ROYAL SOCIETY OF CHEMISTRY | CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP

NEWSLETTER winter 2018-2019

Above: AI3SD speakers and organisers Dr Nathan Brown, Prof Jeremy Frey and Dr Samantha Kanza (see page 11)

CICAG aims to keep its members abreast of the latest activities, services, and developments in all aspects of chemical information, from generation through to archiving, and in the computer applications used in this rapidly changing area through meetings, newsletters and professional networking.

Chemical Information & Computer Applications Group Websites:
http://www.rsccicag.org/
http://www.rsc.org/CICAG

LinkedIn http://www.linkedin.com/groups?gid=1989945

MyRSC http://my.rsc.org/groups/cicag

https://twitter.com/RSC_CICAG

QR Code

Table of Contents

Chemical Information & Computer Applications Group Chair's Report 3
CICAG Planned and Proposed Future Meetings 4
CSD: The Journey to One Million Crystal Structures 4
News from Catalyst Science Discovery Centre 6
Nantwich Museum Celebrates Joseph Priestley and the IYPT 7
Meeting Report: Tony Kent Strix Annual Lecture 2018 8
RSC International Year of the Periodic Table 9
Places of the Periodic Table Interactive Map 10
Clarifying the Cahn-Ingold-Prelog Rules 10
Meeting Report: AI3SD Network+ Launch, SCI London 11
Chemical Information / Cheminformatics and Related Books 15
News from InfoChem and Thieme 16
Other Chemical Information Related News 18

Contributions to the CICAG Newsletter are welcome from all sources - please send to the Newsletter Editor:

Stuart Newbold, email: stuart@psandim.com

Chemical Information & Computer Applications Group Chair's Report

Contributed by RSC CICAG Chair Dr Chris Swain, email: swain@mac.com

We have seen some changes to the CICAG committee over recent months. Alan Tonge, a long-time member of the committee and former secretary, has stepped down. I'd very much like to thank him for his contributions over many years. However, this has given us the opportunity to recruit new members to the committee.

We have three new members, Dr Nathalie Batoux, a Business Analyst with IDBS; Dr Samantha Hughes, Associate Director with AstraZeneca; and Talat Giddings, who worked in chemistry information at Thompson Scientific. I hope you will join me in welcoming all three new members in what looks to be a very interesting time in the chemical information field.

Looking forward to the conferences and events planned for 2019, in April we will be running a Workshop on Computational Tools for Drug Discovery co-organised with the SCI Fine Chemicals Group. This is the first of a number of training events CICAG hope to organise, including workshops on programming for scientists. The tutorials at this workshop come courtesy of some outstanding experts in the field: Christian Lemmen (BioSolveIT), Akos Tarcsay (ChemAxon), Giovanna Tedesco (Cresset), Dan Ormsby (Dotmatics), Greg Landrum (Knime), and Matt Segall (Optibrium). Attendees will able to install software on their laptops along with a license to allow use for a limited period after the event.

After the outstanding success of the first meeting, the organisation of the 2nd RSC-CICAG/RSC-BMCS Artificial Intelligence in Chemistry Meeting is now underway. It will be held at Fitzwilliam College in Cambridge, 2-3 September 2019, so please mark the dates in your calendar. We will shortly be sending out a call for abstracts for what looks like being another very popular event.

We also plan to have an active involvement in the Sheffield Conference on Chemoinformatics, taking place on 17-19 June 2019. This triannual meeting is a major event in the Chemical Information calendar.

It will not have escaped everyone's notice that it is now 20 years since Lipinski published his paper on a guide to developability in drug discovery in which he described the "rule-of-five". Since then there have been an increasing number of other metrics published, and RSC-CICAG and RSC-BMCS plan to co-organise a meeting to mark this.

This newsletter also includes a contribution from Noel O'Boyle on a recent publication clarifying the Cahn-Ingold-Prelog rules. I'd like to thank Noel for this contribution and to also invite other contributions from readers highlighting key publications that would be of interest to the CICAG community.

Finally, the CICAG committee understands that for 2017, overall RSC membership increased 2% (interestingly, ACS membership is falling). During the same period, CICAG membership has increased by 6% - thank you and welcome, to our new CICAG members.

CICAG Planned and Proposed Future Meetings

The table below provides a summary of CICAG's planned and proposed future scientific and educational meetings. For more information, please contact CICAG's Chair, Dr Chris Swain.

Meeting	Date	Location	Further Information
Workshop on Computational Tools for Drug Discovery	10 Apr 2019	The Studio, Birmingham	Registration is now open
Sheffield Conference on Chemoinformatics	17-19 Jun 2019	The Edge, University of Sheffield	MGMS, CSA Trust Meeting
Second AI in Chemistry Meeting	2-3 Sep 2019	Fitzwilliam Museum, Cambridge	2-Day meeting, more details to follow
Celebrate the Centenary of IUPAC	Spring 2019	Burlington House, London	To be organised jointly with the Historical Chemistry Interest Group
Practical Computational Chemistry	Spring 2019	University venue TBC	One-day workshop for small groups
20 Years of the *Rule of 5*	TBD	TBC	To celebrate 20 years of Lipinski's Rule
Structure, Reaction and Patent Information for Small Organisations	TBD	TBD	Proposed joint meeting with RSC Consulting Group
Software Update for Medicinal Chemists	TBD	EBI, Hinxton, Cambridge	Proposed training workshop (Joint workshop with the SCI)
Big Data	TBD	TBD	Proposed joint meeting with the SCI)

CSD: The Journey to One Million Crystal Structures

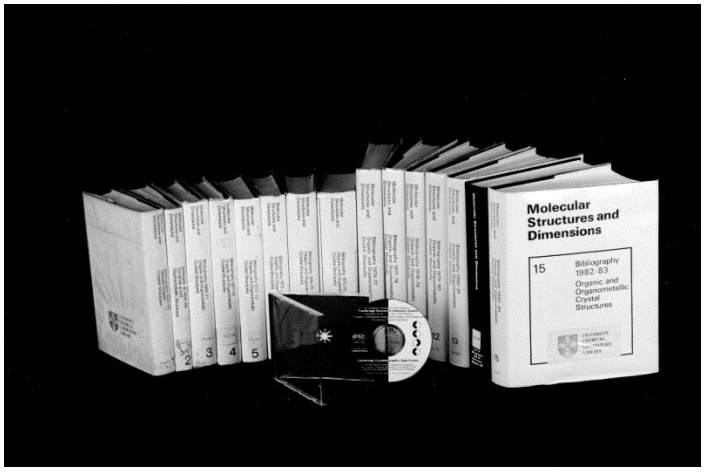
With great thanks to Lucy White: lwhite@ccdc.cam.ac.uk and other members of the Cambridge Crystallographic Data Centre (CCDC) team

Later this year the world of structural chemistry will reach a tremendous milestone, the sharing of one million organic and metal-organic structures through the Cambridge Structural Database (CSD). The CSD was one of the first numerical scientific databases to be established and today enables scientists and educators in over 70 countries to learn from the data.

The journey can be traced back to the 1950s and J.D. Bernal. Bernal believed that the collective use of data would lead to the discovery of new knowledge; transcending the results of individual experiments.[1] Olga Kennard decided with Bernal's support to examine the influence of molecular shape on crystal packing. In 1965 a grant was awarded to allow a small team to start work on the Cambridge Structural Database. There were already nearly 4,000 published structures, so collating these was a daunting task but given the growth of published structures in the following years it was fortunate that efforts started when they did.

Initially, data was released in the form of a book -it took five years to collate the first volumes in the *Molecular Structures and Dimensions* series in 1970. More volumes followed until the early 1980s, when they gave way to a more modern computational database. If the CSD was still released in book form today then it would consist of over 450 volumes! The volumes consisted of bibliographic information and introduced rudimentary ways of searching through permuted name indexes and dividing the entries into classes. It was decided at the outset that the accuracy of the database was paramount. This meant users had numbers they could rely on and so could use the database with confidence, which is certainly still true today. To enable this, an elaborate set of checks were introduced, and authors were contacted if issues with a structure were found. Two volumes of interatomic distances were also created and a database system was constructed. By

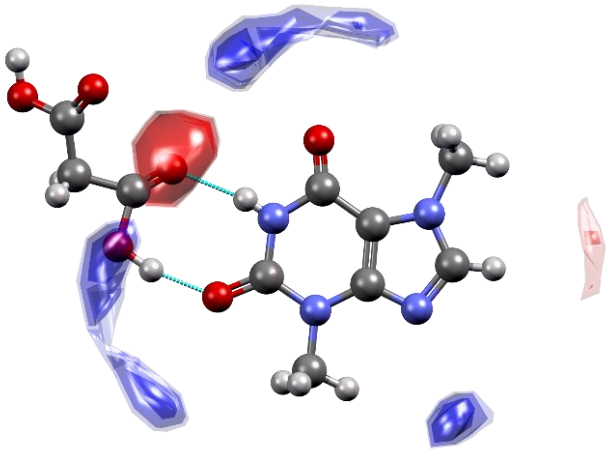
1979 the database contained nearly 25,000 structures, having doubled in size in the five years preceding this. It was recognised that it would be a huge challenge for researchers to analyse and utilize such a large volume of information, and so a computer-based search, retrieval, analysis and display system was designed.[2]



Early versions of the CSD in book form alongside a later version of the CSD on CD

In the 1980s this system, known as the CSD System, was distributed in more than 30 countries worldwide, and interest in the CSD System from pharmaceutical and agrochemical companies increased significantly. Income from these companies, coupled with support from National Affiliated Centres, enabled the organisation responsible for the CSD, the Cambridge Crystallographic Data Centre (CCDC), to become an independent, non-profit charitable centre in 1987.

Today functionality in the CSD-System includes powerful 2D/3D searching, extensive geometry analysis tools, intermolecular interaction analysis, high impact graphics generation, and interconnectivity via the CSD Python API. The ability for users to gain new insights is aided by the addition of two knowledge bases, known as Isostar and Mogul, which provide information about intermolecular interactions and molecular geometries respectively.



A Full Interaction Map showing a molecule's interaction preferences, one of the many ways the data in the CSD can be used today to generate new insights

CCDC software products have diversified to make maximum use of structural data and are used worldwide in industries ranging from pharmaceuticals, agrochemicals, catalysis, and gas storage. The CSD is also used extensively in teaching thanks to the fact that each individual dataset is free to view and retrieve, with a range of educational resources made available alongside the data.

The widespread use of structural data worldwide, the reliance of the CSD and associated software in drug discovery and development, and the thousands of research papers published using the CSD, are testament

to the fact that the collation and curation of individual experiments has enabled new insights and knowledge from the data, making it so much more than the sum of its parts.

Our attention now turns to one million structures, which we expect to be reached this summer. You can see our millionth structure "countdown" on our website, and also sign up to hear the latest news from the CCDC and learn more about how we will be marking this milestone. We will be celebrating at regional crystallographic conferences including the European Crystallographic Meeting (ECM) in Vienna and the American Crystallographic Association (ACA) meeting in Cincinnati. The value of crystallographic data will also be demonstrated at chemical conferences such as the Fall American Chemical Society Meeting, and the Cambridge Science Festival. So keep up to date on our website and social media postings to find out how you can join us in these activities and to learn how you can use the database in your research and educational activities.

1. The Impact of Electronic Publishing on the Academic Community. From private data to public knowledge. Kennard, O. (1996) Portland Press.
2. Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). Acta Cryst. B35, 2331-2339.

News from Catalyst Science Discovery Centre

Contribution by CICAG Treasurer Dr Diana Leitch MBE, FRSC email diana.leitch@googlemail.com

Photo courtesy of Marketing Cheshire

On 4th December 2018, CICAG Treasurer Dr Diana Leitch, became the new Chair of Trustees at the Catalyst Science Centre and Museum in Widnes. Catalyst is the only Science Centre in the UK to concentrate on the promotion of the chemical sciences to young people and its museum, which was created in 1987, is the only one in the UK which celebrates the great history of the chemical industry. Widnes and its sister town Runcorn, which straddle the River Mersey, were the 'crucible' of the chemical industry in the 19th century. The main Catalyst building which houses the Museum section dates from 1856 and was originally the administration block of Gossage's Soap Works, and it was there that Sir John Brunner and Ludwig Mond met while working on the production of soap and bleach using the LeBlanc alkali process. In 1984 an education block was added, and this continues to provide workshops and shows to encourage young people in particular to participate in STEM activities and to follow STEM careers. At the Millennium block, a laboratory specially designed for children and a theatre were created and are still well used. 40,000 visitors were welcomed last year.

Catalyst, like many other Science Centres in the UK, was in need of an upgrade and investment, and last year the UK Department for Business, Energy & Industrial Strategy and Wellcome Trust agreed that it should be one of the Science Centres to receive infrastructure improvement funding. A huge amount of effort went in to writing the funding bid. A grant of over £775,000 has been allocated to Catalyst but over £330,000 must be found by Catalyst in matched funding by mid-2019 to secure this grant, so a very intensive

period of fund raising is in progress. The outcome will be newly refurbished education studios, theatre, cafe and entrance area, and new interactive exhibits in its Scientific Gallery.

One of these new exhibits will be a spectacular interactive Periodic Table to fulfil Catalyst's commitment to celebrating the International Year of the Periodic Table (IYPT) 2019. This exhibit is being partially funded by a grant from the Royal Society of Chemistry - and CICAG has also donated £2449.00 to enable this exhibit to proceed for which Catalyst is very grateful. Further funding is being sought to cover the full cost of the Periodic Table from other RSC Special Interest Groups and Local Sections. The actual date of Mendeleev's creation of the PT in late February 2019 is being celebrated by work with two local primary schools both of which existed in 1869 and the children will be looking at 10-12 elements related to the industries in their area back in 1869.

On 17th May 2019 we hope to have the opening of a new Gallery funded by AIM/Biffa to remember the work of History Maker, Dr Harry Baker. He was the first Works Chemist at Castner Kellner Works in Runcorn and developed the electrolytic process for the production of chlorine which is still used today 121 years on by Inovyn (Ineos). The past history of Harry Baker, one of Roscoe's students, and his family (Professor Wilson Baker worked on penicillin in WWII, Professor Wright Baker opened the Dead Sea Scrolls, Dr Kathleen Drew Baker saved the Japanese seaweed industry), who were all Manchester graduates, will be depicted, as will the current day multiple uses of chlorine. Again, we will be working with year-5 pupils from a local school, where many of the families will have worked at Castner's (ex-ICI), to look at chlorine and to produce a newspaper of their research.

On a day-to-day basis, schools continue to visit to undertake a range of workshops and shows, half term in February 2019 sees our depiction of 'Freezing and matters Frozen' to educate and entertain families, and the ever-popular sleepovers by uniformed groups and schools at weekends and during the week continue apace. All these activities bring much-needed revenue to fund the day-to-day running of Catalyst. Catalyst is also constantly seeking sponsorship from industrial firms as it is a charity with no dedicated public funding. As those CICAG members who are involved with any charity will know - being a Trustee is a huge commitment and Diana is grateful for any support that can be given, particularly in her new role as Chair of Trustees.

More can be found out about Catalyst by visiting the website at www.catalyst.org.uk. The recent UK Government press release is here.

Nantwich Museum Celebrates Joseph Priestley and the IYPT

Contribution from CICAG Committee Member Dr Helen Cooke, email: helen.cooke100@gmail.com

Right: The museum in Nantwich, Cheshire

In October-December 2019, CICAG will be supporting an outreach project, led by Nantwich Museum, which will bring together various themes which we hope will engage and inspire the Museum's visitors to increase their knowledge of chemistry and the periodic table.

Nantwich Museum is popular within the local community and with visitors to the town. With its primary focus on local history, the Museum hosts permanent and temporary exhibitions, schools' events, talks and also produces publications. Experience shows that historical topics are generally more popular with visitors than science, but for this exhibition we will be crossing disciplinary boundaries through the life of Joseph Priestley, who lived in Nantwich from 1758-1761. Priestley was a Unitarian minister and teacher in Nantwich – it has been suggested that his scientific teachings there were the first ever science lessons. Priestley discovered oxygen after leaving Nantwich and his discoveries will be used as a way to introduce the periodic table.

The Museum is fortunate to have two chemists, both of whom are committee members of the North Staffordshire & South Cheshire Local Section, in the Museum's Research Group (Dr Glynn Skerratt, who is also the Environmental Chemistry IG Committee Secretary) and myself, who will be driving this initiative.

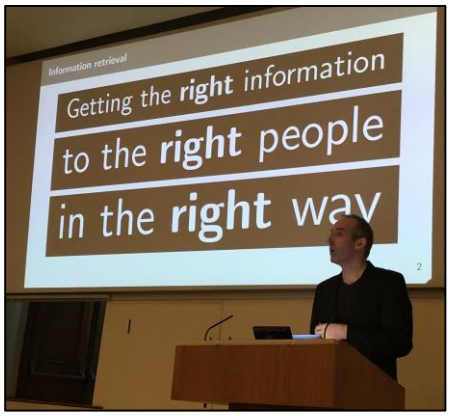
It was heartening to see Roger Highfield's (director of external affairs at the Science Museum Group) recent Chemistry World article, where he describes approaches to public engagement with science, in particular his suggestion to use "hooks" of local interest. We decided to attempt this by blending local history, 18th century religion, education and science – all part of Priestley's story – and from there move on to the periodic table.

Details still have to be finalised, but we'll be reaching out to all age groups, providing talks for adults, information panels and, with help from Heidi Dobbs (RSC's Midlands Education Coordinator), a molecule/element trail for children. We'll even be engaging the Museum's volunteer Craft Group to make us a patchwork periodic table!

As well as CICAG, other partners agreed so far include the North Staffordshire & South Cheshire Local Section, the Environmental Chemistry Interest Group, the Catalyst Museum and Keele University.

Meeting Report: Tony Kent Strix Annual Lecture 2018

Contribution from CICAG member Jane List, email jane@extractinfo.info



Right: Maarten de Rijke presents his lecture

On Friday 23rd November a group of information retrieval enthusiasts packed the lecture hall at the Geological Society in London to find out who would become the recipient of the 2018 Tony Kent Strix Award , and equally importantly to hear from the 2017 award winner.

This year's lecture was given by the 2017 Award winner Professor Maarten de Rijke, Professor of Artificial Intelligence and Information Retrieval at the University of Amsterdam for his work in information retrieval and web searching. Maarten talked about his work under the title 'Retrieval as Interaction'. The talk was also

live streamed to the Dutch-Belgian Information Retrieval Group of 170 people in Leiden. Maarten de Rijke talked about the influences - sentiment, credibility, memory, reputation, and experiences on retrieval of relevant information, and about that most human of approaches - browsing. His work is largely concerned with learning more about how people use the web and its search engines for non-professional search tasks. He talked about the iterative nature of search, and the intent of the searcher (information seeker). He talked about bias of search results, pointing out that bias is both a good as well as bad thing. Indeed utilising bias is necessary to organise results in order to represent them to the information seeker in a manageable quantity.

But before the main lecture Stella Dextre Clarke introduced the afternoon by talking about the history of information retrieval (IR), the award, and its past recipients. The award is now 20 years old, its scope remains broad and this is reflected by the research interests of past recipients. The first Tony Kent Strix winner in 1998 was Professor Stephen Robertson. Stephen lectured at City University, and went on to work at Microsoft Research in Cambridge after winning the award. Stella also took a brave look into the future - postulating that there is now access to infinite content, from anywhere in the world via the web, but noting that the challenge of finding exactly the information you want when you want it still exists. Stella also philosophised about the nature of 'information' itself. And what we mean by 'information retrieval' – are we limiting the concept to textual, computerised, data sets? What will information retrieval mean in an Industrie 4.0 scenario? Is the concept of the 'collection' still valid?

My main IR interest is in patent information retrieval for professional purposes. In our world we remain concerned with 'recall' and 'precision' even as the concept of a 'collection' expands and expands and absolute novelty – anything 'published' anywhere in the world before a certain date - remains a requirement of the patent examination process. This was the first time I'd attended the Tony Strix lecture, and it won't be the last; I thoroughly enjoyed the afternoon reconnecting with fellow information scientists and meeting new researchers and keeping up with IR research.

The 2018 Award went to Professor Pia Borlund of the Department of Archivists, Library and Information Science at Oslo, and Akershus University College of Applied Sciences, in Norway. We look forward to hearing her lecture next year.

RSC International Year of the Periodic Table

Contribution courtesy of the RSC Networks Group: networks@rsc.org

The RSC is inviting members to design and deliver events celebrating the International Year of the Periodic Table (IYPT), in particular to help support creative, diverse and innovative engagement opportunities for communities providing chemistry connections with audiences and science outreach activities. Grants are available for activities throughout 2019 in particular:

- Grants for member networks of up to £1,000 – please discuss with your committees whether you would like to apply for this
- Grants for members of up to £500 – to support members, ChemSocs, and Learn Chemistry Partnerships

There is a rolling application process with new deadlines appearing for each grant consideration stage. Please note that the standard RSC Outreach Fund grants will still be open as usual in addition to these to support wider and/or larger activity.

For more information and to apply please go to rsc.li/iypt or contact Geri Kitley via outreach@rsc.org.

Places of the Periodic Table Interactive Map

With thanks to Carmen Giunta, Professor of Chemistry, Le Moyne College: giunta@lemoyne.edu

Continuing on a theme, readers will no doubt be as delighted as CICAG committee were with this **interactive searchable map** of places associated with the developers of the periodic table of the elements, complete with links to further information. The locations are drawn largely from James and Virginia Marshall's *Rediscovery of the Elements* project, and assembled and supplemented by Carmen Giunta with the encouragement of the ACS Division of the History of Chemistry.

Clarifying the Cahn-Ingold-Prelog Rules

Contribution with thanks to Noel O'Boyle: baoilleach@gmail.com

What happens when four software developers get together and compare their implementations of the CIP rules? This is the background to a recent paper [1] published in Journal of Chemical Information and Modeling.

The CIP system is a series of rules that describe how to assign a stereodescriptor (e.g. R/S, E/Z) to a stereocentre. When Bob Hanson decided to add support for CIP to Jmol, rather than simply read the rules and implement it according to his interpretation as others have done, he decided to work with three other implementations to challenge each other on disagreements and clarify the wording of the rules. The result was published as "Algorithmic Analysis of Cahn–Ingold–Prelog Rules of Stereochemistry: Proposals for Revised Rules and a Guide for Machine Implementation" [1].

Essentially, the issue that the authors are addressing is the fact that existing implementations even in "highly respected software packages" disagree with each other (see the presentation by John Mayfield [2]). By comparing the implementations in Jmol, Centres, Balloon and ChemSketch, they were able to identify cases where "(a) the disagreement was due to different interpretations of CIP rules among software developers, (b) there was a problem with an algorithm or its implementation in code, or (c) the CIP rules themselves were flawed." However, in all cases they were able to come to a consensus, which led to "the discovery of a small number of errors in the Blue Book, two minor flaws in the CIP rules, and a proposal for a new rule".

The paper walks through their discussions of each rule in turn, looking at any issues arising and clarifying any ambiguities. It also includes a validation suite (available online at [3], with an interactive version at [4]) that covers all aspects of the rules and will allow future CIP implementations to avoid the pitfalls that have beset the field in the past.

[1] Hanson, R. M.; Musacchio, S.; Mayfield, J. W.; Vainio, M. J.; Yerin, A.; Redkin, D. Algorithmic Analysis of Cahn–Ingold–Prelog Rules of Stereochemistry: Proposals for Revised Rules and a Guide for Machine Implementation. J. Chem. Inf. Model. 2018, 58 (9), 1755–1765. https://doi.org/10.1021/acs.jcim.8b00324.

[2] Mayfield, J. Comparing Cahn-Ingold-Prelog rule implementations: the need for an open CIP. https://www.nextmovesoftware.com/talks/Mayfield_ChanIngoldPrelog_ACS_201708.pdf

[3] https://github.com/CIPValidationSuite/ValidationSuite

[4] https://cipvalidationsuite.github.io/ValidationSuite/

Meeting Report: AI3SD Network+ Launch, SCI London

Meeting report contributed by RSC CICAG Newsletter Editor Stuart Newbold: stuart@psandim.com

AI3SD: ***Artificial and Augmented Intelligence for Automated Investigations for Scientific Discovery***

Right: Opening speaker Jackie Hunter commences her presentation at the Network Launch event on 5th December 2018

The AI3SD Network+ is a new initiative, funded by EPSRC and hosted by the University of Southampton, which aims to bring together researchers looking to show how cutting edge artificial and augmented intelligence technologies can be used to push the boundaries of scientific discovery.

The inaugural meeting was well-attended (although sadly a significant number of those registered did not attend on the day, doubly-unfortunate given the event was heavily oversubscribed) with quite a diverse flavour from both industry and academia. There was a very good atmosphere, with the breaks and finishing discussion session providing much dialogue.

Jackie Hunter, BenevolentAI

The opening presentation consisted mainly of a top-level look at BenevolentAI, their technologies and methods of working – across the healthcare, physical, chemical and biological disciplines. Jackie started with the somewhat ambitious prediction that "AI will be the saviour of the Drug Discovery Industry", citing the current approach of cost-cutting, M&A activity, and current drug failure-rates (only 6% of molecules currently make it through to market) as otherwise presaging decline. For an evidence-based industry, we are currently not using enough evidence.

AI will impact all drug discovery processes, its adoption is already very rapid, from target selection to molecular design: hypothesis generation (can search hypotheses at scale) -> drug target validation (high-

content analytics) -> lead discovery -> lead optimisation (can reduce LO from 3-years to 1-year and ensure better biological and physicochemical properties) -> preclinical development -> clinical development (better biomarker identification, selection of responsive cohorts, better outcome monitoring).
BenevolentAI has raised over $200m to develop their (now-validated) platform and is the only AI company with activities spanning early discovery to clinical development. Their platform ingests (reads) and contextualises vast quantities of data (both proprietary and freely-available sources; both structured and unstructured). Natural language processing and use of dictionaries allows the creation of relationships which then go on to create systems biology maps.
Jackie showed a video of the system being used – exporting query results to a disease map visualisation tool, which hypotheses were then overlaid, before moving up- and down-stream on pathways. Another example illustrated gefitinib-derived improvement in ALS patients, from which four new targets had apparently been identified. Jackie also illustrated their lead-optimisation tool *Evochem* (a generative chemistry platform to create new molecules). Jackie mentioned two interesting sources - PRO-ACT (an ALS clinical trial database; https://nctu.partners.org/ProACT) and a Nature paper from Wong, Yip, et al on classifying cancer (https://www.nature.com/articles/d41586-018-02881-7).
The take-home message was that AI is already transforming R&D, and we will need new ways of working, funding and organising businesses to harness the greater collaboration and data-sharing inherent in adopting AI.

Adam Prugel-Bennett, University of Southampton
This was a very enjoyable presentation, detailing the advances that have taken place in machine learning ("there is a big buzz about AI, but not quite the same for machine learning"). ImageNet was set-up in 2010, and this proved to be a great test-bed for playing with machine learning. There was something of a revolution in 2012, when Convolutional Neural Networks (CNN) achieved a 10% increase in accuracy over everything else. This trend continued and by 2017, techniques for machine learning were achieving =< 2.5% error rates, compared to the 25% when things commenced. Another staging post was the 2015 achievement of DeepMind in being the first AI programme to beat a professional player at Alpha Go, using Deep CNN (in 2017 DeepMind beat the Alpha Go world number 1 player). DeepMind's technique involved learning the game entirely by self-play. Meanwhile at the same time we also had advances in speech recognition (Alexa), and machine translation (Google Translate). Today, smartphone apps can tell you the name of a bird via a picture or a piece of birdsong.
What is machine learning: Well, it's basically a minimise-a-loss function (define an error). There is, though, some skill in designing the learning system.
Machine learning versus AI: AI is a much larger field. Machine learning is really all about error reduction and recognising patterns in order to make decisions. The first application of machine learning arrived in the areas of fraud detection (banking, finance) and spam detection. There has followed self-drive cars, machine translations, and now applications such as cataloguing the ocean floors and medical diagnostics. But the truth is, machine learning is rather limited. It requires huge quantities of training data. It does work well in computer vision and in language processing, but there are many, many areas where it doesn't work well. There is a push to get machine learning into other domains. But it only succeeds where there is inherent structure. Over the last three years, there has started to be been a focus on:

- Unsupervised learning (i.e. lots of data and no labels)
- Generative Adversarial Networks (GANs) "a solution without a problem"
- Variational Auto Encoders (VAEs) "a poor man's GAN"

A GAN can recognise an image (e.g. a photograph), but can it learn a style (e.g. a van Gogh)? We may be getting good at forging! The current state-of-the-art in machine learning is now 'Image Reasoning' ("which person is wearing the glasses?", "where is the child sitting?"). This is one of the most demanding challenges in AI. But don't panic - machine learning is still pretty rubbish at this, however it is getting better by the tune of 5% per year. When will it end? To make this work, we have developed Automatic Differentiation (AD), Invisible GPU Computing, and so forth. We are now using a large number of big datasets (there is a dataset of 204K images and 1.1M questions), with lots of incredibly motivated young researchers. The Neurex Conference sold out in 6 minutes, and all AI-type conferences are getting 50% and above increases in participation. We are just at the beginning…

Gábor Csányi, University of Cambridge

Gábor talked about Machine Learning for Molecular Dynamics (MD) – many-dimensional function fitting technology, using Gaussian Approximation Potentials and Kernel Ridge Regression. Gábor's team is trying to generate force-fields that are as accurate as high level Quantum Mechanics calculations, all the way from elemental compounds to small molecules. There is lots of complexity. But are the approximations being used sufficient for what we want e.g. "will this drug bind to this protein?" The area, though, is actually an excellent one for machine learning. Gábor showed a table with Representation (e.g. bond length and angle) -> Regression (e.g. parametric' functions, invariant polynomials, artificial NNs, Kernels/ Gaussians) -> Function (e.g. body order expansion, atomic molecular energy) -> Target (e.g. experimental data, quantum chemistry). Data "fills space". Use this approach when fitting functions in many dimensions (e.g. 1000s) – after all, most data is sparse, so it becomes difficult to fit a function well.
Gábor then showed a 'Liner Regression in Many Dimensions' slide – trying to tie the shape in data-space with the function using linear least squares and minimising the loss. Kernel function scales well here in minimising the loss (especially if we introduce weights and regularisers). However, we must pick the correct basis function Kernel to work on. Using methane as an example, we can use the overlap integral, integrate over all 3-D rotations, and use SOAP Kernel to create a basis function.

Gábor's team has developed an open-source gap-framework docker, which is available for download. He quoted Hannah Fry – "there is no revolution in AI, but there is a revolution in computational statistics", and gave a number of examples of work done on different systems –

- Structure and growth mechanics of amorphous carbon films
- Phosphorus (again, has lots of different structures)
- TiO_2 (lots of crystalline forms and amorphous phases)
- GeSbTe storage materials (trying to find the medium range order)
- CH_4 (this is difficult to model without knowing the level of quantum mechanics needed), and he went on to demonstrate various errors in the energy models and used the 'Many-body SOAP-Gap Model'…which when using an atom description, did manage to calculate the density.

Gábor finished with some current work on protein-ligand binding with a brief example of 3-BPA binding to leukotriene hydrolase, using custom force-fields to calculate the binding.

John Overington, Medicines Discovery Catapult
MDC Catapult is a not-for-profit organisation, part of the Catapult Network, charged with helping deliver the UK Government's Industrial Strategy. It is funded by Innovate UK (BIS), with a focus on the SME and translational academic sector, with collaboration as a key methodology. The MDC Informatics Project has several strands –

- The Collaboratory – ELN integration/transfer
- VESPA – variant effect prediction
- CRISPY – their Collaborative Intelligence Platform
- Phi-Scinder – SAR on-demand
- Serendipity-do-dah – Multi-parameter Optimisation tool

They also work on drug cocktails for enhanced efficacy and safety, and on hit-discovery for new target classes e.g. RNA (which John believes is ripe for innovation), transporters and channels.

John's presentation started with a review of the need to get data 'ready for AI'. There are lots of errors in the public data they are involved with – ChEMBL (2.3m compounds, an open-data API is available), SureChEMBL (the public chemical patent resource with 18m structures generated via name-recognition, and available as a client-feed) and Unichem (a single chemical integration source). For instance, errors run to 5% of structures; 2-3% of targets; 1% of activity values. And there is lots of variability in the data too (the underlying 'reproducibility problem') – about the same 10-fold difference for different orthologues as for different labs, a 5-10-fold variance in cell line data, and about 1-2% of compounds from suppliers are probably not the right thing.
Collaborative Intelligence:

- It's often difficult to identify the best collaborators to work with
    - Commercial inertia is often a barrier-to-entry
    - Discoverability problems

- CRISPY – this uses a knowledge graph of UK drug discovery assets and capabilities and a directed graph of assays from targets to clinical trials
- Catapult is using NLP to help collaborators find new partners

Nathan Brown, Head of Chemoinfomatics, BenevolentAI

How can AI help drug discovery? Nathan showed how it is already helping by inventing new reactions, finding new synthetic routes, designing molecules, and finding molecules that interact with particular targets.

Nathan's talk started with a little history, introducing a few papers to take us over the progression in the field since his first 1991 paper on AI by Johnny Gasteiger. Nathan suggested that 'enumeration is wasteful', and as we have more and more data, what do we do with it all? Instead we should use an approach that samples the best data subsets – but to do this, we need sufficiently-good algorithms to find them. Nathan's 2004 J. Chem. Inf. Comput. Sci. paper suggested, given two molecules, that we can identify all the other molecules that lie between them close to a high-dimensional line through descriptor space. This approach can also be used for 'degenerative chemistry' – i.e. starting with two very dissimilar molecules. His 2015 J. Chem. Inf. Model. paper (MOARF) was then illustrated as more recent developments in algorithmic molecular design. But two questions follow – is this all AI? And does computational-design actually work? Nathan's answer to both was yes. Indeed this can all be seen as another rational iteration of what scientists have always done – the last paper is using MOARF as a computational way to do most of work in the ->Design->Make->Test->Design->Make->Test->…cycle.

Nathan then showed us EvoChem, their Multi-Objective Optimisation AI drug design tool with drug-like property prediction. The output ranks a list of AI-designed compounds, along with their metadata. Chemists can then take these molecules and do some live design-work, working alongside the modellers and other scientists (the interactive platform is designed to be used by people from different scientific roles all working together). Once an extracted molecule has been re-designed, this is then put back into EvoChem and re-scored for a read-out of various properties. The next step along this path then becomes 'ok, so now you've told me what to make, show me how to make it', and they are using an AI-powered retrosynthesis to do so – Nathan says this is the first ever retrosynthesis tool which has the full backing of the chemists. The talk finished with a few messages –

- AI is experiencing a renaissance
- Molecular design has been around a long time
- New generative chemistry methods are very powerful, especially when combined with predictive models
- Prospective success and high impact of their methods on BenevolentAI's live drug discovery projects is apparent

Michela Massimi, University of Edinburgh

This was a philosophical perspective on AI. The first question posed by Michela was "how do scientific discoveries happen?" In reality, it takes a long time. There are various philosophical approaches. Reichenbach (Experience and Prediction, 1938) suggested there was a distinction between actually making a discovery, or just making a justification to describe what is being seen. Karl Popper went on to say that the "how" is irrelevant. Thomas Kuhn characterised science as consisting of three phases – those of normality, crises and evolutions. Michela's second questions were (a) "what's it like to discover something inn science?" and (b) "how do we come to discover something in science?" In actual fact, we tend to be ruling things out. This is what happens in AI, and as a result she predicts that progress can and will be achieved in science using AI. Thus the answers to the above are (a) – it's problem-solving; and (b) – there isn't a specific method but a plurality.

However, we do need to be careful not to confuse scientific discovery with the ability of AI to enumerate large numbers of positive and negative results; and there are limits – beware Baconism – deriving an answer by extrapolating what we don't know purely by looking at all the data we do know (which of course may be hugely insufficient). The philosopher David Hume (1748) warned against inductive inference ("nothing guarantees that the unknown will resemble what is already known").

On completing the formal programme and discussion, Prof Jeremy Frey put up a slide with a very impressive list of planned events and activities for AI3SD during 2019. In addition to there being much

planned, the AI3SD website contains quite a lot of information, along with links to receive further information on related events, funding, news and other information.

Chemical Information / Cheminformatics and Related Books

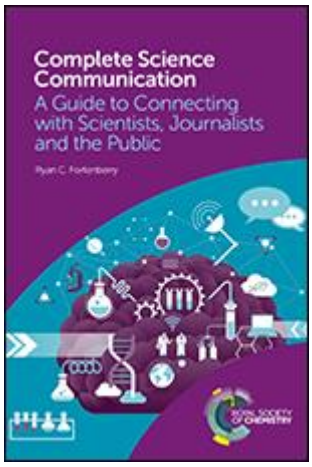
Contributed by RSC CICAG Newsletter Editor Stuart Newbold, email: stuart@psandim.com

Computational Materials Discovery

New technologies are made possible by new materials, and until recently new materials could only be discovered experimentally. Recent advances in solving the crystal structure prediction problem means that the computational design of materials is now a reality.
Computational Materials Discovery provides a comprehensive review of this field covering different computational methodologies as well as specific applications of materials design.
With contributions from pioneers and leaders in the field, this unique and timely book provides a convenient entry point for graduate students, researchers and industrial scientists on both the methodologies and applications of the computational design of materials.

RSC Publishing
Editors: Artem R Oganov, Gabriele Saleh, Alexander G Kvashnin

Complete Science Communication: A Guide to Connecting with Scientists, Journalists and the Public



Science communication is a rapidly expanding area, and a key component of many final year undergraduate and postgraduate courses. Authored by a highly regarded chemist and science communicator, this textbook pulls together all aspects of science communication. Complete Science Communication focusses on four major aspects of science communication: writing for non-technical audiences and science journalism; writing for technical audiences and peer-reviewed journal writing; public speaking of science; and public relations. It first showcases how writing in a journalistic style is done and provides a guide for colloquially communicating science. Then, the art of writing scientific papers is conjoined to this idea to make technical manuscripts more digestible, readable, and, hence, citable. These ideas are next taken into the spoken word so that the scientist can engage in telling their science like that natural human art of campfire stories. Finally, all of these communication concepts are wrapped together in a discussion of public relations, providing the scientist with an appreciation for the marketing directors and news disseminators with whom they will work. Written in an accessible way, this textbook will provide science students with an appreciative understanding of communication, marketing, journalism, and public relations. They can incorporate these aspects into their own practices as scientists, allowing them to liaise with practitioners in the communication field.

RSC Publishing
Author: Ryan C Fortenberry

Data Integrity and Data Governance: Practical Implementation in Regulated Laboratories

The aim of this book is to provide practical and detailed help on how to implement data integrity and data governance for regulated analytical laboratories working in or for the pharmaceutical industry. It provides clarification of the regulatory issues and trends, and gives practical methods for meeting regulatory requirements and guidance. Using a data integrity model as a basis, the principles of data integrity and data governance are expanded into practical steps for regulated laboratories to implement. The author uses case study examples to illustrate his points and provides instructions for applying the principles of data integrity and data governance to individual laboratory needs.
This book is a useful reference for analytical chemists and scientists, management and senior management working in regulated laboratories requiring either an understanding about data integrity or help in implementing practical solutions. Consultants will also benefit from the practical guidance provided.

RSC Publishing
Author: R D McDowall

News from InfoChem and Thieme

Contribution from Dr Josef Eiblmaier, email: je@infochem.de

InfoChem acquired by DeepMatter
DeepMatter has recently announced that it will acquire InfoChem from Springer Nature for a total of approximately €2.0 million. At the same time, the publicly-lited company is looking to raise a further £3.0 million via a new share offer at a price of 2.5 pence per share.

Founded in 1989, Infochem is based in Munich, Germany, and is owned by Springer-Verlag GmbH, part of the Springer Nature, and has 25 employees. It has extensive scientific expertise and a long tradition in developing successful software solutions for handling retrieval, structures and reactions. Its user base and similar focus makes it an ideal match for DeepMatter, who will be able to utilise InfoChem's established data sources and chemical information software tools, assisting it in the development of DeepMatter's DigitalGlassware™ platform, as well as providing specialist staff, revenue streams, and an additional sales channel.

Mark Warne, CEO of DeepMatter, said: "Our aim is to integrate chemistry with technology, thereby enabling a greater use of machine learning and artificial intelligence and reaching a point where chemicals can be autonomously synthesised through robotics. Our first step towards this was the launch of our DigitalGlassware™ Pioneer Programme in 2018 and we have been delighted with its success. The platform is now deployed across seven Pioneer partners in three continents, providing valuable insight into the power of the platform and the benefits it can bring to the scientific community.

"The acquisition of Infochem will accelerate our strategy, bringing additional expertise, software tools and an existing sales channel, strengthening and accelerating the commercialisation of our platform. The net proceeds of the Proposed Placing will enable further technology development, including integration of Infochem's cheminformatics capabilities into DigitalGlasswareTM and an increase in our marketing activities, as we seek to expand our user base in 2019.

"We look forward to being able to welcome the Infochem team to the enlarged group and look to the future with confidence."

Thieme Chemistry deposits nearly 700,000 distinct chemical structures into PubChem

Known for providing high-quality chemical information in the field of organic synthesis, Thieme Chemistry has provided approximately 700,000 distinct structures to PubChem – the open chemistry database at the National Institutes of Health (NIH). In addition to the structural information, the upload includes more than 1,200,000 distinct links from chemicals to the Thieme Chemistry journals and online product portfolio, enlarging the number of chemical structures in PubChem with links to the scientific literature.

PubChem is an open database of chemical molecules at the U.S. National Institutes of Health (NIH) and is used by several million users worldwide. It collects information on chemical structures, identifiers, biological activities, chemical and physical properties as well as patents, health, safety and toxicity data. Since the launch in 2004, PubChem has become a key chemical information resource for scientists, students, and the general public. Now thousands of structures and links to Thieme's chemistry journals SYNTHESIS, SYNLETT, SYNFACTS, as well as to the online reference works Science of Synthesis and Pharmaceutical Substances have been uploaded to the PubChem portal.

Of the 700,000 Thieme Chemistry chemical structures deposited to PubChem, 89% did not have literature links prior to this contribution. In addition, 42% of the 700,000 chemical structures are new to PubChem.

Thieme's contribution, which was facilitated by InfoChem, a longstanding collaboration partner with Thieme Chemistry, improves the discoverability of information about chemicals: "We are proud and very pleased that all the chemical information created during over twenty years collaboration with Thieme now find its way into the world's largest free chemical database. This deposition of approximately 700,000 Thieme compounds makes them even more discoverable to the scientific community", said Dr Josef Eiblmaier, Managing Director at InfoChem.

With its scientific publications, Thieme Chemistry, a part of the Thieme Group, provides chemists in research, science and industry with highly evaluated specialist information on synthetic, organic and general chemistry. All products of Thieme are available in electronic format and predominantly in English language. Full access to current awareness articles from SYNFACTS, primary literature journal articles from SYNTHESIS and SYNLETT, full-text synthetic methodology reviews in Science of Synthesis and encyclopedic information from Pharmaceutical Substances is subject to journal subscription or licensing.

Other Chemical Information Related News

Contributed by RSC CICAG Member Dr Keith White and RSC CICAG Newsletter Editor Stuart Newbold

[*All hyperlinks correct & working as of 15 January 2019*]

A Brief Mention: The new RSC Librarian Hub
Designed to help librarians and information specialists bring all the right elements to their libraries!

The RSC Librarian Hub contains free resources, tools, news, and downloads to help members working in information serve library communities. Many of the resources are freely available, although some do require the user to be registered for a free librarians' portal account prior to access.

Successor to the National Chemical Database Service Announced
The RSC has announced that the National Chemical Database Service (NCDS) will be succeeded from 2019 by the National Physical Sciences Data-science Service (PSDS) as an EPSRC National Research Facility. The new service, run jointly by the School of Chemistry at the University of Southampton and the Scientific Computing Department at the Science and Technology Facilities Council (STFC), will expand on the current database provision to include modern data science techniques such as machine learning and artificial intelligence. The RSC is now working with Southampton and STFC to ensure a smooth transition and continuity of access for users of the data resources, and overlap is likely to extend into 2019 as the PSDS develops its platform.
http://blogs.rsc.org/chemical-database-service/2018/11/16/successor-to-the-national-chemical-database-service/?doing_wp_cron=1546718089.8199260234832763671875
Source: RSC Publishing

Here's what Scientists Searched for in 2018: AI is up, Stress is down
Ever wondered what your colleagues are searching for online? 'Cancer', 'blockchain' and 'big data' were among the top search terms in a major science database in 2018. The search-term data — for 2017 and 2018 — were provided to Nature's news team by the scholarly database Scopus, which indexes abstracts and references from thousands of journals. 'Cancer' takes the top spot for both years. New entries in 2018 hint at what has increasingly been on researchers' minds — 'machine learning' and 'deep learning' broke into the top twenty and 'artificial intelligence' (AI) moved from thirteenth place to fourth.
https://www.nature.com/articles/d41586-018-07879-9
Source: Nature

Google Unveils Search Engine for Open Data
The tool, called Google Dataset Search, should help researchers to find the data they need more easily, and locates files and databases on the basis of how their owners have classified them. It does not read the content of the files themselves in the way search engines do for web pages.

https://www.nature.com/articles/d41586-018-06201-x
Source: Nature

AAIH - A New Platform for Research and Industry Collaborations in Artificial Intelligence
To accelerate the adoption of AI in healthcare, improve transparency and promote collaboration, leading members of the AI community have come together to collaborate and establish the Alliance for Artificial Intelligence in Healthcare (AAIH), an industry-wide research collaboration platform.
https://www.eurekalert.org/pub_releases/2018-09/imi-anp091218.php
Source: EurekAlert! Science News

Clarivate Analytics Announces 2018 Winner of Prestigious Eugene Garfield Award
Dr Orion Penner has received the $25,000 award for his innovative proposal for an indicator of the extent to which a research publication represents a pivot between old and new ideas.
https://clarivate.com/blog/news/clarivate-analytics-announces-2018-winner-of-prestigious-eugene-garfield-award/
Source: Clarivate Analytics

Pharma Companies put Faith in AI for Breakthroughs
The excitement about AI, machine learning, and big data, have all contributed to a boom in health-tech start-ups in a market traditionally dominated by big pharma.
https://www.ft.com/content/e450a688-ddfb-11e8-b173-ebef6ab1374a
Source: Financial Times

How Unpaywall is Transforming Open Science
Unpaywall has become indispensable to many academics, and tie-ins with established scientific search engines could broaden its reach. The free service locates open-access articles and presents paywalled papers that have been legally archived and are freely available on other websites to users who might otherwise have used a paywalled version instead. Since one part of the technology was released in 2016, it has become indispensable for many researchers. And firms that run established scientific search engines are starting to take advantage of Unpaywall. This July, Elsevier announced plans to integrate Unpaywall into Scopus, allowing it to deliver millions more free-to-read papers to users than it does currently. Scopus's embrace of Unpaywall, along with similar moves by other search engines, means that much more open-access content is now at researchers' fingertips. These deals are also enabling funders, librarians and others to study open-access publishing trends comprehensively for the first time.
https://www.nature.com/articles/d41586-018-05968-3
Source: Nature

Is Machine Learning Overhyped?
To explore the space between what some have promised and what machine learning might actually deliver—and to discern among chemists a consensus about the much-ballyhooed tool—C&EN has examined some of the fields where it's generating the most enthusiasm and skepticism.
https://cen.acs.org/physical-chemistry/computational-chemistry/machine-learning-overhyped/96/i34
Source: Chemical & Engineering News

UK Research and Innovation joins Europe-wide Ambition on Open Access
UKRI has helped launch cOAlition S, a collective declaration of a commitment to open access publishing in academia. The statement sets out the ambition that, by 2020, the results from publicly funded research will be open access, meaning that it is free for anyone to access. This is in line with UK's intention that, by 2020, the UK will be publishing almost all of our research output through open access.
https://www.ukri.org/news/uk-research-and-innovation-joins-europe-wide-ambition-on-open-access/
Source: UKRI

European Funders Seek to end Reign of Paywalled Journals
Frustrated with the slow transition toward open access in scientific publishing, 11 national funding organisations in Europe will, as of 2020, require every paper resulting from research funded by its members to be freely available from the moment of publication.
http://science.sciencemag.org/content/361/6406/957
Source: Science

Accenture and Merck Collaborate with Amazon Web Services to lLaunch a Research Platform to Drive Innovation in Drug Discovery and Scientific Research
The cloud-based informatics research platform is designed to help organisations in the life sciences industry improve productivity, efficiency and innovation in the early stages of drug development by creating open, industry-standard application programming interfaces for core research functions, allowing researchers to rapidly adopt new capabilities.
https://newsroom.accenture.com/news/accenture-and-merck-collaborate-with-amazon-web-services-to-launch-a-research-platform-to-drive-innovation-in-drug-discovery-and-scientific-research.htm
Source: Accenture

The Importance of Industrial Publications
Industrial publications are a very valuable and multifaceted tool for the wider catalysis community; they can foster the productive collaboration of university and corporate research laboratories, an essential partnership for the solution of important societal problems.
https://www.nature.com/articles/s41929-018-0119-0
Source: Nature Catalysis

Passing Twitter Follower Threshold lets Scientists Escape Echo Chamber
Social media is a powerful tool for researchers looking to share their findings with a wider audience. However, for academics to reach out beyond their usual audience of colleagues and peers they need to pass a threshold of followers on social media before they stop preaching to the choir and start reaching the general public.
https://www.chemistryworld.com/news/passing-twitter-follower-threshold-lets-scientists-escape-echo-chamber/3009260.article
Source: Chemistry World

Researchers Celebrate Rejection of Controversial Copyright Policy
Proposed changes to EU legislation would introduce barriers to science and innovation, campaigners say.
https://www.chemistryworld.com/news/researchers-celebrate-rejection-of-controversial-copyright-policy/3009261.article
Source: Chemistry World

Science Journals end Open-Access Trial with Gates Foundation
The publisher of Science last month ended a pilot partnership that allowed open-access (OA) publishing for researchers funded by the Bill & Melinda Gates Foundation. The trial was an effort to accommodate a policy clash between the Gates Foundation, which has enforced strict OA demands since 2017, and publishers that run subscription journals which don't comply with those terms.
https://www.nature.com/articles/d41586-018-05729-2
Source: Nature

HighWire Leads Industry Rollout of Universal CASA, in Partnership with Google Scholar
HighWire, the technology partner of choice for leading commercial and scholarly publishers, has become the first systems supplier to roll out the new 'Universal Campus Activated Subscriber Access' service (Universal CASA) in cooperation with Google Scholar. All content published through HighWire's Intelligent Publishing Platform will be compatible with Universal CASA - meaning researchers can access the content any time, anywhere, from any device and any source.

https://www.highwirepress.com/news/highwire-leads-industry-rollout-universal-casa-partnership-google-scholar
Source: HighWire

35,000 Papers in the Biomedical Literature Might need Retracting
A new analysis of the biomedical literature estimates that as many as 35,000 papers may be candidates for retraction due to inappropriate image duplication.
https://www.chemistryworld.com/news/35000-papers-in-the-biomedical-literature-might-need-retracting/3009283.article
Source: Chemistry World

Robot Chemist discovers new Molecules and Reactions
A new type of artificial-intelligence-driven chemistry could revolutionise the way molecules are discovered, scientists claim. Chemists from the University of Glasgow discuss how they have trained an artificially-intelligent organic chemical synthesis robot to automatically explore a very large number of chemical reactions.
https://www.gla.ac.uk/news/headline_598161_en.html
Source: University of Glasgow

Novartis's Cultural Revolution Will Be AI-Enabled, Desk-Free
The AI programme for the sales team will send daily texts and emails helping recipients to plan their time, with 10,000 employees in the first stage of the roll-out.
https://www.bloomberg.com/news/articles/2018-07-24/novartis-s-cultural-revolution-will-be-ai-enabled-and-desk-free
Source: Bloomberg

Glaxo CEO Looks to Silicon Valley in Revamping Drug Research
GSK is acquiring a $300m stake in genetic-testing company 23andMe inc. and revamping its approach to research, betting that the use of genetic data to identifiy new targets plus smarter patient selection criteria for clinical trials recruitment will lead to better productivity outcomes as the company seeks to improve performance.
https://www.bloomberg.com/news/articles/2018-07-25/plotting-a-faster-course-drugmaker-glaxo-teams-up-with-23andme
Source: Bloomberg

The ban-wagon rolls on: YouTube's removal of Chemistry Videos demonstrates a lack of Scientific understanding that will only hurt budding Scientists
What do making chocolate cake and phenylacetic acid have in common? The answer, surprisingly enough, is that both are featured in chemistry videos that have been banned from YouTube. It seems that that the cake may have been just too chocolatey for the censors, but quite what they have against organic synthesis remains a mystery.
https://www.chemistryworld.com/news/the-ban-wagon-rolls-on/3009290.article
Source: Chemistry World

Digital Science Launches Anywhere Access for Fast, One-Click Delivery of Full-Text Scholarly Content
DS's 'Anywhere Access' cloud-based library technology is the first cloud-based solution to bring one-click access to millions of full-text scholarly articles.
https://www.digital-science.com/press-releases/digital-science-launches-anywhere-access-for-fast-one-click-delivery-of-full-text-scholarly-content/
Source: Digital Science

Springer Nature creates new role to lead on Research Integrity
Dr Suzanne Farley has been appointed to the new position of Research Integrity Director to ensure the growing volume of scientific content being published continues to be rigorously assessed, with robust processes in place to prevent and address research misconduct and breaches of publication ethics.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-creates-new-role-to-lead-on-research-integrity/15986566
Source: Springer Nature

Elsevier/Impactstory agreement will make open access articles easier to find on Scopus
With Elsevier's new partnership with Impactstory↗, a nonprofit that creates online tools to make science more open and reusable, researchers will soon be able to find open access content on Scopus more efficiently.
https://www.elsevier.com/connect/elsevier-impactstory-agreement-will-make-open-access-articles-easier-to-find-on-scopus
Source: Elsevier

Figshare Data Now Available in Dimensions, Enhancing Discoverability Of Supplemental Data
The company has announced what they say will be the first of many integrations with Dimensions, part of the Digital Science portfolio of products. Users can now view any open data hosted on Figshare, which is associated with a publication, directly within the Dimensions platform. The integration allows for a more seamless experience for the Dimensions user by displaying associated data directly on the article details page using the Figshare Viewer, without the need to download any files. The technology, which is unique to Figshare, can preview over 1,200 different file types including data, videos, code, images and many other files that are commonly used in research.
https://figshare.com/blog/Figshare_Data_Now_Available_in_Dimensions_Enhancing_Discoverability_Of_Supplemental_Data/416
Source: Figshare

Elsevier to Acquire Aries Systems, a Best-in-class Publication Workflow Solutions Provider
Aries' products are used by journals, books and other publications for manuscript submission, peer review, production tracking and eCommerce.
https://www.prnewswire.com/news-releases/elsevier-to-acquire-aries-systems-a-best-in-class-publication-workflow-solutions-provider-815827594.html
Source: PR Newswire

5000 German Scientists have Published in Predatory Journals
Public broadcasters in Germany are reporting that more than 5000 of the country's scientists have published work in what they call 'pseudo-scientific' or predatory journals. NDR and WDR, together with the newspaper Süddeutsche Zeitung, found papers authored by German researchers from across academia and industry in online publications that charge large upfront fees to publish open access work that does not undergo full peer review or proper editorial checks. In Germany, they report, the number of such publications is five times higher than five years ago and the scale of predatory publishing activity continues to grow. Globally, they estimate 400,000 scientists have been involved.
https://www.chemistryworld.com/news/5000-german-scientists-have-published-in-predatory-journals/3009341.article
Source: Chemistry World

How to pick an Electronic Laboratory Notebook
The ELN market encompasses considerable variety; a study conducted in 2016 by the University of Southampton, UK, identified 72 active products. To help find a software solution that suits particular needs, experienced users suggest taking a number of steps before taking the plunge.
https://www.nature.com/articles/d41586-018-05895-3
Source: Nature

Springer Nature extends its support for ORCID to enable more authors to receive proper credit for their work

Following a successful trial in 2017 across 46 Springer Nature journals (including 14 Nature-branded journals, 10 BMC journals and 22 Springer journals), which saw the use of ORCID identifiers mandated for corresponding authors of primary research manuscripts, Springer Nature continues to avidly support the take up of ORCID, the unique digital identifier for scholarly authors.

https://group.springernature.com/gp/group/media/press-releases/springer-nature-extends-its-support-for-orcid/16066924

Source: Springer Nature

Royal Society of Chemistry Renews Partnership with ACD/Labs to Continue Providing Industry-Leading Data to Worldwide Research Community

The companies have announced an extension of their collaboration to provide ChemSpider with predicted physicochemical properties and chemical nomenclature.

https://www.acdlabs.com/company/media/pr/180726_rsc-renews-partnership-with-acdlabs.php

Source: ACD/Labs

Dimensions Indexes over 1.4 million Patents from Russia

The service has added over 1.4 million patents from Russia into the Dimensions platform, which provides information on scholarly research activity worldwide. The only research insights platform to consistently link Russian patent information to publications from universities and research institutes globally, Dimensions aims to transform the way research is discovered, accessed and analysed. In addition to patents, Dimensions is also the most complete source of funded grant information, with data for projects totalling over $1.3tn funding from more than 250 organisations worldwide, including key funders from G20 and OECD countries and beyond.

https://www.digital-science.com/blog/news/dimensions-indexes-over-1-4-million-patents-from-russia/

Source: Digital Science

Thieme introduces SynOne – a new discovery tool for Synthetic Chemistry Methods

Organised systematically by compound classes to ensure easy access, SynOne links up content from journals and reference works, including Synthesis, Synlett, Synfacts, Science of Synthesis and Pharmaceutical Substances. The innovative interface offers organic chemists fast access to relevant and reliable syntheses in context.

https://www.thieme.com/for-media/1557-thieme-introduces-synone-a-new-discovery-tool-for-synthetic-chemistry-methods

Source: Thieme Medical Publishers

New Genra Module in EPA's Comptox Dashboard will help Predict Potential Chemical Toxicity

To reduce uncertainties and develop a more objective approach, EPA researchers have developed an automated read-across tool called Generalized Read-Across (GenRA), and added it to the newest version of the EPA Computational Toxicology Dashboard.

https://www.epa.gov/sciencematters/new-genra-module-epas-comptox-dashboard-will-help-predict-potential-chemical-toxicity

Source: EPA Science Matters Newsletter

Kudos and BMJ Announce new Partnership

The Kudos Publisher Partner Programme gives publishers the opportunity to support their author communities in copyright-compliant sharing. In return, they benefit from both the additional high quality usage this generates, and the intelligence on which channels and networks are being used for sharing, and to what effect – including scholarly collaboration networks, via the new Kudos shareable PDF.

https://blog.growkudos.com/2018/09/19/kudos-and-bmj-announce-new-partnership/

Source: Kudos

Royal Society of Chemistry Partnership with Publons Expands to Give Reviewers more Recognition
Publons is a service that reviewers can sign up for and record their peer review history, as a way of "getting recognition for peer review". Their mission is to speed up research by harnessing the power of great peer review, working with publishers, institutions and researchers to turn peer review into a measurable output that can be used to demonstrate a researcher's standing, impact and influence in their field.
https://clarivate.com/blog/news/royal-society-of-chemistry-partnership-with-publons-expands-to-give-reviewers-more-recognition/
Source: Clarivate Analytics

Google Search Turns 20, but will we Every Fully Trust it?
Google's biggist challenge is winning over its many sceptics.
https://mashable.com/article/google-search-20-years-integrity/?europe=true
Source: MashableUK

How Pharmaceutical Research is Navigating the Data Lake
A trend in large-scale data storage rigs cloud computing with advanced analytic software.
https://cen.acs.org/business/informatics/pharmaceutical-research-navigating-data-lake/96/i40
Source: Chemical & Engineering News

Major Publishers sue ResearchGate over Copyright Infringement
Elsevier and the American Chemical Society say that the academic-networking website violates US copyright law.
https://www.nature.com/articles/d41586-018-06945-6
Source: Nature

New Platform Unlocks Publisher's Data and Offers a Transparent new Approach
Ingenta, provider of technology solutions to publishers, has officially launched *Aperture*, an intuitive new web-based platform which gives publishers the ability to unlock data and information stored within their systems, and share it with employees and trusted third parties. The application empowers publishers to provide key data which can be accessed on demand in real time, regardless of where the information is stored. Owners can set up access rights so that business insights can be shared securely and on their own terms, while ensuring that sensitive information is shared on a 'need to know' basis.
https://www.ingenta.com/news-article/new-platform-unlocks-publishers-data-and-offers-a-transparent-new-approach/
Source: Ingenta

Design Thinking: Lack of Design Data is Hampering Innovation
Research from BCGDV and Matmatch has found that 80 per cent of materials research starts with Google.
https://www.theengineer.co.uk/design-thinking/?cmpid=tenews_6416001&utm_medium=email&utm_source=newsletter&utm_campaign=tenews&adg=34a63896-c775-4e92-8965-25d6374b7e03
Source: The Engineer

Clarivate Analytics data Powers Annual Reuters' Ranking of the World's Most Innovative Universities
The fourth annual list is based on empirical data including patent filings from Derwent Innovation and research paper citations from the Web of Science to identify the educational institutions most successful in advancing science, inventing new technologies, and powering new markets and industries.
https://clarivate.com/blog/news/clarivate-analytics-data-powers-the-annual-reuters-ranking-of-the-worlds-most-innovative-universities/
Source: Clarivate Analytics

China Awaits Controversial Blacklist of 'Poor Quality' Journals
A proposal by the Chinese government to create a national blacklist of journals is creating much debate among the country's scientists, who are still waiting for the list to be revealed, five months after the plan to create it was announced.
https://www.nature.com/articles/d41586-018-07025-5
Source: Nature

Mobile App Data Sharing 'Out of Control'
Nearly 90% of free apps on the Google Play store share data with Google parent company Alphabet, the Financial Times reported. Google said it had clear policies for how developers could handle data, and that the research had mischaracterised some "ordinary functions" of apps.
https://www.bbc.co.uk/news/technology-45952466
Source: BBC News

One publisher, More than 7000 Retractions
Some 40% of the retracted publications in the Retraction Watch database have a single curious origin: Over the past decade, the Institute of Electrical and Electronics Engineers in New York City has quietly retracted thousands of abstracts of conference presentations. Most were written by authors in China and had been peer reviewed before publication. The episode may reflect the more rapid and less intensive form of peer review that conference submissions often undergo compared with papers submitted to traditional journals. That approach allows for quick sharing of ideas in fast-moving fields such as computer science, but also allows mistakes to slip through.
http://science.sciencemag.org/content/362/6413/393
Source: Science

Citation Capture Report: Challenges and Opportunities with Citing Archives Consistently
Research Libraries UK (RLUK), The National Archives (TNA) and Jisc have published a report to kickstart a discussion about citation practices, and how to standardise references to unique and distinct collections (UDC's) held in repositories across the UK.
https://www.rluk.ac.uk/citation-capture-report/
Source: RLUK

First Angewandte Editor Based in China
Dr Xin Su is Editorial Manager and first Editor of Angewandte Chemie based in Shanghai, China. Angewandte is the flagship journal of the GDCh (German Chemical Society) and one of the most renowned chemistry journals in the world. It is the first time that there is an Angewandte editor based outside of Germany.
https://www.chemistryviews.org/details/ezine/11109607/First_Angewandte_Editor_Based_in_China.html
Source: ChemistryViews

Software Directs Automated Synthesis
"Chemputer" could relieve humans from the busywork of routine syntheses
https://cen.acs.org/physical-chemistry/computational-chemistry/Software-directs-automated-synthesis/96/web/2018/11
Source: Chemical & Engineering News

Dispute over Reaction Prediction puts Machine Learning's Pitfalls in Spotlight
Learning algorithms promise to overhaul drug discovery, synthesis and materials science. 'But since more and more chemists come to the field [of machine learning], unfortunately, sometimes best practices aren't followed,' says Olexandr Isayev, chemist and machine learning expert from University of North Carolina at Chapel Hill, US.
https://www.chemistryworld.com/news/dispute-over-reaction-prediction-puts-machine-learnings-pitfalls-in-spotlight/3009912.article
Source: Chemistry World

Merger of Jisc and Eduserv will create 'UK public sector tech powerhouse'
The new organisation will offer technology advice to higher education, governmental departments and healthcare providers 'without duplication of effort, time and money', said Jisc head Paul Feldman. Overall, Jisc–Eduserv will serve more than 20 million users, covering the UK, as well as Australia and Scandinavia.
https://www.chemistryworld.com/news/merger-of-jisc-and-eduserv-will-create-uk-public-sector-tech-powerhouse/3009923.article
Source: Chemistry World

Machine learning marched forward
New algorithms predict chemical properties and reactions, with scientists showing numerous ways in which machine learning can help explore chemical space.
https://doi.org/10.1021/cen-09649-cover2
Source: Chemical & Engineering News

Elsevier Acquires Science-Metrix Inc., provider of Research Analytics Services and Data
Science-Metrix is known for high-quality and independent bibliometric analysis and research evaluation. Its services enable evidence-based decision-making, strategic planning and outcome assessment processes for governments, international organisations, universities, scientific societies, publishers and technology companies.
https://www.prnewswire.com/news-releases/elsevier-acquires-science-metrix-inc-provider-of-research-analytics-services-and-data-869069769.html
Source: PR Newswire

Digitalisation, innovation and analytics: A year (of blogging) in Review
The new 2018 CAS blog focused on topics that felt to be the most relevant to global business strategies, including:

- Imperatives for success in sci-tech digitalization, AI and machine learning initiatives
- Tools and techniques to enhance innovation across your organization
- Emerging global science and technology trends and opportunities in pharma, specialty chemicals, cosmetics, materials science and more

CAS now highlights three of the most popular posts that best represent these topics.
https://www.cas.org/blog/digitalization-innovation-and-analytics-year-blogging-review
Source: CAS

Egypt and Pakistan had highest rise in research output in 2018
Global production of scientific papers hit an all-time high this year, estimates show, with emerging economies rising fastest.
https://www.nature.com/articles/d41586-018-07841-9
Source: Nature

MDPI announces new platform - Encyclopedia
The online reference is created and curated by active scholars. It aims to highlight the latest research results as well as providing benchmark information for researchers and the general public interested in accurate and advanced knowledge on specific topics.
https://www.mdpi.com/about/announcements/1442
Source: MDPI

Clarivate Analytics Launches Advanced Predictive Analytics Solution to Help Accelerate Drug Development
Cortellis Analytics – Drug Timeline and Success Rates (DTSR), is part of the Cortellis suite of intelligence solutions for drug development and commercialisation.
https://www.prnewswire.com/news-releases/clarivate-analytics-launches-advanced-predictive-analytics-solution-to-help-accelerate-drug-development-300772583.html

Source: PR Newswire

How many Papers do you read a Week?
Everyone needs a strategy to keep up with interesting papers.
https://www.chemistryworld.com/opinion/how-many-papers-do-you-read-a-week/3009892.article
Source: Chemistry World

BMJ and UNSILO partner to create AI-based Subject Collections
UNSILO applies machine learning and AI tools to identify significant concepts from a corpus of text. These concepts form the basis of a wide range of solutions to publishing workflows, including building subject collections, identifying related articles, finding relevant journals, and many other areas.
https://digital.bmj.com/bmj-and-unsilo-partner-to-create-ai-based-subject-collections/
Source: BMJ Publishing