorphisms of a non-orientable compact surface without boundary of genus g (i.e. a connected sum of g real projective planes), where g varies from 1 to 54.

Keywords: periodic points; Morse-Smale diffeomorphisms; minimal sets of Lefschetz periods; Lefschetz number

https://doi.org/10.34808/x55q-sz53_dyr_roz37

Specification table (data records)

Subject area	Dynamical systems
More specific subject area	Topological invariants, Periodic points theory
Type of data	Text
How the data was acquired	The data was generated using the Mathematica program, based on the algorithm published in (Graff and Myszkowski, 2019)
Data format	The lists are in the .txt format
Experimental factors	The data contained in the dataset was not processed
Experimental features	Values of the invariant

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Since 2007, the minimal sets of Lefschetz periods have been considered for Morse-Smale diffeomorphisms f of several classes of spaces, such as orientable and non-orientable compact surfaces without boundary, Cartesian product of spheres and n-dimensional tori. The importance of the set results from the fact that it constitutes a subset of minimal periods for the map f. The definition of is based on the analysis of possible forms of Lefschetz zeta functions for f whose form is rather complicated and thus manual computation of the invariant is usually extremely difficult.

In (Graff and Myszkowski, 2019), the definition of the minimal set of Lefschetz periods was reformulated by the use of so-called periodic expansion (cf. (Jezierski and Marzantowicz, 2006)), and the following theorem was proven: for a given Morse-Smale diffeomorphism f with a given periodic expansion of the Lefschetz numbers , we have:

This statement allows us to build an algorithm for finding minimal sets of Lefschetz periods within the given setting. Let us note that the number of periods grows exponentially as the genus g of the surface increases. As a consequence, the calculations could only be conducted for relatively small values of g.

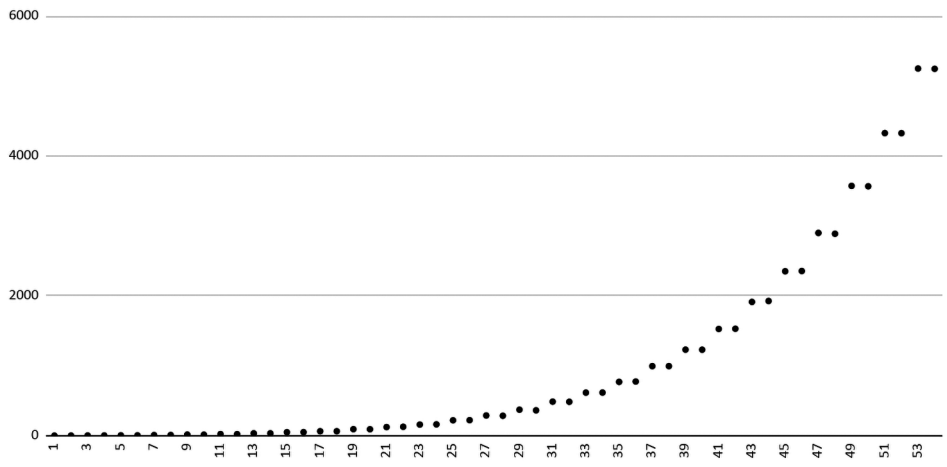


Fig. 37.1. The graph shows the number of all of the different minimal sets of Lefschetz periods for a connected sum of g real projective planes, g = 1, …, 54

Our dataset, Database of the minimal sets of Lefschetz periods for Morse-Smale diffeomorphisms of a connected sum of g real projective planes, may be useful for under-

standing the dynamics of Morse-Smale diffeomorphisms. It can also be used as a tool for checking hypotheses about the behaviour of periods of Morse-Smale diffeomorphisms. Moreover, the datasets allow the omission in (Llibre and Sirvent, 2013) to be eliminated in the calculation of the minimal sets of Lefschetz periods in the case of $g = 9$. Our results are a typical example of generating useful mathematical data: they include a list of some important mathematical quantities that could not be determined in a straightforward way, by manual calculations; however, application of a clever algorithm made it possible to compute them.

Methods

The results were obtained by a program in Mathematica based on the algorithm introduced in (Graff and Myszkowski, 2019). In order to obtain the results in a reasonable amount of computing time, it was necessary to express abstract mathematical theory by means of a computationally efficient framework. The design of the algorithm took into account many numerical properties of Lefschetz numbers and Möbius functions, which allowed the number of necessary calculations to be significantly reduced.

Data quality and availability

Dataset DOI
10.34808/9aj1-1977

Acknowledgements
The generation of this dataset was supported by the National Science Centre, Poland, UMO-2014/15/B/ST1/01710 and Sheng 1 UMO-2018/30/Q/ST1/00228.

References

Graff, G., Lebiedź, M. and Myszkowski, A. (2019) 'Periodic expansion in determining minimal sets of Lefschetz periods for Morse–Smale diffeomorphisms', *Journal of Fixed Point Theory and Applications*, 21, 47. DOI: 10.1007/s11784-019-0680-4.

Jezierski, J. and Marzantowicz, W. (2006) '*Homotopy methods in topological fixed and periodic points theory. Topological fixed point theory and its applications*', 3, Springer, Dordrecht. DOI: 10.1007/1-4020-3931-X.

Llibre, J. and Sirvent, V.F. (2013) 'Minimal sets of periods for Morse–Smale diffeomorphisms on non-orientable compact surfaces without boundary', *Journal of Difference Equations and Applications,* 19(3), pp. 402–417. DOI: 10.1080/10236198.2011.647006

Data Descriptor

On Computing Curlicues Generated by Circle Homeomorphisms

Justyna Signerska-Rynkowska[1*]

[1] Faculty of Applied Physics and Mathematics, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

* Correspondence author: justyna.signerska@pg.edu.pl; ORCID: 0000-0002-9704-0425

Abstract

The dataset entitled Computing dynamical curlicues contains values of consecutive points on a curlicue generated, respectively, by rotation on the circle by different angles, the Arnold circle map (with various parameter values) and an exemplary sequence as well as corresponding diameters and Birkhoff averages of these curves. We additionally provide source codes of the Matlab programs which can be used to generate and plot the first N points of curlicues of these types and calculating related quantities. Illustrative figures are included as well.

Keywords: dynamical systems; curlicues; circle homeomorphisms; rotation on the circle; Arnold circle map

https://doi.org/10.34808/x55q-sz53_dyr_roz38

Specification table (data records)

Subject area	Dynamical systems
More specific subject area	Low dimensional dynamics, Rotation theory
Type of data	Text; Figure; Source code
How the data was acquired	The data was generated using the Matlab program and information contained in (Signerska-Rynkowska, 2020)
Data format	.txt, .m, .eps
Experimental factors	The data contained in the dataset was not processed
Experimental features	Values of the geometric and dynamical quantities

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The notion of a curlicue originates mainly from the visual arts where it stands for a fancy curl that can serve as a decorative motif in architecture or calligraphy. However, in mathematics by a curlicue we shall mean a piece-wise linear curve on the complex plane, where, passing consecutively through the points , and , ,..., where:

A curlicue can be obtained from an arbitrary sequence of real numbers and for some sequences they can indeed form beautiful shapes (see e.g. Berry, Goldberg, 1988; Dekking, Mendès-France, 1981; Sinai, 2008). On the other hand, if the sequence is obtained from an orbit of a given map i.e. , then we can speak about dynamically generated curlicues. For mathematicians it is especially interesting to examine how the dynamical properties of are reflected in the geometric structure of . In particular, the recent work (Signerska-Rynkowska, 2020) studies curlicues generated by lifts of circle homeomorphisms and determines how the properties of the continued fraction expansion of its rotation number are connected with boundedness, the rate of diameter increase and the superficiality of such a curve (for definitions, see the cited paper and references therein).

The three functions Rotation.m, Arnold.m and Sequence.m can be used to plot the first points of a curlicue generated, respectively, by rotation on the circle by the angle of $2\pi\rho$, the Arnold circle map (with different parameter values of and) and the sequence . Additionally, the functions calculate corresponding Birkhoff averages and a diameter of a curlicue depending on number of iterates . A description of the functions and variables involved is provided as comments in the m-files. It is worth pointing out that the function Sequence.m can be easily modified to compute and draw a curlicue generated by an arbitrary sequence given by an explicit formula. Attached txt-files provide exemplary data obtained by these functions. Four figures (eps-files) were generated by the function Sequence.m for, respectively, (Fig. 38.1.eps,), (Fig. 38.2.eps,), (Fig. 38.3.eps,) and (Fig. 38.4.eps,).

Concerning curlicues generated by circle homeomorphisms with an irrational rotation number, it is known that an unbounded Birkhoff average implies that the diameter of a curlicue is unbounded as well. On the other hand, a bounded Birkhoff average can result both in the bounded and unbounded case. The bounded case (for transitive homeomorphisms) is equivalent to the fact that the corresponding cohomological equation has a continuous solution (see: Signerska-Rynkowska, 2020). The intriguing open question is to characterise the unbounded curlicues with a bounded Birkhoff average.

The dataset provided can be used for numerical experiments, perhaps assisting in arriving at new hypotheses concerning this problem. It can also be used for educational purposes (including laboratory classes) for students at the undergraduate level.

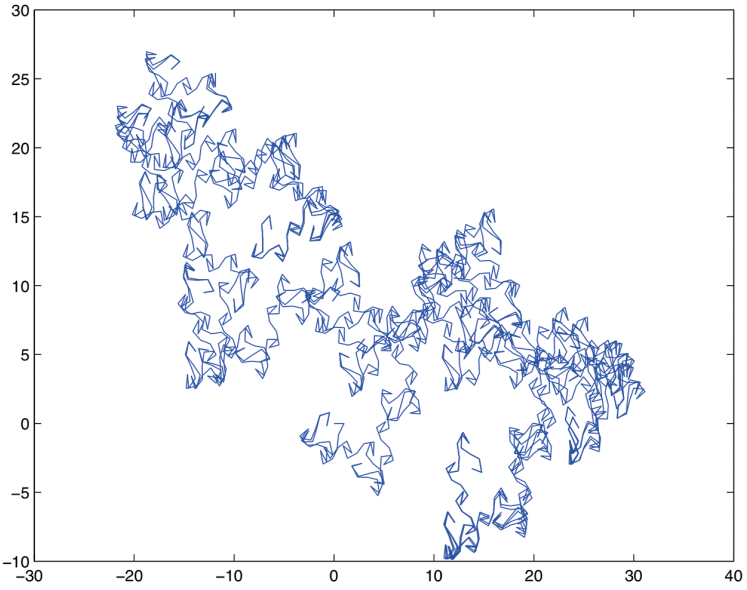


Fig. 38.1. Graph of the curlicue (first 2000 points) on the complex plane

Methods

The figures and datasets were obtained by programs in Matlab based on the definitions and information provided in (Signerska-Rynkowska, 2020). The source-codes for the programs, also attached, can be reused and easily modified to compute other quantities (e.g. the curlicue's diameter).

Data quality and availability

Dataset DOI
10.34808/anjg-q802

Dataset License
CC-BY

References

Berry, M.V. and Goldberg, J. (1988) 'Renormalisation of curlicues', *Nonlinearity*, 1(1), pp. 001–26, DOI: 10.1088/0951-7715/1/1/001.

Dekking, F.M. and Mendès-France, M. (1981) 'Uniform distribution modulo one: a geometrical viewpoint', *Journal für die Reine und Angewandte Mathematik*, 1981(329), pp. 143-153, DOI: 10.1515/crll.1981.329.143.

Signerska-Rynkowska, J. (2022) 'Curlicues generated by circle homeomorphisms', *Geometriae Dedicata*, 216(2), 15, DOI: 10.1007/s10711-022-00676-9.

Sinai, Y.G. (2008) 'Limit theorem for trigonometric sums. Theory of curlicues', *Russian Mathematical Surveys*, 63(6), pp. 1023–1029. DOI: 10.1070/RM2008v063n06ABEH004577.

Data Descriptor

Video of LEGO Bricks on Conveyor Belt Dataset Series

Tomasz Boiński[1*]

[1] Faculty of Electronics, Telecommunication and Informatics, Gdańsk University of Technology (11/12 Gabriela Narutowicza Street, Gdańsk, Poland)

[*] Correspondence author: tomasz.boinski@pg.edu.pl; ORCID: 0000-0001-5928-5782

Abstract

The dataset series titled Video of LEGO bricks on conveyor belt is composed of 14 datasets containing video recordings of a moving white conveyor belt. The recordings were created using a smartphone camera in Full HD resolution. The dataset allows for the preparation of data for neural network training, and building of a LEGO sorting machine that can help builders to organise their collections.

Keywords: LEGO, videos, neural networks

https://doi.org/10.34808/x55q-sz53_dyr_roz39

Specification table (data records)

Subject area	Computer and information, Information and communication technology
More specific subject area	Video recordings
Type of data	Video
How the data was acquired	The data was collected at the Gdańsk University of Technology using a commercially available phone camera (Huawei P20 Pro) using a conveyor belt and stand made of LEGO bricks powered by the LEGO Boost set
Data format	The videos are in .mp4 format
Experimental factors	The data contained in the dataset were not processed
Experimental features	The recordings were made with natural light facing the front of the conveyor belt

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes. Derivative works need to be shared using the same conditions

Background

LEGO bricks are extremely popular and allow almost any type of construction to be built. This can be done thanks to the multiple shapes available. Utilising the full potential of construction possibilities however requires proper arrangement of the bricks. The usual sorting, for average collections, is done by shape, as the colours and decals can be easily distinguished even in a big container of bricks (Alphin, T., 2020). This cannot be said about the shape, and the basic problem is to find a brick of the required shape and size. With over 3700 different LEGO parts (Maren, T., 2018), the ability to find a given part is crucial in everyday constructing, especially for very large LEGO collections, where shuffling through large boxes of mixed bricks is not acceptable.

Unfortunately there is no simple solution for brick sorting. LEGO Group provides only a simple sorting mechanism based on the brick size in the form of LEGO Sort and Store which was released in 2011. Other solutions usually rely on optimising the process of sorting rather than automating it (e.g. Adam, 2017). With the constant development of deep neural networks and the performance of modern computers, it is possible to construct a viable sorting device using the bricks themselves or 3D-printed elements. Independently of the method of building the sorting machine, it requires a well-trained neural network able to distinguish between different, often very similar, bricks, or at least divide them into smaller categories, allowing further manual selection of proper bricks. To date, few projects have taken this approach, with the notable example of (West, D., 2019). The approach is based on images generated using the LDraw library (https://www.ldraw.org/), containing 3D models of every brick available.

Our dataset series, Video of LEGO bricks on conveyor belt, took another approach and has been designed to help gather the data necessary to train a neural network to recognise different LEGO brick shapes using real LEGO pictures. Automatically generated LEGO images, using e.g. the LDraw library, usually are not realistic enough to perform proper training, and usually such images are just too perfect to be used in real applications without the noise introduction and additional augmentation. A neural network trained with such images has difficulties detecting e.g. blurred images taken with a camera. At the time of writing, the datasets in the series contain recordings of the 267 most common LEGO brick shapes in multiple, random positions and in multiple colours.

Methods

The aim of the data gathered was to train a neural network to identify LEGO bricks moving on a conveyor belt. The aim was to use home-available devices during the sorting,

thus for data capturing, we used a smartphone camera and a LEGO-based machine and camera stand was build (Fig. 39.1) to mimic the pre-sorting part of the machine, where bricks will be separated and recorded. All videos were recorded using a Huawei P20 Pro smartphone located 14.5 cm above the belt. The recordings were created with 1920x1080 resolution, H.264 high profile codec, 15 frames per second and natural sunlight coming from a window to limit the amount of shadows and greyness of the background.

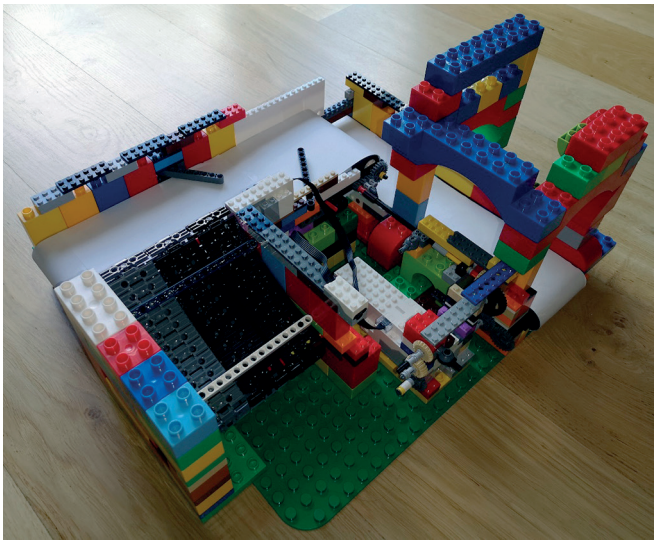


Fig. 39.1. Stand and conveyor belt for video recording. The camera was placed on the stands in the right side of the construction

Data quality and availability

All recordings were created using the default settings of the Huawei P20 Pro smartphone camera, at 1920x1080 resolution and encoded with the H.264 high profile codec at 15 frames per second. The conveyor belt was lit by the natural light coming from a window on the same side that the camera was placed. This way, the recordings tend to be as similar to ones that will be potentially taken during the sorting of LEGO bricks.
Series:
https://mostwiedzy.pl/pl/open-research-data-series/niema,202011132226557715481-0/catalog

Datasets DOI
10.34808/43yh-ck53
10.34808/pw5b-hr02
10.34808/mhdg-9e26
10.34808/5gwb-5y67
10.34808/kn9k-nf31
10.34808/0zhs-wp56

10.34808/9sg4-7992
10.34808/ydqy-jp64
10.34808/qjxa-df20
10.34808/gf9v-q230
10.34808/qt92-qj93
10.34808/f7xv-nz92
10.34808/1v0v-0j40
10.34808/f5nc-7x81
10.34808/43yh-ck53

References

Adam (2017) '*LEGO Sorting chart*'.

Alphin, T. (2020) '*The LEGO Storage Guide*', Available at: https://brickarchitect.com/guide/ (Accessed: 6th January 2021).

Maren, T. (2018) '*60 Fun LEGO Facts Every LEGO Fan Needs to Know*', Mama in The Now, Available at: https://mamainthenow.com/fun-lego-facts/ (Accessed: 5th January 2021).

West, D. (2019) '*LEGO sorting machine*', Available at: https://twitter.com/JustASquid/status/1201959889943154688 (Accessed: 5th January 2021).

Data Descriptor

Regeneration Project of Market Places GOSPOSTRATEG – "Polanki" Market in Gdańsk-Oliwa Pilot Project Monitoring Dataset

Justyna Borucka[1*]

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

* Correspondence author: justyna.borucka@pg.edu.pl; ORCID: 0000-0003-1591-0149

Abstract

The dataset entitled Monitoring of activities carried out as part of prototyping and implementation of the pilot project in the area of the "Polanki" market and its direct neighbourhood, in the Gdańsk-Oliwa district, step1; stage from July 2020 year contains tabular monitoring lists (quantitative and qualitative documentation report in the form of tables) of activities carried out as part of the prototyping and implementation of the pilot project in the area of the Polanki market and its immediate vicinity in the Gdańsk-Oliwa district conducted during the month of July 2020. The data collection was carried out as part of the first step of monitoring in the series: "Monitoring of activities, prototyping and implementation of the pilot project in the area of the "Polanki" market, Gdańsk-Oliwa district, step1". The second step of monitoring is scheduled for 2021. Monitoring is the basis for the subsequent analysis used in the implementation of the pilot project of the "Polanki" Market and supports the proper development of the regeneration project of the marketplace.

Keywords: architecture, urban design, market place, monitoring works, pilot project, project testing, urban renewal, urban regeneration, revitalisation of public space, placemaking

https://doi.org/10.34808/x55q-sz53_dyr_roz40

Specification table (data records)

Subject area	Architecture and Town Planning
More specific subject area	Urban Renewal, Integrated Urban Regeneration, Revitalisation of Public Spaces, Placemaking

Type of data	Set of monitoring cards: incl. text, tables, drawings, photos
How the data was acquired	The data was collected at the Gdańsk University of Technology during regular field visits with direct observation, environmental interviews in the market and documented with notes, freehand drawings and photos
Data format	The data are in .pdf format
Experimental factors	The data contained in the dataset has been processed. The data was reproduced with the use of digital tools available on the market and systematised in the form of monitoring cards
Experimental features	The preliminary, newly designed spatial layout and the organisation of the market were subjected to observations and testing
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Public markets are significant points in the system of public space in modern cities. They are one of the city-defining elements in their traditional sense. In Europe, the understanding of the role of the market in the contemporary city, as well as that of the market communities, has changed. Markets are perceived not only as commercial spaces, but also as cultural spaces. Undoubtedly, public markets play an important role in the socio-economic structure of modern cities. (The Local Enterprise Partnership for London, 2017; Whyte, 2013) What is more, cities, their districts, different public spaces including public markets have become brands (Bennett and Savani, 2003).

The situation is completely different in Poland. Even though markets are important spots of the public sphere and one of the tasks of local governments, according to Polish law, is to maintain public markets as available public spaces, Polish markets are in crisis. Although their number is not decreasing and the economic situation remains stable, their standard and size are weakening and profits are falling. (Bieszk-Stolorz and Felsztyńska, 2018; Czyż and Hanzel, 2018). The goal of the GOSPOSTRATEG project is to develop a comprehensive, operational strategy for socio-economic activation, enriching the cultural offer, transforming the program concept and improving the aesthetic condition of degraded exhibition areas in Poland. The project involves a pilot implementation and monitoring of socio-economic results resulting from the implementation of architectural, artistic and marketing interventions of the existing, representative, operating market facility in Gdańsk. The planned strategic result of the project is the initiation of a new model of revitalisation of facilities of this type by developing the strategy in question. The analysis of the research material obtained from the implemented pilot project will allow a strategy of activities related to the socio-economic activation of these areas to be de-

veloped and the improvement of their functioning in contemporary Polish cities. (Social Catalyst of Entrepreneurship Gospostrateg Project, 2018)

With regard to this project, the role of the research team was to investigate and diagnose the condition of the market case study and monitor the implementation of the proposed, new design of the market as a model study to build a future regeneration strategy.

Our dataset, Monitoring of activities carried out as part of prototyping and implementation of the pilot project in the area of the "Polanki" market and its direct neighbourhood, in the Gdańsk-Oliwa district, step1; stage from July 2020 year has been designed to support testing of the regeneration project of the marketplace and to help to understand how important it is to choose an appropriate spatial and functional organisation and methods of participation due to differences in the behaviour of space users, the location of the public space itself in the historical part of the city and the particular situation of restrictions (due to the COVID-19 pandemic). It also helped to control the implementation and verify all deviations from the initial design plan and to correct it on a regular basis. The dataset contains tabular monitoring lists (quantitative and qualitative documentation report in the form of tables) of activities carried out as part of the prototyping and implementation of the regeneration of the marketplace pilot project.

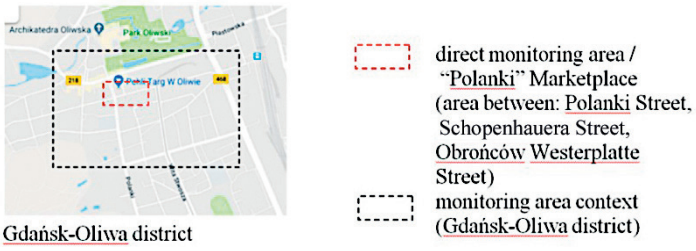


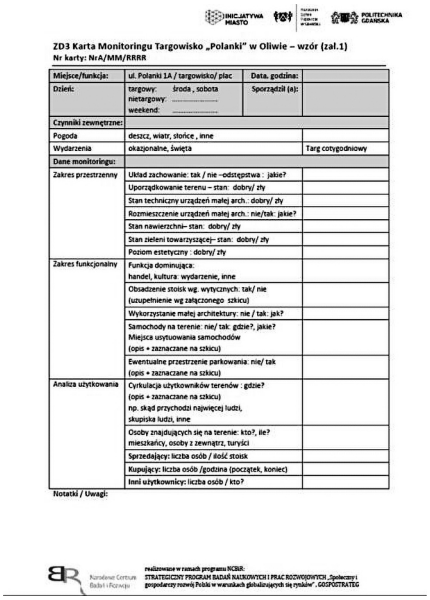
Fig. 40.1. Monitoring area location and its context in the Gdańsk-Oliwa district, source of the map: googlemaps.com

The subject of the study is the monitoring of activities carried out under the prototyping and implementation of the GSOPOSTRATEG pilot project in the area of the "Polanki" market and its immediate vicinity in the Gdańsk-Oliwa district (Fig. 40.1). The data collection was carried out as part of the first step of monitoring in the series: "Monitoring of activities, prototyping and implementation of the pilot project in the area of the "Polanki" market, Gdańsk-Oliwa district, step1". The second step of monitoring is scheduled for 2021. Monitoring is the basis for the subsequent analysis used in the implementation of the pilot project of the "Polanki" Market and supports the proper development of the regeneration project of the market place.

Methods

The monitoring of the Market covers the spatial, functional scope and analysis of use by residents and other users of the space. Quantitative and qualitative monitoring was

carried out systematically from July 2020 to November 2020 (earlier monitoring was not possible due to the introduction of restrictions related to the COVID 19 pandemic). For this particular dataset, the monitoring covers the month of July 2020, when a special organisation and design plan of the market space, respecting the COVID-19 pandemic situation and restrictions, was proposed. Monitoring was carried out through in situ visits (regular field visits with direct observation, environmental interviews in the market and documented with notes, freehand drawings and photos) with the use of the Monitoring Card (Fig. 40.2, 40.3a, 40.3b) for subsequent systematisation and analysis of information collected on their basis.



ZD3 Karta Monitoringu Targowisko „Polanki" w Oliwie – wzór (zał.1)
Nr karty: NrA/MM/RRRR

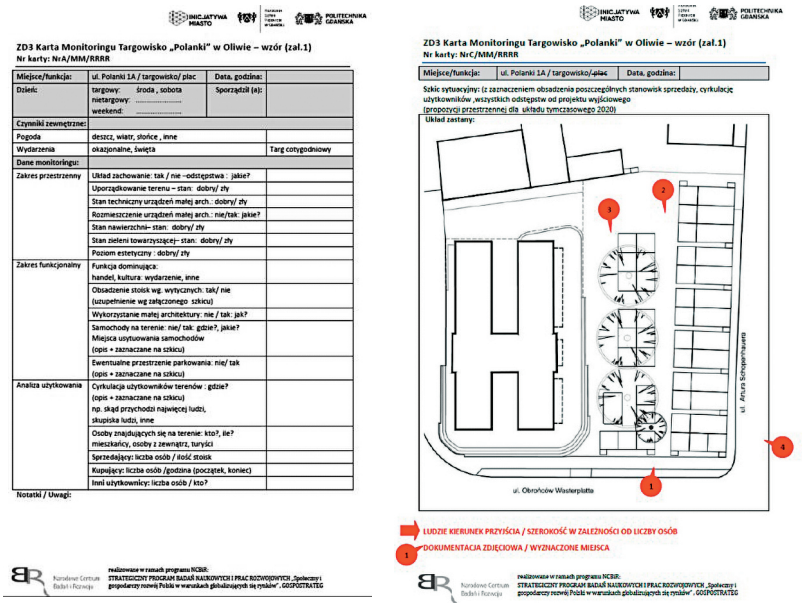
Miejsce/funkcja:	**ul. Polanki 1A / targowisko/ plac**	**Data, godzina:**	
Dzień:	**targowy: środa , sobota** **nietargowy: .....** **weekend: .....**	**Sporządził (a):**	
Czynniki zewnętrzne:			
Pogoda	**deszcz, wiatr, słońce , inne**		
Wydarzenia	**okazjonalne, święta**		**Targ cotygodniowy**
Dane monitoringu:			
Zakres przestrzenny	**Układ zachowanie: tak / nie –odstępstwa : jakie?**		
	Uporządkowanie terenu – stan: dobry/ zły		
	Stan techniczny urządzeń małej arch.: dobry/ zły		
	Rozmieszczenie urządzeń małej arch.: nie/tak: jakie?		
	Stan nawierzchni– stan: dobry/ zły		
	Stan zieleni towarzyszącej– stan: dobry/ zły		
	Poziom estetyczny : dobry/ zły		
Zakres funkcjonalny	**Funkcja dominująca:** **handel, kultura: wydarzenie, inne**		
	Obsadzenie stoisk wg. wytycznych: tak/ nie **(uzupełnienie wg załączonego szkicu)**		
	Wykorzystanie małej architektury: nie / tak: jak?		
	Samochody na terenie: nie/ tak: gdzie?, jakie? **Miejsca usytuowania samochodów** **(opis + zaznaczane na szkicu)**		
	Ewentualne przestrzenie parkowania: nie/ tak **(opis + zaznaczane na szkicu)**		
Analiza użytkowania	**Cyrkulacja użytkowników terenów : gdzie?** **(opis + zaznaczane na szkicu)** **np. skąd przychodzi najwięcej ludzi,** **skupiska ludzi, inne**		
	Osoby znajdujących się na terenie: kto?, ile? **mieszkańcy, osoby z zewnątrz, turyści**		
	Sprzedający: liczba osób / ilość stoisk		
	Kupujący: liczba osób /godzina (początek, koniec)		
	Inni użytkownicy: liczba osób / kto?		
Notatki / Uwagi:			

realizowane w ramach programu NCBiR:
STRATEGICZNY PROGRAM BADAŃ NAUKOWYCH I PRAC ROZWOJOWYCH „Społeczny i gospodarczy rozwój Polski w warunkach globalizujących się rynków". GOSPOSTRATEG

Fig. 40.2. Table for the monitored data from Monitoring Card "Polanki" Market in Gdańsk-Oliwa district – model

The monitoring documentation was supplemented with photo and drawing documentation. Each monitoring card includes the 4 following photographs:

- photographs taken in designated places (4 points described below)
  Photo 1: entrance overlooking the main alley;
  Photo 2: at the end of the alley overlooking the entrance;
  Photo 3: from the side of the passage at the lighthouse between the building and the trees, the view towards Obrońców Westerplatte Street;
  Photo 4: from the corner of Schopenhauera Street + Westerplatte site,
- all places marked on the map on the Monitoring card.
  All photos were inserted into the table and any additional drawings or sketches in the table as a supplement visualization were possible.


ZD3 Karta Monitoringu Targowisko „Polanki" w Oliwie – wzór (zał.1)
Nr karty: Nr A/MM/RRRR

Miejsce/funkcja:	ul. Polanki 1A / targowisko/ plac	Data, godzina:	
Dzień:	targowy: środa , sobota nietargowy: ..... weekend:	Sporządził (a):	
Czynniki zewnętrzne:			
Pogoda	deszcz, wiatr, słońce , inne		
Wydarzenia	okazjonalne, święta	Targ cotygodniowy	
Dane monitoringu:			
Zakres przestrzenny	Układ zachowanie: tak / nie –odstępstwa : jakie?		
	Uporządkowanie terenu – stan: dobry/ zły		
	Stan techniczny urządzeń małej arch.: dobry/ zły		
	Rozmieszczenie urządzeń małej arch.: nie/tak: jakie?		
	Stan nawierzchni– stan: dobry/ zły		
	Stan zieleni towarzyszącej– stan: dobry/ zły		
	Poziom estetyczny : dobry/ zły		
Zakres funkcjonalny	Funkcja dominująca: handel, kultura, wydarzenie, inne		
	Obsadzenie stoisk wg. wytycznych: tak/ nie (uzupełnienie wg załączonego szkicu)		
	Wykorzystanie małej architektury: nie / tak: jak?		
	Samochody na terenie: nie/ tak: gdzie?, jakie? Miejsca usytuowania samochodów (opis + zaznaczane na szkicu)		
	Ewentualne przestrzenie parkowania: nie/ tak (opis + zaznaczane na szkicu)		
Analiza użytkowania	Cyrkulacja użytkowników terenów : gdzie? (opis + zaznaczane na szkicu) np. skąd przychodzi najwięcej ludzi, skupiska ludzi, inne		
	Osoby znajdujących się na terenie: kto?, ile? mieszkańcy, osoby z zewnątrz, turyści		
	Sprzedający: liczba osób / ilość stoisk		
	Kupujący: liczba osób /godzina (początek, koniec)		
	Inni użytkownicy: liczba osób / kto?		

Notatki / Uwagi:

realizowane w ramach programu NCBiR:
STRATEGICZNY PROGRAM BADAŃ NAUKOWYCH I PRAC ROZWOJOWYCH „Społeczny i gospodarczy rozwój Polski w warunkach globalizujących się rynków". GOSPOSTRATEG

ZD3 Karta Monitoringu Targowisko „Polanki" w Oliwie – wzór (zał.1)
Nr karty: Nr C/MM/RRRR

Miejsce/funkcja:	ul. Polanki 1A / targowisko/~~plac~~	Data, godzina:	

Szkic sytuacyjny: (z zaznaczeniem obsadzenia poszczególnych stanowisk sprzedaży, cyrkulację użytkowników ,wszystkich odstępstw od projektu wyjściowego (propozycji przestrzennej dla układu tymczasowego 2020)

Układ zastany:

ul. Obrońców Westerplatte

ul. Artura Schopenhauera

LUDZIE KIERUNEK PRZYJŚCIA / SZEROKOŚĆ W ZALEŻNOŚCI OD LICZBY OSÓB

DOKUMENTACJA ZDJĘCIOWA / WYZNACZONE MIEJSCA

realizowane w ramach programu NCBiR:
STRATEGICZNY PROGRAM BADAŃ NAUKOWYCH I PRAC ROZWOJOWYCH „Społeczny i gospodarczy rozwój Polski w warunkach globalizujących się rynków". GOSPOSTRATEG

Fig. 40.3a, 40.3b. Output drawings for marking important data (observations and deviations from the initial design, plan) from Monitoring Card "Polanki" Market in Gdańsk-Oliwa district – model

The data collection covered individual issues in the field of:

1. Spatial development:
   Behavior arrangement – there is / there is no deviation
   Cleanliness of the area – good / bad condition
   Technical condition of small architecture devices, arrangement
   Condition of the surface and accompanying greenery
   Aesthetic level
2. Functional development
   Stands are allocated according to the guidelines
   Use of street furniture
   Location of cars
   Possible parking spaces
3. Resident usage analysis (counting, photos)
   Circulation of users of the space (where most people come from, people clusters, spaces of gathering and communication)

Number of people in the area divided into: the number of buyers on fair, non-fair and weekend days, the number of visitors to the market on fair days, non-fair days and on weekends in two time spaces (two hours for monitoring at the beginning and end) Counting was conducted for a given hour to check the number of people in the market with the division for buyers and other users;

Archiving and digitisation took place immediately after the monitoring visit, however the scrapbook of cards was left for documentation and archives.

Data quality and availability

All data was collected following the monitoring guidelines / instruction for two separate functions of the public space: Marketplace and Open Public Space Event.

For the function of Marketplace, the monitoring was conducted weekly, on Saturdays and Wednesdays, according to a schedule (following schedule of the monitoring visits) between 9–10 a.m. once during the event (or 1 per month at the beginning and end of the market).

For the function of Open Public Space Event, the monitoring was conducted only for the event on particular weekends (according to the schedule of the events) and twice during the event (at the beginning and end of the event). Unfortunately, due to pandemic situation, not all events programmed in the GOSPOSTRATEG projects schedule were allowed to be organised. Nevertheless, the data contained in the dataset allow for a fair comparison of all situations in the area serving as a marketplace and public open space.

Dataset DOI
10.34808/yn27-gb21

Acknowledgements
The generation of this dataset is strictly linked to and supported by a research project entitled: The public markets revitalisation strategy with the use of the social catalyst entrepreneurship method, brand repositioning and placemaking as a tool for local development policy (in Polish: Strategia rewitalizacji obiektów handlu targowego z wykorzystaniem metody społecznego katalizatora przedsiębiorczości, repozycjonowania marki oraz placemakingu jako narzędzie polityki lokalnej) project number: Gospostrateg1/392278/6/NCBR/2018, conducted within the framework of the GOSPOSTRATEG program financed by the National Centre for Research and Development in Poland. The monitoring was carried out by a group of researchers from the Faculty of Architecture, Gdańsk University of Technology, directly involved in the project: arch. J. Borucka PhD, arch. W. Mazurkiewicz PhD, arch. D. Nalęcz.

References

Bennett, R. and Savani, Sh. (2003) 'The Rebranding of City Places: An International Comparative Investigation', *International Public Management Review*, 4(2), pp. 70–87. Availaible at: https://journals.sfu.ca/ipmr/index.php/ipmr/article/view/217/217 (Accessed: 16th December 2020).

Bieszk-Stolorz, B. and Felsztyńska, I. (2018) 'The analysis of attractiveness of large cities with respect to the development of the marketplace trade' (Polish title: Analiza atrakcyjności

dużych miast pod względem rozwoju handlu targowiskowego), *Studia i Prace WNEiZ US*, 54 (3), pp. 407–418. DOI: 10.18276/SIP.2018.54/3-30.

Czyż, P. and Hanzel, P. (2018) 'Traditional Market Areas in the Context of Revitalization – Issues Overwiev' (in Polish: Przestrzenie Handlu Tradycyjnego w Kontekście Procesu Rewitalizacji – Szkic Problematyki) in Lorens, P. (ed.) *Urban Aspects of Cities` Transformation* (in Polish: Urbanistyczne Aspekty Transformacji Miast), Studia KPZK PAN, 192, Warszawa: KPZK PAN, Wydawnictwo Politechniki Gdańskiej, pp. 315-324. Availaible at: https://journals.pan.pl/dlibra/publication/129845/edition/113348/content/studia-kpzk-2018-tom-192-urbanistyczne-aspekty-transformacji-miast-traditional-market-areas-in-the-context-of-revitalization-issues-overwiev-czyz-piotr-henzel-patrycja?language=en (Accessed: 26th May 2022).

'*Social Catalyst of Entrepreneurship - The public markets revitalisation strategy with the use of the social catalyst entrepreneurship method, brand repositioning and placemaking as a tool for local development Policy Gospostrateg Project*' (Polish title: „Społeczny Katalizator Przedsiębiorczości (SKP) – Strategia rewitalizacji obiektów handlu targowego z wykorzystaniem metody społecznego katalizatora przedsiębiorczości, repozycjonowania marki oraz placemakingu jako narzędzie polityki lokalnej") (2018), Availaible at: https://mostwiedzy.pl/pl/project/strategia-rewitalizacji-obiektow-handlu-targowego-z-wykorzystaniem-metody-spolecznego-katalizatora-p,643-1 (Accessed: 4th January 2021).

'*Social Catalyst of Entrepreneurship Gospostrateg Project*' (Polish title: Społeczny Katalizator Przedsiębiorczości) (2018). Availaible at: http://inicjatywamiasto.pl/portfolio/spoleczny-katalizator-przedsiebiorczosci/ (Accessed: 7th July 2020).

Strategic program 'Social and economic development of Poland in the conditions of globalizing markets - GOSPOSTRATEG' from National Centre for Research and Development in Poland, Availaible at: https://www.ncbr.gov.pl/programy/programy-strategiczne/gospostrateg/ (Accessed: 7th July 2020).

The Local Enterprise Partnership for London (2017) '*Understanding London's Markets*'. London: Greater London Authority City Hall, p.4. Availaible at: https://www.london.gov.uk/sites/default/files/20171219_gla_markets_report_short_web.pdf (Accessed: 14th July 2020).

Whyte, W.H. (2013) '*The Social Life of Small Urban Spaces*', (2nd Edn), New York: Project for Public Spaces.

Data Descriptor

Description of the Dataset Hanow – *Praecepta de Arte Disputandi* – Transcription and Photographs

Jacek Pokrzywnicki[1*]

[1] Institute of Classical and Slavonic Studies, Classical Philology Division, University of Gdańsk (51 Wita Stwosza Street, Gdańsk, Poland)

* Correspondence author: jacek.pokrzywnicki@ug.edu.pl; ORCID: 0000-0002-4765-8577

Abstract

This article briefly characterises the "Hanow – Praecepta de arte disputandi – transcription and photographs" research dataset. The dataset was created based on photographs and transcriptions of the manuscript of the Latin lectures on the rules of effective discussion (the title of the manuscript: Praecepta de arte disputandi) by Michael Christoph Hanow (1695–1773), professor of Gdańsk Academic Gymnasium. The original document is held in the collection of the Gdańsk Library of the Polish Academy of Sciences. The collected research material can be used for studying the practical application of the rhetorical theory in the life of the old Gdańsk urban community.

Keywords: Hanow, Michael Christoph; *Praecepta de arte disputandi*, transcription, Latin, rhetoric

https://doi.org/10.34808/x55q-sz53_dyr_roz41

Specification table (data records) – "Description of the dataset Hanow – *Praecepta de arte disputandi* – transcription and photographs"

Subject area	Humanities
More specific subject area	Neo-Latin studies
Type of data	Text, photographs
How the data was acquired	transcription, photography
Data format	txt, docx, jpg
Experimental factors	–

Experimental features	The original spelling was standardised based on modern Latin dictionaries; the text is a transcription of an original manuscript; the Latin language used by Hanow is typical, both in terms of vocabulary and syntax, for the scholarly milieu from the early modern Republic of Letters
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	16.12.2019 (version 1.0); 30.03.2021 (version 1.1)

Background

The research dataset "Hanow – Praecepta de arte disputandi – transcription and photographs" was created as part of my own research on the application of rhetoric principles in the life of the urban community of eighteenth-century Gdańsk.

"Praecepta de arte disputandi" is a hand-held record of one of the three lectures presented in 1754 at the Gdańsk Academic Gymnasium by Michael Christoph Hanow (1695–1773). This versatile scholar gave lectures on philosophy in this school between the 30s and the 70s of the 18th century.

"Praecepta", together with the other two lectures on logic and metaphysics, form a kind of course of philosophy, by the standards of that era, modern in terms of didactics. In these lectures, Michael Christoph Hanow referred, directly or indirectly, to the textbooks of Christian Wolff (1679–1754), one of the most popular philosophers of the Enlightenment era.

"Precepta de arte disputandi" also presents content important for rhetoric. In these lectures, Michael Christoph Hanow tried to show the audience how to build their own persuasive potential and how to effectively convince opponents during the discussion. This is what made me include this text in my own research on the presence of rhetoric in the life of the urban community of eighteenth-century Gdańsk.

The manuscript of these lectures by Michael Christoph Hanow is part of the collection of manuscripts of the Polish Academy of Sciences Gdańsk Library (reference number: Ms.512).

Methods

Building this dataset required photographing the whole manuscript, which was not created by Hanow himself, but by his student, Constantin Ernst Grodeck (1735–1774), a member of a prominent merchant family from Gdańsk. I also transcribed the photographed manuscript for the dataset. In the transcription, I standardised the spelling to match that recommended by the Latin dictionaries currently in use. I also decided to remove some of the punctuation marks. As a rule, I left those punctuation marks that reflect the syntax boundaries of individual parts of the compound-complex sentences. I did not pay attention to the rhetorical structure of the utterances. I assumed that the

rhetorical system of punctuation, with the division of higher-order structures, that is, compound-complex sentences, into finer parts, i.e. "protasis", "apodosis" or "kolon" and "komma", can be seen in the attached photographs. However, I have to note that the transcription is not a closed matter and should not be treated as an edition of the text which respects the rules of Neo-Latin editing.

All photos have been pre-edited (cutting and sharpening some shots). The photographs were taken with a digital camera. Text files were created using popular text applications (in the Windows environment: MS Word, Notepad).

Data records

The resolution of the photograph (JPG file format) is good enough to get a magnification that allows for fairly comfortable reading of most of the textual original. I also included text files (extension: txt and docx) in the dataset with the transcription of the Latin manuscript.

The dataset includes 50 photo files (resolution from 0.88 MB to 2 MB; total: 63.8 MB) and 2 text files (0.05 MB and 0.06 MB).

Data quality and availability

The verification of the data took place both during the acquisition of the photographic material (the quality of the photograph was verified; it was essential to obtain satisfactory focus). The text files in turn were examined by reading the transcript several times and comparing it with the photo material. In a few passages, the manuscript notation seemed incomprehensible to me. I marked these omissions in the text. Another edition of the text is planned, in which the current gaps will be filled.

Dataset DOI
10.34808/01f6-n909

Acknowledgements
Photographs for the dataset were taken thanks to the courtesy of the Head of the Special Collections Department of the Gdańsk Library of the Polish Academy of Sciences.

References

'*Allgemeine Deutsche Biographie: Auf Veranlassung Seiner Majestät des Königs von Bayern hrsg. durch die historische Commission bei der Königl. Akademie der Wissenschaften*' (1875). Leipzig: Verlag von Duncker & Humblot.

Brodnicki, M. (2012) '*Nauczanie filozofii w Gdańskim Gimnazjum Akademickim do połowy XVII wieku*'. Gdańsk: Wydawnictwo Athenae Gedanenses.

Grześkowiak-Krwawicz, A. (1998) '*Gdańsk oświecony: Szkice o kulturze literackiej Gdańska w dobie Oświecenia = Die Aufklärung in Danzig*'. Warszawa: IBL.

Hanow, M.C. (1765) '*Potiores Philosophiae Recentioris Controversiae, publice expositae in Athenaeo Gedanensi*', Leipzig: Junius, Joachim Friedrich. Available at: https://books.google.pl/books?id=8G5NAAAAcAAJ&hl=pl&pg=PP3#v=onepage&q&f=false (Accessed: 13th January 2021).

Krollmann, C. (ed.) (1941-) '*Altpreußische Biographie im Auftrage der historischen Kommission für ost- und westpreußische Landesforschung*'. Königsberg (Pr.): Gräfe und Unzer Verlag, Marburg/Lahn.

Nowak, Z. (ed.) (1994). Słownik biograficzny Pomorza nadwiślańskiego. Gdańsk: Gdańskie Tow. Naukowe (2).

Weichbrodt-Tiedemann von, D. (1993) '*Patrizier, Bürger, Einwohner der Freien und Hansestadt*', Danzig. Klingberg.

Data Descriptor

Description of the Dataset Rhetoric at School – a Selection of the Syllabi from the Academic Gymnasium in Gdańsk – Transcription and Photographs

Jacek Pokrzywnicki[1*]

[1] Institute of Classical and Slavonic Studies, Classical Philology Division, University of Gdańsk (51 Wita Stwosza Street, Gdańsk, Poland)

* Correspondence author: jacek.pokrzywnicki@ug.edu.pl; ORCID: 0000-0002-4765-8577

Abstract

The research dataset described in the article was based on photographs and transcription of a textual record from Latin syllabi for classes at the Gdańsk Academic Gymnasium. The syllabi concern the years 1645/1648/1652/1653. The original document is held in the collection of the Gdańsk Library of the Polish Academy of Sciences [reference number: Ma 3920 8o]. The collected research material can be used for studying the practical use of persuasion art by the members of the old Gdańsk urban community.

Keywords: Syllabi, Catalogus lectionum, transcription, Latin language, rhetoric, Academic Gymnasium, Gdańsk

https://doi.org/10.34808/x55q-sz53_dyr_roz42

Specification table (data records)

Subject area	Humanities
More specific subject area	Neo-Latin studies
Type of data	Text, photographs
How the data was acquired	transcription, photography
Data format	txt, docx, jpg, xlsx

Experimental factors	–
Experimental features	The original spelling was unified based on modern Latin dictionaries; the text is a transcription of printed syllabi published in the 17th century; the Latin from the syllabi is supposedly the one used by the professors of the Academic Gymnasium in Gdańsk; the language is typical, both in terms of vocabulary and syntax, for the Gdańsk scholarly milieu of the time
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	31.12.2020

Background

The research dataset "Rhetoric at school – selection of syllabi from the Academic Gymnasium in Gdańsk – transcription and photographs" was created as part of my research on rhetoric in the urban community of eighteenth-century Gdańsk. All of the described syllabi were published in one binding and form part of the old prints collection of the Gdańsk Library of the Polish Academy of Sciences (reference number: Ma 3920 8o; order of prints in the binding: 106, 107, 108, 109).

Methods

While the texts of the syllabi included in this dataset date back to the mid-17th century (more precisely the years: 1645, 1648, 1652, 1653), I collected them (i.e. the texts) as a comparative material for my research on the practical dimension of the art of persuasion in the life of the urban community of Gdańsk in the eighteenth century.

I built this dataset photographing particular entries by specific professors of the Gdańsk Academic Gymnasium (e.g. Peter Oelhaf, Johann Mochinger, Heinrich Nicolai, Johann Raue, Georg Neufeld, Johann Maukisch, Daniel Lagus, Michael Falck, Johann Peter Titz). I also transcribed photographed fragments of the syllabi. The spelling was standardised in the transcription. I applied words from the Latin dictionaries currently in use (e.g. A Latin Dictionary, ed. by Ch. Lewis, Ch. Short, and Słownik łacińsko-polski, ed. by J. Korpanty, etc.). I tried to eliminate some punctuation marks. I left those that reflect the syntax boundaries of particular parts in compound-complex sentences, or facilitate the reading process thanks to separation of e.g. the construction of ablativus absolutus. I did not pay too much attention to the rhetorical structure of the utterances. The rhetorical punctuation system can be seen in the attached photographs. Importantly, this transcription should not be regarded as an edition of a text, since it (i.e. the transcription) does not respect all of the rules used in the editing of Neo-Latin texts.

All photos have been pre-edited (cutting and sharpening some shots). The photographs were taken with a digital camera. Text files were created using popular text ap-

plications (in the Windows environment: MS Word, Notepad). A special folder contains files with individual entries by specific professors. The dataset also contains a file containing all of the transcribed entries. In addition, the user will find an information file (formats: csv, xlsx), which facilitates navigating among the transcribed professorial entries. It contains: the number of the print in the volume, the number of the photograph file, the year of the publication, printed or default signature of the sheet or the catchword, and references to the name of the author of the entry or title page and number of the file (all title pages and professor entries separately).

Data records

The resolution of the photograph (JPG file format) is good enough to get a magnification that allows for fairly comfortable reading of most of the textual original. I also included text files (extension: txt and docx) in this dataset with the transcription of the Latin manuscript and the information file in csv, xlsx and txt formats. This amounts to: 30 txt files (0.04 MB), 18 photograph files (resolution of 1.48 MB to 3 MB; total: 44.5 MB) and 2 all-in-one text files, i.e. the transcription of all entries (0.05 MB and 0.04 MB), as well as 3 information files (names: readme_a_Ma3920_8o.txt, readme_b_Ma3920_8o.csv, readme_b_Ma3920_8o.xlsx).

Data quality and availability

I was verifying the data during the acquisition of the photo material, as well as during its development. I evaluated the quality of the photographs: the purpose being to obtain a satisfactory focus of the photograph. The text files in turn were verified by reading the transcript several times and comparing it with the photo material.

Dataset DOI
10.34808/823g-1y16

Acknowledgements
Photographs for the dataset were taken thanks to the courtesy of the Head of the Special Collections Department of the Gdańsk Library of the Polish Academy of Sciences.

References

Gierszewski, S. and Nowak, Z. (eds.) (1997) '*Słownik biograficzny Pomorza Nadwiślańskiego*' (4 vols., 3 supp.). Gdańsk: Wydawnictwo Gdańskie.

Korpanty, J. (ed.) (2001-2003) '*Słownik łacińsko-polski*' (2 vols). Warszawa: Wydawnictwo Szkolne PWN.

Krollmann, C. (ed.) (1967) '*Altpreußische Biographie, hrsg. im Auftrage der Historischen Kommision* für *ost-und werstpreußische Landesforschung*' (4 vols). Marburg-Lahn: N.G. Elwert Verlag.

Lewis, C.T. and Short, C. (1987) '*A Latin dictionary: Founded on Andrews' ed. of Freund's Latin dictionary*'. Rev., enl., and in great part rewritten. Oxford: Clarendon Press.

Oelhaf, P. et .al. (1645-1653) '*Typus et Syllabi lectionum publicarum et operarum*' Athenaei Gedanensis. Danzig.

Praetorius, E. et al. (2016) *'Athenae Gedanenses Ephraima Praetoriusa*'. Gdańsk: Wydawnictwo Athenae Gedanenses.

Praetorius, E. (1713) '*Athenae Gedanenses, sive commentarius historico-chronologicus succinctus, originem et constitutionem Gymnasii Dantiscani Illustris itemque recensionem superiorum eius antistitum, seu protoscholarcharum, nec non vitas et scripta rectorum ac professorum eiusdem continens.: Accedit series I. rectorum scholarum reliquarum publicarum Gedanensium*', II. rectorum Gymnasiorum tum Thoruniensis, tum Elbingensis. Leipzig: Johann Friedrich Gleditsch und Sohn. Available at: https://books.google.pl/books?id=vm5ALijSHlQC&dq=Athenae%20Gedanenses%20Praetorius&hl=pl&pg=PP7#v=onepage&q=Athenae%20Gedanenses%20Praetorius&f=false (Accessed: 15th January 2021).

Data Descriptor

Legislation and Practice of Selected State Aid Issues, According to EU and Polish Law

Anna Dobaczewska[1*]

[1] Faculty of Law and Administration; Chair of Public Economic Law and Environment Protection; University of Gdańsk (6 Jana Bażyńskiego Street, Gdańsk, Poland)

[*] Correspondence author: anna.dobaczewska@ug.edu.pl; ORCID: 0000-0001-8918-9847

Abstract

The dataset encompasses several tables, each consisting of three elements: legislation, jurisprudence and scientific articles on numerous subjects and economic activities receiving public financial support in the form of state aid instruments. The set includes a subjective list of the most commonly used and/or disputable examples of granting aid, such as for (local) airports and airlines, steel production, shipyards, and coalmines. Less controversial are: regional aid, broadband infrastructure, small entrepreneurs, and research & development which should all help to achieve the goal of the "Europe 2020" Strategy and other EC politics.

Keywords: state aid, internal market, competition law

https://doi.org/10.34808/x55q-sz53_dyr_roz43

Specification table (data records)

Subject area	Competition Law
More specific subject area	State aid law
Type of data	Text
How the data was acquired	The data was collected at the University of Gdańsk using freely available paper and online non-commercial (library) databases, in order to allow the reader to receive immediate access to the elements included in the dataset and further study the state aid cases in the EU member states
Data format	The tables are in .pdf format

Experimental factors	The data contained in the dataset were not processed
Data source location	MOST Wiedzy Open Research Catalog
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The reason to prepare the dataset on numerous state aid forms, displaying the purposes of the state aid granted under the EU law in the Member States was to give the reader a quick reference set; enabling them to get acquainted with the topic. Granting state aid and using it responsibly is a matter of fair competition between entrepreneurs, which should be guarded by public authorities at the national and EU levels. Therefore, in each part of the dataset, a ready-to-use list of the most important EU and Polish legislation has been presented. Most sets refer to the budgeting period 2014–2020; some however also present previous legislation (2007–2013). At the time of research, no provisions for the 2021–2027 were available, so a comparative study has not been undertaken. The EC legislation encompasses secondary law and soft law. The second element of the dataset, which is important for the practice, are EC decisions which appeared to become important milestones of the state aid practise. Where applicable, examples of Polish court decisions have also been presented. The aim of this was to explain to the reader whether and how the EU law has been implemented at the national level. In order to show how the national authorities comply with the EU legal provisions, each topic contains cases from different EU member states. This should enable the reader to get an overall view of the specific problem in a short time. The third component of the set is a representative list of literature describing and developing the problems arising while using state aid. Again, the assumption was mainly to show the Polish practise and research. The intention lying behind this research is to supply students, practitioners and authorities with data on separate topics within a broad field of state aid towards entrepreneurs. The year 2020 showed how important such financial support to entrepreneurs can be.

Methods

The elements of the dataset were prepared solely using the dogmatic and legal research method, according to the contemporary legislation at the EU level and its transformation into Polish law. The best method to disclose how the legal provisions work in practise is to disclose how the courts and public authorities have acted to the given law provisions. Therefore, the Author has compiled a selection of EC and court cases representative of each topic, and based on her own opinion in this respect. The usual research workshop for lawyers also includes a verification of the scientific doctrine, therefor a list of literature references has been included in the dataset.

Data quality and availability

Datasets DOI
10.34808/6jrs-4z71
10.34808/sfqa-1280
10.34808/ktm3-jk08
10.34808/03df-w284
10.34808/wwev-e426
10.34808/gb06-wn56
10.34808/28yt-gm46
10.34808/emkv-wn28
10.34808/z5k9-7x70
10.34808/n2ep-6t24
10.34808/jq38-xj88
10.34808/7fd8-7k02
10.34808/w7c6-gj69
10.34808/bavs-3596
10.34808/srd1-3q96
10.34808/kk4k-qn72
10.34808/s1ys-7c02

Datasets License:
CC-BY

References

Each of the dataset elements (separate tables) is an individualised list of legal acts in force (sometimes additionally also those of historical value), literature references (mainly articles in scientific journals) and quotations of EC decision or court judgements, personally selected by the Author; each set referring to another topic or another economic activity where state aid is or has been granted to entrepreneurs. Therefore no references, other than those being the actual content of the dataset have been used.

Data Descriptor

Dataset Relating Collective Angst, Identifications, Essentialist Continuity and Collective Action for Progressive City Policy among Gdańsk Residents

Michał Jaśkiewicz[1*], Tomasz Besta[1], Judyta Borchet[1]

[1] Faculty of Social Sciences, Institute of Psychology, University of Gdańsk (4 Jana Bażyńskiego Street, Gdańsk, Poland)

* Correspondence author: michal.jaskiewicz@ug.edu.pl; ORCID: 0000-0003-4358-1071

Abstract

This dataset contains the individual responses of 456 residents of Gdańsk who participated in the study. The study was conducted before the second term of the presidential election in Poland in 2020. Demographic variables as well as psychological measures of angst, place attachment, identification in-group continuity and willingness to engage in collective action were collected. We also measured the perception of the risk of politically motivated violence and perceived motives for the assassination of Mayor Adamowicz.

Keywords: collective angst, collective action, identification

https://doi.org/10.34808/x55q-sz53_dyr_roz44

Specification table (data records)

Subject area	Psychology
More specific subject area	Social psychology
Type of data	quantitative
How the data was acquired	Survey Monkey platform
Data format	.sav
Experimental factors	online survey
Experimental features	residents of Gdańsk

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	creative commons, attribution, non commercial

Background

Collective angst as the result of perceived threats to a group to which one belongs has usually been investigated in relation to in-group protective action (Wohl et al, 2011). We decided to explore if collective action for progressive city policy may also serve this function. We investigated this relation in the local Gdańsk context, where the assassination of mayor Adamowicz may create the feeling that Gdańsk and its values are being attacked. Our main goal was to investigate the relationship among collective angst, city identification, essentialist group continuity, and willingness to engage in collective actions for the progressive policy initiated by Adamowicz (supporting the Model of Equal Treatment, helping refugees). The second goal of this study was to explore whether collective angst related to President Andrzej Duda's reelection or Rafał Trzaskowski victory will provide the same predictions as collective angst related to the assassination of Mayor Adamowicz. In this study, we also measured the perception of the risk of a politically motivated attack.

Methods

Participants were recruited through the official website of the city of Gdańsk, where short information about the survey and invitation to take part in the study were placed. Gdańsk.pl is an important portal with a local reach that allowed us to spread the information about the survey to the local community. We used the SurveyMonkey platform to collect the data from the respondents. Each participant declared informed consent to participate in the study. The research was in line with the ethical standards of the American Psychological Association and the Polish Psychologists' Association.

Data records

The participants responded to several psychological measures as well as demographic questions (age, gender, place of living). The quantitative data were collected. In the case of measures consisting of several items, responses for each item were aggregated. Respondents used a 7-point response format (with 1 – strongly disagree, and 7 – strongly agree) for all scales. A detailed description of the numeric variables related to the scales are presented in Tab. 44.1.

The dataset contains perception of the motives for the assassination: verbal attacks by PiS politicians, verbal attacks by PO politicians, verbal attacks by the media, political views of the assassin, mental instability of the assassin, common use of hate speech and calls for violent action in the public debate, passivity of the state services in relation to previous cases of incitement to killings, personal conflict, poor psychiatric healthcare,

poor prison system. The dataset also includes respondents' answers as to what extent they think that aggression in public debate and the political situation in the country can lead to attempts of physical attacks on: LGBT community, refugees, economic immigrants, Catholics, Polish President Andrzej Duda, Mayor of Gdańsk Aleksandra Dulkiewicz, politicians of the Law and Justice Party, and politicians of opposition parties. A 7-point response format (anchored 1 – strongly disagree, and 7 – strongly agree) was used by the respondents.

Tab. 44.1

Variable description

Variable	Scale	Number of aggregated items
inherit	Place attachment inherited	2
discover	Place attachment discovered	2
Ident	Group identification	12
pci	Politicised identity	2
collAct	Collective action	12
angst	Collective angst	3
essent	Essentialist in-group continuity	2
narrat	Narrativist in-group continuity	2
Duda_angst	Collective angst_Duda reelection	3
Trzaskowski_angst	Collective angst – Trzaskowski victory	3

Data availability

Dataset DOI
10.34808/ypz9-mr46

Dataset License
CC-BY-NC

References

Wohl, M. J. A. et al. (2011) 'One day we might be no more: Collective angst and protective action from potential distinctiveness loss', *European Journal of Social Psychology*, 41(3), pp. 289–300. DOI:10.1002/ejsp.773.

Data Descriptor

Multiple Group Membership and Collective Action Intention

Tomasz Besta[1*], Paweł Jurek[2]

[1] Faculty of Social Sciences, University of Gdańsk (8 Jana Bażyńskiego Street, Gdańsk, Poland)

[*] Correspondence author: tomasz.besta@ug.edu.pl; ORCID: 0000-0001-6209-3677

Abstract

Datasets from two studies conducted in Poland on the relation between identity fusion, group identification, multiple group membership, perceived injustice, and collective action intention. The presented studies, in the context of protests against attempts to restrict abortion law, were conducted to examine the link between belonging to multiple groups, group efficacy & identification, perceived injustice and collective action intention. The dataset includes responses from 181 participants in Study 1, and 262 participants in Study 2.

Keywords: opinion-based group; multiple group membership; collective action; Facebook; identity fusion

https://doi.org/10.34808/x55q-sz53_dyr_roz45

Specification table (data records)

Subject area	Psychology
More specific subject area	Social psychology; Collective action research
Type of data	Quantitative data
How the data was acquired	The data was collected online via the Survey Monkey platform
Data format	SPSS .sav format
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes; CC BY Attribution

Background

Data from both studies were gathered in the context of collective action directed to stop the law that would introduce a total ban on abortion in Poland. So-called Black Protest took place in Poland in 2016. The offline street demonstrations were linked to a huge social media campaign and activist actions. Protests appeared in the biggest cities in Poland, as well as in smaller towns. Eventually, after the large demonstrations, lawmakers voted the bill down. This decision was seen by many as persuaded by civic resistance. The datasets consist of participants' responses in online surveys grounded in the social identity research on opinion-based collective action (McGarty, Bliuc, Thomas, & Bongiorno, 2009). Analyses conducted on the presented data could broaden the knowledge on the issues of online activism by examining the link between multiple group membership and Collective Action (CA) as well as by outlining the role of group efficacy belief as a mediator of the relationship between belonging to many social groups and CA.

Methods

181 Facebook users participated in Study 1; 262 Facebook users participated in Study 2. The studies were conducted online via the Survey Monkey platform, where all questions were asked. A link to the questionnaire was posted on various Facebook sites, and undergraduate students were asked to circulate the link to the studies among their friends. Participants were asked about their attitudes towards abortion law in Poland, following items included in the identity fusion scale (Gómez et al., 2011; for use in Polish, see Besta, Gómez, & Vázquez, 2014), group identification (Leach et al., 2008), multiple group membership scale (Iyer et al., 2009), group efficacy scale and perceived injustice scale (both scales based on items used by Van Zomeren et al., 2004), and collective action intention scale (both normative and radical non-normative CA).

Data quality and availability

All measurements used a Likert-type scale (e.g. 7-point scale with 1=not at all to 7=very much). The reliability of scales were good, with the lowest alpha coefficients of .80. The dataset is linked to the publication under the title "Multiple group membership and collective actions of opinion-based internet groups: The case of protests against abortion law restriction in Poland" with the doi address https://doi.org/10.1016/j.chb.2019.01.025

Dataset DOI
10.34808/ass9-pk50

Acknowledgements
The generation of this dataset was supported by a research grant from National Science Centre, Poland; project acronym: 2014/14/E/HS6/00587. A full description of the data analyses could be found in the published paper by Besta et al. (2019).

References

Besta, T., Gómez, Á., and Vázquez, A. (2014) 'Readiness to deny group's wrongdoing and willingness to fight for its members: The role of the Poles' identity fusion with the country and religious group'. *Current Issues in Personality Psychology*, 2(1), pp. 49–55. DOI: 10.5114/cipp.2014.43101.

Besta, T. et al. (2019) 'Multiple group membership and collective actions of opinion-based internet groups: The case of protests against abortion law restriction in Poland', *Computers In Human Behavior*, 95, pp. 94–100. DOI: 10.1016/j.chb.2019.01.025.

Bliuc, A.-M. et al. (2007) 'Opinion-based group membership as a predictor to commitment to political action', *European Journal of Social Psychology*, 37(1), pp. 19–32. DOI: 10.1002/ejsp.334.

Gómez, A. et al. (2011) 'On the nature of identity fusion: Insights into the construct and a new measure', *Journal of Personality and Social Psychology*, 100 (5), pp. 918–933. DOI: 10.1037/a0022642.

Iyer, A. et al. (2009) 'The more (and the more compatible) the merrier: Multiple group memberships and identity compatibility as predictors of adjustment after life transitions', *British Journal of Social Psychology*, 48 (4), pp. 707–733. DOI: 10.1348/014466608X397628.

Leach, C.W. et al. (2008) 'Group-level self-definition and self-investment: A hierarchical (multicomponent) model of in-group identification', *Journal of Personality and Social Psychology*, 95(1), pp. 144–165. DOI: 10.1037/0022-3514.95.1.144.

McGarty, C. et al. (2009) 'Collective action as the material expression of opinion-based group membership', *Journal of Social Issues*, 65(4), pp. 839-857, DOI: 10.1111/j.1540-4560.2009.01627.x.

Van Zomeren, M. et al. (2004) 'Put your money where your mouth is! Explaining collective action tendencies through group-based anger and group efficacy', *Journal of Personality and Social Psychology*, 87(5), pp. 649–664. DOI: 10.1037/0022-3514.87.5.649.

Data Descriptor

Data from the Survey on Entrepreneurs' Opinions on Factors Determining the Employment of the Gdańsk University of Technology Graduates

Michał T. Tomczak[1*]

[1] Faculty of Management and Economics; Gdańsk University of Technology (79 Traugutta Street, Gdańsk, Poland)

[*] Correspondence author: michal.tomczak@pg.edu.pl; ORCID: 0000-0002-2916-5015

Abstract

The dataset includes data from a survey on factors determining the employment of the Gdańsk University of Technology (Gdańsk Tech) graduates' in the opinion of entrepreneurs. The survey was conducted in 2017. The research sample included 102 respondents representing various firms from the Pomeranian Voivodeship, Poland. The study concerned i.a. factors determining the decision to hire a candidate, methods of recruiting employees, methods for diagnosing the competencies of candidates, and opinions on recruiting Gdańsk Tech graduates who do not have extensive professional experience, considering variables such as the company size and business sector. To summarise, the entrepreneurs focus on the diagnosis of candidates' competencies during a job interview, and have nothing against employing inexperienced graduates.

Keywords: entrepreneurs; employment; university graduates

https://doi.org/10.34808/x55q-sz53_dyr_roz46

Specification table (data records)

Subject area	University graduates, Employment, Labour market, Entrepreneurs
More specific subject area	Employment of university graduates
Type of data	Text
How the data was acquired	The data was collected by an online questionnaire (Computer Assisted Web Interview – CAWI)

Data format	The data is in .csv format
Experimental factors	The data contained in the dataset has been processed.
Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

Higher education institutions (HEIs) contribute to shaping the competencies of their graduates (Gawrycka, Kujawska and Tomczak, 2021, Gawrycka, Kujawska and Tomczak, 2020; Stankiewicz et al., 2020; Krawczyk-Bryłka et al., 2020), which has a huge impact on their future professional careers, translating into their situation on the labour market (Tomczak 2018). Therefore, it is very important to discover the opinions of entrepreneurs employing graduates about the factors determining the decision to hire a candidate. The data was created as a result of research conducted by the Research Team in the field of Gdańsk University of Technology Graduates Professional Situation Monitoring. The study aimed to obtain data on factors determining the employment of the Gdańsk University of Technology (Gdańsk Tech) graduates in the opinion of entrepreneurs who employ such people.

Methods

The quantitative study was conducted in 2017 using the Computer Assisted Web Interview (CAWI) method. The research sample was non-random, purposive and included 102 respondents representing various firms from the Pomeranian Voivodeship, Poland. The dataset was developed with the use of the IBM SPSS Statistics 25 software.

Data records

The data included in the dataset is anonymous and constitutes the respondents' responses to the questions contained in the questionnaire.

Data quality and availability
Dataset has been verified for data quality and completeness.

Dataset DOI
10.34808/t0c6-jz49

Dataset Licence
CC-BY

Acknowledgments
The generation of this dataset was performed within a Research Team in the field of Gdańsk University Technology Graduates Professional Situation Monitoring and was supported by Gdańsk University of Technology statutory funds.

References

Gawrycka, M., Kujawska, J., and Tomczak, M.T. (2021) 'Self-assessment of competencies of students and graduates participating in didactic projects – Case study', *International Review of Economics Education*, 36, 100204. DOI: 10.1016/j.iree.2020.100204.

Gawrycka, M., Kujawska, J., and Tomczak, M.T. (2020) 'Competencies of graduates as future labour market participants – preliminary study', *Economic research-Ekonomska istraživanja*, 33(1), pp. 1095-1107. DOI: 10.1080/1331677X.2019.1631200.

Krawczyk-Bryłka, B. et al. (2020) 'Effective Collaboration of Entrepreneurial Teams—Implications for Entrepreneurial Education', *Education Sciences*, 10(12), 364. DOI: 10.3390/educsci10120364.

Stankiewicz, K. et al. (2020) 'The Structure of Entrepreneurial Team Members, Competencies: Between Effectuation and Causation', *Education Sciences*, 10(11), 337. DOI: 10.3390/educsci10110337.

Tomczak, M.T. (2018) 'Gender and labor market situation among technical university graduates in Poland', *Horyzonty Wychowania*, 17 (41), pp. 179–189. DOI: 10.17399/HW.2018.174112.

Data Descriptor

Data from the Survey on Gdańsk University of Technology Graduates' Professional Careers

Michał T. Tomczak[1*]

[1] Faculty of Management and Economics; Gdańsk University of Technology (79 Traugutta Street, Gdańsk, Poland)

[*] Correspondence author: michal.tomczak@pg.edu.pl; ORCID: 0000-0002-2916-5015

Abstract

The dataset titled Data from the survey on Gdańsk University of Technology graduates' professional careers includes data from a survey of Gdańsk University of Technology (Gdańsk Tech) graduates' professional careers. The survey was conducted in 2017, two years after the respondents obtained graduate status. The research sample included 2553 respondents. The study concerned, i.a. the percentage of people working among graduates of the Gdańsk Tech, length of the job search period, job position, economic sector, company size, amount of remuneration, and also considers variables such as the faculty graduated. Summarising, the graduates' general situation within the labour market was very good.

Keywords: labour market; university graduates; professional careers

https://doi.org/10.34808/x55q-sz53_dyr_roz47

Specification table (data records)

Subject area	University education, University graduates, Professional competencies, Professional careers, Labour market
More specific subject area	University graduates' professional career monitoring
Type of data	Text
How the data was acquired	The data was collected by an online questionnaire (Computer Assisted Web Interview – CAWI)
Data format	The data is in .csv format
Experimental factors	The data contained in the dataset has been processed

Data source location	MOST Wiedzy Open Research Catalog, Gdańsk University of Technology, Gdańsk, Poland
Data accessibility	The dataset is accessible and is publicly and freely available for any research or educational purposes

Background

The education is an important factor influencing the further professional careers of university graduates (Ziemiański and Golik, 2020, Tomczak, 2018), and shaping their professional competencies (Gawrycka, Kujawska and Tomczak, 2021, Gawrycka, Kujawska and Tomczak, 2020 Stankiewicz et al., 2020). The data was created as a result of research conducted by the Research Team in the field of Gdańsk University of Technology Graduates Professional Situation Monitoring. The aim of the study was to obtain data on the professional situation and careers of graduates of the Gdańsk University of Technology.

Methods

The quantitative study was conducted in 2017 using the Computer Assisted Web Interview (CAWI) method. The research sample included 2553 respondents, was non-random, purposive, and composed of individuals who graduated from Gdańsk University of Technology in 2015. The dataset was developed with the use of the IBM SPSS Statistics 25 software.

Data records

The data included in the dataset is anonymous and constitutes the respondents' responses to the questions contained in the questionnaire.

Data quality and availability
Dataset has been verified for data quality and completeness.

Dataset DOI
10.34808/xz4h-ag48

Dataset License
CC-BY

Acknowledgements
This dataset was generated by a Research Team in the field of Gdańsk University Technology Graduates Professional Situation Monitoring and was supported by Gdańsk University of Technology statutory funds.

References

Gawrycka, M., Kujawska, J. and Tomczak, M.T. (2021) 'Self-assessment of competencies of students and graduates participating in didactic projects – Case study', *International Review of Economics Education*, 36, 100204. DOI: 10.1016/j.iree.2020.100204.

Gawrycka, M., Kujawska, J., and Tomczak, M.T. (2020) 'Competencies of graduates as future labour market participants – preliminary study', *Economic research-Ekonomska istraživanja*, 33(1), pp. 1095–1107. DOI: 10.1080/1331677X.2019.1631200.

Stankiewicz, K. et al. (2020) 'The Structure of Entrepreneurial Team Members, Competencies: Between Effectuation and Causation', *Education Sciences*, 10(11), 337. DOI: 10.3390/educsci10110337.

Tomczak, M.T. (2018) 'Gender and labor market situation among technical university graduates in Poland', *Horyzonty Wychowania*, 17(41), pp. 179–189. DOI: 10.17399/HW.2018.174112.

Ziemianski, P. and Golik, J. (2020) 'Including the Dark Side of Entrepreneurship in the Entrepreneurship Education', *Education Sciences*, 10(8), 211. DOI: 10.3390/educsci10080211.

Summary

This monograph is a collection of experiences gathered by the team implementing the Bridge of Data project. However, it is not just a simple summary of the project implementation. It shows and systematizes the substantive and technical works performed by the teams and several issues related to data management itself in various disciplines, represented by members of the scientific team and other researchers from partner universities.

The aim of an attempt to describe as much data as possible and the diversity thereof was to approximate to the reader of this publication both the complexity of research processes and data acquisition methods in different disciplines. Thus, it also shows various data types, formats, and methods – collection, description, and sharing.

The importance of ensuring that the results of research in the form of data are not lost is increasingly emphasized. There is also a growing awareness among researchers themselves, though not everywhere and not at the same pace.

Some funding bodies have defined requirements that plans for managing the data (Data Management Plans) produced during a research project should be included in grant proposals.

These plans show how research data will be collected, organized, managed, and preserved during the project and after. What such a plan looks like depends on the specificity and conditions of the project.

In general, these plans require a description of the data that will be produced or used, formats, metadata standards that will be used to store and organize data, the how and where of the data storage and accessibility.

To help researchers in data management, libraries and other departments responsible for research output, created data services.

The Bridge of Data project started in 2018 and focuses on implementing the Open Research Data repository. The data repository was established in accordance with the most important principles and trends in creating this type of repositories. It is used to collect and share research data from three universities in Gdańsk: the University of Gdańsk, Gdańsk University of Technology and the Medical University. It is indexed in Google Data, Web of Science Data Citation Index and other services.

The repository also allows researchers to perform Big Data Analysis on the supercomputer Tryton.

The project's key objective was to create substantive support in the Open Science Competence Center at the Gdańsk University of Technology Library. It was a completely

innovative initiative at Polish universities. In 2019, the national funder – the National Science Centre, which has already signed up to Plan S, set up the obligation to attach the Data Management Plan to all grant applications starting from September 2019.

Being the only university in Poland with a team of professionals who can support research teams in the preparation of DMP's for the needs of project applications, Gdańsk University of Technology has grown into the role of a national leader.

In February 2020, representatives of the Gdańsk University of Technology Library were invited as experts by the National Science Center 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 Center trained several hundred people from all over Poland – researchers, librarians, and administration employees. Special individual training was also prepared at the request of the authorities of several Polish universities.

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

Meanwhile, the involvement of researchers allowed not only to collect a wide catalog of data, but also to develop the functionalities of the platform itself, which meets the needs of scientists from many disciplines.

People involved in the implementation of the project constitute the basis of scientific structures, at departments, in institutes, and can act as data champions, supporting other colleagues who want to share their data in the best possible way with their knowledge and experience.

We hope that the materials and experiences we have collected will prove useful to both researchers and teams supporting research processes, including research data management, both in libraries, university administration and IT departments.

Anna Wałek, PhD
Scientific editor

Gdańsk University of Technology Publishing House

Edition I. Ark. ed. 14,9, ark. print 17,0, 263/1177

Printing and binding: Volumina.pl Daniel Krzanowski
Księcia Witolda 7-9, 71-063 Szczecin, tel. 91 812 09 08

WYDAWNICTWO

ISBN 978-83-7348-864-9

ISBN 978-83-7348-864-9
9 788373 488649

mostwiedzy.pl

'Bridge of Data – Multidisciplinary Open System Transferring Knowledge. Stage II Open Research Data' project is co-financed by the European Regional Development Fund under the Operational Program Digital Poland 2014-2020

European Funds Digital Poland

Republic of Poland

BRIDGE OF KNOWLEDGE

European Union
European Regional Development Fund