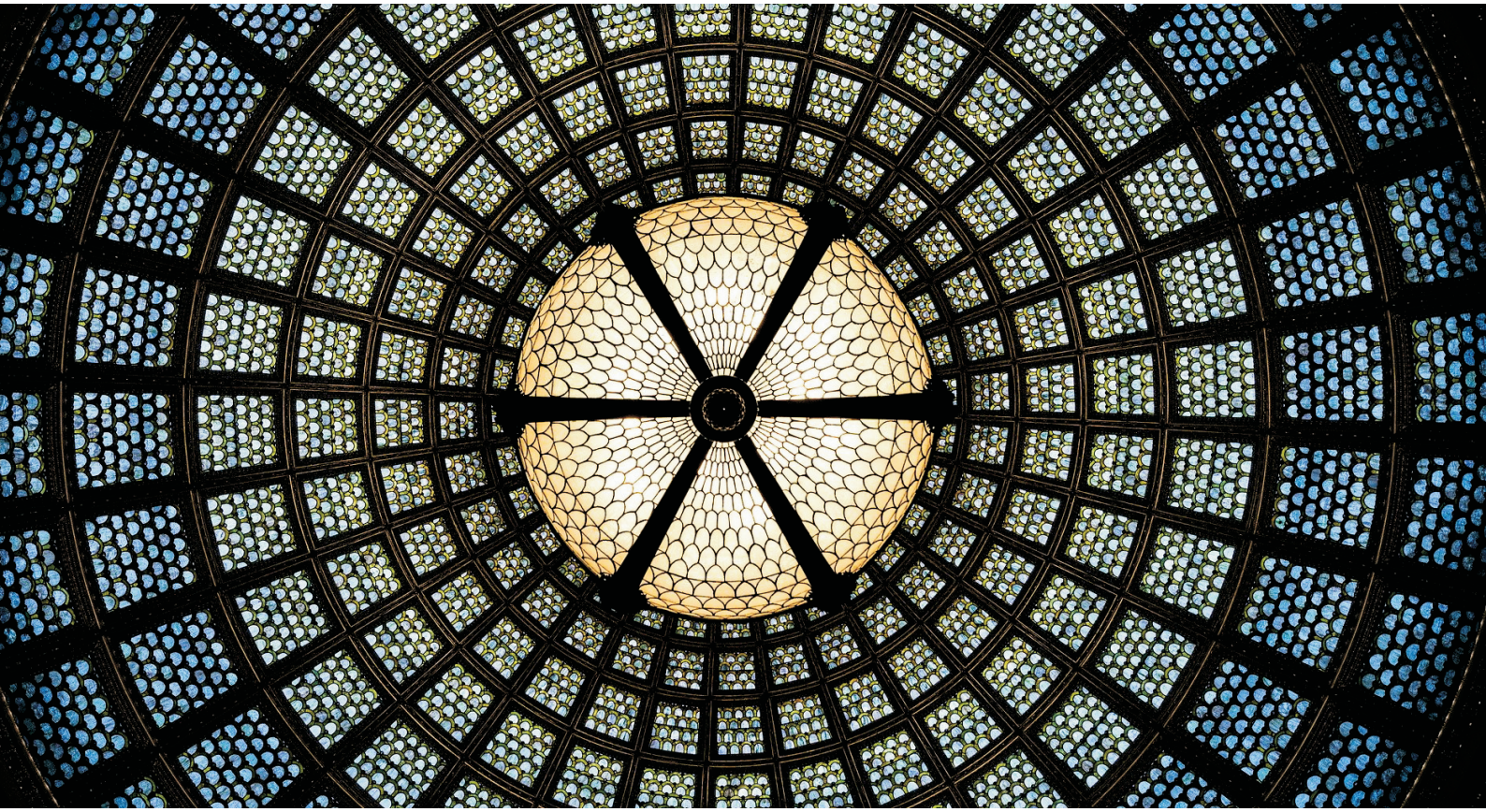
BEILSTEIN SYMPOSIUM

Transparent Data for Effective Research



Beilstein Open Science Symposium 2021

October 5-7, 2021
Online

The Beilstein-Institut and Open Science

The non-profit Beilstein-Institut is one of the most respected organizations in the communication and dissemination of high-quality information in chemistry. Since 1951, when the foundation was established by the Max Planck Society, we have been fulfilling our mission to support the scientific community by providing high-quality information that is essential for research.

Our role has evolved over the years: from the production of the Beilstein Handbook and Database, to being one of the first open access journal publishers in chemistry, to host of interdisciplinary symposia and supporter of open data initiatives. We believe that free access to scientific research results, giving everyone in the world an equal chance to participate in the exchange of experimental findings and data, is the best way to advance science.

Open Science is a new paradigm to scientific research. It is based on cooperation and creates new ways to disseminate information and broaden knowledge through digital technologies and new collaborative tools. It aims to make the primary outputs of publicly funded research results – publications (open access) and the research data (open data) – publicly accessible in digital format with no or minimal restrictions.

The Beilstein-Institut supports open science and makes the results of its projects freely available to the scientific community, which is an essential contribution to the foundation's mission to advance the chemical and related sciences. All journal articles, conference proceedings and videos are open access to allow the worldwide, unhindered sharing and exchange of ideas. This allows scientists, students, educators and the public the opportunity to inform themselves of the latest developments in research and to build on these ideas to further advance scientific knowledge.

Our two platinum open access journals, the *Beilstein Journal of Organic Chemistry* and the *Beilstein Journal of Nanotechnology*, which we fully fund, have no fees for authors or readers. Both journals are produced and managed by the Beilstein Editorial Office team, who work together with a global scientific network of experts that are responsible for the peer review, ensuring high quality science is published. In 2015, the Beilstein Journals were awarded the DOAJ Seal which recognizes the exceptionally high level of publishing standards and best practices adhering to these journals. Both journals publish thematic issues on subjects of high contemporary interest; these are often edited by guest editors, further expanding our network and outreach.

Launched in April 2019, a new addition to our publishing platform is the Beilstein Archives. This is the preprint server for the Beilstein journals. During the manuscript submission process, authors have the option to request that the manuscript be posted as a preprint, which is the version of the manuscript before peer review. Our preprints are posted on average within 2 days, allowing authors to rapidly disseminate their research results and ensuring that they claim priority for their work. Currently, over 30% of our authors select the preprint option.

The Beilstein-Institut runs two data standards projects: STRENDA which is concerned with the reporting of enzymology data and MIRAGE which is working on guidelines for the reporting of glycomics experimental results. Both of which are now widely accepted and acknowledged by the scientific community.

The direct interaction and the exchange of thoughts and ideas between scientists are supported by a program of regularly hosted symposia. These international meetings are organized by the Beilstein-Institut and cover a variety of topics ranging from organic chemistry and biochemistry, to nanotechnology and open science, as well as interdisciplinary meetings on contemporary topics. The Beilstein-Institut has been hosting symposia since 1988. Each meeting is always a unique event, and the lively and intense exchange of thoughts and ideas of the participants turn it into a memorable and lasting experience. The number of participants is usually limited to around 50 and the program is designed specifically to allow sufficient time for discussions. The talks also provide a framework and catalyse discussions which often go on into the night and have led to subsequent cooperation projects. The resulting exchange between researchers, at all stages of their careers, is the underlying goal of the meeting and gives the Beilstein Symposium their unique character.

More information can be found on our website:
www.beilstein-institut.de

Upcoming symposia in 2022:

Beilstein Organic Chemistry Symposium 2022 *Stereoselective Alkene Functionalizations*	Scientific Program: Thomas Wirth
April 26-28, 2022, Rüdesheim, Germany	https://www.beilstein-institut.de/en/symposia/org-chem-stereo-alkene-functions

Beilstein Bozen Symposium 2022

Simulation and AI in the Future of Science

May 17-19, 2022, Rüdesheim, Germany

Scientific Program:
Lee Cronin, Tim Clark, Christian Kramer and Carsten Kettner

http://www.bozen.beilstein-symposia.org

Book of Abstracts

Table of Contents

Overview

Rich digital data underlie modern research work at every step, and how these data are handled, structured and shared has a major impact on the pace of scientific progress. New technologies that generate massive datasets, interconnect lab devices, and enable automated experiments, challenge traditional methods for disseminating scholarly knowledge. The FAIR and open data science movements present complementary visions for more transparent, reproducible and digitally-savvy future for our research. These movements recognize that for complex research data to be shared and used effectively, we need powerful infrastructures, as well as community-supported reporting standards and well-designed exchange formats for data and metadata. At the same time, success requires not only an investment in infrastructure and formats, but also a change in the culture of data generation and sharing among researchers.

This symposium will address the many ways that data transparency contributes to the research progress. Attendees will learn from leaders in the FAIR and open data movements who are advocating for better practices in their fields and demonstrating their commitment in their own work.

Talks will present tools that can help researchers better handle and share data, and methods to combat common statistical abuses that undermine reliability and reproducibility. Key topics on the reproducible sharing of methods and code will also be covered, since both are essential for data to be made usable and transparent.

A series of talks will also be presented by influential publishers, funders and others who are developing policies that incentivize and support data management and sharing. In addition, we will look at the essential services and infrastructure needed to support modern data-driven science, from repositories to high-performance computing, and the necessary role of international cooperation in maintaining these resources.

The symposium will cover a wide range of research fields, including biomedical research, physics and social science, and will also explore how open data practices are transforming sectors outside academia, like health care and journalism.

Enjoy the Symposium!

Andrew Hufton & Carsten Kettner

Scientific Committee

Andrew Hufton
Leibniz Institute
DSMZ-German Collection of
Microorganisms and Cell Cultures
Braunschweig, Germany

andrewlee.hufton@yahoo.com

Carsten Kettner
Beilstein-Institut
Frankfurt am Main, Germany

ckettner@beilstein-institut.de

The Symposium

The symposium will be held online from 5 to 7 October, 2021 on the DGM Inventum platform. Lectures will be life streamed and we will have Q&A sessions after the life talks. These life talks will take place by using Zoom. All talks will be recorded and made available to only those participants who are registered with the platform.

This platform is open for registered participants only.

We will also use the DGM Inventum platform for the display of the posters, Q&A at the posters in both via written chats and live chats in dedicated rooms with up to 25 participants.

In addition, there will be dedicated meeting rooms that allow all registered participants to meet the speakers, to meet at the virtual coffee bar for informal chats in dedicated meeting rooms with up to 25 participants, and to talk in 1:1 chats. This networking is carried out using Jitsi Meet which runs on DGM Inventum servers.

For data protection details, please refer to the Privacy Policy.

Presentations of Posters

Poster Exhibition:
Throughout the conference

Location of the posters

There will be a poster exhibition on the DGM platform.

Posters may be uploaded on the platform on request by the conference service provider.

The size of your poster board is 120 x 90 cm (height x width). Hanging material for the poster boards will be provided on-site.

Presentations

The oral poster presentations will take place as indicated in the scientific program. The presentations should not exceed 10 min. You will have 1 min in addition for questions.

There is also any opportunity to chat with visitors on Jitsi Meet, the inbuilt audio-video conferencing system on the platform

Scientific Program

Tuesday, 5th October

14.00pm CEST	Opening and Introductory Remarks	Carsten Kettner and Andrew Hufton
	Session Chair: Beth Plale	
14.20	Transparent Data and Effective Research both Depend on Software!	Andrew Treloar
14.50	Reproducibility: Foundation of Science – or – Invitation to Test Trivial Hypothesis?	Ulrich Dirnagl
15.20	CODECHECK: an Open System for Checking the Reproduction of Computations Underlying Scientific Articles	Stephen Eglen
15.50	Resolving the Paradox: Incentivising Detailed High-quality Metadata Through Metadata-driven Applications	Michael Hanke
16.20	Discussion and Networking	
17.00	*Break*	
	Session Chair: Stuart Chalk	
17.20	Building Standards and Communities to Boost your Research	Melanie Courtot
17.50	Making Research Data Transparent: Making Data Citation Count	Rachael Lammey
18.20	MassBank on Ice: Spectral Data Meets Git	Steffen Neumann
18.50	Methods Matter for Reproducible Research	Emma Ganley
19.20	Discussion and Networking	
20.00	*Close*	

Wednesday, 6th October

14.00	Opening	Andrew Hufton and Carsten Kettner
	Session Chair: Stephanie Dawson	
14.10	Data Science Meets Open Science – What Comes Next?	Philip E. Bourne
14.40	Open Science – Trusted Results	Beth A. Plale
15.10	NFDI4Chem: Shaping Digital and Cultural Change in Chemistry	Sonja Herren-Pawlis
15.40	Lessons from the UK: Data Access, Patient Trust and Real World Impact with Health Data Science	Varsha Khodiyar
16.10	Discussion and Networking	
16.50	*Break*	
17.10	*Poster Flash Talk Session*	
	Session Chair: Ian Bruno	
17.40	Making Research Data Ready for Open Science: Universities Responding to Challenges in a Heterogenous Environment	Heidi Imker
18.10	The Global Biodata Coalition	Chuck Cook
18.40	Going Beyond FAIR to Create a Connected Data Ecosystem	Susan K. Gregurick
19.10	Discussion and Networking	
19.50	*Close*	

Thursday, 7th October

14.00	Opening	Carsten Kettner and Andrew Hufton
	Session Chair: Heidi Imker	
14.10	Open Science at the Swiss National Science Foundation: from Open Access to Open Data	Matthias Egger
14.40	Societies in Transition	Stuart Taylor
15.10	Game Over: We must urgently Agree on How to Reform Academic Evaluation and Reward Systems	Toma Susi
15.40	Incentivizing FAIR Practices Through Policies and Rigor and Transparency Initiatives. The TOP Guidelines and RTI Initiatives for Funders and Publishers	David Mellor
16.10	Discussion and Networking	
16.50	*Break*	
	Session Chair: Matthias Egger	
17.10	Open and Transparent Research in the COVID-19 Pandemic and Beyond	Rima-Maria Rahal
17.40	Leveraging Open Data for Journalism and Civic Impact in Africa	Tricia Govindasamy
18.10	Panel Discussion and Conclusion	
19.00	*Close*	

List of Posters

The poster presentation includes a short (10 min) oral presentation on Wednesday, 6th October, and the poster sessions during the breaks. The posters will be displayed on the DGM platform throughout the entire symposium from Tuesday, 5th October, to Thursday, 7th October.

Please find the poster abstracts at the end of this abstract book.

Wednesday, 6 October,
17.10 – 18.00 (CEST)

#1	Book Metadata for an Open Access World	Stephanie Dawson
#2	OpenCitations, an Open Infrastructure to Provide Access to Global Scholarly Bibliographic and Citation Data	Silvio Peroni
#3	Global.health: a Data Science Initiative for the Curation of Detailed Public Health Data	Felix Jackson
#4	Open myDNA: Empowering Citizens to Become their own Genome Hackers	Inga Patarcic
#5	Global Disparities in SARS-CoV-2 Genomic Surveillance	Anderson F. Brito

Abstracts

Tuesday

14.20pm CEST

Research Software is the Necessary Lens for Transparent Data

Andrew Treloar

Australian Research Data Commons
Malvern East, VIC
Australia

Software is pervasive in modern life, and research is no exception to this. In a survey conducted by the UK Software Sustainability Institute, around 90% of researchers acknowledged the importance of software, with about 70% saying that their work would not be possible without the use of software. In contemporary research, generation, handling and analysis of data almost always involves software. Digital data always involves software. Highly specialised digital research data almost always requires highly specialised research software. These “research data processes” are the lens through which research data become valuable, but in looking through it, this value is more commonly attributed to the research data rather than the combination of the two. While the role of software engineer and value of software in the modern economy is broadly recognised, the value of the highly skilled coder-scientist, the impact of their research software outputs, and how much research depends on their work, is less recognised.

This presentation will consider the critical role that software plays in providing transparent access to data and ensuring effective research. In order to address these challenges, the Australian Research Data Commons is advocating for a national research software agenda. The presentation will argue the need to view research software as first-class research output, consider the key pillars of a national agenda and provide an implementation update of what is happening in Australia.

Back to Program

Tuesday

14.50

Reproducibility: Foundation of Science – or – Invitation to Test Trivial Hypothesis?

Ulrich Dirnagl

Berlin Institute of Health at Charité
QUEST Center
Berlin, Germany

It appears that we are in the midst of a "replication crisis". At the same time, meta-research has exposed numerous validity threats in the research literature. Researchers enjoy substantial latitude in selecting measures and models for hypothesis testing. Together with publication and related biases this leads to a unreliable body of published research results, which turn out to be not reproducible.

- But what do we actually mean when we say that a study was 'replicated', or that 'it could not be reproduced'? How can we properly interpret a failed replication? How informative is a replication? Most researchers overestimate the additional evidence generated by strict (exact).
- Tacit knowledge and 'unknown unknowns' of hidden moderators confound replication success.
- Confirmatory research carries a stigma: Isn't it done be unoriginal scientists, with a affinity to embarrass fellow researchers?
- The more original initial findings are, the less likely it is that they are true and can subsequently be confirmed. This is a necessary consequence of the fact that cutting edge research must operate with low base rates ('prior probabilities'). On the flip side, the more mainstream and less novel research is, the likelier it will be that it will find its hypotheses to be true.
- Since discovery is unavoidably linked to high false positive rates and cannot support confirmatory inference, dedicated confirmatory investigation is needed for pivotal results.

Reproducibility is not an end in itself. Trivial findings may be highly reproducible. But the new knowledge mapped through exploration is only useful if it is robust and reproducible. To overcome the reproducibility crisis, we need to rethink research reproducibility in education, scientific practice, publishing, funding, and how we incentivize researchers.

Back to Program

Tuesday

15.20

CODECHECK: an Open System for Checking the Reproduction of Computations Underlying Scientific Articles

Stephen Eglen

University of Cambridge
Department of Applied Mathematics
Cambridge, United Kingdom

Computation, in the form of analysis of data or simulations, plays an increasing part of most scientific research and papers. As this computation becomes more complex, it becomes increasingly difficult to summarise it in research articles. Sharing the code and data underlying research articles can therefore help with the communication and understanding of our work. We have developed a workflow called CODECHECK that provides certificates to verify that an independent codechecker has been able to reproduce key findings from a paper using data and code provided by the paper's authors.

In this talk, I will highlight this workflow and how it can integrate with publishing workflows and improve transparency of research reporting.

Back to Program

Tuesday

15.50

Resolving the Paradox: Incentivising Detailed High-quality Metadata through Metadata-driven Applications

Michael Hanke

Research Center Jülich and Heinrich Heine University Düsseldorf
INM-7 / Institute of Systems Neuroscience
Jülich, Germany

The general lack of suitable metadata for most research data is a well-known issue and, today, nobody can escape the call for FAIR research outputs. In my talk I claim that the widespread practice of "data FAIRification" for and at the time of publication is not appropriate for achieving the goals of FAIR, because data expertise and incentives are ill-aligned with the demands and motifs that dominate this stage of the research process.

I propose that the production of high-quality metadata can only be sufficiently incentivized by making the consumption of such metadata an integral part of an initial investigation performed by the original data producers. This conclusion is one outcome of running the open science and open data project studyforrest.org for eight years, and implementing solutions for the research data management and communication requirements of such projects in the DataLad software.

Back to Program

Tuesday

17.20

Building Standards and Communities to Boost your Research

Melanie Courtot

EMBL – European Bioinformatics Institute
Active Infrastructure Team
Hinxton, United Kingdom

In her talk, Dr Courtot highlights the need for metadata standards and data management and transparency to support open, reproducible science. She will present some use cases based on her experience with international consortia such as the Global Alliance for Genomics and Health (https://www.ga4gh.org), CINECA (https://www.cineca-project.eu) and the International HundredK+ Cohorts Consortium (https://ihccglobal.org) and showcase how different communities can align to enable effective research.

Back to Program

Tuesday

17.50

Making Research Data Transparent: Making Data Citation Count

Rachael Lammey

CrossRef
Strategic Initiatives
Oxford, United Kingdom

The increased focus on recognising data as a key research output is a positive development. Repositories, publishers, societies, funders and industry organizations are incentivising or supporting researchers in sharing their data. Part of this support involves providing structure and standards around data, from creation to curation through to submission, deposit and citation.

At Crossref, we've been working with DataCite, the Research Data Alliance (specifically the Scholix Working Group), OASPA and STM to make data citation as standard as citing other works in article reference lists, and in the associated metadata. This provides a mechanism to recognise and reward researchers by showing them how their data is being cited and therefore used. This usage is also key however, and initiatives like Make Data Count are exploring how to make sure that there are responsible and meaningful approaches to research data assessment.

None of these are easy tasks, but progressing this work as a community is key to creating a transparent ecosystem around research data.

Back to Program

Tuesday

18.20

MassBank on Ice: Spectral Data Meets Git

Steffen Neumann

Leibniz Institute of Plant Biochemistry
Bioinformatics and Scientific Data
Halle, Germany

This talk is about the spectral library system MassBank, where it came from and how much Open Data is in there today, and how we got there. We use approaches like version control and continuous integration known from software engeneering for the data itself.

Spoiler: there is even a copy of MassBank on ice!

Back to Program

Tuesday

18.50

Methods Matter for Reproducible Research

Emma Ganley

Protocols.io
Berkeley, CA, United States of America

The power of the internet has pushed us much closer to achieving Open Science and Open Research; Open access to research articles and preservation and availability of research data are both important in moving towards this goal. However, often research articles do not contain sufficient methodological information for the experiments and results to be reproduced. A lack of reproducibility diminishes the value of the research; results, interpretation and conclusions can no longer be properly verified/validated and any resulting data will be missing the context of how it was generated. Although great advances have been made to ensure we have better access to articles and data, the same has not happened to ensure adequate reporting of methodological approaches and protocols.

For Open Research to live up to its potential, mechanisms to improve on methods reporting and availability need to be explored, one option to address this are implementing modular approaches to making research outputs available in a FAIR (Findable, Accessible, Interoperable, and Reusable) manner—using dynamic online tools like protocols.io can help to address this current underappreciated challenge to ensure research is as reproducible as possible.

Back to Program

Wednesday

14.10

Data Science Meets Open Science – What Comes Next?

Philip E. Bourne

University of Virginia
School of Data Science & Department of Biomedical Engineering
Charlottesville, VA, United States of America

The explosion of academic data science initiatives around the world would arguably not have happened if not for open science – the availability of large amounts of data, methods, protocols all intended, at least, to be FAIR. Is there a virtuous cycle developing? In other words, is data science further driving open science or is it having a different impact?

We will address these questions from the vantage point of our own, newly launched, School of Data Science.

Back to Program

Wednesday

14.40

Open Science – Trusted Results

Beth A. Plale

Indiana University Bloomington
Department of Intelligent Systems
Bloomington, IN, United States of America

Open science is a principle of openness that, when applied across the entire scientific research enterprise, holds significant promise to advance the frontiers of knowledge and help ensure a nation's prosperity. Gains in open science in the US for data, software and other products of research are complicated by a highly distributed landscape of solution providers across the landscape of providers supporting US National Science Foundation funded science. Lightweight harmonization to effective practices for data is a path forward.

The reuse of the products of research (the "R" in FAIR principles where products are Findable, Accessible, Interoperable, and Reusable) has an element of trust. A researcher or end-user trusts the dataset, software, research result, or decision. A research result that is packaged for reproducibility or replicability carries with it a higher degree of trustworthiness. But reproducibility at publication submission time, we argue, is neither meaningful nor scalable, especially for research in high performance computing (HPC).

Trust has a whole new import for artificial intelligence (AI). *"On artificial intelligence, trust is a must, not a nice-to-have"* states M. Vestager, EU commission Exec VP overseeing digital policy. An AI research infrastructure that democratizes access to resources that fuel AI development is one that draws on earlier scientific research in containerized workflow systems, synthetic data generation, persistent IDs, model commons, and secure computing environments. We discuss how these work together and our effort to measure AI accountability in accordance with properties that are exhibited vertically throughout the software stack of the AI research infrastructure.

Back to Program

Wednesday

15.10

NFDI4Chem: Shaping Digital and Cultural Change in Chemistry

Sonja Herres-Pawlis

RWTH Aachen University
Department of Chemistry
Aachen, Germany

More and more chemical digital research data are generated. So, new concepts are essential: In which data formats can data be stored in the long term? How and where can data be stored? Which information of the experiment / the simulation should be noted in the metadata? How can these data be made accessible for group members and other researchers? How can these data be made findable for researchers and AI algorithms?

In chemistry, NFDI4Chem tackles the challenges of research data management nationally but also internationally (e.g in the context of IUPAC, EOSC und RDA).[1-3] NFDI4Chem is a consortium consisting of representatives from universities and non-university research institutions, infrastructure facilities and the German Chemical Society, the Bunsen Society and the German Pharmaceutical Society. The focus of the NFDI4Chem consortium is on the molecule itself, its properties and reactions. The project started on 1st October 2020 and this talk will recapitulate the challenges and first steps already done. Moreover, it highlights best practice examples for the chemical community.

I report on a workshop we have designed and executed at Utrecht University to raise awareness about the notion of the commons and its potential for developing open science both internationally and inside our university.

References

[1] S. Herres-Pawlis, O. Koepler, C. Steinbeck, *Angew. Chem. Int. Ed.* **2019**, *58*, 10766-10768

[2] Steinbeck C, Koepler O, Bach F, Herres-Pawlis S, Jung N, Liermann JC, Neumann S, Razum M, Baldauf C, Biedermann F, Bocklitz TW, Boehm F, Broda F, Czodrowski P, Engel T, Hicks MG, Kast SM, Kettner C, Koch W, Lanza G, Link A, Mata RA, Nagel WE, Porzel A, Schlörer N, Schulze T, Weinig H-G, Wenzel W, Wessjohann LA, Wulle S, *Research Ideas and Outcomes*, **2020**, *6*: e55852.

[3] https://www.nfdi4chem.de/

Back to Program

Wednesday

15.40

Lessons from the UK: Data Access, Patient Trust & Real World Impact with Health Data Science

Varsha Khodiyar

Health Data Research UK – HDR Alliance
Data & Connectivity
London, United Kingdom

Health Data Research UK's mission is to unite the UK's health data to enable discoveries that improve people's lives. Our 20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health. A key part of HDR UK's vision is our data portal, the Innovation Gateway. The Gateway facilitates discovery of healthcare data and simplifies data request procedures across multiple data custodians. The Gateway contains metadata on a variety of datasets, including those related to COVID-19, cardiovascular, maternal health, emergency care, primary care, secondary care, acute care, palliative care, biobanks, research cohorts and deeply phenotyped patient cohorts.

From the outset HDR UK has sought the voices, views and experiences of patient and lay-public groups to ensure there is transparency and clear public benefit in the use of the UK's health data. Patient and public involvement is key to making the Gateway accessible, transparent and to ensure public confidence in research access to health data. The importance of public outreach combined with providing research access to data is illustrated with HDR UK's contribution to the UK's coronavirus pandemic response. HDR UK was tasked by the UK's Chief Scientific Office to build and facilitate the infrastructure to support the National Core Studies, providing key insights on the evolving situation to UK policy makers during the course of the pandemic.

In this talk, I will show how HDR UK is enabling open science by facilitating the discovery of health data, and simplifying the process of requesting access to multiple datasets. I'll discuss HDR UK's approach to embedding transparency on research data usage for patients and public, and summarise some of the key ways in which HDR UK has contributed to the coronavirus pandemic.

Back to Program

Wednesday

18.00

Making Research Data Ready for Open Science: Universities Responding to Challenges in a Heterogenous Environment

Heidi J. Imker

University of Illinois at Urbana Champaign
University Library
Urbana, IL, United States of America

As Open Science matures, activities at research universities move from being focused on motivating buy-in to enabling practice. In this talk, I'll present three different examples of how US universities are addressing the practical challenges associated with data sharing, especially given the volume and variety of data created at any single university. To show the breadth of strategies, the examples will cover efforts at three levels: by a single university (nudging better data management practices), several universities (curating data across institutional boundaries), and many universities (guiding conversations and frameworks for action).

These examples are deceptively simple, but each relies on coordination and cooperation across boundaries, both geographic and functional. Furthermore, they put the principles of Open Science into practice by creating resources primed for adoption, adaptation, and reuse by others. As Open Science practices become more common, there are opportunities for universities as a whole to leverage Open Science for the creation of institutional objectives and products that have broad appeal and utility.

Back to Program

Wednesday

18.30

The Global Biodata Coalition

Chuck Cook

Global Biodata Coalition
Cambridge, United Kingdom

Research in the life sciences is data-driven and dependent on the data integration and analysis enabled by open-access biodata resources. These resources form a major global infrastructure that is crucial to current and future biomedical and life sciences research, but the infrastructure has developed organically without systematic coordination among funders or data resource managers, and there is no comprehensive understanding of the infrastructure at a global level. Many key biodata resources rely on short-term funding from a small set of funders and are at risk of losing their funding.

The Global Biodata Coalition (GBC) is supported by, and works with, public and charitable research funders to address the challenges in maintaining the biodata infrastructure. The GBC is a forum for funders of data resources to coordinate and share approaches for the efficient management and growth of the infrastructure, and will work with funders to stabilize and ensure support for the global biodata infrastructure.

I will provide a brief overview of the Global Biodata Coalition and current activities. Specifically I will describe ongoing work to undertake a global inventory of biodata resources that will provide baseline data for the size and scope of the publicly available infrastructure of biodata resources. Additionally, I will describe our work to identify the set of Global Core Biodata Resources that are most essential for biomedical and life sciences research. The goal of this process is to ensure long-term support for these key components that underpin the entire research infrastructure for life sciences.

Back to Program

Wednesday

19.00

Going Beyond FAIR to Create a Connected Data Ecosystem

Susan K. Gregurick

National Institute of Health
Data Science Strategy
Bethesda, MD, United States of America

In this talk the myriad of activities is discussed the NIH is implementing and supporting to go beyond FAIR to create a connected biomedical data ecosystem at the National Institutes of Health. The activities Dr. Gregurick will share are aligned with the NIH Strategic Plan for Data Science, which aims to create an integrated, FAIR biomedical data ecosystem and promote open science.

Back to Program

Thursday

14.10

Open Science at the Swiss National Science Foundation: from Open Access to Open Data

Matthias Egger

Swiss Naational Science Foundation
University of Bern,
Bern, Switzerland

Open science can increase the quality and impact of research. Public research funders like the Swiss National Science Foundation (SNSF) must, therefore, support such developments. Currently, the SNSF, through its funding, creates incentives and rewards to share data and publish open access. In 2001, data sharing was the first open science aspect introduced at the SNSF. Since fostering open science represents a change in research culture, implementing open science measures requires getting researchers, research institutions, and many other stakeholders on board. This talk will present the steps the SNSF went through in the past 20 years, from observing the evolution of open science practices, revising its own requirements, and explaining and justifying them to current and future grantees and research institutions in Switzerland. It will provide examples of how sustainable open science practices and infrastructures can be established through joined-up action of funders, higher education institutions, and government: the Swiss National Strategies on Open Access and Open Research Data.

Back to Program

Thursday

14.40

Societies in Transition

Stuart Taylor

The Royal Society
London, United Kingdom

The Royal Society's publishing operation is committed to the principles of open science. In 2020 almost half of our total published articles were open access. We permit authors to self-archive their accepted manuscript version (AAM) in a repository with no embargo and with an open licence. All journals have a strict open data policy with authors being required to provide a data availability statement as a condition of publication and we cover any deposit charges they may incur to deposit their data sets in repositories. All our references are open (in accordance with I4OC) and all our abstracts are open (in accordance with I4OA). We permit text and data mining for both commercial and non-commercial purposes and four of our journals operate open peer review.

We have recently made the decision to transition our hybrid research journals to full open access over the next five years. Of course, learned societies face considerable challenges in making the transition to open access and this is the reason behind our co-founding of the Society Publishers' Coalition in 2018, which now has almost 100 member societies working together to make open access a reality.

Back to Program

Thursday

15.10

Game Over: We must urgently Agree on How to Reform Academic Evaluation and Reward Systems

Susi Toma

University of Vienna
Faculty of Physics
Vienna, Austria

Processes of research evaluation and rewards are coming under increasing scrutiny, with critics convincingly arguing that they have adverse effects on research quality and support a research culture of competition to the detriment of collaboration. Based on my personal perspective, I consider how current systems of research evaluation lock both early career researchers and their supervisors into practices that are deemed necessary to progress academic careers within the current evaluation frameworks [1].

Accordingly, many have called for a wide-ranging reform of research assessment, which almost certainly should include Open Science practices as a core component. The overarching aim would be to enable a positive change in research culture that not only helps improve the transparency, reproducibility and reliability of research, but by rewarding excellent science on its own merits, also de-emphasize those forms of evaluation that contribute to toxic competition.

A prerequisite for a successful reform would arguably be to increase the involvement of researchers themselves in designing reward and evaluation systems. However, during my work on the Open Science task force of the Initiative for Science in Europe (https://initiative-se.eu/), it has become clear that although there is increasing recognition of the need and desire to move away from flawed prestige-based metrics, there is little consensus on what should replace them.

Considering the increasingly strong drive from policy-makers, it is vital that different research communities both urgently and concretely consider how they wish evaluation systems to be adapted to reward OS practices in each of their individual circumstances. If we cannot agree amongst ourselves what we wish to replace impact factors and journal rank with, true change will either not take place – or someone else will decide for us.

Reference

[1] De Herde, Véronique, Mattias Björnmalm, and Toma Susi. 2021. “Game Over: Empower Early Career Researchers to Improve Research Quality”. Insights 34 (1): 15.
DOI: http://doi.org/10.1629/uksg.548

Back to Program

Thursday

15.40

Incentivizing FAIR Practices Through Policies and Rigor and Transparency Initiatives. The TOP Guidelines and RTI Initiatives for Funders and Publishers

David Mellor

Center of Open Science
Charlottesville, VA, United States of America

The Transparency and Openness Promotion (TOP) Guidelines provide journals, publishers, and funders specific, modular, tiered recommendations for implementing open and reproducible research practices into their published articles. TOP Factor evaluates the degree to which journals' policies align with the standards recommended by TOP. TOP Factor currently evaluates over 1,000 journals preliminary evidence shows that comparisons between ones' own journals and a peer group of similar journals encourages policy change and implementation of better, more open practices.

In this session, David Mellor from the Center for Open Science will give the past, present, and future of TOP Factor.

Back to Program

Thursday

17.10

Open and Transparent Research in the COVID-19 Pandemic and Beyond

Rima-Maria Rahal

Max Planck Institute for Research on Collective Goods
Department of Behavioural Law and Economics
Bonn, Germany

During the current COVID-19 pandemic, science needs to move fast and it can seem like the traditional way we do science cannot keep up with the rapid pace at which the crisis is spreading. After working on a research idea, it usually takes years to publish the results. Now, however, things need to move faster. One more reason to embrace open scientific practices. At the same time, the public is talking about science more than ever, from discussing vaccine efficacy to incidence rates and clinical trials.

In the light of this intense interest in scientific output and in the scientific process, I discuss the role of open and transparent research practices in the COVID-19 pandemic, and beyond.

Back to Program

Thursday

17.40

Leveraging Open Data for Journalism and Civic Impact in Africa

Tricia Govindasamy

Code for Africa
Department of Knowledge
Durban, South Africa

The lack of open data is a major challenge hindering decision-making in Africa. Data is very difficult to access, many datasets aren't available online and if it is, it's stuck on websites or pdf reports that need to be scrapped into a machine-readable format to generate insights. Code for Africa (CfA) is the continent's largest network of civic technology and data journalism labs. CfA builds digital democracy solutions that give citizens unfettered access to actionable information that empowers them to make informed decisions, and which strengthens civic engagement for improved public governance and accountability. This includes building infrastructures like the continent's largest open data portals, openAFRICA, and sourceAFRICA.

openAFRICA is Africa's largest Volunteer Driven Open Data Portal. Users can upload datasets which is accessible to anyone for free, this allows for collaborative sharing and reuse. sourceAFRICA is Africa's premier repository for documents that the public needs to help make more informed decisions or that help civic watchdogs such as the media or social justice NGOs substantiate their reportage or documents that have been suppressed or censored elsewhere. It allows you to turn your documents into machine-readable data, for better analysis. It also makes it easier for you to share documents and/or collaborate with your audience and colleagues.

Code for Africa has various other initiatives promoting open data and the reuse of datasets through data-driven projects, networks, fellowships, and training. academy.AFRICA offers free training on data literacy for journalists, researchers, government officials, civil society organisations, and decision-makers. These trainings are available as scheduled webinars or can be taken self-paced on academy.AFRICA's MOOC.

Back to Program

Poster

#1

Book Metadata for an Open Access World

Stephanie Dawson

ScienceOpen
Berlin, Germany

ScienceOpen is a discovery environment and citation index with over 73 million records – including 5 million book and chapter records. But the advantages and requirements of digital aggregation often run into a wall with book metadata. Book metadata systems such as MARC records and ONIX were developed to serve a different purpose – describing books for catalogues, but not linking back to the publisher website, as opposed to JATS XML metadata and Crossref DOIs for journal articles which always assume a single digital version of record. Only a small fraction of academic titles has a DOI, fewer still deposit rich, reusable metadata with Crossref. Simultaneously, open access books are often assigned multiple DOIs by a range of repositories. With a BMBF grant, ScienceOpen has begun to map the range of book metadata challenges and develop a platform to support book publishers, with the goal of better integrating books and chapters into to the digital knowledge graph.

Back to Program

Poster

#2

OpenCitations, an Open Infrastructure to Provide Access to Global Scholarly Bibliographic and Citation Data

Silvio Peroni

University of Bologna
Department of Classicalo Philology and Italian Studies
Bologna, Italy

OpenCitations has been established as a fully free and open infrastructure to provide access to global scholarly bibliographic and citation data as Linked Open Data using Semantic Web technologies, thereby providing a disruptive alternative to traditional proprietary citation indexes. Bibliographic citation data have a fundamental value for the world of scholarship, and their open availability is a crucial requirement for the bibliometrics and scientometrics domain. Besides providing open citation metadata, OpenCitations increases the creation of reproducible metrics for research assessment exercises, develops and share open-source software and open services. OpenCitations, in fact, fully espouses the UNESCO principles of Open Science. It also complies with the Principles of Open Scholarly Infrastructure, the FAIR data principles (data should be findable, accessible, interoperable, and reusable), and with the recommendations of I4OC that citation data, in particular, should be structured, separable, and open. In compliance with these values, one of OpenCitations' main priorities is to keep its services, software and data always without charge under open licenses (CCo for data ISC for software) for fostering maximum reuse. This undeniable aspect leads to an acknowledged sustainability issue, principally in terms of salaries and technical infrastructure costs. However, OpenCitations can rely on an international network of generous supporters, and in 2019 has been selected by SCOSS as an open infrastructure worthy of financial support by research organisations, academic libraries, funders and scholars worldwide. Thanks to this international endorsement, OpenCitations is now looking to the future from two perspectives: firstly, enhance its partnerships and dialogue with scholarly community; secondly, developing new services (such as the database OpenCitations Meta) and expand the citation coverage.

Back to Program

Poster

#3

Global.health: a Data Science Initiative for the Curation of Detailed Public Health Data

Felix Jackson

University of Oxford
Computer Science Department
Oxford, United Kingdom

As the COVID-19 pandemic emerged, global governments turned to epidemiologists for insights to guide response strategies. These insights relied on open-access, high-quality, international and real-time epidemiological data. However, detailed public health data were rarely reported in a standardised form making comparisons across regions difficult. Considerable time and effort was therefore needed to identify, process and standardise epidemiological data for each country, before any multi-country analysis could be conducted.

Global.health (https://www.global.health/) changed this by collating disparate linelist datasets from around the world into a single online platform, which allows both visualisation and easy access to real-time epidemiological data. Our COVID-19 dataset currently contains rich information on over 50 million anonymized cases with over a billion data points. Our platform is reproducible and adaptable to any emerging infectious disease in the future.

We present the challenges of building this open-access data platform, including the tools for designing standards and assuring data quality, scaling the database as case counts exploded, and how to provide open-access data to encourage global use whilst preserving privacy. We'll also discuss research opportunities enabled by this multi-country database and the emerging need for data and analytics to be decentralised.

Back to Program

Poster

#4

Open myDNA: Empowering Citizens to Become their own Genome Hackers

Inga Patarcic

MDC Berlin
Department of Communitcation and Outreach
Berlin, Germany

More than 27 million people have had their genome sequenced or purchased direct-to-consumer (DTC) tests to genotype their genomes (MIT Technology Reviews). The growing interest of citizens to understand their genetic background doubled in 2018 - as many DTC tests were purchased as in all other years before. However, most consumers of DTC rely on the analysis that was provided by the company and do not understand that they could get more insight from their data.

I have purchased a myHeritage DTC and have designed an R package to see how much information I could extract from my genotyping data. This has now turned into the "myDNA" project including a Github blog.

I realized that working on DTC data is a very good way to educate citizens about the advances and limitations of the DTC tests, encourage them to think critically about genome analyses and learn programming. I have now tested how the myDNA works with high-school students and scientist colleagues. Interestingly, myDNA has the potential to convert scientists into citizen scientists, as it encourages them to donate their genotyping data for research and forces them to reflect on all implications of such data donations.

In this talk, I will present how myDNA can be used for education and citizen science projects and introduce the audience to the possibilities and limitations of DTC.

Back to Program

Poster

#5

Global Disparities in SARS-CoV-2 Genomic Surveillance

Anderson F. Brito

Instituto Todos pela Saúde
Saõ Paulo, Brazil

The COVID-19 pandemic has revealed the importance of virus genome sequencing to guide public health interventions to control virus transmission and understand SARS-CoV-2 evolution. As of July 20th, 2021, >2 million SARS-CoV-2 genomes have been submitted to GISAID, 94% from high income and 6% from low and middle income countries. In this study we investigated the spatial and temporal heterogeneity in SARS-CoV-2 global genomic surveillance efforts. We report the genomic surveillance strategies adopted globally, by analysing the number of sequenced genomes, percentage of sequenced cases, and the turnaround (time between sample collection and genomic data submission), a factor that is influenced by data ownership concerns.

Our study provides a perspective on the global disparities surrounding SARS-CoV-2 genomic surveillance, their causes and consequences, and possible solutions to further maximize the impact of pathogen genome sequencing on public health.

Back to Program