Invited paper

Christopher J. Swain*, Jeremy G. Frey and Jonathan M. Goodman

RSC CICAG Open Chemical Science meeting: integrating chemical data from two symposia and a series of workshops

https://doi.org/10.1515/pac-2021-1003

Abstract: In November 2020 the Royal Society of Chemistry Chemical Information and Computer Applications interest group (RSC CICAG) ran a five-day meeting entitled Open Chemical Science (https://www.rsc.org/events/detail/42090/open-chemical-science). This event had three intertwined themes, Open Data, Open Access publishing and a series of workshops highlighting a variety of Open-Source tools for chemistry. The online event proved to be enormously popular, with attendees from 45 different countries. The challenges involved in converting what was planned as a three-day physical event into a five day virtual event with three intertwined strands was recognised by the RSC with the award of the "2021 Inspirational Committee Award" (https://www.rsc.org/prizes-funding/prizes/2021-winners/rsc-chemical-information-and-computer-applications-group/). The workshops in particular proved to be enormously popular and spawned a year long series of further workshops.

Keywords: Cheminformatics.

Article

The five-day online meeting *Open Chemical Science*, which was run by RSC CICAG [1, 2] in November 2020 was an innovative event with intertwining themes of *Open Chemical Data*, *Open Chemical Publishing* and *Open Chemical Software*. Our aim was to investigate open science and highlight good practice in the areas cheminformatics and chemical science, as a first step towards developing ideas for the standardisation of the best approaches (Fig. 1).

The COVID-19 pandemic meant that our original plans, for a traditional meeting at which people travelled to a single location and would have been able to talk to each other as well as listen to the presentations, became impossible. As a result, we were forced to invent different ways of presenting the information. Indeed, removing the need for people to travel together had a series of important advantages.

Firstly, we were able to run three interweaving conferences over five days. Some people were able to attend all the themes, but others could attend just one stream and benefit from the material by committing just a couple of hours a day, rather than incurring long and expensive days of travelling.

Secondly, we were able to attract international speakers and session chairs who did not have to give up many days of time to travel around the world, with its associated costs and climate impact. By running the conference

Article note: A collection of invited papers on Cheminformatics: Data and Standards.

***Corresponding author: Christopher J. Swain**, Cambridge MedChem Consulting, 8 Mangers Lane, Duxford, Cambridge, CB22 4RN, UK, e-mail: Swain@mac.com. https://orcid.org/0000-0002-9636-4567
Jeremy G. Frey, School of Chemistry, University of Southampton, Southampton, SO17 1BJ, UK.
https://orcid.org/0000-0003-0842-4302
Jonathan M. Goodman, Yusuf Hamied Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK.
https://orcid.org/0000-0002-8693-9136

from late morning to early evening in the UK, it was accessible to participants from the US and throughout Europe and Africa. Participants from Asia and Australasia required rather more determination to attend.

Additionally, by making all aspects of the meeting on-line, it was possible to keep a greater proportion of the information available after the conference, and the video archive is a major asset. During the meeting, social media, particularly via the twitter hashtag #OpenChem20 was an important feature of the engagement at the meeting.

Overall, the meeting was inexpensive to run, as there was no requirement to pay for travel or food. However, online meetings do incur some costs, in additional to training time for the organisers. Generous support from the RSC and sponsoring companies made it possible to make the meeting free to attend, although we did require registration (Fig. 1).

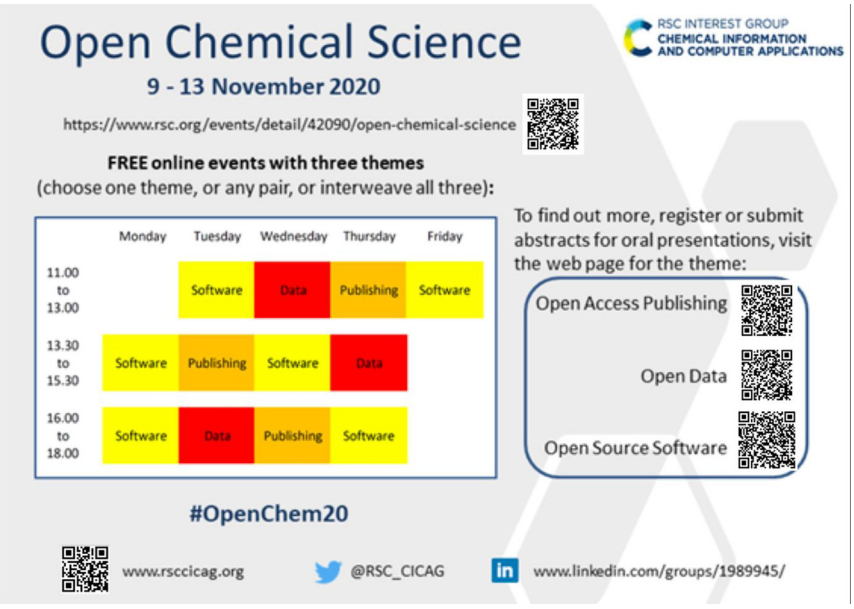

Fig. 1: The intertwined conference themes.

The three streams ran as separate meetings, weaving together to make a large whole, with a broad range of open science in chemical science. The three streams were: *Software*, *Data*, and *Publishing*. Detailed meeting reports have been written by Wendy Warr and are available in the *CICAG Newsletter Winter 2020–21* [3].

Open software

All scientists working in chemistry need software tools for accessing, handling, and storing chemical information, or performing molecular modelling and computational chemistry. There is now a wealth of open-source tools to help in these activities, however many are not as well-known as commercial offerings. These workshops were intended to offer a unique opportunity for attendees to try out a range of open-source software packages for themselves with expert tuition in different aspects of chemistry. All software and training materials required for the workshops were provided for attendees to install and run on their own computers.

The success of the workshops in the Open Chemical Science week led to the creation of the monthly workshop series, each month a workshop highlighting a particular package/resource.

Initial motivation for workshops

In April 2019 RSC CICAG [1] in partnership with the Society of Chemical Industry held a one-day workshop on Computational Tools for Drug Discovery featuring a series of commercial vendors. One of the key elements

from the feedback was that whilst the commercial tools are superb, they are not accessible to all and there are open-source tools that are good alternatives. This was particularly true for teaching where classes may need access for limited periods. On further investigation it was apparent that whilst the commercial vendors spend much time and effort into producing excellent training materials the open-source alternatives often had little more than a readme on GitHub that was impenetrable to a novice user. Open-source computational drug design software is usually written by experts, for experts, with little time or resource to invest in optimization. Thus, user interfaces are often minimal and not written with the inexperienced user in mind and supporting documentation can be limited. With this in mind a series of workshops were envisaged to showcase some of the Open-Source software and to provide a "get you started" tutorial.

The Open Chemical Science week workshops

The original six workshops were chosen to illustrate several different features that were felt to be key aspects of using computers in chemistry. These were online resources that are available, data analysis, molecular visualisation, workflow tools and accessing online computational resources. The six workshops:

Cheminformatics and data analysis using DataWarrior (Isabelle Giraud)

DataWarrior combines dynamic graphical views and interactive row filtering with chemical intelligence. Scatter plots, box plots, bar charts and pie charts not only visualize numerical or category data, but also show trends of multiple scaffolds or compound substitution patterns. This workshop was an introductory tutorial, DataWarrior can be downloaded here.

http://www.openmolecules.org/datawarrior/download.html.

Molecular visualisation using pymol (Garrett Morris)

PyMOL is a comprehensive software package for rendering and animating 3D structures, in particular biomolecules. PyMOL is available from GitHub https://github.com/schrodinger/pymol-open-source.

Chemistry in the cloud: leveraging Google Colab for quantum chemistry (Jan Jensen)

Colaboratory, or "Colab" for short, allows you to write and execute Python in your browser, with Zero configuration required Free access to GPUs Easy sharing. This workshop was an introductory tutorial.

Accessing biological and chemical data in ChEMBL (Anna Gaulton)

ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. This workshop was an introductory tutorial. Website https://www.ebi.ac.uk/chembl/.

Fragment based screening, XChem at diamond (Rachel Skyner)

Fragalysis (fragment analysis) is a web-based platform for fragment-based drug discovery (https://fragalysis.diamond.ac.uk/viewer/react/landing). Its initial use case is focussed around the fragment screening experiment at Diamond. This workshop was an introductory tutorial.

An introduction to KNIME workflows (Greg Landrum)

KNIME Analytics Platform is the open-source software for creating data science. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone. This workshop is an introductory tutorial. KNIME can be downloaded here https://www.knime.com/downloads.

All presenters provided download links and workshop materials prior to the event, and we allowed a couple of minutes at the start to ensure everyone had everything installed. Most of the workshop presenters also had an expert helper. During the workshop only typed questions were allowed, some questions could be answered by a typed reply from the expert helper, others were relayed to the workshop presenter to give an oral answer or live demonstration. The feedback from both presenters and attendees was this worked well, having the expert helper ensured the flow of the workshop was not interrupted by minor queries.

Thoughts and feedback from the Open Chemical Science week workshops

The workshops were hugely popular! By having the workshops online, they were accessible to many, many more participants (over 300 for some sessions) than we could have accommodated at a physical venue. It also meant that they were attendees from 45 different countries. The 2 h sessions (with a 5 min break after 1 h) seems to have hit the sweet spot with folks willing to dedicate the time to attend. All workshops were recorded and made available on the RSC CICAG YouTube channel (https://www.youtube.com/c/RSCCICAG). Many of these videos have now been watched over 1000 times.

The Chemistry in the cloud: leveraging Google Colab for quantum chemistry session by Jan Jensen was particularly interesting. We had some reservations about a workshop that required programming skills however it was apparent that the use of Jupyter notebooks (https://jupyter.org) has made it much more accessible to novice users. The ability to mix code blocks, with comments and images of chemical structures all within a familiar web browser interface was clearly not as intimidating as a command line interface.

The ongoing monthly workshops

The success of the workshops in the Open Chemical Science week led to the creation of the monthly workshop series, each month a workshop highlighting a particular package/resource. As before the aim was to mix desktop tools, online resources, and coding tutorials, a couple of presenters were invited back to give more advanced workshops based on requests from attendees.

ChimeraX with Tom Goddard

This session provided an introduction to ChimeraX for visualizing proteins, ligands and X-ray and electron microscopy maps (https://www.rbvi.ucsf.edu/chimerax/).

Chemical structure validation/standardisation with Greg Landrum

Possibly the most important step in model or database building is data curation. This workshop aimed to deal with chemical structure validation and standardisation using RDKit.

GNINA 1.0: molecular docking with deep learning with David Koes

The use of docking to predict ligand binding to a receptor is now well established. This workshop covered docking and structure-based virtual screening, with an introduction to the theory followed by practical examples: https://chemrxiv.org/articles/preprint/GNINA_1_0_Molecular_Docking_with_Deep_Learning/13578140.

Advanced DataWarrior with Isabelle Giraud

The previous very popular introductory workshop brought DataWarrior to a new, wider audience. This workshop seeks to highlight advanced features, macros and other topics that were brought up by users.

OPIG (Oxford Protein Informatics Group) antibody modelling tools with Charlotte Deane

The OPIG Antibody Suite V2.0 is a collection of informatics tools and databases, including CoV-AbDab, the coronavirus antibody database: https://academic.oup.com/bioinformatics/article/37/5/734/5893556. More information about the antibody modelling tools can be found here: https://www.blopig.com/blog/2020/09/what-can-you-do-with-the-opig-antibody-suite-v2-0/.

An introduction to cheminformatics, data analysis and machine learning with Pat Walters

This was a hands-on workshop on building and validating ML models, including:
- Initial exploratory data analysis
- ML model building
- Model evaluation
- Making predictions on a larger data set

Web apps for fragment-based drug discovery, fragalysis and more with Rachael Skyner

More information about fragalysis and the other online tools at Diamond can be found here: https://fragalysis.diamond.ac.uk/viewer/react/landing.

KNIME workshop with Daria Goldmann

KNIME Analytics Platform is the open-source software for data science. It makes it possible to design data science workflows with reusable components: https://www.knime.com/downloads.

Protein Data Bank workshop

The Protein Data Bank (https://www.wwpdb.org) archives information about the 3D shapes of proteins, nucleic acids, and complex assemblies that help students and researchers understand all aspects of

biomedicine and agriculture, from protein synthesis to health and disease. This workshop coincides with the PDB's 50th anniversary.

Mol* workshop

Mol* is a modern web-based open-source toolkit for visualisation and analysis of large-scale molecular data. More information can be found here: https://molstar.org.

What have we learnt so far?

Whilst physical workshops have some attractions, the virtual platform is so much more widely and easily accessible. The world-wide audience does mean that time-zones must be considered. Holding the workshops in the afternoon UK time, allowed late evening viewing in Asia and early morning in America. Aiming to have the workshop on the same day each month allowed regular attendees to plan their work. Based on email addresses the audience seemed to be a combination of academics, students, biotechs/startups and pharma companies.

Two hours is a relatively short period, so it was essential that download instructions and workshop material was distributed well in advance. A 2 h workshop is not enough time to demonstrate every feature a package may have, but hopefully was sufficient to make users confident in getting started and willing to explore.

Having an additional expert helper certainly helps with the questions; they can answer simple questions directly using the Q&A box and only forward more complex or questions that are likely to be of general interest to the presenter. We did discuss having a parallel Slack channel for further discussion but felt it might be intimidating for new users. As folks become more accustomed to remote working and the various communication options this could be revisited, and be developed into a FAQ for future users.

Whilst most of the questions were of a technical nature, it was apparent that a number of questions were more strategic, pharmacophore screening vs. docking, size of training sets, which algorithm to use when? Perhaps we need an overview workshop/discussion putting all the tools into perspective? The focus of the current workshops was very much novice users, however workshops for more advanced users should be part of the future plans.

Having examples the attendees could work through was absolutely invaluable, a lot of thought and effort went into these. All the examples and training materials have been available and are linked from the YouTube video.

As mentioned previously, Jupyter notebooks have made python coding more accessible, however, there is a need to train bench scientists in good coding practices Python seems to be the *de facto* standard programming language in chemistry at the moment and CICAG are actively developing a suitable Python course.

The YouTube videos [11] (https://www.youtube.com/c/RSCCICAG), will be a valuable and regularly accessed resource judging by the number of views and likes. A number of universities are now pointing their students towards these workshop videos.

These workshops were free to attend, and we thank the presenters giving their time free of charge. There are of course various videoconferencing license fees that must be paid, and these were covered by generous contributions from KNIME, Liverpool ChiroChem and RSC.

Open data strand

The open data strand of the conference had nine speakers from four countries, exploring many different aspects of open data.

Professor John Overington, from the Medicines Discovery Catapult, discussed how large datasets of bioactivity can be used in drug discovery. The ChEMBL database [4] is a key resource in this area. Dr Samantha Kanza, an Enterprise Fellow at the University of Southampton, described how the semantic web is relevant for chemistry and suggested best practices. AI is the next great adventure for well-organised datasets. Professor Emma Schymanski, Head of Environmental Cheminformatics at the University of Luxembourg, talked about her quest to identify molecules in the environment using cheminformatics and mass spectrometry. Her digital detective work connects expert knowledge to environmental observations [5]. Suzanna Ward from the Cambridge Crystallographic Data Centre, discussed the way that reliable chemical data is shared through the CSD and the impact of FAIR data principles [6].

Dr Nicole Jung, a group leader from the Karlsruhe Institute of Technology, described the Chemotion project, which comprises an ELN and repository for research data [7]. Dr Ed Griffen, the co-Founder of MedChemica, described how large ADMET databases could be analysed and emphasised the importance of good data. Dr Barbara Zdrazil from the University of Vienna showed how it was possible to model transporter-mediated toxicity data from the open domain. Professor Matthew Todd from UCL summarised six laws of operation for open source drug discovery [8]. Finally, Dr Kim Jelfs, from Imperial College, discussed the tools she and her group are building to help explore chemical space for molecular material discovery [9].

The open data strand of the conference showed how much data is available, how the quality of these data is critical for automated analysis, and the challenges of making use of it. Cheminformatics standards for formats, metadata and nomenclature are needed to bring together information from distinct databases.

Open publishing strand

There are diverse views on the best approaches to open-access publishing in chemistry and even debates on the value of the approach. The symposium brought together publishers, funders, academic librarians, academics, editors, and innovators to investigate this sometimes-contentious subject.

Dr Martin Hicks from the Beilstein Institute gave an opening keynote lecture, emphasising that open science works requires people to work collectively. Rachel Bruce is the head of open science for UK Research and Innovation (UKRI), the organisation which brings together publicly funded research in the UK. She looked forward to the UKRI's new Open Access Policy, which was subsequently released in August 2021 [10]. This was followed by a presentation from Clair Castle, the Librarian at the Department of Chemistry of the University of Cambridge, who explored how she is able to support chemistry researchers who are exploring open access publications.

ChemRxiv is a preprint server for chemistry, supported by the American, Chinese, German and Japanese Chemical Societies, as well as the Royal Society of Chemistry. Dr Marshall Brennan, the Publishing Manager of ChemRxiv, discussed the dramatic growth of the resource since its launch in 2017 and how this is changing the way that chemistry papers are published. Processes of peer review are also being re-examined, and Dr Tony Ross-Hellauer from the Graz University of Technology reported on a systematic review of the effects of open peer review. Professor Cameron Neylon from Curtin University then investigated how different open access policies in the UK and the US are having a dramatic effect on publishing patterns.

Open access publication is important in the commercial sector as well as in academia. Helen Malone from GSK discussed its effect on the pharmaceutical industry. Both paid and open access resources are important, but open access is an increasingly important part of the chemical information landscape. Professor Egon Willighagen from Maastricht University, who is also Editor-in-Chief of the open access Journal of Cheminformatics, emphasised that open science is distinct from open access. The adoption of open access publishing in journals is incomplete, but growing steadily. The open publishing thread of the conference closed with Dr Neil Hammond, the Publisher for Open Access Journals at the Royal Society of Chemistry, discussed the challenges facing the society that open access brings, and how the RSC is responding.

The open publishing strand of the conference demonstrated how open access is affecting diverse organisations, and the possibilities and potential for open science as they react to this new phenomenon.

Conclusions

This meeting was a very successful scientific meeting, but we also view it as a successful research project addressing key questions for chemical information:

(i) What is the state of open science in chemistry?
(ii) What are the scope and limitations of the current chemistry situation in data, in publishing and in scientific software?
(iii) How is it possible to run meetings without meeting in-person?

The data we gathered by running this meeting enables us to address these questions.

(i) What is the state of open science in chemistry?

Open science is flourishing in chemistry. The interest in the meeting, the large number of contributions from around the world and the continuing access to meeting materials now the meeting has finished all provide evidence that this is an area of great and continuing interest, that there is a large amount of activity in developing open science in chemistry, that much material and resources are available, and that there is worldwide enthusiasm to develop chemistry in this way. The situation is not completely positive, however. So much data is available only under restrictive or expensive license conditions that using a mixture of open and closed science is inevitable for most researchers. Ways of funding and assessing open science are still being established, and researchers are reasonably concerned that their work may not receive all of the recognition that it deserves. The current infrastructure of academic and commercial research developed mainly with closed science in mind and a transition to an open infrastructure for open science would be a significant shift.

(ii) What are the scope and limitations of the current chemistry situation in data, in publishing and in scientific software?

Data: There are many examples of major open data resources which are critical to open chemistry, including PubChem, Chembl, PDB, and others. There is no doubt that open data is central to current chemical science. However, the situation is not a settled one. Open resources require funding, and mechanisms for this are less well-established that might be hoped. The other side of open data is every researcher making their data open. This is a much more significant challenge. Despite support from some funding agencies, making data not only open but also navigable, tractable and readily reusable is a major issue in addition to researchers' everyday challenge of imagining, implementing and achieving ground-breaking advances in chemistry. The benefits of making data available and navigable are further in the future than the next grant application or performance review and the mechanisms for assigning credit equitably are also a subject for forward-looking hopefulness rather than well-established and reliable procedures.

Publishing: Open access publishing is becoming more accepted as a normal way of publishing research. However, a large number of important publishing resources are not open-access and it is not clear when or if this will change. Publishers are moving towards transformative agreements which are transitional arrangements whilst the open-publishing landscape becomes clearer. They are not guarantees that the landscape will change. It is not obvious why open-publishing should cost more, in total, than traditional routes, but this is the current trend. Financial limitations are likely, therefore, to continue to drive and to restrict the future of publishing in chemistry. The Declaration on Research Assessment (DORA https://sfdora.org) has been accepted by many institutions. However, there is significant and well-founded concern that these principles are not always applied by every organisation and every person who is involved with the assessment of research and of researchers. Indeed, members of institutions that have committed to DORA are not always aware of the obligations that this entails.

Software: Open software for chemistry is thriving with a large number of powerful programs available to the community. The rise of Python as a lingua franca for data handling has helped facilitate this. This aspect of the meeting/research-project generated the most interaction, probably because of the scale of the programs.

People learned about examples and best-practice in open data from the meeting, but did not generally restructure their own research as the meeting progressed. Open publishing is central to open chemistry, but currently largely controlled by publishers who were present to report on their approaches rather than make immediate changes as a result of the discussion. Open software, however, could be presented by experts or the authors to people who were able to learn how to use it, either for the first time or with greater expertise, as the meeting progressed. We discovered, therefore, that this section of the meeting generated more immediate data and feedback than the other two sections. Open software is a key component of open chemistry, but it is not dramatically more important than either open data or open publishing. Indeed, the distinction between these different areas is a fuzzy one. Chembl and the PDB are both major resources for open data, but programs to access and analyse these data were covered in the open software section rather than the open data section.

The experience of the meeting and the data gathered from it, demonstrated both the excitement and the rapid changes in open data, open software and open publishing for chemistry. Simply keeping up with these developments is a challenge in itself and the meeting was an important contribution to making it possible for as many scientists as possible to improve their understanding of the areas. There was no unanimity in views on the best pathways forwards, nor even on identifying the major issues that need to be addressed. However, so many ideas and projects are advancing that we are confident that the field will move forward rapidly. We expect that a new workshop on chemical open science will be required before too long.

(iii) How is it possible to run meetings without meeting in-person?

Our initial apprehension about the success of the meeting as the data approached and the pandemic did not subside turned out to be unfounded. The meeting almost certainly had a higher and more diverse attendance than it would have done had it been a traditional in-person meeting as originally planned. More countries were represented both by the speakers and in the audience than would otherwise have been possible. Attendance was much cheaper for most people and the total resources required for the meeting were a tiny fraction of those that would have been needed for an in-person meeting. The impossibility of attendees leaving home counter-intuitively led to a much more global interchange of ideas and results. The effects were not all positive, however. Informal interactions over a video conference are possible but more stilted than actually meeting people. Remembering a person you have met in person and had a sense of their three-dimensional presence is much easier that recalling someone you have encountered only as a low-quality two-dimensional image. Attendees were more likely to be distracted by their surroundings at home than they would have been if gathered together in a meeting room. Despite these issues, however, the meeting was a significant success. In planning future meetings, instead of just thinking about in-person meetings, the CICAG committee will have to balance the benefits of a more international and more economical on-line meeting with a more intense but smaller in-person meeting. It may well be that both in-person and on-line meetings are organised in the future, based on the best choice for the subject of the meeting rather than the pressure of a pandemic. Hybrid meetings present different challenges, but ones for which good approaches may well be found.

Conclusions to conclusions

Viewed as a meeting, the event was a striking success, enabling a large number of scientists from around the world to get together virtually to share knowledge and views on open science in chemistry and to learn more about this fast-developing field. Viewed as a research project, the data confounded our expectations, as so often happens with innovative research. The operational challenges were overcome and found to have significant advantages. The anticipated balance between publishing, data and software tilted more towards software than we had expected. The reasons for this are both the interactivity that is possible in this area and also the choice of the dividing line between software and data, which was placed closer to software than might be desirable for a future meeting. We hope to build on these results by running another meeting in the area even better tuned to the needs of the participants. The most significant impact of this research is the enhancement of the skills in open science of a large number of chemical scientists distributed over the whole world.

References

[1] http://www.rsccicag.org (last checked March 2022).
[2] https://www.rsc.org/membership-and-community/connect-with-others/through-interests/interest-groups/cicag/ (last checked September 2021).
[3] CICAG newsletters: http://www.rsccicag.org/newsletters.htm (last checked September 2021).
[4] D. Mendez, A. Gaulton, A. P. Bento, J. Chambers, M. De Veij, E. Félix, M. P. Magariños, J. F. Mosquera, P. Mutowo, M. Nowotka, M. Gordillo-Marañón, F. Hunter, L. Junco, G. Mugumbate, M. Rodriguez-Lopez, F. Atkinson, N. Bosc, C. J. Radoux, A. Segura-Cabrera, A. Hersey, A. R. Leach. *Nucleic Acids Res* **47**, D930 (2019).
[5] E. L. Schymanski, N. C. Baker, A. J. Williams, R. R. Singh, J.-P. Trezzi, P. Wilmes, P. L. Kolber, R. Kruger, N. Paczia, C. L. Linster, R. Balling. *Environ. Sci: Process. Impacts* **21**, 1426 (2019).
[6] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons. *Sci. Data* **3**, 160018 (2016).
[7] https://www.chemotion.net/ (last checked September 2021).
[8] M. H. Todd. *ChemMedChem* **14**, 1804 (2019).
[9] L. Turcani, E. Berardo, K. E. stk. Jelfs. *J. Comput. Chem.* **39**, 1931 (2018).
[10] https://www.ukri.org/news/ukri-announces-new-open-access-policy/ (last checked September 2021).
[11] https://www.youtube.com/c/RSCCICAG (last checked March 2022).