BEILSTEIN SYMPOSIUM

The What, How and Why of Open Science

Beilstein Open Science Symposium 2019

15 – 17 October, 2019
Hotel Jagdschloss Niederwald
Rüdesheim, Germany

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 read and reuse experimental findings and data, is the best way to advance science.

Open Science is a new approach to scientific research. It is based on cooperation and uses 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 restriction.

The Beilstein-Institut supports open science and makes the results of its projects freely available to the scientific community as open access publications. This 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. 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.

An essential prerequisite for open science data is reporting guidelines and technical standards that provide the framework for the exchange of data from one laboratory to another without technical and textual barriers.

The Beilstein-Institut runs two data standards projects: STRENDA which is concerned with the reporting of enzymology data and MIRAGE with 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 an interesting event with an open result: the Beilstein-Institut provides the framework 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. In some ways the talks can be seen as providing a catalyst for these discussions which often go on into the night and have led to subsequent cooperation projects. The resulting exchange between researchers is the underlying goal of the meeting and gives the Beilstein Symposium their unique character.

Regularly updated information about our symposia is available at www.beilstein-symposia.org.

Upcoming symposia in 2019/2020:

Beilstein Nanotechnology Symposium 2019
MXene at the Frontier of the 2D Materials World
15 – 17 October, 2019, Mainz, Germany

Scientific Program:
Yuri Gogotsi, Xinliang Feng, Johanna Rosén
https://www.beilstein-institut.de/en/symposia/nano-2d-materials

Beilstein Nanotechnology Symposium 2019
New Directions for Nanoporous Materials
12 – 14 November 2019, Rüdesheim, Germany

Scientific Program:
Sir Fraser Stoddard, Cafer T. Yavuz
https://www.beilstein-institut.de/en/symposia/nano-porous

Beilstein Organic Chemistry Symposium 2020
Stereoselective Alkene Functionalizations

21 – 23 April, 2020, Rüdesheim, Germany

Scientific Program:
Thomas Wirth

https://www.beilstein-institut.de/en/symposia/org-chem-stereo-alkene-functions

Beilstein Bozen Symposium 2020
Models of Convenience

16 – 18 June, 2020, Rüdesheim, Germany

Scientific Program:
Lee Cronin, Tim Clark, Martin G. Hicks and Carsten Kettner

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

Beilstein Organic Chemistry Symposium 2020
Earth-abundant 3d Metal Catalysis

22 – 24 July, 2020, Mainz, Germany

Scientific Program:
Lutz Ackermann

https://www.beilstein-institut.de/en/symposia/org-chem-3d-metal-catalysis

Book of Abstracts

Table of Contents

Overview

A new science eco-system is growing: Open Science. This is based on the conviction that free access to research publications is not only a moral right of citizens, but a necessity to allow the maximum use and impact of research. The current publishing system is no longer fit for purpose; too much emphasis is being placed on using publications for evaluation and not enough on dissemination of new research results. Redressing the balance will not be easy; the publish-or-perish paradigm is detrimental, and could productively be replaced with a more transparent, effective system based on quality and not quantity.

Change is coming, but some scientists are feeling that they will be restricted in their freedom by being mandated to make data available or ensure that their publications are open. Others welcome change, but are hindered by the lack of a framework for structured and standardized data reporting. Open Science aims to make research and development more effective by better supporting collaboration. This can be between research groups, but also between academia and industry. Adoption of the FAIR data principles is an important step to support this, but much needs to be done to ensure that sufficient tools are available so that making data open is not an onerous task for scientists.

In chemistry, biochemistry and neighbouring areas, funding agencies as well as national and supranational bodies are strongly advocating the sharing and depositing of data. To make this work the incentive structures for academics need to be realigned, investment in infrastructure and new technologies increased, and the awareness of the advantages of making data available for AI and similar technologies heightened. New technologies are diffusing into the lab allowing devices to be interconnected, data automatically recorded, and experiments to be automated.

This symposium addresses the interfaces between the laboratory and the new infrastructures currently being set up. The advantages of making data open will be critically reviewed and the development of highly interconnected, collaborative research in data driven laboratories of the future will be discussed.

The symposium will bring together research scientists, data scientists, publishers, funders and other interested parties to review critically their needs and concerns and discuss how they see the future of Open Science developing.

Enjoy the Symposium!

Scientific Committee

Martin G. Hicks and Carsten Kettner
Beilstein-Institut
Frankfurt am Main, Germany
mhicks@beilstein-institut.de
ckettner@beilstein-institut.de

Registration

All participants must be registered to have access to the conference area.

Participants can ask the organizers for a confirmation of the payment of the conference registration fee. Insurance of participants against accidents, sickness, cancellation, theft, property damage or loss is not covered. Participants are advised to take out adequate personal insurance (see also „Liability and Insurance“).

Participants are responsible for settling their hotel bills directly with the hotel on departure. The total price for participants staying at the Hotel Jagdschloss Niederwald is 567 EUR and includes both accommodation for four nights and the conference package that covers lunches, dinners and coffee breaks as well as admittance to the conference room.

> **Participants not staying at the Hotel Jagdschloss Niederwald are requested to register with the hotel for booking and paying the conference package, i.e. 291 EUR per person.**

Extras, such as drinks, telephone calls etc. are not included in the price.

The Symposium

The symposium will be held from 15 to 17 October, 2019, with the 14th and the 17th (after lunch) for travelling.

The setting and the limited number of participants (max. 50 persons) provide a very convivial atmosphere for the ready exchange of thoughts and ideas.

The scientific program will take place over three days and will

start at 9.00 am on Tuesday, the 15th, and

end around noon (ca. 2.00 pm) on Thursday, the 17th.

If you wish to extend your stay, please contact the hotel directly.

For the length of the individual talks, please refer to the program. Speakers should allow sufficient time for discussion at the end of their talks (e.g. a 40 min slot includes 30 min talk + 10 min for questions). A Windows PC connected to an LCD projector will be available.

Presentations of Posters

Poster Exhibition:
Tuesday, 15th and Wednesday, 16th October during the coffee breaks

Location of the posters

The Poster exhibition will be placed in the coffee room which is close to the conference room. Your poster board will be marked with your poster number which is the same in the abstract book.

Poster mounting

Please mount your poster on the 14th October from 4 pm or on the 15th latest by 8.30 am. Your poster will be on display throughout the Symposium. You are asked to remove all poster materials from the board at the end of the meeting otherwise it will be taken down at the end and disposed by the organizers. The organizers cannot take any responsibility for this material.

Poster Material

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 5 min. You will have 1 min in addition for questions.

Please make sure that you are using the power point template sent out and that you have delivered your final presentation to the organizers in time.

A Windows PC connected to an LCD projector will be available.

Liability and Insurance

The Beilstein-Institut will not be liable for any accident, theft or damage to property, nor for any delays or modification in the program due to unforeseen circumstances.

Participants and accompanying persons are advised to arrange personal travel and health insurance.

Scientific Program

Monday, 14th October

19.00 Welcome reception

19.30 Dinner

Tuesday, 15th October

09.00	Opening and Introductory Remarks	Martin G. Hicks
	Session Chair: Matthew H.Todd	
09.20	Open Science and the Rebuilding of the Publishing Functions	Jean-Claude Guédon
10.00	Building a National Research Data Commons – Transforming Scholarship Through Transformative Infrastructure	Andrew Treloar
10.40	Poster Presentations	F. Schuhmacher, R. Bruno
11.00	*Coffee Break and Poster Session*	
11.30	Implementing Open Science in Academic Biomedicine: a Report from the Trenches	Ulrich Dirnagl
12.10	Distributed Drug Discovery and its Application in Neglected Tropical Diseases	Lori Ferrins
12.50	*Lunch*	
	Session Chair: Evan Bolton	
14.00	Towards Knowledge Graph-based Representation, Augmentation and Exploration of Scholarly Communication	Sören Auer
14.40	Accelerating Biomedical Discovery with an Interest of FAIR Data and Services	Michel Dumontier
15.20	*Tea Break and Conference Photo*	
15.50	Watch out for Influencers! When Scientific Reasoning Relies on a Single Data Point	Andrej-Nikolai Spiess
16.30	Stories from the “Open Science Revolution”: How Scientists Talk about Openness	Rosalind Attenborough

17.10	"The University Cooperative", Learning to Manage Academic Resources as Common Property	Sanli Faez
17.50	*Close*	
19.30	*Dinner*	

Wednesday, 16th October

	Session Chair: Andrew L. Hufton	
09.00	A Universal Chemical Synthesis Text to Code Translator that Enables the Autonomous Synthesis of Organic Molecules	Cole Mathis
09.40	Open Science Platform for Materials Science: AiiDA and Materials Cloud	Giovanni Pizzi
10.20	*Coffee Break and Poster Session*	
10.50	Simulation Foundry: Repeatable, Replicable, Reproducible, Open and FAIR Molecular Modelling	Gudrun Gygli
11.30	Towards an Improved Data Ecosystem for Scientists	Evan Bolton
12.10	Wikidata and Scholia as a Hub Linking Chemical Knowledge	Egon Willighagen
12.50	*Lunch*	
	Session Chair: Marta Teperek	
14.00	Towards a National Research Data Infrastructure for Chemistry in Germany	Christoph Steinbeck
14.40	The GO FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-related Resources	Erik Schultes
15.20	Discussion: The Why, How and What of Open Science	Martin G. Hicks
16.00	*Tea Break*	
16.30	Supporting Ireland's Open Research Agenda – HRB's Open Publishing Platform and FAIR Data Stewardship Pilot	Aileen Sheehy
17.10	FAIR Data? Not Without Code! But How to Get There? - Case Study from TU Delft	Marta Teperek
19.30	*Dinner*	

Thursday, 17th October

	Session Chair: Lori Ferrins	
09.00	Open Science is Accelerating Early Target Discovery and Validation, and Facilitates the Generation of More Novel Medicines for Patients	Chas Bountra
09.40	Can Openness Pay?	Matthew H. Todd
10.20	DataSTAGE: Improving Access to FAIR Data to Accelerate Scientific Discovery for Heart, Lung, Blood, and Sleep Research	Rebecca R. Boyles
11.00	*Coffee Break*	
11.20	Behind der FAIR Brand: Thinkers, Doers and Dreamers	Susanna-Assunta Sansone
12.00	Five Years of Data Sharing at PLOS: Challenges and Opportunities	Iratxe Puebla
12.40	Bridging the Gap Between the Scholarly Literature and Public Data Repositories	Andrew L. Hufton
13.20	Closing Remarks	Martin G. Hicks
13.30	*Lunch*	

List of Posters

The poster presentation includes a short (5 min) oral presentation on Tuesday, 15th October, and the poster sessions during the coffee breaks on Tuesday and Wednesday. The posters will be displayed throughout the entire symposium from Tuesday, 15th October, to Thursday, 17th October.

Tuesday, 15th October

#1	OSALF Concept: An Interface between IoT 4.0, Chemistry and FAIR Realization	Frank Schuhmacher
#2	Science Reproducibility and Reusability with FutureGateway and a Zenodo-like Repository: the PALMS Experiment	Riccardo Bruno
#3	MIRAGE – Minimum Information Required for A Glycomics Experiment	Carsten Kettner
#4	STRENDA DB – The 'PDB' for Enzyme Function Data?	Carsten Kettner

Abstracts

Tuesday

09.20

Open Science and the Rebuilding of the Publishing Functions

Jean-Claude Guédon

Université de Montréal
Department de Littératures et langes du monde
Montréal, Canada

The recent report of the Expert Group on the future of scholarly communication and publishing foregrounds two important elements that deeply shape the communication and publication systems at work in the present knowledge production systems: publishing is made up of several functions which, up to now, have been largely organized around publishers (in the familiar sense of the word), and an evaluation system which simultaneously deals with the intellectual dimensions of research, and the economic dimensions of the publishing world. If it is to be changed and improved, the communication and publication system of knowledge production must clearly separate the economic parameters from the intellectual parameters. Evaluation must directly deal with the publication units (e.g. journal articles).

Open science and open access provide means to achieve this objective, if the registration and certification phases of "publishing" are kept separate from the rest of publishing (dissemination and preservation). Research sites (universities, national laboratories) as well as funding agencies may well play a crucial role in reorganizing scholarly and scientific publishing in this fashion. Learned societies can also find a renewed vocation within this perspective.

Back to Program

Tuesday

10.00

Building a National Research Data Commons – Transforming Scholarship Through Transformative Infrastructure

Andrew Treloar

Australian Research Data Commons
Platforms and Engagements
Malvern East, Vic., Australia

The Australian Research Data Commons (ARDC) project formally began in July 2019, building on the legacy of the Australian National Data Services (ANDS), National eResearch Collaboration Tools and Resources (Nectar), and Research Data Services (RDS) projects. The ARDC is a core element in the Digital Data and eResearch Platforms (DDeRP) national capability identified in the Australian 2016 National Research Infrastructure Roadmap. The ARDC Strategic Plan is based on four strategic pillars and a capstone. The pillars are People and Policy, Software and Platforms, Data and Services, and Storage and Compute. The capstone is Coordination and Coherence, which builds on and connects the pillars. The ARDC is a transformational, sector-wide initiative, working with sector, government, and industry partners to build a coherent national and collaborative research data commons. This will deliver a world-leading data advantage, facilitate innovation, foster collaboration and enhance research translation. This talk will explore the role a national data commons can play in the global move towards more open and innovative scholarship.

Back to Program

Tuesday

Poster #1

OSALF Concept – An Interface between IoT 4.0 and FAIR Realization

Frank Schuhmacher

Free University Berlin
Berlin, Germany

Modern IoT developments enable the networking of laboratory equipment with a reasonable effort. With the implementation of Makerboards, functions can be realized that are not provided by the device manufacturer. This allows hardware from different providers to communicate with each other without restriction. These self-sufficient state machines are the backbone of OSALF. The **O**pen **S**ource & **A**ccess **L**aboratory **F**ramework makes it quick and efficient to assemble different devices for a test setup. The respective functions can be parameterized via graphical user interfaces. Programmable access is possible by defined API functions. These functions are also clustered and parallel available. The realized interfaces to different chemistry editors allow an evaluation of information documented there for direct use in the experimental procedure.

However, the direct and intensive use of electronic laboratory books is recommended. This results in synergy effects regarding the documentation and during the chemical synthesis.

The data generated during the lab work can be merged from different workplaces via cloud implementation. As a unique identifier, a unique hash is generated from the ORCID, the local workstation information and the requested test identifier, under which the respective DOIs are also requested. Since all data is permanently digitized, a complete FAIR implementation is possible. Through this concept, the researcher is relieved of a multitude of time-consuming, boring, unpopular and therefore error-prone activities.

The user can concentrate on the tasks for which a person is needed: the sensible conception of experiments and the creation of new scientific questions.

Back to Program

Tuesday

Poster #2

Science Reproducibility and Reusability with FutureGateway and a Zenodo-like Repository: the PALMS Experiment

Riccardo Bruno[1], R. Barbera[1,2], M. Fargetta[1], R. Rotondo[1], A. Anagnostou[3], S.J. Taylor[3]

[1] Italian National Institute of Nuclear Physics, Divison of Catania, Italy
[2] University of Catania, Department of Physics and Astronomy, Italy
[3] Brunel University, London, United Kingdom

OpenScience (OS) is a powerful and novel paradigm to share knowledge across multidisciplinary scientific communities with the aim to improve the quality of science [1]. One of the most important OS enablers are the FAIR principles, which involves the way to Find, Access, Interoperate and Reuse research data [2]. In most of the cases published scholarship materials are not linked with computed datasets, open source software and/or virtualized computing environments and OS currently lacks of means helping to reproduce and eventually reuse cited results exploiting public or private distributed computing infrastructures (DCIs). Moreover from the final user point of view, the best option would be the use of graphical user interfaces (GUIs) normally hosted by a ScienceGateway (SG) [3] built for a specific scientific community.

The FutureGateway Framework (FGF) [4] consists of a complete software toolkit made of different parts such as: source codes, utilities, libraries and APIs capable to comfortably build reliable Science Gateways and link them to one or more DCIs avoiding any usage complexity from the final user perspective. Moreover, since SGs based on the FGF are capable to keep track of who is accessing the DCIs, not only its usage ensures OS-compliant reproducibility and reusability but also provides a possible answer in protecting or simply tracking people who are accessing data and this is one of the aspects that today still makes the adoption of the OS a delicate matter.

This work presents and explains how the use of the EGI's Science Software on Demand (SSOD) service, built using the FutureGateway Framework in conjunction with the INFN Open Access

Repository (OAR)[5], based on Zenodo [6] software, to reproduce/reuse the outputs of the agent-based Physical Activity Lifelong Modelling & Simulations (PALMS) experiment [7].

[1] https://www.fosteropenscience.eu/content/what-open-science-introduction
[2] https://www.force11.org/group/fairgroup/fairprinciples
[3] https://www.xsede.org/ecosystem/science-gateways
[4] https://futuregatewayframework.github.io
[5] https://openaccessrepository.it
[6] https://zenodo.org/
[7] https://bit.ly/2ksoyD2

Back to Program

Tuesday

11.30

Implementing Open Science in Academic Biomedicine: a Report from the Trenches

Ulrich Dirnagl

Charité Universitätsmedizin Berlin
Department of Experimental Neurology

Berlin Institute of Health
QUEST Center for Transforming Biomedical Research
Berlin, Germany

Since 2017, the QUEST Center at the Berlin Institute of Health aims at increasing value and reducing waste in the biomedical research of Europe's largest university hospital, Charité - Universitätsmedizin Berlin, as well as a large Berlin biomedical research institute, the Max Delbrück Center for Molecular Medicine (MDC). QUEST is an emerging institutional governance action which is internationally unique in scale and scope, offering services and education and aiming to incentivize robust research mechanisms as well as innovative and relevant research by introducing responsible indicators to the institutional decision making in funding, hiring, tenure and promotion.

With a multitude of interlinked activities QUEST is fostering the accessibility and transparency of BIH research through Open Science (OS) and promoting high quality research data management. With awards and reimbursement of Open Access (OA) article processing charges (APC) we have popularized OA at Charité, which is now governed by our institutional library. We have now shifted our activities to promoting Open Data (OD), again by a combination of awards, web tools (e.g. FIDDLE), workshops, and development of an institutional research data policy. Through a text-mining and machine-learning based process developed by QUEST (ODDPub), as part of intramural performance oriented funding publications containing OD are automatically rewarded with a financial incentive earmarked for research.

This is complemented by introducing OS into our structured, science-based, iterative policy development and implementation program which frames criteria for robust, innovative and translational research.

For example, we have implemented a novel application portal for professorships at the Charité. Applicants now need to enter short narratives, e.g. introduced by questions like: 'The Charité attaches great importance to transparent, replicable research and supports the objectives of Open Science (Open Access, Open Data). This includes the registration of studies in registries (clinicaltrials.gov, DRKS, etc.), the preregistration of studies, and the publication of negative and zero results.

How have you been pursuing these goals so far and what are your plans for the future?'. Through meta-research we aim at identifying measures for improving research practice and obtaining evidence for the impact of our activities. For example, we have tracked and benchmarked all German university medical centers with regard to their performance in results dissemination (Wieschowski S et al 2018).

In my talk I will provide insights into this large scale behavioural change intervention in which OS practices are important intended outcomes, and in particular highlight challenges as well as opportunities. (Further info: http://quest.bihealth.org)

Back to Program

Tuesday

12.10

Distributed Drug Discovery and its Application in Neglected Tropical Diseases

Lori Ferrins[1], Rosario Diaz[2], Miguel Navarro[2], and Michael P. Pollastri[1]

[1] Northeastern University
Department of Chemistry and Chemical Biology
Boston, MA, United States of American

[2] Consejo Superior de Investigaciones Científicas (CSIC)
Instituto de Parasitología y Biomedicina "López-Neyra"
Granada, Spain

Neglected tropical diseases (NTDs) are a group of parasitic diseases that disproportionately affect those living in poverty. Consequently, there is no economic incentive to develop new drugs that meet needs related to efficacy and safety. To address this, our lab focuses on repurposing both early and late stage chemical matter as inhibitors of parasite proliferation, and then performing hit-to-lead medicinal chemistry optimization to fine-tune the hits to be effective at killing parasitic protozoans. This is a proven strategy for identifying anti-parasitic lead compounds, potentially resulting in a shorter overall timeline for drug discovery.

One of the diseases for which we have employed this strategy is human African trypanosomiasis (HAT), an NTD that affects approximately 8,000 people in the remotest parts of Africa and is devastating to those afflicted. Although treatments are available, they are not effective in all cases and can cause severe side effects, and new drugs are needed. To identify starting points for HAT drug discovery, we employed a lead repurposing strategy which led to the identification of 59 clusters of structurally-related compounds and 53 singletons. Multiple clusters and singletons were then taken forward for further optimization.

Throughout our optimization efforts we have engaged collaborators across multiple disciplines. In addition to the progression of several compounds to preclinical evaluation, this highly collaborative approach has enabled us to expand our drug discovery effort to include optimization against *T. cruzi*, *Leishmania donovani*, and *Cryptosporidium parvum*, the causative agents of Chagas Disease, Leishmaniasis and Cryptosporidiosis, respectively.

This presentation will highlight the distributed drug discovery paradigm we employed in the hit-to-lead optimization of HTS hits against multiple disease-causing pathogens and describe the results of this optimization.

Back to Program

Tuesday

14.00

Towards Knowledge Graph-based Representation, Augmentation and Exploration of Scholarly Communication

Sören Auer

TIB Leibniz Information Centre for Science and Technology
Department of Data Science and Digital Libraries
Hannover, Germany

Despite an improved digital access to scientific publications in the last decades, the fundamental principles of scholarly communication remain unchanged and continue to be largely document-based. The document-oriented workflows in science have reached the limits of adequacy as highlighted by recent discussions on the increasing proliferation of scientific literature, the deficiency of peer-review and the reproducibility crisis. We need to represent, analyse, augment and exploit scholarly communication in a knowledge-based way by expressing and linking scientific contributions and related artefacts through semantically rich, interlinked knowledge graphs. This should be based on deep semantic representation of scientific contributions, their manual, crowd-sourced and automatic augmentation and finally the intuitive exploration and interaction employing question answering on the resulting scientific knowledge base. We need to synergistically combine automated extraction and augmentation techniques, with large-scale collaboration to reach an unprecedented level of knowledge graph breadth and depth.

As a result, knowledge-based information flows can facilitate completely new ways of search and exploration. The efficiency and effectiveness of scholarly communication will significant increase, since ambiguities are reduced, reproducibility is facilitated, redundancy is avoided, provenance and contributions can be better traced and the interconnections of research contributions are made more explicit and transparent.

In this talk we will present first steps in this direction in the context of our Open Research Knowledge Graph initiative and the ScienceGRAPH project.

Back to Program

Tuesday

14.40

Accelerating Biomedical Discovery with an Interest of FAIR Data and Services

Michel Dumontier

Maastricht University
Institute of Data Science
Maastricht, The Netherlands

With its focus on improving the health and well being of people, biomedicine has always been a fertile, if not challenging domain for computational discovery science. Indeed, the existence of millions of scientific articles, thousands of databases, and hundreds of ontologies, offer exciting opportunities to reuse our collective knowledge, were we not stymied by incompatible formats, overlapping and incomplete vocabularies, unclear licensing, and heterogeneous access points.

In this talk, I will discuss our work to create computational standards, platforms, and methods to wrangle knowledge into simple, but effective representations based on semantic web technologies that are maximally FAIR - Findable, Accessible, Interoperable, and Reuseable - and to further use these for biomedical knowledge discovery. But only with additional crucial developments will this emerging Internet of FAIR data and services enable automated scientific discovery on a global scale.

Back to Program

Tuesday

15.50

Watch out for Influencers! When Scientific Reasoning Relies on a Single Data Point

Andrej-Nikolai Spiess[1], Stefan Rödiger[2], Michal Burdukiewicz[3], Andreas Ziegler[4], and Joel Tellinghuisen[5]

[1] University Hospital Hamburg-Eppendorf, Department of Andrology Hamburg, Germany

[2] Brandenburg University of Biotechnology, Institute of Biotechnology Cottbus-Senftenberg, Germany

[3] University of Wroclaw, Faculty of Biotechnology Wroclaw, Poland

[4] Universitätsklinikum Schleswig-Holstein, Institute for Medical Biometry and Statistics Lübeck, Germany

[5] Vanderbilt University, Department of Chemistry, Nashville, TN, United States of America

There are no doubts that science is facing a reproducibility crisis. A survey with 1500 scientists revealed that these were unable to replicate 40-80% of their own and others' results (https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970), while a large-scale data replication study in the psychological sciences demonstrated that 30% of results were not reproducible or even led to contrary conclusions (https://doi.org/10.1126/science.aac4716). In these studies, the authors merely replicated the statistical analysis steps based on the original raw data, demonstrating that by just repeating simple statistical steps, one can obtain completely different analytical (and hence interpretable) outcomes. **We are strongly convinced that the lack of understanding statistical prerequisites frequently results in highly unstable and irreproducible results.**

In this respect, one of the most widely applied methods in the life/chemical/physical sciences for testing the association between two variables is linear regression. The slope of the linear regression β_1 can be tested against the null hypothesis $\beta_1 = 0$ (no association) by using a one-sample *t*-test, delivering the usual *p*-value.

This *p*-value is then frequently used for a dichotomous statement of significant association (when $p < 0.05$) or not (when $p > 0.05$), often in a way that may drive research in a certain direction.

We frequently observed the depiction of linear regressions in high impact journals (*Nature*, *Science*, PNAS) in which single, outlying data points may render the regression significant and where their omission might revert significance, i.e. result in the regression's *p*-value crossing the $\alpha = 0.05$ border. Although many influence measures exist (Cook's distance, leverage, covratio, studentized residuals, dfbetas), none of these has a direct relation to the regression's *p*-value. To estimate the impact of this effect on cutting-edge research, we specifically reanalyzed linear regressions published in these three journals in 2016, by i) digitizing applicable graphs and ii) subjecting the acquired raw data to a leave-one-out p-value analysis. To do so, we developed the R-package "reverseR", which is specifically tailored to identify *p*-value influencers.

Interestingly, 30-40% of all analyzable linear regression graphs from *Nature*, *Science* and PNAS could be reverted in their significance statement (significant => insignificant, and vice versa) by omission of a single data point, uncovering the high statistical instability of these publications. The *S* value is introduced that provides a measure of a linear regression's stability in terms of significance reversal, and we will run through several examples using "reverseR"s built-in Shiny GUI.

Back to Program

Tuesday

16.30

Stories from the “Open Science Revolution”: How Scientists Talk about Openness

Rosalind Attenborough

University of Edinburgh
Department of Science, Technology and Innovation Studies
Edinburgh, United Kingdom

Movements advocating “open” science and research have been gaining momentum since the turn of the twenty-first century. From open access publishing and open data archiving, to open peer review and citizen science, practices under the “open” umbrella carry hopes of fixing or revolutionising science communication, and science itself. Some scientists have led movements towards openness from within their communities, but many others seem ambivalent. Now that funders and universities increasingly require open practices, questions arise about “cultural” resistance to openness amongst scientists.

Ros aims to step into this cultural sphere and listen to scientists' understandings - including their views on the current open science agenda, and other meanings of openness that are important to them.

Back to Program

Tuesday

17.10

“The University Cooperative”, Learning to Manage Academic Resources as Common Property

Sanli Faez

Utrecht University
Department of Physics
Utrecht, The Netherlands

Nearly thirty years ago, Elinor Ostrom published her groundbreaking book Governing the Commons, in which she showed that users of natural and agricultural resources can and do govern such resources themselves. By making the rules together and complying, 'commoners' can manage their own resources. They don't have to rely on government regulation or corporate policy. They can face challenges and make decisions collectively. Ostrom also extended her work on commons governance to knowledge commons.

Sharing resources is one of the foundations of university education and development. For example, the curriculum, the literature, and the scientific method in the most general sense all form common resources that are transferred from generation to generation. These common resources are so close to us that we sometimes do not see them, and do not realize that they should be properly organized. Meanwhile, some of these resources have been subject to forces of privatization and commercialization. It is therefore, an imperative of modern academia to organize its common resources in such a way that they can stay accessible for ourselves and for future generations. Recent activities on open science could be seen as important steps to achieve this goal.

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.

Back to Program

Wednesday

09.00

A Universal Chemical Synthesis Text to Code Translator that Enables the Autonomous Synthesis of Organic Molecules

Cole Mathis and Lee Cronin

University of Glasgow
School of Chemistry
Glasgow, United Kingdom

As the number of new compounds expands, so does our capacity to access these molecules. However, problems with the reliability of the protocols due to errors, ambiguity, or the expertise required, mean that many molecules are still not available. The automatic synthesis of molecules is growing, but commercial systems only cover a limited amount of chemistry, and a universal programming language for all of organic synthesis does not exist. We present a synthesis-reader (SynthReader), that standardises and ties the abstraction of organic synthesis to our modular organic synthesis robot (the Chemputer). We show that SynthReader can automatically convert protocols found in the literature to code which is then interpretable via production of a chemical mark-up language, XDL. Not only have we validated the system for text to robot driven synthesis, we can show the universality of SynthReader by converting protocols written in English to XDL, and then back to any other language, see Figure. This means that we can interconvert chemical protocols into different languages without introducing errors, as well as code to automatically drive the Chemputer, establishing a new standard for synthesis reporting and route to instruct the automatic synthesis of molecules without programming skills.

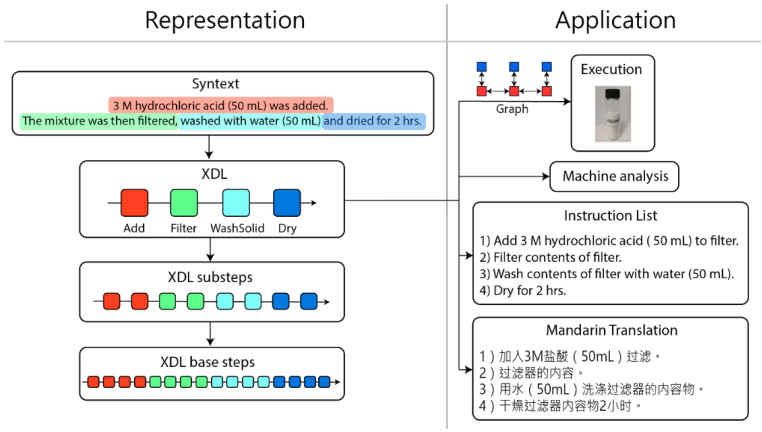

Figure: Left: Different representations of chemical synthesis made possible by the *XDL* and *SynthReader* software. The colour coding shows the way in which one high-level step gives rise to multiple low-level steps. Right: Multiple uses of the *XDL* representation. Going from top to bottom; execution on a *Chemputer* platform, machine analysis, e.g. process complexity, cost, safety etc, English instruction list, and translation to Mandarin.

References

[1] J. Granda, L. Donina, V. Dragone, D. –L. Long, L. Cronin 'Controlling an organic synthesis robot with machine learning to search for new reactivity', *Nature*, **2018**, 559, 377-381.

[2] P. Kitson, G. Marie, J. –P. Francoia, S. Zalesskiy, R. Sigerson, J. S. Mathieson, L. Cronin 'Digitization of multistep organic synthesis in reactionware for on-demand pharmaceuticals', *Science*, **2018**, 359, 314-319.

[3] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin 'Organic synthesis in a modular robotic system driven by a chemical programming language', *Science*, **2019**, 363, 144-152

Back to Program

Wednesday

09.40

Open Science Platform for Materials Science: AiiDA and Materials Cloud

Giovanni Pizzi

École Polytechnique Fédérale Lausanne
Department of Materials Science and Engineering
Lausanne, Switzerland

In this talk I will present first our vision for an Open Science Platform for Computational Materials Science. Such an platform should be based on four cornerstones: 1) support for open data generation tools, including in particular the (classical and quantum) simulation codes; 2) an open infrastructure for the automated tracking of the provenance of simulations, in order to guarantee reproducibility of computational research; 3) an open engine to implement and execute scientific workflows (consisting of sophisticated sequences of calculations), encouraging at the same time the reuse of these workflows by sharing their code under open licenses; 4) support and fostering of code and research data sharing (with full provenance) through portals that make data FAIR (Findable, Accessible, Interoperable, Reusable) and are compliant with the requirements of scientific Data Management Plans.

I will then discuss the implementation of this platform concepts into our tools: the AiiDA automation and provenance-tracking software [1] and the Materials Cloud web platform [2], designed to fulfil the requirements outlined above.

[1] http://www.aiida.net

[2] http://www.materialscloud.org

Back to Program

Wednesday

10.50

Simulation Foundry: Repeatable, Replicable, Reproducible, Open and FAIR Molecular Modelling

Gudrun Gygli and Jürgen Pleiss

University of Stuttgart
Institute of Biochemistry and Technical Biochemistry
Stuttgart, Germany

Molecular dynamics simulations of enzyme-substrate-solvent systems have become a valuable modelling method to investigate the impact of protein sequence, substrate structure, and solvent composition on the interaction between enzyme and substrate, and to predict the substrate profile or the stereo- and regioselectivity of an enzyme. In analogy to wet lab experiments, series of simulations are performed to compare mutants, substrates, or reaction conditions, and multiple replicates enable a reliable statistical analysis.

Workflows allow scientists to store and execute all steps (setup, production, and analysis) as one process, making molecular dynamics simulations repeatable, easy to use and to debug. Large-scale parameter studies require a workflow to ensure quality and replicability of simulations. The description the workflow and the simulation data must be F.A.I.R. and open to ensure data quality, to make data replicable and reproducible, and to facilitate exchange of methodologies between groups.

We have built a workflow based on widely used bash scripts to perform systematic parameter studies with molecular dynamics simulations as computational equivalent of a 96 well plate experiment. We demonstrate its feasibility by simulating binary solvent mixtures and systematically study directly measurable thermophysical properties.

Back to Program

Wednesday

11.30

Towards an Improved Data Ecosystem for Scientists

Evan Bolton

National Center for Biotechnology
Department of Chemistry
Bethesda, MD, United States of America

Scientists are more data aware than ever. Data science requires access to an ever increasing fraction of available information. FAIR principles emphasize the findability, accessibility, interoperability, and reusability of data sets. Machine readable identifiers, including ORCID, DOI, and InChI, are increasingly utilized in (meta)data. Data licensing schemes such as creative commons are becoming the norm. Integration of diverse sources of information complicates clouds data use as it crosses international boundaries. Publishing workflows are increasingly focused on capturing sufficient metadata to enable reproducibility of the science. All of these factors and more demand changes in how scientists generate and use data.

This talk will help to outline ongoing efforts towards an improved data ecosystem for scientists. PubChem efforts towards this end will be highlighted. In addition, the role of standards and metadata in this data ecosystem will be explored.

Back to Program

Wednesday

12.10

Wikidata and Scholia as a Hub Linking Chemical Knowledge

Egon L. Willighagen[1], D. Slenter[1], L. Rasberry[2], D. Mietchen[2], C. Evelo[1,3], F.Å. Nielsen[4]

[1] Maastricht University, Department of Bioinformatics – BiGCaT, Maastricht, The Netherlands

[2] University of Virginia, Data Science Institute, Charlottesville, VA, United States of America

[3] Maastricht University, Maastricht Centre for Systems Biology – MaCSBio, Maastricht, The Netherlands

[4] Technical University of Denmark, Cognitive Systems, DTU Compute, Lyngby, Denmark

Making chemical databases more FAIR (findable, accessible, interoperable, and reusable) benefits computational chemistry and cheminformatics. We here discuss Wikidata, a sister project of Wikipedia but with one big difference: it is a machine readable database, making it far more useful for interoperability of molecular databases in systems biology [1]. Thanks to the Wikidata:WikiProject Chemistry community, there is a growing amount of information about chemical compounds: Wikidata currently has over 160 thousand chemical compounds, of which more than 96% is associated with InChIKeys and has more than 73thousand CAS registry numbers. Ongoing work by this WikiProject includes capturing chemical classes and chemical compounds in the various Wikipedias as machine readable data. Other projects include covering human drugs [2], MeSH Chemicals and Drugs, and volatile organic compounds. This work is supported by the many tools around Wikidata.

We here introduce our contributions to the WikiProject Chemistry to support FAIR-ification of open chemical knowledge. For example, we proposed new Wikidata properties to annotate compounds with external database identifiers for the EPA CompTox Dashboard [3], the SPLASH [4], MassBank, and LIPID MAPS. Furthermore, we used a combination of Bioclipse (via the Bacting project) and QuickStatements to add missing chemical compounds for biological pathways from WikiPathways [5].

Finally, we introduce an extension of Scholia [6], visualizing data about compounds and compound classes, including external identifiers, physicochemical properties, and an overview of the literature from which the knowledge is derived.

Scholia has received funding from the Alfred P. Sloan Foundation under grant number G-2019-11458.

[1] Mietchen D, et al. Enabling Open Science: Wikidata for Research (Wiki4R). *Research Ideas and Outcomes*. **2015** Dec 22;1:e7573.
[2] Putman TE, et al. WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata. *Database*. **2017** Jan;2017(1).
[3] Williams, AJ, et al. The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminform*. **2017** Nov 18;9:61.
[4] Wohlgemuth G, et al. SPLASH, a hashed identifier for mass spectra. *Nature Biotechnology*. **2016** Nov 8;34(11):1099–101.
[5] Slenter DN, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Research*. **2018** Jan 4;46(D1):D661–D667.
[6] Nielsen, FÅ, et al. Scholia, Scientometrics and Wikidata. The Semantic Web: ESWC 2017 Satellite Events, **2017**.

Back to Program

Wednesday

14.00

Towards a National Research Data Infrastructure for Chemistry in Germany

Christoph Steinbeck (on behalf of the NFDI4Chem Consortium)

Friedrich-Schiller-University Jena
Institute for Inorganic and Analytical Chemistry
Jena, Germany

Germany is preparing to develop a national infrastructure for research data. As part of this, NFDI4Chem is an effort by the German chemistry community to develop and maintain a national research data infrastructure for the research domain of chemistry in Germany. NFDI4Chem aims to represent all disciplines of chemistry in academia and industry. In the initial phase, NFDI4Chem focuses on handling data of molecules and data of their characterization and reactions, both experimental and theoretical.

This overarching goal is achieved by working towards a number of key objectives:

1. Connect existing data repositories and, based on a requirements analysis, establish one or multiple domain-specific research data repositories for the national research community in Germany, and to link them to international repositories.
2. Together with IUPAC and other international partners, initiate community processes to establish minimum information (MI) standards for data and metadata, where missing, in key areas of chemistry, as well as missing open data standards, to support the FAIR principles for research data.
3. Foster the development and adoption of Electronic Laboratory Notebooks (ELN), software, tools and Application Programming Interfaces (APIs) between commonly used instrumentation and software towards an embedded, digital information architecture to help researchers to capture research data in well-annotated electronic form at the earliest possible point in time in the research process.

4. With the support of the German Chemical Society (GDCh), engage with the chemistry research community in Germany through a wide range of measures to create awareness for and foster the adoption of FAIR research data management. Initiate processes to integrate research data management into curricula. Offer a wide range of training opportunities.
5. Maintain a close relationship with neighbouring NFDI consortia to avoid duplicate development and exploit synergies.
6. Engage with experts to explore the legal aspect of FAIR research data management, design and develop the NFDI4Chem accordingly, and to offer advice for the research community.

In this talk, we will outline the concepts of NFDI4Chem to make open science in chemistry a reality as well as its integration into the national and international context of research data management and open science in general.

Reference

Oliver Koepler, Nicole Jung, Angelina Kraft, Janna Neumann, Thesenpapier Nationale Forschungsdateninfrastruktur für die Chemie (NFDI4Chem), DOI: 10.5281/zenodo.1404201

Back to Program

Wednesday

14.40

The GO FAIR Convergence Matrix: Optimizing the Reuse of Existing FAIR-related Resources

Erik Schultes

GO FAIR International Support and Coordination Office
Leiden, The Netherlands

The FAIR Principles articulate the behaviours expected from digital resources that are Findable, Accessible, Interoperable and Reusable by machines. However, as originally published, the FAIR Principles do not consider implementation options. Coordinating a broadly accepted, widely used set of implementations remains a major challenge as numerous interlocking standards and technologies are needed in order to realise a global Internet of FAIR Data and Services (some already existing, others in need of development). Furthermore, communities have their own implementation preferences and priorities when reusing data. In an effort to accelerate community convergence on FAIR implementations, we have launched the development of the FAIR Convergence Matrix.

The Matrix is a platform that compiles for any community of practice, an inventory of their self-declared FAIR implementation choices and challenges. The Convergence Matrix is itself a FAIR resource, openly available, and encourages voluntary participation by any self-identified community of practice (not only the GO FAIR Implementation Networks). Based on patterns of use and reuse of existing resources, the Convergence Matrix supports the transparent derivation of optimal strategies for coordinating convergence onto FAIR standards and technologies.

Back to Program

Wednesday

15.20

Discussion:
The Why, How and What of Open Science

Martin G. Hicks

Beilstein-Institut
Frankfurt am Main, Germany

Moving from an analogue to digital world in research is by no means straightforward. When it comes to sharing research data most scientists have an idea of how they would go about doing it. Pre-requisites are workflow tools in smart laboratories, such as ELNs and LIMS capturing analytics data directly and feeding into domain specific repositories. However, the questions of why open science is needed and what do we envisage being able to accomplish through it have not yet been addressed in sufficient detail. To avoid building a bigger and bigger haystack of data, answering the why helps us work out the what.

Back to Program

Wednesday

16.30

Supporting Ireland's Open Research Agenda – HRB's Open Publishing Platform and FAIR Data Stewardship Pilot

Aileen Sheehy

Health Research Board
Post-Award and Evaluation
Dublin, Ireland

The Health Research Board is committed to ensuring that its funded research is open, accessible and usable, so it can have the greatest possible impact.

HRB Open Research is an open publishing platform that enables our researchers to immediately publish and share their findings in a fast, open and transparent way. The platform has many benefits over traditional academic journals – a completely open peer review process, immediate and transparent publication, quick timelines, open data policies and is free to use for HRB grant holders. The range of alternative article types that can be published allows researchers credit for all the different research outputs they are producing, as well as reducing research waste by allowing results to be published that often face barriers at other journals.

The HRB sees access to and sharing of research data as central pillars of Open Research. The benefits from opening up research data for scrutiny and reuse are potentially very significant; including economic growth, increased resource efficiency, securing public support for research funding and increasing public trust in research. The HRB has partnered with the GoFAIR International team to build awareness amongst the research community of the FAIR Data principles (Findable, Accessible, Interoperable, Reusable) and to support effective data stewardship.

Back to Program

Wednesday

17.10

FAIR Data? Not Without Code! But How to Get There? - Case Study from TU Delft

Marta Teperek

Technical University Delft
Library
Delft, The Netherlands

Who hasn't heard about FAIR Data? FAIR data has become the new buzzword within research communities: funders, publishers, research institutions... all require that researchers make their research data FAIR: Findable, Accessible, Interoperable and Re-usable. However, what's the value of data, if the accompanying code is not available?

Most researchers are now regularly use code in their analyses. Therefore, if institutions really care about supporting research reproducibility and helping their researchers to make their data FAIR, it is key that alongside support for good data management, researchers are provided with adequate support for coding. But how to do that?

In this talk I will describe a case study from TU Delft: the key drivers that led the university to extend its data management support to coding as well, and explain how this is provided in practice.

Back to Program

Thursday

09.00

Open Science is Accelerating Early Target Discovery and Validation, and Facilitates the Generation of More Novel Medicines for Patients

Chas Bountra

University of Oxford
Structural Genomics Consortium (SGC)
Oxford, United Kingdom

The validation of pioneer targets for drug discovery, remains a major challenge. In my lecture I will describe how we:

- work with a large number of pharmaceutical companies, to develop high quality, freely available, small molecule inhibitors
- focus only on novel targets, or those deemed to be 'difficult' or intractable
- give these inhibitors to a large and growing international network of academic collaborators, to crowd source new biology, disease understanding and 'target discovery'
- have facilitated proprietary efforts in pharma, and catalysed the creation of new biotechs, and
- have accelerated numerous clinical studies

We are now

- generating 'Target Enabling Packages' (comprising purified protein, biophysical or biochemical assays to assess function, the three dimensional X ray structure and chemical starting points) for novel, high priority, disease linked genes
- building platforms of primary human cells, for screening novel inhibitors, in order to identify new 'better' targets for drug discovery
- building major collaborations with patient groups and hospitals, in order to catalyse these studies and enhance dissemination into the best disease labs across the world

Together, we are creating a new ecosystem for drug discovery. One which we believe will accelerate the generation of more novel medicines, more quickly. We hope these will also be more affordable.

Back to Program

Thursday

09.40

Can Openness Pay?

Matthew H. Todd

University College London
School of Pharmacy
London, United Kingdom

Carrying out R&D openly provides major advantages of speed, robust peer review and the construction of nimble, motivated teams. However, if the R&D has potential monetary value it must be performed behind closed doors in order to permit the ownership of IP through patenting.

Is there a way to combine open, inclusive ways of working with a secure financial route to market? Yes.

This will be illustrated in the area of drug discovery and development, where secrecy is typically aggressively deployed.

Back to Program

Thursday

10.20

DataSTAGE: Improving Access to FAIR Data to Accelerate Scientific Discovery for Heart, Lung, Blood and Sleep Research

Rebecca R. Boyles

RTI International
Research Computing Division
London, United Kingdom

Biomedical and genomics research represents an exponentially growing source of data that far outstrip individual researchers' ability to store and analyze. Recognizing this challenge and the need to democratize data and tool access for the scientific community, NHLBI initiated DataSTAGE (**S**torage, **T**oolspace, **A**ccess and analytics for bi**G** data **E**mpowerment) in 2018. DataSTAGE's goal is to enable heart, lung, blood, and sleep research investigators to find, access, share, store, and compute on large data sets. DataSTAGE is a cloud-based platform providing improved access to data, tools, applications, and workflows to enable these capabilities in secure workspaces. DataSTAGE initially focused on opening access to TOPMed datasets, representing a high value, high return for NHLBI investigators.

DataSTAGE is seeking to democratize data access while respecting data use and consent limitations. Within the Consortia solutions are agile and iterative while being embedded within initial scientific use cases to accelerate infrastructure and policy development. DataSTAGE focused on two initial use cases: 1) enabling large scale genomic analyses and 2) enabling investigators to conduct Deep Learning analyses on image data. Both require addressing extensive challenges, such as handling big data, computing on a large scale, and enabling easy data access. DataSTAGE addresses these by developing solutions on a cloud-based platform utilizing technology and expertise from multiple teams.

In summary, DataSTAGE has developed many solutions to address NHLBI's challenges. DataSTAGE is piloting novel data access approaches in collaboration with NIH and GA4GH. DataSTAGE utilizes cloud technology to empower big data and its computation. DataSTAGE utilizes federated, centralized storage to facilitate secure data access.

DataSTAGE worked with TOPMed investigators to provide state-of-the-art, sophisticated tools for genomics and other analyses. DataSTAGE facilitates team science, collaboration, and the ability for investigators to bring-your-own data to the platform. Thus, DataSTAGE significantly accelerates research discovery.

Back to Program

Thursday

11.20

Behind the FAIR Brand: Thinkers, Doers and Dreamers

Susanna-Assunta Sansone

University of Oxford
Department of Engineering Science
Oxford, United Kingdom

The reuse of other people's data is providing useful insights for new research questions and products, and driving new scientific discoveries. For this purpose, data management and stewardship, data quality and readiness for (re)use are prerequisites for success. The FAIR Principles are about high-quality, machine-readable data. And better data means better science and more efficiently. Therefore, if the FAIR Principles have made it big should come as no surprise.

To turn FAIR into reality, however, we need to provide researchers with FAIR-enabling tools and services that make invisible the (potentially complex) technical machinery of (meta)data standards and identifiers, which underpin FAIR. Service providers, librarians, journal publishers and funders, among others, are actively working to deliver the next generation framework for FAIR data, which is collaborative, interdisciplinary and sustainable. FAIR has also become the (core) mission of a growing number of initiatives - especially in Europe, USA and Australia - encompassing R&D projects and programmes, national and global service provisions, alliances and societies, training and educational efforts.

Where are we now? Is the road to FAIRness understandably or unnecessarily congested? Do bottom-up (discipline-specific) initiatives interact and interoperate with top-down (cross-discipline) efforts? Do the expectations match with the reality of the tools and services at hand? In my presentation I will summarize my journey in this FAIR ecosystem, providing my experience as an author of the FAIR principles (thinker), as a R&D scientist and resource provider (doer), and will reflect on the lessons learned that may help us transforming the vision of FAIR data into a powerful toolkit at the researchers' fingertips (dreamer).

References

- Data Readiness Group, University of Oxford: sansonegroup.eng.ox.ac.uk
- Wilkinson et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016. doi.org/10.1038/sdata.2016.18
- Sansone et al. FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology*. 2019. doi.org/10.1038/s41587-019-0080-8
- Wilkinson, Dumontier, Sansone et al. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Scientific Data*. 2019 doi.org/10.1038/s41597-019-0184-5

Back to Program

Thursday

12.00

Five Years of Data Sharing at PLOS: Challenges and Opportunities

Iratxe Puebla

Public Library of Science (PLOS)
PLOS ONE
Cambridge, United Kingdom

The value of data sharing in research is being increasingly recognized, many funders have developed policies that mandate data sharing and institutions are building support mechanisms so that researchers can make data management and accessibility part of their research process. The PLOS journals require that authors of published articles make the underlying data available at the time of publication.

In the five years since the implementation of the current PLOS data policy, we have seen encouraging progress in the availability of data and the use of repositories, while facing a number of challenges associated with the need to account for a wide range of research disciplines and differing views on what constitutes policy compliance. As we seek to build on our experience and support FAIR data sharing, we need to look into strengthening mechanisms for author credit and supporting additional technology and frameworks that support the re-usability of data.

Back to Program

Thursday

12.40

Bridging the Gap Between the Scholarly Literature and Public Data Repositories

Andrew L. Hufton

Nature Research
Scientific Data
Berlin, Germany

The Nature Research journals understand that data sharing supports reproducibility and increases the impact of published works. The increasing complexity and size of research datasets, however, pose challenges for scientists who wish to share their data in a reusable and transparent manner. *Scientific Data* (https://www.nature.com/sdata/), an open-access journal from Nature Research, works closely with public data repositories to help researchers share their data FAIR-ly, in manner that supports the credibility of their research and ensures they get proper credit. Here, I will explain *Scientific Data's* editorial policies and share some of our experiences peer-reviewing and publishing data so far.

I will also give an overview of broader efforts across Nature Research that are pushing more data into public repositories, giving a more transparent record of data associated with our publications and promoting a culture of robust data citation.

Back to Program