ELIXIR CZ Annual Conference 2024: Data management

9-11 October 2024, Prague, Czech Republic

Organised by ELIXIR CZ

ELIXIR CZ Annual Conference 2024: Data management

Organising committee
Jiří Vondrášek, Anna Strachotová,
Tereza Votrubová

www.elixir-czech.cz
ISBN 978-80-86241-72-2

No professional language editing of the abstract content was performed.
Scientific and legal responsibility for the abstracts belongs to the authors.

ISBN 978-80-86241-72-2

Summary of content

About ELIXIR CZ infrastructure

The Czech National Infrastructure for Biological Data, abbreviated ELIXIR Czech Republic or ELIXIR CZ, is a distributed research infrastructure for bioinformatics that has arisen from an advanced computational environment. We are dedicated to organisation, storage, sharing and facilitation of interoperability of life science data for further processing and analysis. We respond to the needs of national scientific community, but we are also a proud member of the pan-European infrastructure for biological data ELIXIR, which brings together life science resources from throughout Europe.

ELIXIR CZ is comprised of 14 research performing organisations across the Czech Republic, with its headquarters in the Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences.

Institute of Organic Chemistry and Biochemistry of the CAS (IOCB)
Coordinating body of ELIXIR CZ and administrator of computational resources for bioinformatics research. IOCB develops proteomics resources and a database of small molecules, which are the flagships of ELIXIR CZ.

CESNET
CESNET is one of providers of a large national e-infrastructure for research and development, more specifically, provides communication, computing and storage facilities. CESNET acts as an ambassador of the Czech Republic in GÉANT Project, EGI Federation, and TERENA Association.

Masaryk University: CEITEC, CERIT-SC
CEITEC dedicates its' services to molecular medicine and structural biology, it is also a member of INSTRUCT infrastructure. CERIT-SC is one of providers of an e-infrastructure that provides advanced IT services.

Palacký University Olomouc (UP)
Provider of structural bioinformatics tools. UPOL acts as a liaison point to infrastructure EATRIS.

Charles University (CU)
Developer of tools for diagnostics and prognosis in medicine. UK also provides tools for high-throughput analysis of genomic, proteomic and structural data. UK is active in education and training on aspects of work with biological data.

Institute of Molecular Genetics of the CAS (IMG)

UMG provides DNA and RNA sequence analysis and tools. UMG represents the liaison point to infrastructures INFRAFRONTIER and EU-OPENSCREEN.

Institute of Microbiology of the CAS (IMIC)

MBÚ provides tools for computational biology and bioinformatics as well as models of biological networks.

Institute of Biotechnology of the CAS (IBT)

Provider of bioinformatics tools for structural biology. IBT provides database of DNA structural families.

Biology Centre of the CAS (BC)

BC is dedicated to sequence composition, molecular organisation and evolution of plant genomes and chromosomes.

University of Chemistry and Technology, Prague (UCT)

UCT provides training in the use of tools in cheminformatics and bioinformatics. UCT also provides structural bioinformatics computing tools.

Czech Technical University in Prague – Faculty of Information Technology (CTU)

CTU is dedicated to conceptual modelling and software implementation of conceptual models and development of modelling tools.

University of West Bohemia (UWB)

IT provider for marrow donor analysis and search applications. UWB operates a synthetic biology laboratory that supports tools for efficient assembly protocols and tools for hybrid biochemical reaction simulation.

University of South Bohemia in České Budějovice (USB)

USB represents a genomic centre for plants and microorganisms and applied informatics.

International Clinical Research Center of St. Anne's University Hospital in Brno (FNUSA ICRC)

Developer and provider of novel bioinformatics tools for protein structure analysis and prediction of the effect of mutations on human health.

Scientific Programme

Wednesday, 9 October

DM in EU context
Chairman: Jiří Vondrášek

12:00 – 13:00	Registration, Vila Lanna
13:00 – 13:10	Welcome word by Jiří Vondrášek, Head of ELIXIR CZ
13:10 – 13:50	Data Management Community – towards standardized approach across EU countries, infrastructures and project Niclas Jareborg, National Bioinformatics Infrastructure Sweden
13:50 – 14:30	Data management in national and EU context from funding providers point of view Petr Baldrian, Czech Science Foundation
14:30 – 15:10	Data Management Support through the ELIXIR Compute Platform Matej Antol, Masaryk University
15:10 – 15:45	Coffee break
15:45 – 16:25	Data management from HPC point of view – towards AI solution Jan Martinovič, VSB – Technical University of Ostrava, IT4Innovations
16:25 – 17:05	EIRENE CZ & ELIXIR CZ cooperations for data management Elliott Price, Recetox Brno, Masaryk University
17:05 – 17:45	Implementing Data Management in Academic Institutions Using ELIXIR Services and Resources Jiří Vondrášek, Institute of Organic Chemistry and Biochemistry of the CAS
17:45 – 19:00	Dinner
19:00 – 22:00	Poster session

Thursday, 10 October

Tools, Workflows
Chairman: Radka Svobodová

9:00 – 9:35	A Comprehensive Approach to Data Management Planning using DSW Marek Suchánek, Robert Pergl, Czech Technical University in Prague
9:35 – 9:50	Software Management Planning with DSW Marek Suchánek, Czech Technical University in Prague
9:50 – 10:10	Onedata for comprehensive management of distributed scientific data Łukasz Opioła, ACC Cyfronet AGH, Krakow
10:10 – 10:40	Coffee break
10:40 – 10:55	iRODS as a solution for distributed data management Martin Golasowski, VSB – Technical University of Ostrava
10:55 – 11:15	Electronic laboratory notebooks in theory and praxis Marek Cebecauer, J.Heyrovský Institute of Physical Chemistry
11:15 – 11:30	Tools for Public Data Management in Galaxy Project's Ecosystem Martin Čech, Institute of Organic Chemistry and Biochemistry of the CAS
11:30 – 13:00	Lunch time

Community, Best Practice
Chairman: Karel Berka

13:00 – 13:05	Introduction to community best practices Karel Berka, Palacký University Olomouc
13:05 – 13:25	Molecular Biophysics Database of raw data – MBDB Jan Dohnálek, Institute of Biotechnology of the CAS
13:25 – 13:45	From Need to Practice: Data Stewardship at UCT Prague Martin Schätz, University of Chemistry and Technology, Prague
13:45 – 14:05	A Few Glimpses into Data Management at Masaryk University Michal Růžička, Masaryk University
14:05 – 14:25	Sequencing Data Management: Specifics, Applications, and European Initiatives Vojtěch Bystrý, Masaryk University, CEITEC
14:25 – 14:45	SIP – platform for management of the scientific data life-cycle and its implementation in mid-sized cryo-EM service laboratory Jiří Nováček, Masaryk University, CEITEC
14:45 – 15:00	FAIR Molecular Dynamics Adrian Rošinec, Masaryk University
15:00 – 15:30	Coffee break

Thursday, 10 October

Community, Best Practice
Chairman: Karel Berka

15:30 – 15:45	Data management in plant phenotyping and Agri-Food system Michal Stočes, Czech University of Life Sciences Prague
15:45 – 16:00	Data managenet at CENAKVA LRI. Jaromír Kovárník, University of South Bohemia in České Budějovice
16:00 – 16:20	Clinical data towards multiomics Petr Pavliš, Palacký University Olomouc
16:20 – 16:40	GlobalFungi and GlobalAMFungi Databases: Powerful tools for in-depth exploration of fungal ecology and biogeography Tomáš Větrovský, Institute of Microbiology of the CAS
16:40 – 17:00	Data Management of Biological Imaging Data Tomáš Svoboda, Masaryk University
17:00 – 18:00	Panel discussion on DM best practices Karel Berka, Palacký University Olomouc
18:00 – 22:00	Social event – Bowling Dejvice

Friday, 11 October

NRP, EOSC, OS II
Chairman: Marek Suchánek

09:00 – 09:25	EOSC CZ & Open Science Activities in the Czech Republic Jiří Marek, Masaryk University
09:25 – 09:45	National Repository Platform as the data infrastructure for FAIR data management Luděk Matyska, Masaryk University
09:45 – 10:05	NRP: Services for Users – Academic Institutions and Individual Scientists and Researchers Michal Růžička, Masaryk University, CERIT-SC
10:05 – 10:15	Data Stewardship Wizard: from Pilot's Checklist to a Full Cockpit Marek Suchánek, Robert Pergl, Czech Technical University in Prague
10:15 – 10:45	Coffee break
10:50 – 11:15	What's a Data Steward, Precious? Dagmar Hanzlíková, Charles University
11:15 – 11:30	Open Science II Jitka Baťková, Charles University
12:00	Farewell


GDPR
PPT
HTML
CSV
API
SSL

Lectures

Data Management Community – towards standardized approach across EU countries, infrastructures and project

Niclas Jareborg - National Bioinformatics Infrastructure Sweden

Research data management (RDM) is central to the implementation of the FAIR (Findable Accessible, Interoperable, Reusable) and Open Science principles. Recognising the importance of RDM, ELIXIR Platforms and Nodes have invested in RDM and launched various projects and initiatives to ensure good data management practices for scientific excellence. These projects have resulted in a rich set of tools and resources highly valuable for FAIR data management. The ELIXIR RDM Community brings together RDM experts to develop ELIXIR's vision and coordinate its activities, taking advantage of the available assets. It aims to coordinate RDM best practices and illustrate how to use the existing ELIXIR RDM services. The Community is built around three integral pillars, namely, a network of RDM professionals, RDM knowledge management and RDM training expertise and resources. It will also engage with external stakeholders to leverage benefits and provide a forum to RDM professionals for regular knowledge exchange, capacity building and development of harmonised RDM practices, keeping in line with the overall scope of the RDM Community. In the short term, the Community aims to build upon the existing resources and ensure that the content of these remain up to date and fit for purpose. In the long run, the Community will aim to strengthen the skills and knowledge of its RDM professionals to support the emerging needs of the scientific community. The Community will also devise an effective strategy to engage with other ELIXIR structures and international stakeholders to influence and align with developments and solutions in the RDM field.

Data management in national and EU context from funding providers point of view

Petr Baldrian - Czech Science Foundation

Data represent the fundament on which science is able to observe nature and society and to test its hypotheses. As such, data management is an integral part of research and data are a valuable output that science generates. In the ongoing debate about the role of science in the society, there is a consensus that science funded from public sources should serve the society and its results, including correctly managed data, should be widely available to the scientific community and/or public society. In the European Union countries, this view is backed by the legal requirement to make the data from publicly funded research open. In order to make this happen, European research funders associated in the Science Europe organization formulated a set of recommendations. While the basic principle, that data should be well preserved and FAIR – i.e., findable, accessible, interoperable, and reusable, as much as possible are beyond doubt, there is some flexibility among the funders in the definition of data to be shared this way and about the approaches to Open Access. In the Czech research environment, there is a consensus that data management plans should be unified across all funders to make it easier for scientists to fulfil their requirements. The Czech Science Foundation wants to further support the scientific community by providing tools that help to prepare DMP and manage data, such as, e.g., the “Data Stewardship Wizard”. It also leaves important decisions about publication and data sharing on their funded researchers provided that they follow the best practices of their research field.

Data Management Support through the ELIXIR Compute Platform

Matej Antol - Masaryk University

The ELIXIR Compute Platform aims to establish a robust, distributed infrastructure that plans to integrate cloud services, high-performance computing (HPC), and artificial intelligence (AI) resources to support life sciences research. A key component of the platform is to enhance data management, supporting and promoting the adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) data principles. By collaborating across multiple European countries, the platform seeks to facilitate seamless data processing, analysis, and the sharing of both computational and data resources. It is designed to ensure that researchers have access to the tools and environments necessary for managing and processing large-scale biological datasets and workflows.

The platform aspires to play a key role in fostering collaboration, supporting open science, and enabling data-driven discoveries in bioinformatics, genomics, and other areas of life sciences.

Data management from HPC point of view – towards AI solution

Jan Martinovič - VSB – Technical University of Ostrava, IT4Innovations

Data management is an open challenge that is becoming increasingly important, especially with the rise of AI. Training neural networks requires large amounts of data to be physically located in places with powerful computing resources, such as HPC centres. The landscape of data transfer protocols and storage solutions is diverse, and one of the challenges is to connect remote sites with a common solution for data and metadata transfer. The presentation will introduce the LEXIS Platform for distributed data management and complex workflow orchestration.

This platform provides an iRODS-based solution for federated data transfer and management, as well as orchestration of complex computing tasks on HPC infrastructures. In addition, the Horizon Europe EXA4MIND project, which builds on this concept, will explore concepts for handling extreme data volumes in various AI and machine learning application cases. The project aims to provide a unified solution that enables extreme data analysis across different data storage solutions and computing infrastructures.

EIRENE CZ & ELIXIR CZ cooperations for data management

Elliott Price - Recetox Brno, Masaryk University

Ongoing, planned, and potential future collaborations of the Czech nodes of the European research infrastructure for exposome research (EIRENE-CZ) and the Life-Science Infrastructure for Biological Information (ELIXIR-CZ) will be discussed. Particular focus will be paid to activities surrounding small molecule mass spectrometry (MS) data, databases, repositories, tools, and processing pipelines. EIRENE-CZ interactions with existing ELIXIR services (e.g., Galaxy Europe, MetaboLights, IDSM) and communities will be outlined, and areas for enhanced national capacity prioritised.

Implementing Data Management in Academic Institutions Using ELIXIR Services and Resources

Jiří Vondrášek - Institute of Organic Chemistry and Biochemistry of the CAS

Effective data management (DM) is essential for enhancing research integrity, collaboration, and compliance with FAIR (Findable, Accessible, Interoperable, Reusable) principles. This presentation outlines a structured workflow for implementing professional DM in academic institutions, leveraging ELIXIR's comprehensive suite of tools and services. The talk covers a step-by-step approach starting from needs assessment and policy development to utilizing ELIXIR's technical infrastructure, such as the Data Stewardship Wizard and RDMkit. A key focus is placed on the role of training in embedding robust DM practices. By utilizing resources like the Training e-Support System (TeSS) and engaging with the RDM Trainer Network, institutions can ensure their staff are equipped with the necessary skills and knowledge. Attendees will gain insights into developing a sustainable DM strategy that integrates training at every stage, fostering a culture of continuous learning and alignment with international standards. The presentation concludes with practical recommendations for monitoring and refining the DM strategy using ELIXIR's Training Metrics Database (TMD) and other impact assessment tools.

This session is ideal for research managers, data stewards, and academic professionals seeking to implement or enhance DM practices within their organizations using ELIXIR's resources.

A Comprehensive Approach to Data Management Planning using DSW

Marek Suchánek, Robert Pergl - Czech Technical University in Prague, Faculty of Information Technology

Data management planning is a demanding endeavour that encompasses knowledge of the whole research data lifecycle including practices, tools, repositories, policies. The Data Stewardship Wizard (https://ds-wizard.org) has proven to be a versatile and effective solution for making high-value DM plans for FAIR and open science. In this walkthrough talk, we explain the importance of planning within the context of the entire data life cycle. Then, we focus on the most notable aspects of DSW that have shown practical value over the years. As such, the audience will gain better insight into the possibilities, scope of features and effective use of the tool. The talk is directed mainly for researchers and data stewards.

Software Management Planning with DSW

Marek Suchánek - Czech Technical University in Prague, Faculty of Information Technology

This short talk will cover Software Management Plans (SMPs) and their growing role in the research world. We will start by explaining what SMPs are and how they connect to Data Management Plans (DMPs), focusing on how they help plan, maintain, and sustain research software. We will also discuss current efforts to create and use SMPs and how they link with other activities and tools like CodeMeta, Software Heritage, and the Research Software Alliance (ReSA). By the end, the participants will have a clear idea of how SMPs help make research software more reliable and long-lasting.

Onedata for comprehensive management of distributed scientific data

Łukasz Opioła - ACC Cyfronet AGH, Krakow

Onedata is a versatile platform designed to manage large, distributed scientific datasets efficiently. It provides a unified user experience, allowing seamless access to data stored across organizationally and geographically distributed environments, including autonomous data centers. Thanks to built-in, powerful data management tools and integration with IAM services, Onedata enables collaborative data sharing between users and working groups of different affiliations.

From the data provider perspective, Onedata virtualizes heterogeneous storage systems into a POSIX-compliant logical namespace and handles synchronization of datasets shared with peer organizations. In an era of multinational projects run by big consortia, seamless integration between different providers on the data access level is paramount for scientific innovation. Onedata is designed specifically for these needs while being an open-source software stack, easy to run on any infrastructure.

This talk will provide a cross-cutting overview of the Onedata platform, highlighting its capabilities in scientific data management and its support for high-performance access to distributed data.

iRODS as a solution for distributed data management

Martin Golasowski - VSB – Technical University of Ostrava, IT4Innovations

Distributed data management for scientific data faces several challenges. Storage technologies vary widely from site to site, and to enable seamless movement of data and metadata, these variations need to be levelled out. iRODS stands for Integrated Rule-Oriented Data Systems and provides a virtual file system as an abstraction of local storage resources, parallel data transfers, a programmable rules engine, and federation of independent deployments. In our talk we present iRODS as a solution for distributed data management and storage in conjunction with the Persistent Identifier Infrastructure (ePIC) and the EUDAT B2SAFE service. We will show the basic concepts of iRODS, such as zones, collections and objects, and its main advantages, which lie mainly in the fact that there is no central point for data storage and that the control of the data is left to the local administrators, who can choose other locations for federation. We will also show how iRODS can be used for data staging close to the computing infrastructure using the LEXIS Platform and present several data preservation use cases from different domains.

Electronic laboratory notebooks in theory and praxis

Marek Cebecauer - J.Heyrovský Institute of Physical Chemistry

The open sharing of data with other scientists and the public is a key aspect of modern science. Electronic laboratory notebooks (ELNs) act as hubs of data and help with the daily tasks of research data management (RDM). In this talk I will share our experience with the implementation of ELNs at the J. Heyrovsky Institute of Physical Chemistry, explain what is the difference between an ELN and a basic e-notebook (e.g. OneNote), why to use ELNs, how to choose an ELN for your lab, and how some ELNs can help to automate RDM processes. I hope to help colleagues to choose their way of handling data and to facilitate engagement in Open Science.

Tools for Public Data Management in Galaxy Project's Ecosystem

Martin Čech - Institute of Organic Chemistry and Biochemistry of the CAS

Recent large-scale collaborative efforts in genomics like the Vertebrate Genome Project (VGP) and the European Reference Genome Atlas are designed to facilitate a new era of scientific discovery and democratize access to an abundant amount of data.TheVGP effort alone publishes genome assemblies of 12 species every week until all the ~70,000 extant vertebrate species genomes are known – which is expected within a decade. Effectively sharing this trove of data with scientists worldwide is a known challenge.

Galaxy Project is an open platform for accessible, reproducible, and transparent computational research and in this talk we will present some of the approaches and tools that communities around Galaxy use to address the need for ubiquitous dataset access.This includes the CernVM File System (CVMFS) for large data distribution on a file system level – mainly used as a global reference data distribution network that includes sharing of tool containers directly enabling computational reproducibility. Another piece of the puzzle - which the Galaxy Training Network uses in a successful symbiosis - is a general-purpose open data repository Zenodo. Experience gathered from using these tools are portable and can be leveraged in other communities and ecosystems.

Molecular Biophysics Database of raw data - MBDB

Jan Dohnálek - Institute of Biotechnology of the Czech Academy of Sciences

Biomolecular research profits from a plethora of biophysical methods that enable characterization of molecular properties, stability assessment and optimization, interaction analysis, etc. However, a standardized solution for deposition and public access to raw measurement data from such methods is missing.

Under the MOSBRI project we have taken the endeavor to define a standard format of metadata for selected techniques of molecular biophysics and create a public database to store raw data files together with metadata descriptions and enable them to be Findable, Accessible, Interoperable and Reusable. The metadata for individual data sets will consist of a general part and a method-specific part. The general part defines descriptors for key parameters common for different experimental techniques, e.g. source organism, identity of individual molecules, including chemicals, with reference to external databases and unique identifiers for the most relevant types. The method-specific part is devoted to the metadata special for a particular technique (such as MST, BLI, etc.). This approach will eventually enable searching across different techniques and allow direct comparisons of results e.g. interaction parameters for the same molecular system measured by different techniques.

The Molecular Biophysics Database (MBDB) is being built using the Invenio repository platform technology (https://inveniosoftware.org/) and JSON as the key representation format of metadata, in collaboration with the CESNET data storage team and their hardware resources. Based on the previous survey on the use of biophysical methods and the need for databases [1] the first set of covered techniques includes microscale thermophoresis, bio-layer interferometry and surface plasmon resonance.

This work is supported by the EU project MOSBRI – Molecular Scale Biophysics Research Infrastructure of the Horizon 2020 research and innovation program of the European Union, no. 101004806.

References

[1] Dohnálek J, Stránský J, Malý M, Černý J. MOSBRI survey - Biophysical Data Standards and Accessibility, https://doi.org/10.5281/zenodo.6604159.

From Need to Practice: Data Stewardship at UCT Prague

Martin Schätz - UCT Prague

The journey of establishing data stewardship at UCT Prague began not with a pressing requirement for a data management plan (DMP) under a European grant but through our involvement in the DocEnhance grant. Piloting the Data Stewardship course in this project was a pivotal moment. Shortly after, we encountered our first Horizon 2020 DMP request, underscoring the need for systematic development of rules and a supportive community.

Introduced during the course, the Data Stewardship Wizard (DSW) tool proved invaluable in addressing the first DMP challenge and became our first link to the DSW team. Graduates of the course soon became teachers and assistants in subsequent pilots, creating a cycle of knowledge-sharing and skill-building. This foundation allowed us to respond quickly and efficiently to evolving research data management (RDM) requirements.

Following the conclusion of the DocEnhance project, a small but dedicated community of researchers formed. They regularly participated in our "Breakfast with Data Stewards" sessions, where we discussed their needs and challenges. These gatherings offered invaluable insights into improving our data stewardship (DS) and IT services and helped shape the future of research data management at UCT Prague. This community-driven approach allowed us to address immediate concerns and long-term strategic goals, including planning and implementing internal storage solutions and refined processes for managing data effectively.

By building on these experiences, UCT Prague has continued to evolve its data stewardship services, fostering a culture of collaboration and continuous improvement in research data management.

A Few Glimpses into Data Management at Masaryk University

Michal Růžička - Masaryk University, CERIT-SC

In this talk, we would like to provide a few glimpses into some aspects of data management at Masaryk University, which is representative of a mid-size university with faculties covering topics from arts through computer science to medicine. We will quickly look at preparations for the university's data strategy, centralised support of projects data management consultations and Data Management Plans preparations, and particular use cases of data management technical support on selected university centres.

Sequencing Data Management: Specifics, Applications, and European Initiatives

Vojtěch Bystrý - Bioinformatics Core Facility, CEITEC Masaryk University

How do we harness the full potential of genomic data while safeguarding sensitive, identifying information? Sequencing data are revolutionizing a wide range of fields—from basic biology and environmental science to agriculture and, most notably, clinical medicine. Yet, with great promise comes great complexity, particularly when it comes to managing clinical genomic data where issues of data protection and ethics loom large.

In this talk, we will uncover the broad applications of sequencing data, focusing on the clinical sector, where the stakes are highest. We'll dive into the specifics of managing these datasets and explore the current landscape of global and European solutions, like the European Genome-phenome Archive (EGA), the Federated European Genome-phenome Archive (FEGA), and projects such as the Global Data Infrastructure (GDI) and the European Open Science Cloud (EOSC).

SIP - platform for management of the scientific data life-cycle and its implementation in mid-sized cryo-EM service laboratory

Jiří Nováček - Masaryk University, CEITEC - Central European Institute of Technology

We present a comprehensive workflow that facilitates the management of raw scientific data. Our approach has been tested using electron cryo-microscopy (cryo-EM), which, like other imaging methods, generates substantial amounts of raw data per dataset. The workflow is built on the iRODS federated cloud system and incorporates previously developed tools for data publication. These components are orchestrated by a Python-based system to automate the indexing of data to remote cloud storage and, alternatively, its archival for long-term preservation. Additionally, the workflow includes the deposition of metadata into a public database, with release following the expiration of an embargo period. Furthermore, we have implemented real-time data analysis of single-particle cryo-EM data at a remote high-performance computing (HPC) center, eliminating the need for computational resources within the service laboratory.

FAIR Molecular Dynamics

Adrián Rošinec - Masaryk University

Molecular Dynamics (MD) simulations have become an essential tool in various fields, including biophysics, chemistry, and materials science. By providing atomic-level insights into molecular behavior, MD enables scientists to study dynamic processes such as protein folding, ligand binding, and material properties.

However, despite their extensive use and necessity, molecular dynamics (MD) simulations are facing a significant challenge in the field of data organisation and management. There is a pressing need to adopt the FAIR data principles (Findable, Accessible, Interoperable, Reusable) within MD workflows. As MD outputs generate vast amounts of data, ensuring that this data is accessible and reusable by the broader scientific community is of paramount importance.

This presentation addresses key challenges in the efficient execution of FAIRification of MD data and highlights the need for adopting FAIR data principles within MD by providing curated data repositories, FAIRification tools for automated metadata collection significantly boosting the efficiency of data sharing and reproducibility, while preventing unnecessary reruns and laborious manual work with annotating datasets.

Data management in plant phenotyping and Agri-Food system

Michal Stočes - Czech University of Life Sciences Prague

The main objective of this contribution is to introduce a platform designed for the storage and management of plant phenotyping data. This platform intents to play a crucial role in efficiently storing, organizing, and sharing data gathered from plant phenotyping experiments. A key focus of the contribution is on the technical aspects of the platform, including the implemented standards such as the BrAPI (Breeding API) and MIAPPE (Minimum Information About Plant Phenotyping Experiment), which ensure interoperability and data consistency. Additionally, the talk will provide practical examples that illustrate how tabular (2D - two-dimensional) phenotyping data can be transformed and structured into a multi-table environment, showcasing the flexibility and adaptability of the platform for various research scenarios. This contribution presents the importance of adopting standardized data management practices in advancing plant science and breeding research.

Data managenet at CENAKVA LRI.

Jaromír Kovárník - CENAKVA; Faculty of Fisheries and Protection of Waters, University of South Bohemia Ceske Budejovice

The research centre South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses (CENAKVA) is one of the large research infrastructures of the Czech Republic. Open science is, therefore, the basic concept of the centre. We provide open access to the research infrastructure and to the research papers and scientific datasets. To be able to provide open access, we need to care about the data management produced by the researchers of CENAKVA. We are at the beginning of the implementation of the data management tools and services, but we are already involved in the EOSC and EOSC CZ initiatives. We also use our experience from international infrastructure projects, where we learned the principles of infrastructure open access and data exchange.

The data managed by CENAKVA covers a wide range of research areas, including morphometric measurements and the weighing of aquatic fauna, feeding experiments, toxicological studies, growth curve analysis, and biochemical data. High-quality imaging from microscopy and videos are also integral to the research, particularly for studying cell growth and development, as well as the behaviour of fish and crayfish. More advanced research uses specific scientific data formats such as .fsa, .fcs, .fastq, .bw, .bed, and .bam for DNA sequencing and cytometric measurements. The specific data repositories exist for several types of data produced by our laboratories and are used for open access. However, we are also involved in the development of specific repositories for complex data, such as non-target screening of environmental pollutants.

Clinical data towards multiomics

Petr Pavliš - Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University Olomouc.

The cornerstone of clinical and genomic data management at UMTM is the ClinData system. The ClinData is software solution for collecting of clinical and laboratory data on patients included in various research projects and clinical studies. ClinData software features exceptional flexibility in design of research and clinical projects, ability to store data from several different, unrelated projects in parallel and secure access in multicenter studies. The data is stored in a parameterized format and ready for further statistical analysis or visualization and processing. ClinData can harmonize clinical terminology according to Observational Medical Outcomes Partnership (OMOP) standard data model which is required for implementation of EHDS legislation. The OMOP is an open community data standard, designed to standardize the structure and content of observational data and to enable efficient analyses that can produce reliable evidence. ClinData software has web-based client/server architecture, and all data transfers are secured by SSL encryption. Users can have multiple distinct roles; each role being fundamentally tailored to specific study related tasks and responsibilities (authorized access). ClinData software is connected to IMTM CEPH Object storage which allows GDPR compliant storage of big scientific and clinical data (genomic, proteomics, metabolomics, imaging, etc.). Objects in the storage can be directly linked to ClinData forms to allow accessing all data from one place. ClinData software collaborates with public LifeData Portal, platform where users can obtain or visualize selected anonymized data from clinical studies and registries, adhering to the core principles of Findable, Accessible, Interoperable, and Reusable (FAIR).

GlobalFungi and GlobalAMFungi Databases: Powerful tools for in-depth exploration of fungal ecology and biogeography

Tomáš Větrovský - Institute of Microbiology of the Czech Academy of Sciences

The advent of high-throughput sequencing has dramatically enhanced our capacity to monitor microbial distribution across various ecosystems and geographic regions. The GlobalFungi database (Větrovský et al., 2020) was developed to provide unparalleled access to global data on fungal occurrences, supporting research into fungal community composition and biogeography on a broad scale. This resource is pivotal in studying fungal diversity, compiling over 4.5 billion observations of ITS1 and ITS2 marker sequences from 846 studies, covering nearly 85,000 samples worldwide. GlobalFungi allows for extensive exploration of fungal communities across diverse terrestrial environments using general fungal primers, thus offering invaluable insights into the distribution of various fungal groups in natural ecosystems.

However, due to a known bias of general fungal primers against arbuscular mycorrhizal (AM) fungi in the GlobalFungi database, the need for a specialized platform arose. This led to the creation of GlobalAMFungi (Větrovský et al., 2023), a complementary database focused on key barcoding regions (SSU, ITS2, and LSU) for AM fungi. GlobalAMFungi compiles nearly 50 million observations of AM fungal DNA sequences from approximately 8,500 samples, integrating geographical metadata from 100 studies.

Despite this distinction, GlobalFungi remains a crucial tool in fungal biogeography, providing a comprehensive overview of fungal distribution. Both databases feature web interfaces for data searching and visualization, and they promote community contributions to further enrich their collections. However, managing such massive datasets (especially in the case of GlobalFungi) poses significant challenges for data management, including issues of storage, accessibility, and data integrity. As the volume of data continues to grow, the need for robust infrastructure and sophisticated tools to handle, process, and maintain the quality of these datasets becomes increasingly important. The integration of these resources marks a significant advancement in mapping global fungal diversity and distribution, offering researchers a powerful platform to study fungal ecology and the environmental factors that influence fungal distribution.

Větrovský T. et al. (2020) GlobalFungi. Scientific Data 7, 228.
Větrovský T. et al. (2023) GlobalAMFungi. New Phytologist 240, 2151-2163.

Data Management of Biological Imaging Data

Tomáš Svoboda - Masaryk University

In many scientific disciplines, expensive equipment is now often shared through centralized facilities where researchers request specific experiments. The outcome of such experiments is a dataset, which can be quite large in many cases. Researchers then process these datasets, draw scientific conclusions through their interpretation, and publish the results. Recently, there has been growing importance placed on the availability and interoperability of such data to ensure that scientific results can be independently verified and reproduced.

The Cellular Imaging Laboratory (CELLIM) at CEITEC Masaryk University faced challenges in managing image data acquired by optical microscopes. The process previously relied on a combination of local storage and manual data-sharing methods (like USB flash disk), which has limitations in terms of data accessibility, integrity, and long-term preservation. To improve this situation, we developed a data management solution based on the Onedata data management system and our automation application.

Our solution streamlines the entire data lifecycle: from parsing technical metadata from raw imaging data to transferring processed data from the laboratory server to a centralized storage system. The Onedata integration enables seamless sharing through publicly accessible URLs, ensuring that researchers can easily and securely access their data. To further enhance the user experience, we are integrating the OMERO platform, allowing data visualization and processing in a user-friendly way in a web browser.

EOSC CZ & Open Science Activities in the Czech Republic

Jiří Marek - Institute of Computer Science, Masaryk University / Secretariat EOSC CZ

Since 2020, Open Science activities have grown within the Czech Republic. Universities are building their Open Science support centers and creating institutional strategies, and the data is becoming more and more critical in the daily lives of individual researchers. The year 2021 marks the start of the European Open Science Cloud (EOSC) implementation in our country, which aims to create a national node for this European initiative and promote good practice in research data management across scientific communities. Implementing National Data Infrastructures (NDI) will make a common platform for sharing, managing, and accessing data and computing resources for research purposes. NDI will support both scientific and multidisciplinary research activities. It will cover a wide range of scientific fields and disciplines. The talk will dive in and present the introductory lecture on the history, current state, and the way forward within the Open Science and EOSC activities in the Czech Republic.

National Repository Platform as the data infrastructure for FAIR data management

Luděk Matyska - Masaryk University

This presentation will provide an in-depth overview of the National Repository Platform (NRP), focusing on its key components, services, and role in supporting FAIR (Findable, Accessible, Interoperable, Reusable) data management within the Czech research ecosystem. We will discuss the platform's current state of implementation, highlighting its capabilities in ensuring the secure, long-term accessibility of research data. The NRP's planned suite of tools and services will be outlined, demonstrating how they will aid researchers in managing, sharing, and preserving their data in alignment with FAIR principles. The presentation will conclude with a look at future developments and milestones as NRP continues to evolve as a critical infrastructure for research data management in Czechia.

NRP: Services for Users – Academic Institutions and Individual Scientists and Researchers

Michal Růžička - Masaryk University, CERIT-SC

The main objective of the National Repository Platform for Research Data (NRP) project (https://www.eosc.cz/en/about-eosc-cz/national-support/national-repository-platform) is to provide various research data-related services for NRP users: academic institutions and individual scientists and researchers. In this talk, we will focus on services for the end users, i.e. what NRP will provide to end researchers and their institutions to enhance their abilities to fulfil FAIR principles on their datasets and data management requirements in research project calls through the life cycle of research data.

Data Stewardship Wizard: from Pilot's Checklist to a Full Cockpit

Marek Suchánek, Robert Pergl - Czech Technical University in Prague, Faculty of Information Technology

The Data Stewardship Wizard (DSW) is known as a data management planning tool with a metaphor of being a "pilot's checklist" – similarly to pre-flight checks, researchers are able to make sure that nothing important was forgotten to carry out their data-intensive project. Now in the context of the National Repository Platform (NRP), DSW is evolving into another aviation metaphor: the pilot's cockpit. By thorough integration in the NRP ecosystem, researchers will also be able to obtain navigation and monitoring of their progress throughout the project. In this short talk, we present the plans for this future.

What's a Data Steward, Precious?

Dagmar Hanzlíková - Charles University

When you are implementing a new infrastructure, such as the National Repository Platform (NRP), you obviously need resources and people who build it. However, having the infrastructure is not enough. You also need people who know about the infrastructure and who are willing and able to use it. You need trained researchers who will often need to learn to organise their work differently. And they will need support to do that.

Data stewards play a crucial role by providing direct support to researchers. They help ensure compliance with data security standards, assist in the organization and management of research data, and facilitate the adoption of best practices in data handling.

Apart from the role of data stewards, the talk will also introduce the Key Activity 7 of the NRP project, titled *"Training and Awareness in NRP Functionalities and Services"*. The aim of the Activity, is to employ the train the trainers approach, engage the community of data stewards and prepare training materials on NRP functionalities. As a result, data stewards will be better equipped to help researchers utilize the NRP infrastructure more effectively.

Open Science II

Jitka Baťková - Charles University

This presentation is aimed at presenting the steps taken so far in the preparation of the project application for the OP JAK OpenScience II call. The aim of the call is to support field-specific and interdisciplinary activities within the framework of the implementation of the European Open Science Cloud (EOSC) initiative in the Czech Republic in accordance with the conceptual document "Architecture of EOSC implementation in the Czech Republic". It supports the development of disciplinary and interdisciplinary repositories, their integration into NDI and ensuring the transnational interoperability of research data. The basic goal is to ensure the availability and reuse of research data in accordance with FAIR principles, while the challenge emphasizes the specifics of sensitive data.

Posters

List of posters

1. **Inferring transcription factor regulatory features using an interpretable unsupervised algorithm**
   Kateřina Balážová Faltejsková - Institute of Organic Chemistry and Biochemistry of the CAS
2. **MolMeDB - Molecules on Membranes Database**
   Václav Bazgier - Palacký University Olomouc
3. **Data behind the conformational analysis of nucleic acids at dnatco.datmos.org**
   Jiří Černý - Institute of Biotechnology of the CAS
4. **Challenges in data exchange as a part of donor search for transplantation needs**
   Jiří Fatka - University of West Bohemia
5. **Insights on protein conformational diversity through apo-holo binding site exploration**
   Christos Feidakis - Charles University
6. **Multi-class predictions of intracellular locations of proteins in organisms with complex plastids**
   Ansgar Gruber - Biology Centre of the CAS
7. **Managing Data Provenance and Versioning in Protein-Ligand Binding Site Predictions: A Case Study of PrankWeb**
   David Hoksza - Charles University
8. **Managing genomic data across Europe - European Genomic Data Infrastructure (GDI)**
   Jaroslav Juráček - Masaryk University
9. **usegalaxy.cz**
   Aleš Křenek - Masaryk University
10. **Minimal Alphabet for Protein Design**
    Kseniia Kushnir - University of Chemistry Technology, Prague
11. **The Pea Pangenome: Data Management and Sharing within the Framework of the International Genome Sequencing Initiative**
    Jiří Macas - Biology Centre of the CAS

12. **Supporting Research Data Management in the Czech Republic: The EOSC-CZ Project**
    Jiří Marek - Masaryk University
13. **AlphaErector: Visualization of AlphaFold models of multi-domain proteins**
    Veronika Milatová - University of Chemistry and Technology, Prague
14. **Scientific dataset management system for the research institute based on Onedata**
    Adrián Rošinec - Masaryk University
15. **Look, a Concept! The Research Data Infrastructure Roadmap at IOCB**
    Marie Šafner - Institute of Organic Chemistry and Biochemistry of the CAS
16. **In silico Assessment of Primer Bias in the Fungal Kingdom**
    Johannes Schweichhart - University of South Bohemia; Institute of Hydrobiology, Biology Centre CAS
17. **Fast, structure-based searching in a large-scale protein data repository**
    Terézia Slanináková - Masaryk University
18. **ChannelsDB 2.0: A Comprehensive Database of Protein Tunnels and Pores in AlphaFold Era**
    Anna Špačková - Palacký University Olomouc
19. **Data Management and FAIRification in MAFIL**
    Tomáš Svoboda - Masaryk University
20. **Scientific dataset management system for the research institute based on Onedata**
    Tomáš Svoboda - Masaryk University
21. **Data issues in HLA-KIR interaction assessment workflow**
    Kateřina Wolf - University of West Bohemia

Inferring transcription factor regulatory features using an interpretable unsupervised algorithm

Kateřina Balážová Faltejsková, Jiří Vondrášek - Institute of Organic Chemistry and Biochemistry of the CAS

The transcription factor binding in the context of gene regulation is a widely studied issue. Multiple studies suggest that other factors influence the binding (apart from the transcription factor affinity to its binding site), such as flanks of the binding motif, interactions with other regulatory elements, and chromatin modifications. For future usage in e.g. personalized medicine, it would be beneficial to identify and understand these features only from the DNA sequence.

To move towards this goal, we have developed bindmate – a software tool for quantifying the similarity between DNA loci based on the similarity of individual k-mers in the locus combined as the best pairing. The k-mer similarity can be calculated from a set of arbitrary functions of the DNA sequence.

In our experiments, we worked with GC content, affinity to known transcription factors, and k-mer sequence similarity. By searching for the best pairing of k-mers between two compared sequences, we operate without the assumption that the shorter regulatory elements are arranged linearly (as is expected in sequence alignment).

Using our software, we have probed ChIP-seq experiments for 15 distinct transcription factors to see whether there are sequence features around the binding sites unique for the particular experiment, showing that the binding sites of some transcription factors (e.g., MYC and NFκB) can be clustered together using our approach. We also proceed with ex-post analysis, identifying and examining the k-mers important for the distinction. That way, we provide information on the size and composition of the regulatory element.

MolMeDB - Molecules on Membranes Database

Kateřina Storchmannová, Jakub Juračka, Dominik Martinát, Václav Bazgier, Karel Berka - Palacký University Olomouc

Biological membranes are natural barriers to cells. They play a key role in cell life and the pharmacokinetics of drug-like small molecules. A small molecule can pass through the membranes in two ways: via passive diffusion or actively via membrane transport proteins. There is a huge amount of data available about interactions among the small molecules and the membranes and also about interactions among the small molecules and the transporters.

MolMeDB (molmedb.upol.cz) is a comprehensive and interactive database of interactions of small molecules with membranes.[1] From the start, we have collected data about partitioning and penetration of the small molecules crossing the membranes. Recently, we have expanded our area of interest to include interactions of small molecules with transporters and ion channels. Nowadays, more than 930,000 interactions for almost 500,000 molecules are available in MolMeDB.

The data within the MolMeDB is collected from scientific papers, our in-house calculations (COSMOmic/COSMOperm[2]), and obtained by data mining from several databases (e.g. ChEMBL[3], PubChem[4], The IUPHAR/BPS Guide to PHARMACOLOGY[5]). Data in the MolMeDB are fully searchable and browsable by name, SMILES, membrane, method, transporter, or dataset, and we offer collected data openly for further reuse. Also, the content of the database is available via REST API and the RDF model of MolMeDB (docs.molmedb.upol.cz).

References

(1) Juračka, J.; Šrejber, M.; Melíková, M.; Bazgier, V.; Berka, K. MolMeDB: Molecules on Membranes Database. Database 2019, 2019, baz078. https://doi.org/10.1093/database/baz078.

(2) Schwöbel, J. A. H.; Ebert, A.; Bittermann, K.; Huniar, U.; Goss, K.-U.; Klamt, A. COSMO Perm: Mechanistic Prediction of Passive Membrane Permeability for Neutral Compounds and Ions and Its pH Dependence. J. Phys. Chem. B 2020, 124 (16), 3343–3354. https://doi.org/10.1021/acs.jpcb.9b11728.

(3) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Marañón, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. ChEMBL: Towards Direct Deposition of Bioassay Data. Nucleic Acids Research 2019, 47 (D1), D930–D940. https://doi.org/10.1093/nar/gky1075.

(4) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 Update. Nucleic Acids Research 2023, 51 (D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956.

(5) Harding, S. D.; Armstrong, J. F.; Faccenda, E.; Southan, C.; Alexander, S. P. H.; Davenport, A. P.; Spedding, M.; Davies, J. A. The IUPHAR/BPS Guide to PHARMACOLOGY in 2024. Nucleic Acids Research 2024, 52 (D1), D1438–D1449. https://doi.org/10.1093/nar/gkad944.

Data behind the conformational analysis of nucleic acids at dnatco.datmos.org

Jiří Černý, Paulína Božíková, Michal Malý, Terezie Prchalová, Jakub Svoboda, Lada Biedermannová, and Bohdan Schneider - Institute of Biotechnology of the CAS

We will present the new features of the DNATCO v5.0 web service and discuss the data formats and dictionaries we developed to make the DNATCO results easily available. For compatibility with the structural biology ecosystem we designed a DNATCO-based extension of the current mmCIF dictionary (https://mmcif.wwpdb.org/dictionaries/mmcif_pdbx_v50.dic/Index) introducing categories specific for the annotation, validation, and refinement of backbone conformers in structures of nucleic acids. The dictionary extension is available at https://mmcif.wwpdb.org/dictionaries/mmcif_ndb_ntc.dic/Index allowing external tools to use the DNATCO data in a consistent way.

Challenges in data exchange as a part of donor search for transplantation needs

Jiří Fatka[1], Filip Jani[1], Kateřina Wolf[1], Pavel Jindra[2], Kateřina Steinerová[2] a Lucie Houdová[1] - [1]Faculty of Applied Sciences, University of West Bohemia, [2]The Czech National Marrow Donors Registry

Search for suitable donor of hematopoietic stem cells (HSC) is a critical and time-consuming process needed for haemato-oncology patients. Effective data exchange between unrelated HSC donor registries within the transplantation process is essential to identify suitable donors and enable timely transplantation to the patient.

The WMDA (World Marrow Donor Association) community, which globally covers donor registries, still uses outdated technology in electronic communication between donor registries. This technologically greatly limits the possibilities of data transfer, especially in terms of speed, quantity, and type of data. Therefore, the development of a modern replacement has been underway, but its creation and subsequent implementation into practice is a very time-consuming and implementation-intensive operation.

In terms of development, it must cover the architecture of the communication system, the method of data transfer and the structure of the transferred data. It also means, each WMDA member who wants to communicate electronically with other registers will have to implement this technology in some form into his own information infrastructure. And this creates specific challenges arising from the interconnection of the proposed WMDA technology and registries' specific internal information systems and standard operating procedures, that need to be dealt with.

Insights on protein conformational diversity through apo-holo binding site exploration

Christos Feidakis - Charles University, Faculty of Science, Department of Cell Biology

A single protein structure rarely captures the conformational variability of a protein. Both the bound and unbound (holo and apo) forms of a protein are essential for understanding its geometry and making meaningful comparisons. Nevertheless, docking or drug design studies, often just consider a single, rigid protein structure in its holo form. With the recent explosion in the field of structural biology, there is an urgent need for large, curated datasets.

We developed AHoJ to match holo structures to their apo and holo counterparts by searching the Protein Data Bank (PDB) for alternative versions of a given binding site and annotating each version as holo or apo, depending on whether it binds a ligand or not. We then built AHoJ-DB (www.apoholo.cz/db), by searching for apo-holo pairs for the binding sites of 515,467 biologically relevant ligands in 29,464 proteins and annotating each pocket with several metrics.

Analysis of AHoJ-DB reveals cryptic pockets, disordered apo sites, and a high variance in the availability of apo forms between binding sites of different ligands. Overall, less than half of the binding sites we search have an apo form in the entire PDB. We investigate and discuss possible reasons for these inherent biases in the PDB. AHoJ-DB can be used to train and evaluate ligand binding site predictors, discover potentially druggable proteins, and reveal protein- and ligand-specific relationships that were previously obscured by intermittent or partial data.

Multi-class predictions of intracellular locations of proteins in organisms with complex plastids

Marta Vohnoutová[1,2], Miroslav Oborník[1,2], Ansgar Gruber[1,2] - [1]Faculty of Science, University of South Bohemia; [2]Laboratory of Evolutionary Protistology, Institute of Parasitology, Biology Centre, Czech Academy of Sciences

The intracellular location of a protein is an important aspect of its function within the cell. Therefore, annotations or predictions of protein locations are necessary for analyses of all types of 'omics data. However, the signals and mechanisms for intracellular protein targeting depend on the ultrastructure and phylogenetic origin of the cell, with high variability between the eukaryotic super groups. Algae with complex plastids are groups of organisms with high importance for the global oceans and water bodies, however, in comparison to animals (including humans) or plants (including crops), they are so far poorly studied with respect to intracellular protein targeting.

Cells of diatoms and related algae with complex plastids of red algal origin are highly compartmentalized. These plastids are surrounded by four envelope membranes, which also define the periplastidic compartment (PPC), the space between the second and third membranes. The PPC corresponds to the cytosol of the eukaryotic alga that was the ancestor of the complex plastid. Metabolic reactions as well as cell biological processes take place in this compartment; however, its exact function remains elusive. Automated predictions of protein locations proved useful for genome wide explorations of metabolism in the case of plastid proteins, but until now, no automated method for the prediction of PPC proteins was available. We present an updated version of the plastid protein predictor ASAFind, which includes optional prediction of PPC proteins. Furthermore, we release a Python script to calculate custom scoring matrices for adjustment of the ASAFind method to other groups of algae. This way, the method can be extended to other algae with related types of plastids, for example it has been applied to analyze genomes of the cryptophyte *Guillardia theta*, or *Chromera velia*, the closest known free living and photosynthetic relative of apicomplexan parasites.

Further information:
https://doi.org/10.48550/arXiv.2303.02488

Original publication of the method:
https://doi.org/10.1111/tpj.12734

Managing Data Provenance and Versioning in Protein-Ligand Binding Site Predictions: A Case Study of PrankWeb

David Hoksza, Petr Škoda - Faculty of Mathematics and Physics, Charles University

In the era of large-scale biological data, ensuring accurate, reproducible, and transparent data management is a critical challenge. This abstract presents PrankWeb, a web-based tool for predicting protein-ligand binding sites from a submitted protein structure. PrankWeb utilizes both real-time and precomputed data, leveraging P2Rank software for binding site predictions and a conservation pipeline for computing sequence conservation. The precomputed data are added daily, reflecting new PDB structures. In addition, PrankWeb allows users to execute docking of user given molecule within a predicted pocket using Autodock Vina.

We describe our current strategies for versioning and provenance tracking, detailing how the system aligns with FAIR principles. Additionally, we highlight weak points in the workflow, including issues related to updating precomputed data, handling dependencies between software versions, dealing with data-related issues, and ensuring long-term data and computational service accessibility.

We aim to spark a broader discussion on how best to address these challenges in bioinformatics. Key questions include: How can we improve version control for complex workflows involving multiple tools and data sources? How should we handle the continuous integration of new data into existing systems? How to handle data or computational pipeline depreciation? How to address computationally expensive data migration?

By sharing our approach and identifying its potential weaknesses, we hope to engage the community in a conversation on improving data management strategies for bioinformatics services.

Managing genomic data across Europe - European Genomic Data Infrastructure (GDI)

Jaroslav Juracek - Institute of Computer Science, Masaryk University, Brno, Czech Republic

Genomic data helps researchers and clinicians better understand diseases and personalize medical procedures. Advances in technologies, combined with global initiatives, have made genomic data more accessible and impactful, paving the way for breakthroughs in medicine, public health, and biotechnology. However, ethical concerns and data privacy challenges remain crucial considerations as this field grows.

The European Genomic Data Infrastructure (GDI) is a pioneering project, coordinated by ELIXIR, that builds on the 1+ Million Genomes (1+MG) initiative effort to enable secure access to human genomics and corresponding clinical data across Europe by creating unified data infrastructure. The GDI provides a robust framework of components that ensures five core functionalities: data discovery, access management tools, data processing, data reception, and storage and interface. The current state of the art is represented by GDI Starter Kit, a set of software applications and components based on FAIR principles and open community standards like those from the Global Alliance for Genomics and Health (GA4GH). These products have been tailored for GDI nodes to deploy, giving them standardized technical capabilities.

The creation of the Starter Kit products has been led by several member countries, including the Czech Republic, which has contributed to the development of two of the core functionalities, LS AAI (Authentication and Authorization Infrastructure) and Containerised Computation. In cooperation with the other pillars of the GDI project, these technical components will be further fine-tuned to achieve a production deployment that ensures the technical interoperability of the connected nodes.

Another challenge is maintaining the legal, ethical, and economic sustainability of the GDI. Based on the GDI experience, the Genomic European Digital Infrastructure Consortium (EDIC)—a formal structure similar to ERICs—is currently under preparation, coordinated by Luxembourg with strong support from the Czech Republic.

Such a collaborative and federative approach not only accelerates research but also ensures that Europe remains at the forefront of global genomic research. By building a stronger infrastructure for genomic data management, the GDI sets the foundation for a future where genomic discoveries can be rapidly translated into real-world applications across the continent.

usegalaxy.cz

Aleš Křenek - Masaryk University

Galaxy is a widespread, web based solution to deal with large sets of scientific data, their computational processing, and tracking thorough provenance of the secondary datasets. Having originated in the bioinformatics community, it spans over numerous scientific disciplines nowadays, thousands of computational tools are available, and it is supported by dozens of scientific and software development teams worldwide.

Besides the “main” installations of Galaxy, there is an emerging network of nationally based “usegalaxy.*” installations, loosely coordinated by the community to meet certain common requirements. usegalaxy.cz, operated by the e-Infra CZ team, is expected to receive an official “usegalaxy.*” approval by the end of 2024.

We will introduce the installation from both user perspective (how it can be accessed, what tools are available, …) and technical/implementation details (where the data are stored, how the computations are executed, …). Custom developments, with which we contribute to the community, will be also mentioned.

Minimal Alphabet for Protein Design

Kseniia Kushnir, Vojtěch Spiwok - Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague

Recent developments of machine learning algorithms makes it possible to design new proteins that fold into compact 3D structures. We explorer ESMfold algorithm to ask the question what is the minimal amino acid alphabet for design structured proteins. We tried to remove some amino acids from the standard alphabet of 20 proteinogenic amino acids to design proteins of size 100 amino acid residues. The designed proteins have been tested by computational methods. In future we plan to test some of the designed proteins experimentally by recombinant expression and biophysical and possibly also structural biology methods.

The work was supported by Ministry of Education, Youth and Sports (LM2023055).

The Pea Pangenome: Data Management and Sharing within the Framework of the International Genome Sequencing Initiative

Jiří Macas, Petr Novák - Biology Centre of the Czech Academy of Sciences, České Budějovice

Recent advances in genome sequencing technologies have ushered in a new era in biology, enabling construction of complete genome assemblies for even the most complex organisms. In addition, a parallel assembly of multiple genomes representing related species or their populations provides ultimate information about their genetic variation. This approach, also known as pangenomics, is invaluable for deciphering the genetic basis of phenotypic variation, which is of great interest for plant genetics and breeding. Therefore, pangenome projects are being launched for a growing number of crop species. However, the ability to generate huge amounts of sequencing data brings a number of challenges in terms of their efficient storage, sharing and exploration. Here we present an example of the data management workflow set up for the International Pea Pangenome Consortium, which aims to assemble and analyze the pangenomes of several accessions of the garden pea (*Pisum sativum*) and its wild relatives.

Supporting Research Data Management in the Czech Republic: The EOSC-CZ Project

Jiri Marek - Institute of Computer Science, Masaryk University

The European Open Science Cloud (EOSC) initiative provides an infrastructure for publishing, searching, and reusing research data, tools, and services across Europe. In the Czech Republic, the implementation of the EOSC initiative aims to establish the National Data Infrastructure (NDI) – a unified platform for sharing, managing, and accessing data and computing resources for research purposes. Thanks to funding from the Ministry of Education, Youth, and Sports, several large national projects have been supported, collectively forming the aforementioned infrastructure.

The EOSC-CZ project plays a crucial and strategic role in the Czech National Programme. The project is built on three pillars: a) Promotion of Cooperation (the EOSC-CZ Secretariat serves an administrative and support role), b) Technological Foundation (the project will ensure the pilot operation of technical and software components and services, such as NMD, AAI, PID, etc.), and c) Competence Development (the Training Centre provides comprehensive support for training, educational, and related activities). The technological foundation and facilities developed through the EOSC-CZ project form the cornerstone of the National Data Infrastructure (NDI), on which subsequent projects like the National Repository Platform (NRP) can build. By integrating thematic repositories and advancing the NDI, EOSC-CZ fosters collaboration and ensures that Czech research institutions actively contribute to the broader European research community.

AlphaErector: Visualization of AlphaFold models of multi-domain proteins

Veronika Milatová, Vojtěch Spiwok - Department of Biochemistry and Microbiology, University of Chemistry and Technology, Prague

Proteins, especially eukaryotic, are often composed of multiple structured domains interconnected by unstructured regions of differing lengths. AlphaFold and related machine learning tools for protein structure prediction can usually correctly distinguish between structured and unstructured parts and predict 3D structures of structured parts. However, predicted structures of multi-domain proteins usually contain structured domains in the center. These are wrapped by unstructured regions. This structural arrangement is difficult to interpret. Furthermore, it is not realistic. We developed a tool based on the Modeller package that dissects the AlphaFold model into individual domains, arranges them in the space and connects them by unstructured regions.

The work was supported by Ministry of Education, Youth and Sports (LM2023055, LUC2413).

Scientific dataset management system for the research institute based on Onedata

Tomáš Svoboda, Adrián Rošinec, Tomáš Raček, Josef Handl, Aleš Křenek, Radka Svobodová - Masaryk University

As the volume and complexity of scientific data continue to grow, the efficient management of data across its entire lifecycle has become paramount. In this context, we have decided to create a system for CEITEC Research Institute, which would allow emerging data sets to be registered and managed, using the existing Onedata system as the data layer.

At its core, Onedata oversees the entire data lifecycle, commencing with the acquisition of data from various connected instruments (cryo-EM, NMR, light microscopy) at the moment of data generation. The automated processes employed by the system enable the organisation of acquired data into coherent datasets, enriched with metadata harvested directly from the instruments themselves and the execution of workflows designed to generate data-aware metadata annotations where feasible, in accordance with defined metadata schemas established in specific fields. This facilitates the creation of FAIR datasets which are ready for publication in thematic data repositories, as and when required.

The ability to integrate heterogeneous storage capacity with heterogeneous high-performance computing (HPC) platforms, such as Jupyter notebooks and Kubernetes container clouds, is a significant advantage. By facilitating the connection between storage capacity and direct access to compute resources, Onedata enables access to compute resources for data analysis, thereby accelerating scientific discovery.

Finally, the ability to share live data via Onedata enables data sharing within and beyond the research group. Once the analysis has been completed, the system is prepared to allow scientists to easily complete and publish the final dataset to the thematic data repositories.

The objective of this poster is to illustrate the development of tools that will facilitate and streamline data sharing among scientific communities at the national and international levels. These tools are intended to support the principles of FAIR and Open Science.

Look, a Concept! The Research Data Infrastructure Roadmap at IOCB

Jiri Vondrasek, Anna Strachotova, Marie Safner, Matus Drexler - Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Prague

With European and national initiatives towards Open Science and globally widespread shifts from traditional research data practices, often characterized by rather fragmented, old-fashioned approaches in data management, to an adoption of modern, comprehensive, open, and FAIR data management, many research institutions have been investing time and effort in developing strategies to stay ahead. At IOCB, we are in the process of testing tools, exploring possibilities, devising solution concepts addressing needs of our researchers, taking the first steps towards implementing an in-house data management infrastructure, and asking questions you may know the answers to!

In silico Assessment of Primer Bias in the Fungal Kingdom

Johannes Schweichhart - University of South Bohemia; Institute of Hydrobiology, Biology Centre CAS

Sequence databases are central information stores in biological sciences which commonly serve as reference for newly obtained sequencing data. As many of these databases grow exponentially, the information stored within them is primarily accessed through a few prevalent workflows while other important aspects and properties of the reference data remain unexplored. In this study we aim to target an important but experimentally elusive knowledge gap by performing a comprehensive in silico analysis of local alignments of 150 fungal primers and 830 primer pair combinations with over 725.000 curated fungal long-read rRNA reference sequences covering SSU, ITS1, 5.8S, ITS2, LSU. Based on a weighted alignment metric and a currently theoretical threshold we infer amplification biases of the tested primers in terms of overall coverage for fungi and specific fungal taxa as well as specificity for fungi in plant and animal backgrounds. Our results indicate that primer bias is indeed an important and currently unquantified confounding factor in metabarcoding studies which should be taken into account in the interpretation of metabarcoding results. In order to make good use of these findings we aim to provide an analytical interface linked with the EUKARYOME database to assist researchers in making informed decisions when it comes to primer selection. We demonstrate here how sequence databases, when integrated with simple analysis and visualisation pipelines, can serve as reference points for important aspects of data interpretation and study design in fungal ecology.

Fast, structure-based searching in a large-scale protein data repository

Terézia Slanináková - Masaryk university

Since the 1970s, the Protein Data Bank (PDB) has been meticulously curating protein structure data, establishing a comprehensive foundation for structural biology research. Later joined by UniProt, SCOP, and CATH databases, these resources collectively provided a set of well-structured and rigorously maintained repositories of protein information. In 2021, the confluence of these repositories and advancements in deep learning culminated in the solution to the long-standing protein-folding problem. The AlphaFold system, capable of predicting protein structures based solely on their amino acid sequences, has expanded our structural knowledge, increasing the number of known protein structures from approximately 180,000 in the PDB (as of 2021) to an unprecedented scale. Current databases contain between 214 million (AlphaFold DB) and 700 million (ESM Atlas) predicted protein structures.

This sudden expansion of available protein structure data unlocks numerous opportunities for advancing biological research, such as understanding protein families, evolution, and interactions, while accelerating drug discovery and advancing protein engineering and systems biology. However, while this data repository is technically open, the sheer volume and complexity of protein structure data present significant challenges regarding their practical use, well captured by the FAIR data principles. The computational resources required to navigate and analyze this vast dataset often exceed those available to typical researchers and biologists.

To address the accessibility issues and enhance the findability of relevant proteins, we have developed AlphaFind [1, 2], a web-based similarity search system designed to bridge the gap between the expansive AlphaFold DB and the structural biology research community. AlphaFind allows users to search through structurally similar proteins within the AlphaFold DB efficiently. The system's architecture is designed to handle the substantial volume (214 million structures, 21 TB) as well as the inherent complexity of protein structure data, managing to find relevant proteins in a matter of seconds.

In contrast to metadata-based approaches, AlphaFind employs a function-based search strategy, extracting semantic information directly from structural features. AlphaFind thus allows realizing the potential of well-structured, large-scale biological data contained in a single, specialized repository by enhancing its accessibility and applicability for the scientific community.

[1] https://doi.org/10.1093/nar/gkae397
[2] https://alphafind.fi.muni.cz

ChannelsDB 2.0: A Comprehensive Database of Protein Tunnels and Pores in AlphaFold Era

Anna Špačková[1], Ondřej Vávra[2,3], Tomáš Raček[4,5], Václav Bazgier[1], David Sehnal[4,5], Jiří Damborský[2,3], Radka Svobodová[4,5], David Bednář[2,3], Karel Berka[1] - [1]Department of Physical Chemistry, Faculty of Science, Palacký University; [2]Loschmidt Laboratories, Department of Experimental Biology and RECETOX, Faculty of Science, Masaryk University; [3]International Clinical Research Center, St. Anne's University Hospital Brno; [4]CEITEC – Central European Institute of Technology, Masaryk University Brno; [5]National Centre for Biomolecular Research, Faculty of Science

ChannelsDB 2.0 has been upgraded to offer a comprehensive insight into protein channels' structural characteristics, geometry, and physicochemical attributes, encompassing both tunnels and pores. These channels are computed on deposited biomacromolecular structures originating from the PDBe and AlphaFoldDB databases. In the new version of the ChannelsDB database, we have incorporated data generated through the widely used CAVER tool, augmenting the insights previously acquired only through the original MOLE tool. Additionally, we have extended the database's coverage by introducing tunnels originating from cofactors localised within the AlphaFill database or from cognate ligands within PDB structures. This expansion has increased the available channel annotations by almost five times. ChannelsDB 2.0 houses information concerning geometric properties such as length and radius and physicochemical attributes based on the amino acids lining the channels. These stored data are intricately linked with the existing UniProt mutation annotation data, facilitating in-depth investigations into the functional roles of biomacromolecular tunnels and pores.

In summary, ChannelsDB 2.0 represents an invaluable resource for conducting in-depth analyses of the significance of biomacromolecular channels. The database is freely accessible to the public at the address https://channelsdb2.biodata.ceitec.cz.

Data Management and FAIRification in MAFIL

Tomáš Svoboda - Masaryk University

The Multimodal and Functional Imaging Laboratory (MAFIL) at CEITEC Masaryk University specializes in neuroimaging and neuroscience research. It utilizes advanced imaging technology and manages complex datasets from sources like MR and EEG. The lab aims to ensure that all data complies with FAIR principles for easier sharing and collaboration.

MAFILDB is a platform designed to manage data workflows. It securely collects personal data from volunteers, operators, and MR devices and pairs them with data from other modalities. Operators can annotate data with notes, e.g., on quality or interruptions. Annotations link to instruments performing simultaneous measurements.

Researchers have a clear overview of the data and can export them in various formats. The system makes sharing data with the community-established repositories much easier. The solution allows volunteers to receive their brain scan results electronically after the session.

Scientific dataset management system for the research institute based on Onedata

Tomáš Svoboda, Adrián Rošinec, Tomáš Raček, Josef Handl, Aleš Křenek, Radka Svobodová - Masaryk University

As the volume and complexity of scientific data continue to grow, the efficient management of data across its entire lifecycle has become paramount. In this context, we have decided to create a system for CEITEC Research Institute, which would allow emerging data sets to be registered and managed, using the existing Onedata system as the data layer.

At its core, Onedata oversees the entire data lifecycle, commencing with the acquisition of data from various connected instruments (cryo-EM, NMR, light microscopy) at the moment of data generation. The automated processes employed by the system enable the organisation of acquired data into coherent datasets, enriched with metadata harvested directly from the instruments themselves and the execution of workflows designed to generate data-aware metadata annotations where feasible, in accordance with defined metadata schemas established in specific fields. This facilitates the creation of FAIR datasets which are ready for publication in thematic data repositories, as and when required.

The ability to integrate heterogeneous storage capacity with heterogeneous high-performance computing (HPC) platforms, such as Jupyter notebooks and Kubernetes container clouds, is a significant advantage. By facilitating the connection between storage capacity and direct access to compute resources, Onedata enables access to compute resources for data analysis, thereby accelerating scientific discovery.

Finally, the ability to share live data via Onedata enables data sharing within and beyond the research group. Once the analysis has been completed, the system is prepared to allow scientists to easily complete and publish the final dataset to the thematic data repositories.

The objective of this poster is to illustrate the development of tools that will facilitate and streamline data sharing among scientific communities at the national and international levels. These tools are intended to support the principles of FAIR and Open Science.

Data issues in HLA-KIR interaction assessment workflow

Katerina Wolf[1], Monika Holubova[2,3], Pavel Jindra[3,4], Robin Klieber[2,3] and Lucie Houdova[1] - [1.]Faculty of Applied Sciences, University of West Bohemia; [2.]Faculty of Medicine in Pilsen, Charles University; [3.]Department of Haematology and Oncology, University Hospital Pilsen; [4.]Czech National Marrow Donors Registry

Even though creating an analytical pipeline for the investigation of any biological markers is becoming standard the main weakness remains the same as in the workflow itself and that is dependency on data. To make the workflow of data processing and results interpretation run smoothly it is necessary to deal with different data issues. These can be divided into two parts: 1. input data and 2. data avability. Input data can be sequenced data or already preprocessing data from another pipeline. Data availability represents reference data or availability database containing knowledge about biological markers' relationship. These data issues influence not only the correctness of results but also their precision. In the case of investigating HLA-KIR interaction is no different.

Donor selection is one of the key factor influencing the outcome of haematopoietic stem cell transplantation (HSCT). The donor is being selected based on a Human Leukocyte Antigen (HLA) match followed by other non- immunogenetic and immunogenetic parameters mostly based on Killer-cell immunoglobulin-like receptor (KIR). HLA-KIR interaction regulates the function of natural killer cells (most critical cells in the early post-transplant period) and it is rationally considered as the next parameter for donor selection. However, the HLA-KIR interaction has been investigated for two decades with contradictory results. So it is time to consider data influence.

For evaluated HLA-KIR interaction it is necessary that each input sample had to contain information about the patient's HLA and donor KIR typing. The first step was to convert the obtained HLA gene sequence into antigen groups binding KIR (C1, C2, Bw4) and Bw6. The next step was to evaluate interaction based on KIRs and their ligands's presence/absence. During pipeline preparation, possible forms of data on input (such as the low resolution of HLA typing) were analyzed as well as data availability (such as unknown ligands for some KIRs). It was also considered an achievable solution for this issues.

This work was funded by Ministry of Education, Youth and Sports, Czech Republic grant No. LM2023055 Czech National Infrastructure for Biological data (ELIXIR CZ) and specific university research project SGS-2022-022

List of participants

Zahra Aliakbartehrani
Institute of Biotechnology of the CAS

Matěj Antol
Masaryk University

Kateřina Balážová Faltejsková
Institute of Organic Chemistry
and Biochemistry of the CAS

Petr Baldrian
Czech Science Foundation

Jitka Baťková
Charles University

Václav Bazgier
Palacký University Olomouc

Karel Berka
Palacký University Olomouc

Lada Biedermannová
Institute of Biotechnology of the CAS

Dominika Bohuslavová
Czech Technical University in Prague

Jana Borůvková
Masaryk University

Paulína Božíková
Institute of Biotechnology of the CAS

Vojtěch Bystrý
Masaryk University

Arzuv Caryjeva
Czech Academy of Sciences, Institute
of Microbiology

Marek Cebecauer
J. Heyrovský Institute of Physical Chemistry
of the CAS

Martin Čech
Institute of Organic Chemistry and
Biochemistry of the CAS

Jiří Černý
Institute of Biotechnology of the CAS

Jan Dohnálek
Institute of Biotechnology of the CAS

Kristýna Dostálová
Institute of Physics of the ASCR

Jiří Fatka
University of West Bohemia

Christos Feidakis
Charles University

Věra Franková
Charles University

Barbora Gergelová
Institute of Ethnology of the CAS

Martin Golasowski
VSB-TU Ostrava, IT4Innovations

Ansgar Gruber
Biology Centre of the CAS

Jiří Grulich
Charles University

Dagmar Hanzlíková
Charles University

Matyáš Hiřman
Charles University

David Hoksza
Charles University

Kristýna Holzerová
Institute of Physiology CAS

Milan Janíček
Charles University

Niclas Jareborg
National Bioinformatics Infrastructure Sweden

Jan Jarolímek
Czech University of Life Sciences Prague

Jan Jelínek
Institute of Microbiology of the CAS

Jaroslav Juráček
Masaryk University

Soňa Kehmová
Mendel Unievrsity in Brno

Jana Klánová
Recetox Brno, Masaryk University

Vojtěch Knaisl
Czech Technical University in Prague

Michal Kolář
Institute of Molecular Genetics of the CAS

Martin Kolisko
Biology Centre of the CAS

Kryštof Komanec
Czech Technical University in Prague

Georgia Koutentaki
Czech Technical University in Prague

Jaromír Kovárník
University of South Bohemia in České Budějovice

Vladimír Krylov
Charles University

Aleš Křenek
Masaryk University

Jan Kubovčiak
Institute of Molecular Genetics of the CAS

Kseniia Kushnir
University of Chemistry and Technology, Prague

Hana Litavská
Czech Technical University in Prague

Radka Lukášová
Charles University

Jiří Macas
Biology Centre of the CAS

Jan Macháň
Palacký University Olomouc

Jiří Marek
Masaryk University

Jana Martínková
Czech Technical University in Prague

Jan Martinovič
VSB-TU Ostrava, IT4Innovations

Luděk Matyska
Masaryk University

Anna Mauci
Institute of Computer Science of the CAS

Veronika Milatová
University of Chemistry and Technology, Prague

Martin Mokrejš
Institute of Biotechnology of the CAS

Pavla Navratilová
Institute of Experimental Botany of the CAS

Pavel Neumann
Biology Centre of the CAS

Jiří Nováček
Masaryk University

Petr Novák
Biology Centre of the CAS

Vendula Novotná
Institute of Ethnology of the CAS

Petr Novotný
Charles University

Łukasz Opioła
ACC Cyfronet AGH, Krakow

Josef Pánek
Institute of Microbiology of the CAS

Petr Pavliš
Palacký University Olomouc

Šimon Pavlů
Palacký University Olomouc

Robert Pergl
Czech Technical University in Prague

Karolina Podloucká
National Library of Technology

Elliott Price
Recetox Brno, Masaryk University

Katarína Řiháčková
Masaryk University

Adrián Rošinec
Masaryk University

Miroslav Ruda
CESNET, z.s.p.o.

Michal Růžička
Masaryk University

Marie Šafner
Institute of Organic Chemistry and Biochemistry of the CAS

Martin Schätz
University of Chemistry and Technology, Prague

Marek Schwarz
Institute of Microbiology of the CAS

Johannes Schweichhart
Biology Centre CAS, HBU

Tereza Šírová
Institute of Computer Science of the CAS

Petr Škoda
Charles University

Terézia Slanináková
Masaryk University

Jan Slifka
Czech Technical University in Prague

Martina Šmardová
Mendel University in Brno

Anna Špačková
Palacký University Olomouc

Vojtěch Spiwok
University of Chemistry and Technology, Prague

Michal Stočes
Czech University of Life Sciences Prague

Anna Strachotová
Institute of Organic Chemistry and Biochemistry of the CAS

Marek Suchánek
Czech Technical University in Prague

Tomáš Svoboda
Masaryk University

Radka Svobodová
Masaryk University

Tomáš Talášek
Palacký University Olomouc

Libuše Vaněčková
Charles University

Tomáš Větrovský
Institute of Microbiology of the CAS

Rudolf Vohnout
University of South Bohemia in České Budějovice

Marta Vohnoutová
Biology Centre of the CAS

Jiří Vondrášek
Institute of Organic Chemistry and Biochemistry of the CAS

Tereza Votrubová
Institute of Organic Chemistry and Biochemistry of the CAS

Kateřina Wolf
University of West Bohemia

Veronika Zemanová
Institute of Czech Literature of the CAS

Notes


GDPR
PPT
HTML
CSV
API
SSL


GDPR
PPT
HTML
CSV
API
SSL

Published by Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences and ELIXIR CZ.

Editors: Jiří Vondrášek, Anna Strachotová, Tereza Votrubová
Print: VENICE Praha s.r.o.
Year: 2024
Pages: 73

Organised with the support of MŠMT – large infrastructure project ELIXIR-CZ (Grant LM2023055)