

Sharing research data across disciplines



Edited by Anna Watek

Gdańsk 2022

GDAŃSK UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE
CHAIRMAN OF EDITORIAL BOARD

Dariusz Mikielawicz

EDITOR OF SCIENTIFIC PUBLICATIONS

Michał Szydłowski

REVIEWERS

Henryk Krawczyk

Ignasi Labastida

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>

The monograph edited by Anna Wałek, Phd "Sharing research data across disciplines" provides immediate open access to its content under the Creative Commons BY 4.0 license.

DOI

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

ISBN 978-83-7348-864-9

Table of Contents

Introduction	7
Part 1. The Bridge of Data Project Assumptions	13
The Bridge of Data Project Objectives <i>Anna Wałek, Michał Nowacki, Paweł Lubomski</i>	15
Data Analysis in Bridge of Data <i>Wojciech Artichowicz, Krzysztof Drypczewski, Jerzy Proficz</i>	46
Part 2. Data Management Practices	59
The Digital Tissue and Cell Atlas and the Virtual Microscope <i>Jarosław Skokowski, Marika Bolcewicz, Kamila Kreft, Thierry van de Wetering, Jacek Gulczyński, Anna Lewandowska, Leszek Kalinowski</i>	61
Towards Open Research Data in the Economics Discipline <i>Piotr Kasprzak, Magdalena Szufliata-Żurawska</i>	70
A Text as a Set of Research Data. A Number of Aspects of Data Acquisition and Creation of Datasets in Neo-Latin Studies <i>Grzegorz Kotłowski, Jacek Pokrzywnicki, Elżbieta Starek</i>	79
Surf Zone Currents in the Coastal Zone of the Southern Baltic Sea – a Modelling Approach <i>Aleksandra Dudkowska, Aleksandra Cupiał</i>	92
Using Synchronously Registered Biosignals Dataset for Teaching Basics of Medical Data Analysis – Case Study <i>Tomasz Kocejko</i>	100
Part 3. Data Descriptors	107
Educational Dataset of Handheld Doppler Blood Flow Recordings <i>Tomasz Kocejko</i>	109
Testing the Diagnostic Utility of Recombinant <i>Toxoplasma Gondii</i> Chimeric Antigens – Generated Datasets <i>Bartłomiej Ferra</i>	113
Long-Term Measurement of Physiological Parameters – Child Dataset <i>Małgorzata Szczerska</i>	119

Porous Phantoms Mimicking Tissues – Investigation of Optical Parameter Stability Over Time <i>Monika Ronowska</i>	123
Non-Contact Temperature Measurements Dataset <i>Aleksander Mroziński</i>	127
Determination of Changes in Viscosity of Hydrogel Depending on Shear Rates <i>Szymon Mania, Robert Tylingo, Katarzyna Kozłowska-Tylingo</i>	130
Mechanical Properties of Human Stomach Tissue <i>Szymon Grymek</i>	134
Measurement of the Temporal and Spatial Temperature Distribution on the Surface of PVCPT Tissue Phantom Illuminated by Laser Dataset <i>Anna Sękowska-Namiołko</i>	138
Herbarium of Division of Marine Biology and Ecology as the Primary Basis for Conservation Status Assessments in the Gulf of Gdańsk <i>Ilona Złoch</i>	142
Macrophytobenthos in the Puck Bay in 2010–2018 Dataset <i>Aleksandra Zgrundo</i>	146
Cyanobacterial and Algal Strains in the Culture Collection of Baltic Algae (CCBA) <i>Filip Pniewski</i>	151
Occurrence of Cyanobacteria in the Gulf of Gdańsk (2008–2009) <i>Justyna Kobos, Anna Krakowiak, Agata Błaszczyk, Anna Toruńska-Sitarz</i>	154
Mercury in various components of the environment (Gulf of Gdańsk, southern Baltic Sea) <i>Magdalena Beldowska</i>	158
Methodology of the Morphometry Study on Baltic Herring <i>Jakub Idczak</i>	162
Early Oceanographical Data Collected by the Institute of Oceanography University of Gdańsk <i>Maciej Matciak, Adam Krężel</i>	168
High-Resolution Wind Wave Parameters in the Area of the Gulf of Gdańsk During 21 Extreme Storms <i>Gabriela Gic-Grusza, Aleksandra Dudkowska</i>	171
Long-term Hindcast Simulation of Currents, Sea Level, Water Temperature and Salinity in the Baltic Sea <i>Marek Kowalewski</i>	175
Areas of Updraft Air Motion in an Idealised Weather Research and Forecasting Model Simulation of Atmospheric Boundary Layer Response to Different Floe Size Distributions <i>Marta Wenta</i>	180

High Resolution Sea Ice Floe Size and Shape Data from Knox Coast, East Antarctica <i>Agnieszka Herman</i>	184
Long-Term GNSS Tropospheric Parameters for the Tropics (2001-2018) derived from selected IGS stations <i>Zofia Baldysz, Grzegorz Nykiel, Mariusz Figurski</i>	187
Simulations of the Derecho Event in Poland of 11 th August 2017 Using WRF Model <i>Mariusz Figurski, Grzegorz Nykiel</i>	193
Split-Beam Echosounder Data from Puck Bay <i>Patryk Damian Pezacki</i>	199
Vehicle Detection and Speed Estimation Using Millimetre Wave Radar <i>Piotr Ody</i>	203
AC Motor Voltage and Audible Noise Dataset <i>Beata Pałczyńska</i>	207
Impedance Spectra of RC Model as a Result of Testing Pulse Excitation Measurement Method Dataset <i>Grzegorz Lentka</i>	211
X-ray Photoelectron Spectroscopy of Carboxylic Acids as Corrosion Inhibitors of Aluminium Alloys <i>Jacek Ryl, Joanna Wysocka</i>	215
Crack Mouth Opening Displacement for EH36 Shipbuilding Steel Measurements Dataset <i>Jakub Kowalski, Janusz Kozak</i>	219
Tribological Properties of Thermoplastic Materials Formed by 3D Printing by FDM Process <i>Jacek Lubiński</i>	222
Potential Energy Curves of Diatomic Alkali Molecules Datasets <i>Patryk Jasik, Tymon Kilich</i>	226
Minimal Sets of Lefschetz Periods for Morse-Smale Diffeomorphisms of a Connected Sum of g Real Projective Planes <i>Grzegorz Graff, Adrian Myszkowski</i>	230
On Computing Curlicues Generated by Circle Homeomorphisms <i>Justyna Signerska-Rynkowska</i>	233
Video of LEGO Bricks on Conveyor Belt Dataset Series <i>Tomasz Boiński</i>	236
Regeneration Project of Market Places GOSPOSTRATEG – “Polanki” Market in Gdańsk-Oliwa Pilot Project Monitoring Dataset <i>Justyna Borucka</i>	240
Description of the Dataset Hanow – <i>Praecepta de Arte Disputandi</i> – Transcription and Photographs <i>Jacek Pokrzywnicki</i>	247

Description of the Dataset Rhetoric at School – a Selection of the Syllabi from the Academic Gymnasium in Gdańsk – Transcription and Photographs <i>Jacek Pokrzywnicki</i>	251
Legislation and Practice of Selected State Aid Issues, According to EU and Polish Law <i>Anna Dobaczewska</i>	255
Dataset Relating Collective Angst, Identifications, Essentialist Continuity and Collective Action for Progressive City Policy among Gdańsk Residents <i>Michał Jaśkiewicz, Tomasz Besta, Judyta Borchet</i>	258
Multiple Group Membership and Collective Action Intention <i>Tomasz Besta, Paweł Jurek</i>	261
Data from the Survey on Entrepreneurs’ Opinions on Factors Determining the Employment of the Gdańsk University of Technology Graduates <i>Michał T. Tomczak</i>	264
Data from the Survey on Gdańsk University of Technology Graduates’ Professional Careers <i>Michał T. Tomczak</i>	267
Summary	270

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.