ROYAL SOCIETY OF CHEMISTRY | CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP

NEWSLETTER summer 2018

Above: Peter Willett (left) & Johann Gasteiger; see pages 12, 18

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.rsc.org/CICAG
https://www.liverpool.ac.uk/~ngberry/cicag/index.htm

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

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

This has been a busy few months for CICAG with a number of changes behind the scenes. Frank Longford, an EPSRC Doctoral Prize Research Fellow at the University of Southampton, has been co-opted onto the committee to co-ordinate the social media channels. Both the Twitter feed (https://twitter.com/RSC_CICAG) and LinkedIn page (https://www.linkedin.com/groups/1989945) now have more regular submissions and I encourage you to join in the conversations. We have also used the Twitter account for CICAG committee members attending conferences to provide live Twitter updates, with Nathan Brown providing updates from the recent International Conference on Chemical Structures at Noordwijkerhout. This provided a conduit for scientists not at the meeting to ask questions from the "virtual floor".

CICAG Committee is delighted to announce that Dr Samantha Hughes has agreed to join the Committee as a co-opted member. Samantha is currently an Associate Director at AstraZeneca in Cambridge and brings a wealth of experience in computational chemistry and cheminformatics. She is particularly keen to support some of the upcoming scientific meetings.

The new CICAG website developed under the direction of Neil Berry is now live (http://www.rsccicag.org) and includes a CICAG twitter feed section. We plan to add additional content based on member feedback, so please visit the site and have a browse.

CICAG supported a meeting on Predicting Reaction Outcomes organised by the Dial-a-Molecule network at the Bradfield Centre on the Cambridge Science Park on March 26th. This meeting generated a very active discussion on the need for greater data access to support the development of the new machine learning tools for reaction prediction.

CICAG, together with BMCS, organised a meeting entitled "Artificial Intelligence in Chemistry" (http://www.maggichurchouseevents.co.uk/bmcs/AI-2018.htm). This was an extremely successful one day meeting. Registration for the event had to be closed and the event could have been sold out many times over. There were more oral abstract submissions that we had time for and we had more poster abstracts than we had space for. I'm also delighted to be able to report that we were also able to support double the number of bursaries than we had originally budgeted for. The meeting report is also in this newsletter. The feedback from the meeting has been extremely positive and the organising committee have already started work on a 2-day follow up meeting in 2019, probably in Cambridge. A significant number of people have already indicated that they will be attending, so we anticipate this will also be another very popular meeting. As part of the feedback form attendees were asked to suggest topics for the 2019 meeting which met with an enthusiastic response! However if you did not manage to get to the meeting, we would still like to hear about topics you would like to see included in this second event.

Finally, CICAG together with BMCS, is organising a workshop on computational tools for drug discovery on 10th April 2019, at The Studio, Birmingham, UK. We aim to open registration in October. The workshop will provide attendees with "get you started" tutorials with a variety of desktop tools together with demo licenses to allow users to try them out on their own data.

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
Spectroscopic Data to Chemical Knowledge	2018	AstraZeneca, Macclesfield	To be organised jointly with the Molecular Spectroscopy Interest Group and Dial-a-Molecule Grand Challenge Network
Second AI in Chemistry Meeting	2018	Cambridge	2-Day meeting, date TBC
Tony Kent Strix Award and Annual Lecture	23 Nov 2018	Geological Society, London	See also page 8
Workshop on Computational Tools for Drug Discovery	10 Apr 2019	The Studio, Birmingham	Registration will be open October 2018
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
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)

The Formation of the CICAG

Recently we have been sorting out and scanning all the historical documents of the group. During this process we came across the notes on the discussion surrounding the formation of CICAG.

By the middle of 2005 following the retirement of one of its key officers and the untimely death of another, the Computer Applications Subject Group (CASG) had become inactive and the RSC was increasing worried about this lack of activity. At one of the regular meetings for subject group secretaries, Doug Veal the secretary of the Chemical Information Group (CIG), noted that CIG was probably the closest group to the CASG in terms of interest.

At a meeting between Bob McDowall chair (CASG) and Doug Veal it was agreed there was substantial overlap and this lead to a proposal from the RSC.

i) Discussions should take place between the Computer Applications Group and the Chemical Information Group to explore options. Members of the Computer Applications Group could be invited to join the Chemical Information Group or the two groups could be merged.

ii) If a merger is not an appropriate option the Computer Applications Group should be wound up.

At the 2006 CIG Annual General Meeting the merger was proposed.

To agree that the Chemical Information Group merges with the Computer Applications Subject Group and that the CIG constitution be amended by modifying clauses 1 and 2 to read as follows:

1. Name

The Subject Group shall be called the "Chemical Information and Computer Applications Group" of The Royal Society of Chemistry, reporting to the Science and Technology Board. The Group shall be deemed a Charitable Trust.

2. Aims

To keep chemists and other users of chemical information and of computer applications for gathering, storing, retrieving and using chemical information aware of relevant activities, services and developments by means of organising meetings and lectures, promoting newsletters and publications of interest, and representing members' views to information and computer applications providers.

On 18 January 2007 the deed was approved by the RSC, and CICAG was established.

CICAG Committee News

Contributions by RSC CICAG Committee Members Dr Chris Swain, email: swain@mac.com,Neil Berry, email: ngberry@liverpool.ac.uk, and Dr Frank Longford, email: f.longford@soton.ac.uk

New CICAG Committee Members

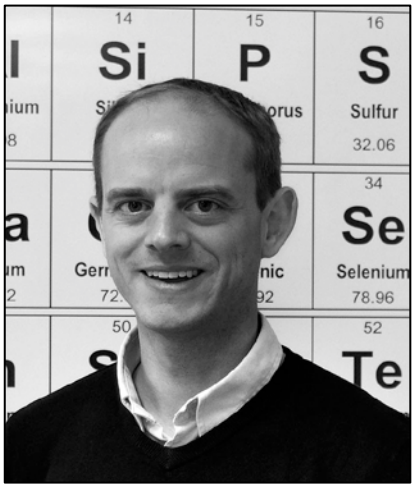
CICAG Committee is delighted to announce that Dr Samantha Hughes has agreed to join the Committee as a co-opted member. Samantha is currently an Associate Director at AstraZeneca in Cambridge and brings a wealth of experience in computational chemistry including SBDD, LBDD, virtual screening, library design, HTS triage and data mining. Samantha also has experience in the design of novel proprietary tools for fragment based drug discovery and is on the organising committee for the RSC Fragments Meetings.

Welcome, Samantha.

CICAG is similarly delighted to have Dr Frank Longford join the committee as a co-opted member. Frank is currently a postdoctoral researcher at the University of Southampton specialising in applying computational simulation techniques to model chemical and biophysical systems. Frank obtained his PhD from Southampton's Institute of Complex Systems Simulation under the supervision of Prof. Jeremy G. Frey, and is currently supported by an award from EPSRC, investigating how to track prostate cancer progression by studying the behaviour of collagen fibres in the extracellular matrix. As part of his committee work Frank will be co-ordinating CICAG social media channels.

Welcome, Frank.

CICAG Committee Member Dr Neil Berry – New Acting Head of Department of Chemistry, University of Liverpool



It was announced on 15th March 2018 that Dr Neil Berry would be interim Head of Department of Chemistry at the University of Liverpool. Dr Berry said: "It is a great honour to be asked to be Head of Department of Chemistry at Liverpool University. I have been in the department for over 15 years and have seen it going from strength to strength. As a Russell Group institution, the University of Liverpool is one of the most exciting places in the UK to study or research Chemistry. Our research is highly rated in the most recent 2014 Research Excellence Framework - 1st in the UK for 4* and 3* research (world leading and internationally excellent). Our teaching and student experience is also very highly regarded as shown by 2nd position in the 2019 Guardian league table. I look forward to working with colleagues in order to maintain our standards of excellence."

Congratulations to Neil on behalf of CICAG.

Catalyst to create new gallery - Harry Baker and 120 years of Chlorine!

Contribution by CICAG Treasurer Dr Diana Leitch MBE, FRSC email diana.leitch@googlemail.com and Meryl Jameson, Marketing Manager at Catalyst Science Discovery Centre, email meryl@catalyst.org.uk

catalyst science discovery centre

The Catalyst Science Discovery Centre and Museum in Widnes, Cheshire, the only UK museum celebrating the history of the chemical industry, has been awarded a grant of £36,159 to create a new permanent exhibition entitled '*Harry Baker and 120 years of chlorine*' which is being funded by an AIM Biffa Award, through their History Makers Project. The new exhibition in a special Baker Gallery in Catalyst will celebrate the life and work of little-known local man but distinguished chemist, Dr Harry Baker (1859-1935), who was a student of two internationally famous chemists Professor Sir Henry Roscoe FRS in Manchester and Professor Robert Bunsen in Heidelberg. It will concentrate on his amazing work from 1897 onwards, (when he came with his family to Runcorn), on the development of chlorine production by the electrolysis of brine pumped from central Cheshire to the newly-opened factory of the Castner Kellner Alkali Company (later ICI, now Inovyn) at the site in Weston Point, Runcorn using the famous mercury rocking cells. The exhibition will depict how this work links with present day science, the many current and vitally-important uses of chlorine and chlorine products and the continued production of chlorine to this day, 120 years on, at the same site in Weston Point. Chlorine has saved millions of lives worldwide in key areas of disinfection and still does.

The life and work of this fascinating man will be interpreted and brought to life for the first time, using items, photographs and information from the Catalyst archives that are currently not accessible or used. The development of this exhibition and gallery will create new resources which will include panels, interactives, activities and talks to inspire families and children to understand about the continued importance of chlorine and its uses and to relive through oral history the working lives of many local people. His legacy and influence locally and nationally were extensive as was his inspiration and educational assistance to his own children, two of whom, Professor Wilson Baker FRS and Professor Wright Baker FREng, went on to great scientific and technological achievements themselves.

The History Makers scheme funds museums to create new exhibitions that will inspire the public through the lives and achievements of extraordinary, historical figures. The scheme is managed by the Association of Independent Museums (AIM) and Biffa Award, a multi-million pound fund that helps to build communities and transform lives through awarding grants to communities and environmental projects across the UK.

Project Manager, Catalyst Trustee and chemist, Dr Diana Leitch said "It has been the goal of Catalyst for some time to depict the work and influence of local industrial chemist, Harry Baker, and to highlight the importance of this local achiever and his achievements in chlorine production to both the local and wider world and to inspire young people to follow his example. The AIM Biffa Award grant will now enable us to do that and we are extremely grateful to be receiving this money. I was born and brought up in Runcorn and came from a family who brought the salt to Runcorn and worked in the chemical industry in the area for generations as did many local people. Chlorine has benefitted humanity in numerous and diverse ways and still does and it will be a great pleasure to see this project unfold, the work of Harry Baker to be acknowledged and celebrated and the new Baker Gallery open in early 2019."

Emma Chaplin, AIM Director said "We are delighted that Biffa Award History Makers has been able to give Catalyst the opportunity to fascinate and inspire a new generation of visitors with the story of Dr Harry Baker and his huge impact on science and the world we live in. We can't wait to see the new gallery!"

More can be found out about Catalyst by visiting the website at www.catalyst.org.uk.

The 8th Capps Green Zomaya Award for Medicinal and Computational Chemistry

With thanks to RSC Awards Team

Background

Nigel Capps, Rich Green and Alex Zomaya died in a tragic air crash in 2001. They each had distinguished careers as medicinal chemists in the pharmaceutical industry and between them they contributed to the development of new drugs to treat cardiovascular and infectious diseases, cancer, asthma and arthritis. Their

network of colleagues and contacts spanned the scientific communities of both the industrial and academic sectors. Following their deaths, a group of colleagues and friends decided to set up a lasting memorial to their lives. The four companies with whom the chemists were most closely associated, GSK, Celltech, Vernalis and NicOx, generously contributed to an Award Fund under the auspices of the Royal Society of Chemistry (RSC). In 2006, a Charitable Trust was created to administer the Award. In 2013 the Award was incorporated into the Biological and Medicinal Chemical Sector (BMCS) of the RSC and is now run in tandem with the Malcolm Campbell Award.

The biennial Award acknowledges outstanding contributions made by young scientists (up to the age of 40) to the fields of medicinal or computational chemistry. Winners of the Award receive a Royal Society of Chemistry Certificate, a commemorative medal and a cheque for £2000. They are also invited to give a Memorial Lecture at the East of England Medicinal Chemistry Symposium at the University of Hertfordshire.

Biographies

Nigel Kenneth Capps (1960 – 2001)

Nigel gained his BSc in chemistry from Imperial College and his PhD on novel synthetic approaches to bicyclic lactam structures under the direction of Professor Doug Young at the University of Sussex. After completing a post-doctoral fellowship with Professor Tony Barrett at Northwestern University, Illinois he joined Glaxo UK in 1986 and worked on 5HT-related drugs for migraine. He was appointed a Group Leader at Celltech in 1991, working on novel immunomodulators targeting signal transduction in T-cells. In 1994, he joined the Business Development group at Celltech for three years before moving to Cerebrus, which merged with Vanguard to form Vernalis plc in 1999. In 2000 he became Head of Business Development at the French Biotech Company, NicOx.

Richard Howard Green (1947 – 2001)

Rich was awarded his BSc in chemistry from the University of Sheffield in 1969, and his PhD on acid catalysed rearrangements of cyclo-octyl and anisyl systems from the University of Leicester in 1973. He joined Glaxo as a medicinal chemist in 1973 where he worked on a number of projects including cephalosporin, prostaglandins and Beta-3-adrenoceptor agonists. Rich identified several novel synthetic routes in pursuit of these targets, which greatly aided the generation of structure activity (SAR) relationships. He developed a particular interest in natural product chemistry and was the author of several outstanding reviews in the area. In 1997, he became a full-time computational chemist at Glaxo with interests in intelligent compound selection. He achieved success in identifying several lead series for G-protein-coupled receptor targets. He also had interests in utilisation of protein crystal structure for compound selection, pharmacophore compound design and the creation of virtual and real libraries. He brought these skills to Celltech in 2000, where he was appointed as a Principal Scientist in computational chemistry. He had a significant impact in enabling Celltech's structure based drug design approaches in the short time before his death in 2001. Throughout his career, Rich was an excellent teacher and mentor to his younger colleagues.

Alexander Ishu Zomaya (1946 – 2001)

After his schooling in Baghdad, Alex came to England in 1962 to continue his education. He obtained his BSc in chemistry at the London South Bank University before joining Beechams in1970 where he stayed, working as a medicinal chemist, for 23 years. His work focussed mainly on antibiotic research and he made major contributions to therapies based on the chemistry of erythromycin and clavulanic acid; programmes which ultimately led to the discovery of augmentin. In 1993 he joined Celltech where he worked on novel anti-inflammatory therapeutics. He contributed to the discovery of CDP840, an inhibitor of phosphodiesterase type 4 and CDP323, an antagonist of $\alpha 4$ integrin function, both of which progressed to clinical trials. Alex was named as an author and inventor on 17 scientific articles and patents.

The 2018 8th Capps Green Zomaya Award was jointly awarded to:

Dr James Crawford received the Award for his work on drug discovery at both AstraZeneca UK and at Genentech USA, and in particular for his contribution to the development of Cathepsin K inhibitors for the treatment of osteoarthritis, and of tyrosine kinase inhibitors currently in clinical trials for the treatment of

rheumatoid arthritis, lupus and chronic spontaneous urticaria. Dr Crawford was presented with the Award on 26th April and delivered the 9th Memorial Lecture entitled 'Discovery of GDC-0853: a highly potent, selective, and non-covalent BTK inhibitor'.

Dr Richard Ward received the Award for his work at AstraZeneca UK on small molecule drug discovery, and in particular for his contribution to the discovery and development of Tagrisso (Osimertinib), a new treatment for EGFR T790M-positive non-small cell lung cancer. Dr Ward was also presented with the Award on 26th April and will present the 10th Memorial Lecture at the 30th Symposium on Medicinal Chemistry in April 2019.

For more information on the award and previous winners, see
http://www.rsc.org/ScienceAndTechnology/Awards/CappsGreenZoyama/

Tony Kent Strix Award and Annual Lecture 2018

Contributed by RSC CICAG Member Dr Doug Veal, email dougveal@waitrose.com

1. UKeiG - 2018 Call for Tony Kent Strix Award Nominations

The UK electronic information Group (UKeiG) is now seeking 2018 nominations for this prestigious award.

The Tony Kent Strix Award was inaugurated in 1998 by the Institute of Information Scientists. It is now presented by UKeiG in partnership with the International Society for Knowledge Organisation UK (ISKO UK), the Royal Society of Chemistry Chemical Information and Computer Applications Group (RSC CICAG) and the British Computer Society Information Retrieval Specialist Group (BCS IRSG).

The Award is given in recognition of an outstanding practical innovation or achievement in the field of information retrieval in its widest sense, including search and data mining, for example. This could take the form of an application or service, or an overall appreciation of past achievements from which significant advances have emanated. The award is open to individuals or groups from anywhere in the world.

The deadline for nominations is Friday 28th September.

Nominations should be for achievement that meets one or more of the following criteria:

- A major and/or sustained contribution to the theoretical or experimental understanding of the information retrieval process
- Development of, or significant improvement in, mechanisms, a product or service for the retrieval of information, either generally or in a specialised field
- Development of, or significant improvement in, easy access to an information service
- Development and/or exploitation of new technologies to enhance information retrieval
- A sustained contribution over a period of years to the field of information retrieval for example, by running an information service or by contributing at national or international level to organisations active in the field.

Key characteristics that the judges will look for in nominations are innovation, initiative, originality and practicality. The information to be supplied in the nomination should comprise:

- The name, institutional address and qualifications of the nominee
- A brief biography (not more than one page of A4)
- A relevant bibliography (i.e. not comprehensive but including the key publications relevant to the nomination)
- A justification for the nomination, of not more than one page of A4, showing clearly which of the Strix award criteria the nominee meets and how the criteria are met

- Additional material (e.g. letters of support - letters from past winners would be especially valuable).

It is possible that the Award Committee will request additional information from the nominators for those nominees considered suitable candidates for the award. Nominations for the 2018 award should reach the judges by Friday 28th September 2018 and be emailed to:

John Wickenden – Hon. Secretary UKeiG secretary.ukeig@cilip.org.uk
cc-ed to Gary Horrocks - UKeiG administrator info.ukeig@cilip.org.uk
cc-ed to Sue Silcocks – Hon. Treasurer UKeiG treasurer.ukeig@cilip.org.uk
For more information about UKeiG and the Tony Kent Strix Award visit: https://tinyurl.com/ybjfjotu

The Strix panel hope to announce the winner at the 2018 Strix Annual Lecture in London on the afternoon of Friday, 23rd November. Booking are open for the event at: https://tinyurl.com/y8pag2rr

2. The 2018 Annual Lecture

The 4th Tony Kent Strix Annual Memorial Lecture 2018 is to be delivered by Maarten de Rijke, Professor of Computer Science at the University of Amsterdam, and will take place on the afternoon of Friday, 23rd November 2018 at The Geological Society, Burlington House, Piccadilly, London.

Book here: https://tinyurl.com/y8pag2rr

Last year UKeiG, in partnership with the International Society for Knowledge Organisation UK (ISKO UK), the Royal Society of Chemistry Chemical Information and Computer Applications Group (RSC CICAG) and the British Computer Society Information Retrieval Specialist Group (BCS IRSG) was delighted to announce that the winner of the prestigious Tony Kent Strix Award for 2017 was Professor de Rijke. The Award was presented to him by Doug Veal (Strix Chair) and David Ball (UKeiG Chair) in London on Friday October 20th 2017 in recognition of his major and sustained contributions to the field of information retrieval and web searching.

Professor de Rijke is a well-known and highly respected member of the international information retrieval community having made considerable and widely recognised contributions to the field. He has an impressive and formidable high impact publications record in a range of areas including semantic search, semi-structured retrieval and social media. He has produced influential research outputs on the large-scale semantic analysis of online content and on the analysis of subjective aspects of information (sentiment, credibility, memory, reputation and experiences). His contributions to information retrieval, in particular to the fast evolving areas of computational methods for analysing, understanding and enabling effective human interaction with information sources, have been profound.

He leads the Information and Language Processing Systems Group at the Informatics Institute of the University of Amsterdam. It is one of the world's leading academic research groups in information retrieval and intelligent information access, with projects on self-learning search engines, semantic search and the interface between information retrieval and artificial intelligence.

His 2018 Strix Lecture is entitled: '*Retrieval as Interaction*'

Abstract: Information retrieval systems, such as search engines, recommender systems and conversational agents, may well be the prime example of interactive systems to which people are exposed. Their development is best thought of as a two-stage process: off-line development followed by continued online adaptation based on interactions with users. Off-line development, which involves evaluation and learning from annotated datasets or from logged interactions, is risk free in the sense that the system does not require the ability to make interactive interventions. In contrast, in online development retrieval systems use interactions and interventions for evaluation and for learning.

In the lecture Professor de Rijke will compare the off-line and online development phases. How much can a retrieval system learn off-line from historical interaction data? How much can it gain by being able to make interventions and explore new actions? Can we give bounds on the risks a retrieval system takes when performing online interventions? He will also discuss another set of questions that come up in this space and

that have to do with how we should design systems that learn and adapt online and are respectful of their users.

Please note, although this is a free event, open to everyone, advance bookings are required.

Please book a ticket online at: https://tinyurl.com/y8pag2rr

Full programme details:

1.30 Registration
2.00 Douglas Veal - Chairman's welcome
2.10 - Stella Dextre Clarke, 2006 Winner of the Tony Kent Strix Award, reflects on 'Then and Now: Contrasts in the retrieval environment.' Now retired from consultancy in information management, she is still active as Vice-Chair of the UK Chapter of ISKO (International society for knowledge Organization).
2.45 - Questions & Discussion
3.00 - Tea & coffee
3.45 - The Tony Kent Strix Annual Memorial Lecture
4.30 Questions & discussion
5.00 Meeting closes

The Tony Kent Strix Award was inaugurated in 1998 by the Institute of Information Scientists. The 2018 Tony Kent Strix Award winner will also be announced during the afternoon.

For more information about UKeiG and the Tony Kent Strix Award visit: https://tinyurl.com/ybjfjotu

UKeiG is a Special Interest Group of CILIP, Registered Charity Number 313014.

Undergraduate Research Bursaries – CICAG Sponsorship in 2018

Contributed by RSC CICAG Committee Member Neil Berry, email: ngberry@liverpool.ac.uk

A number of Undergraduate Research Bursaries, funded by the Royal Society of Chemistry are usually available to students in UK and the Republic of Ireland Chemistry and related departments from June-September each year. The purpose of the awards is to give experience of research to undergraduates with research potential in the middle years (i.e. 2/3, 2/4 or 3/4) of their degree and to encourage them to consider a career in scientific research. Applications relevant to the interests of the CICAG will be welcomed in the areas of cheminformatics, chemical information, chemical data management, chemistry data analytics, chemistry IT solutions and applications. The deadline for applications is usually late February each year.

As mentioned in the summer 2017 Newsletter, CICAG sponsored an Undergraduate Research Bursary which was awarded to Grzegorz Winter, who is studying chemistry at the University of Newcastle. Grzegorz's research project was also at the University of Newcastle, where he was working in Dr Daniel Cole's group on re-writing the functional form of classical force fields for molecular simulations. Grzegorz has summarised his experience:

"*The most valuable part about my placement was working independently. During my placement, I had to be completely self-reliant in how I was working. I was given ideas to help me proceed but ultimately, it was my decision to choose the method in which they would be applied. Part of my role was to test new software, OpenMM, to see its capabilities. I thoroughly enjoyed this as I was given an end-goal, testing the new functional form of classical force fields, and had a deadline to abide by. My programming skills were very useful as it meant I could begin working straight away to test OpenMM and think about how I would achieve implementing the new equation. I felt that I could trust myself when obtaining results. However, there were a few moments that were difficult. I would then try to find the cause of the issue and rectify it. My supervisor was very helpful when I couldn't fix the issue myself. I believe that this experience has made me more confident in myself.*"

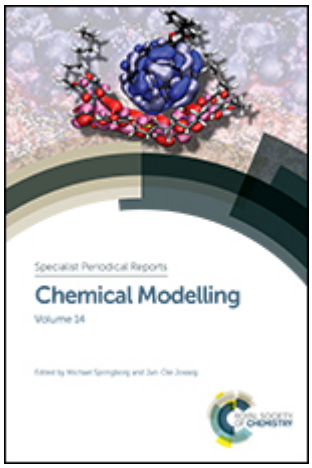
Further information, including guidelines and the application form can be found at:
http://www.rsc.org/ScienceAndTechnology/Funding/undergraduate-bursary.asp

---

Chemical Information / Cheminformatics and Related Books

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

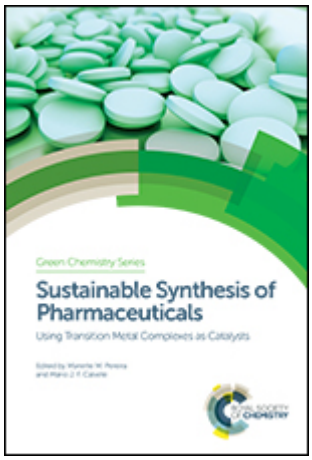
Chemical Modelling



Chemical modelling covers a wide range of hot topics and active areas in computational chemistry and related fields. With the increase in volume, velocity and variety of information, researchers can find it difficult to keep up to date with the literature in these areas. Containing both comprehensive and critical reviews, this book is the first stop for any materials scientist, biochemist, chemist or molecular physicist wishing to acquaint themselves with major developments in the applications and theory of chemical modelling.

RSC Publishing
Editors: Michael Springborg, Jan-Ole Joswig

Sustainable Synthesis of Pharmaceuticals



There is a growing interest in the development of sustainable processes for the synthesis of pharmaceuticals and this book bridges the divide between industrial examples and the fundamental chemistry. It explains the basic principles of using transition metal catalysis with several green approaches for the synthesis of pharmaceuticals. The topic is an important one for green chemistry and the chapters in this book on hydroformylation, green oxidation and olefin metathesis will also be of interest to both medicinal and organic chemists.

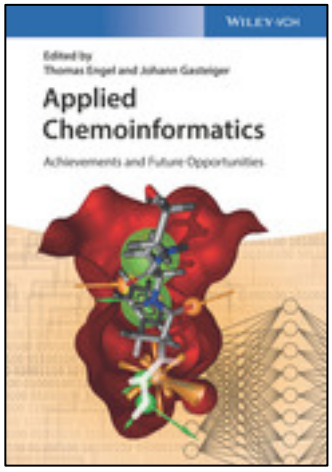
RSC Publishing
Editors: Mariette M Pereira, Mário J F Calvete

Applied Chemoinformatics

Achievements and Future Opportunities

Edited by world-famous pioneers in chemoinformatics, this is a clearly structured and applications-oriented approach to the topic, providing up-to-date and focused information on the wide range of applications in this exciting field.
The authors explain methods and software tools, such that the reader will not only learn the basics but also how to use the different software packages available. Experts describe applications in such different fields

as structure-spectra correlations, virtual screening, prediction of active sites, library design, the prediction of the properties of chemicals, the development of new cosmetics products, quality control in food, the design of new materials with improved properties, toxicity modeling, assessment of the risk of chemicals, and the control of chemical processes.
The book is aimed at advanced students as well as lectures but also at scientists that want to learn how chemoinformatics could assist them in solving their daily scientific tasks.
Together with the corresponding textbook Chemoinformatics - Basic Concepts and Methods on the fundamentals of chemoinformatics readers will have a comprehensive overview of the field.



Wiley Publishing
Editors: Thomas Engel, Johann Gasteiger

Meeting Report: Dial-a-Molecule Conference - Predicting Reaction Outcomes

Contributed by RSC CICAG Committee Member Prof Jonathan Goodman, email: jmg11@cam.ac.uk

The Dial-a-Molecule conference on Predicting Reaction Outcomes was held at the Bradfield Centre on the Cambridge Science Park on Monday, 26 March, 2018. It was organised by Jonathan Goodman, Richard Whitby, Gill Smith and Jayshree Mistry. Sponsorship from RSC-CICAG and MestreLab is gratefully acknowledged.

What is the state of the art of predicting reaction outcomes? Seven diverse speakers discussed different aspects of the problem.

Peter Johnson (University of Leeds)
- *Reaction Databases in Automated Synthesis Planning and Reaction Product Prediction: Scope and Limitations*

Kris Ermanis (University of Cambridge)
- *Chiral Phosphoric Acid Catalysis: Reaction Prediction and Tool Development*

Arthur Smith (Office of Scholarly Communication, University Library, Cambridge)
- *The future of research in an open world*

Jonathan Goodman (University of Cambridge)
- *Present Chemistry; Future Reactions*

Robert Phipps (University of Cambridge)
- *Harnessing Non-Covalent Interactions for Control of Selectivity in Catalysis*

Elizabeth Krenske (University of Queensland, Australia)
- *The KOtBu-Catalyzed C–H Silylation of Heteroarenes: Theoretical Insights into a Complex Reaction Mechanism*

Ola Engkvist (AZ)
- *Predicting chemical reactions with machine learning and large data sets*

Jeremy G Frey (University of Southampton)
- *Chemical Discovery & AI: The EPSRC Network Artificial & Augmented Intelligence for Automated Investigation and Discovery (AI3D)*

Peter Johnson (University of Leeds) spoke about the development of synthesis planning tools and how data-driven reaction prediction is a key part of the process. Kris Ermanis (University of Cambridge) then explained how DFT methods can be used to explain the outcomes of enantioselective catalytic oxy-Michael reactions, and how machine learning can make conformation analysis more tractable. Arthur Smith, from

the Office of Scholarly Communication at the University of Cambridge, showed how open data and open research is changing the way chemical discoveries are being used, and how they have the potential to revolutionise research in the near future. Jonathan Goodman (University of Cambridge) discussed how specific insights into structural behaviour can be verified by calculation and used to explain and to predict reaction outcomes. In addition, knowledge from outside the immediate realm of study may be needed in order to best interpret chemical data and so predict future measurements, particularly if it is possible to unite diverse databases using a canonical index such as the RInChI.

Robert Phipps (University of Cambridge) gave an account of how an experimentalist discovers new reactions, with examples from his work on non-covalent interaction controlled catalysis and enantioselective Minisci-type additions. Elizabeth Krenske (University of Queensland) then described how a superficially simple reaction, a base-catalysed silylation, required a worldwide consortium to investigate the mechanism, and then the mechanisms, of this highly complex process, and illustrated the care which is needed to draw conclusions about detailed reaction mechanisms. Ola Engkvist, from AstraZeneca, reported on the possibilities of automating reaction development and the opportunities that this gives for rapid and consistent data acquisition; these data, combined with data from publishers and open sources of information, make it possible to predict reaction outcomes using machine learning. Jeremy Frey (University of Southampton) then highlighted how a new EPSRC network, AI3D, should be able to advance scientific discovery in chemistry by helping people and computers to work effectively together.

The panel discussion covered the reliability of DFT and other simulation methods. Ideally, these should be validated against experimental results. Availability of data is a major issue. We can never have enough, and there are issues both with gathering data and making it available. Effective indexing can make information retrieval more tractable. With plentiful, high-quality data, and appropriate contextual information, machine learning could soon change the way we think about reactions.

Meeting Report: ICCS 2018

Contributed by RSC CICAG Committee Member Nathan Brown, email: nathan.brown@benevolent.ai

11th International Conference on Chemical Structures
27th-31st May 2018
Noordwijkerhout, the Netherlands

The 11th ICCS conference was held in Noordwijkerhout in the last week of May. The weather was notable this time due to this being by far the hottest of all of the six ICCS meetings that I have attended. The weather also had adverse effects for many of the delegates due to lightning strikes at Stansted Airport causing major delays and cancellations. The conference was covered extensively on Twitter with the hashtag #11thICCS, more coverage here: #11thICCS.

My first ICCS was in 2002 as a PhD student, and the periodicity of the ICCS has always offered a high quality bellwether of current trends in the field, and this meeting was no exception. Enumeration of chemistry space was a definite theme; and more importantly how you handle search and retrieval in such vast spaces. Another theme I picked up on was a prevailing honesty and openness on expectations of our modelling approaches and reflecting on what this may mean for our data sources now and how we may influence these in future. This may be a reflection of the availability of data and algorithms, such as ChEMBL and the RDKit, respectively, empowering a great amount of this work to be tackled rapidly and learnings gained much more quickly than before.

This year, the CSA Trust Mike Lynch Award was presented to Dr Rudy Potenzone for his outstanding work over 45 years: "*From Teletype Structure Input to Biology and Chemistry Intelligent Knowledge Graphs: My 45 Years in Cheminformatics*". Rudy's career is vast and varied, his work encompassing key players in cheminformatics software including: Accelrys, MDL, CAS, Lion Bioscience, CambridgeSoft, Microsoft and PerkinElmer. One of Rudy's most important contributions to the field was the initiation of the project that became SciFinder.

A highlight for me was the talk from Dr Greg Landrum (KNIME, but also the lead developer of RDKit), talking on QSAR modelling en-masse: "*How do you build and validate 1500 models, and what can you learn from them: An automated and reproducible system for building predictive models for bioassay data*." [slides]. Greg used the KNIME Model Factory to test out many different combinations of modelling algorithms and molecular descriptors. In total, Greg generated 310,000 models in his workflow and reported key learnings from his work. Many of his models appeared very good in terms of model statistics, but on further introspection he noted that it is likely that these were either through overfitting or simply learning to discriminate the papers from which the data were derived. Greg's mantras throughout the talk were:

- Automate all the things!
- Question all the things!
- Analyse all the things!

A highlight from the conference for me was the touching celebration to Prof. Peter Willett of the University of Sheffield on reaching forty years of research in the field of Chemoinformatics from longstanding collaborators and friends Prof. Val Gillet and Dr Bob Clark. I may be biased as one of Peter's many PhD students (current count is 75!), but I believe it is fair to say that if you do not know who Peter is, or have never read one of his papers, then you probably have to wonder what it is you actually do. Thankfully, Bob was able to stitch together his video tribute with Val's live speech in Peter's honour, and made the video available here.

The conference ended as it started, with bad weather affecting flights back to the UK. Most were okay, but a special mention to Dr Roger Sayle (NextMove Software) who, along with his colleagues Dr Noel O'Boyle and Dr John Mayfield, were diverted from Stansted Airport to Edinburgh. Roger nobly stood up and drove his colleagues through the night in a hire car from Edinburgh back down to Cambridge. Good job, Roger!

As always, the ICCS was a terrific conference bringing together some two hundred scientists from academia and industry. There are already plans afoot for the 12th ICCS to be held from 6th to 10th June 2021. Save the date!

Meeting Report: Artificial Intelligence in Chemistry Meeting

A joint meeting, organised by RSC-BMCS & RSC-CICAG
15th June 2018
Royal Society of Chemistry, Burlington House, London
Twitter Hashtag: #RSC_AIChem

Summaries contributed by RSC-BMCS and RSC-CICAG student bursary award winners

Introduction from the Organising Committee

Our first conference on aspects of Artificial Intelligence in Chemistry was held on Friday, 15 June, at Burlington House. The conference was a year in the planning and surprised us by being sold out fully two months before the event with 115 delegates. The day itself was great for the quality of scientific talks and poster presentations, and also the chance to network with those interested in this highly exciting and fast-

moving field. However, as highlighted in the first talk from Marwin Segler (BenevolentAI), artificial intelligence has been a part of chemistry going back more than half a century.

The success of our first Artificial Intelligence in Chemistry conference has led us to begin work on a second conference to take place in 2019, which we hope will be even bigger and better than our event this year. Watch this space!

Organising Committee: Nathan Brown, Phil Jones, Chris Swain

Presentation Summaries from Bursary Awardees

Speaker: **Marwin Segler, BenevolentAI**
Presentation: **Self-Driving Synthesis Planning**
Report by: **Po-Chang Shih, University College London**
The great talk given by Dr Marwin Segler clearly gives a broad view of current progress of AI technology in the field of retrosynthesis. In the beginning of the talk, a story about the challenge of developing AI in retrosynthesis was introduced – encoding ~20,000 synthetic chemistry rules took 10 years. This is absolutely startling, as there are apparently more than such number of synthetic rules. To begin his journey of AI application in retrosynthesis to a wonderful land, he and his team picked up few factors that could make it work, such as focusing on more promising reactions and filtering infeasible reactions, although getting the rules into the machine and having an efficient search was initially challenging. Their current findings are that applying deep neural networks and Monte Carlo Tree Search to the machine is able to create synthetic route(s) to a molecule, of which route(s) are not inferior to those found in the literature. The success of machine-generated retrosynthesis sheds a light on AI's potential in drug discovery, although the wrap-ups of the talk lists current challenges that AI cannot do so far – no conditions of reactions; no yield prediction; that it does not work well for natural products; and cannot invent new reactions.

Speaker: **Nadine Schneider, Novartis**
Presentation: **Chemical topic modelling - an Unsupervised Approach to Organize and Explore Chemical Information**
Written by: **Robert Shaw, University of Sheffield**
Topic modelling is a useful AI technique for determining the co-occurrence of words, and from this building a probability distribution of words in topics. Schneider's talk demonstrated how this approach could be used instead for chemicals, classifying molecules by the common (or rare) fragments they contain. In an age where the amount of data, especially in chemical space, is unmanageably large, this approach provides a robust, unsupervised way of identifying chemically "interesting" compounds. Schneider demonstrated the efficacy of chemical topic modelling by showing how a 100-topic model could be built in around an hour for a database of 1.6 million molecules. This led to sensible and interpretable distinctions, highlighted by intuitive visualization tools.

Speaker: **Bob Sheridan, formerly Merck**
Presentation: **What I Learned About Machine Learning - Revisited**
Written by: **Shuzhe Wang, ETH Zurich**
Bob gave an insightful review of his experience in QSAR for the past 30 years. In that time, he made important contributions to the field, such as the promotion of random forest models and replacement of random train-test set, split by split, based on time. He showcased the performance comparison of random forest against a gradient boosting tree and deep neural nets, based on their computational cost, accuracy and model size, and concluded that gradient boosting tree is the most cost-efficient approach to date. Following domain of applicability analysis aiming to assign error bars to predictions, he moved on to some remaining challenges on the interpretability of model outputs. These include assigning input descriptor's importance and more so, atoms importance – by mapping of activity readout back to a subset of atoms in the input molecule. I really enjoyed Bob's talk, especially his emphasis on the importance of data quality over the complexity of model architecture.

Speaker: **Bob Sheridan, formerly Merck**
Presentation: **What I learned About Machine Learning - Revisited**
Written by: **Miha Skalic, University Pompeu Fabra**

Dr. Sheridan provided insight into data science and predictive model development at Merck Sharp & Dohme. The talk was focused on the domain of QSAR and the presenter pointed out that the methods are somewhere between worthless and sort of working, and that more important than anything in the pipeline, is the data itself. If one has insufficient data, no model optimization will help you. The three key points of the talk were:
1. Model development
2. Estimation of model reliability (assigning error bars)
3. Interpretability of the models
In terms of feature engineering, MSD use features extracted from substructure (e.g. ECFP4/MACCS fingerprints), molecular properties or properties such as atom pairs descriptions. As predictive models they usually use one - either random forests, XGB-boost, or neural networks, which if they are trained as multitask networks, perform better. For the task determining confidence they rely on DA metrics - metrics that evaluate differences between observations and predictions. Finally, Bob described two methods used for interpretability of models: (i) Universal descriptor importance generator (UDIG) and (ii) colouring by atoms. UDIG uses permutations of descriptors, while colouring methods hide atoms in computation of substructure descriptions.

Speaker: **Ola Engkvist, AstraZeneca**
Presentation: **Molecular de Novo Design through Deep Learning**
Written by: **Fergus Imrie, University of Oxford**
Ola's talk provided a fantastic summary of the current capabilities of molecular de novo design, whilst also hinting at what might come next. Ola began his talk by motivating de novo design and discussing its potential impact on industry. It was made clear that attempts to enumerate chemical space would not succeed and thus developing methods to sample or search this space is of crucial importance. The speed with which the field was advancing towards these goals was made apparent through a brief outline of some of the recent literature, a recurring theme throughout the day. After a primer on neural networks, Ola described how methods from natural language processing could intuitively be applied to molecular structure generation by using SMILES strings. A compelling demonstration of this technique showed exactly how a trained model produced a new molecule. Ideally the generated molecules should have desirable properties, in particular ones that the user can specify. Through the use of reinforcement learning, Ola described how a generative model can be incentivised to produce molecules with specific properties. Ola then addressed some of the criticisms of generative models, two majors ones being that molecules are not sufficiently diverse and are not synthetically feasible. Ola showed that the generated molecules followed the properties of the dataset used for training in both cases, but conceded that evaluating generative algorithms was challenging and needed thought. Finally, Ola presented his thoughts for the future of chemistry, and how this would be influenced by automation and AI. A very engaging talk, and exciting times lie ahead!

Speaker: **Willem van Hoorn, Exscientia**
Presentation: **Scaling de Novo Design, from Single Target to Disease Portfolio**
Written by: **Lee Steinberg, University of Southampton**
Willem Van Hoorn spoke about the technology that Exscientia applies in the automatic design of patentable compounds. In the words of Willem, "You would not go looking for mountains in The Netherlands". Equivalently, when looking for potential candidates, we should first search the regions of chemical space that we would expect to be most fruitful (maybe we should try looking in Switzerland for our mountains). Furthermore, Willem introduced his idea of the "Centaur Chemist", where humans design strategy and assess progress, but computers do the running. This leads to an environment where the scientist and computer work in tandem, and this methodology was shown to reduce the number of molecules assessed in a normal target search from 500 to 250, with the time taken reduced by more than half. Lastly, Willem ended with a variant of a quotation that will likely remain with the conference attendees: "*Artificial intelligence will not replace chemists – but chemists who don't use AI will be replaced by those who do*".

Speaker: **Ella Gale, Bristol University**
Presentation: **Investigating Clusters in Solvent Data using K-means**
Written by: **José Jiménez, University Pompeu Fabra**
The need for investigating solvent data arises from several factors. Historically, solvent choice was based on history or previous best practice at the lab bench. However, changes may be needed in order either to save money, be more environmentally friendly, or be safer. In this study, the speaker focused in the application of Organic Electrolyte Solutions (OES), that is, the best co-solvent choice for the dissolution of cellulose in

paper and plastic processing. In order to tackle the problem, there is a clear need to map the solvent landscape, and previous attempts have tried to do so by Principal Component Analysis, currently the most used method. There is always a discrepancy between the problems unique to chemistry when applying machine-learning algorithms, namely few labelled and incomplete examples available. The speaker then focused on the chemical properties that are the best predictors of solvent properties, using solvent, chemical group, and a structure database, each with different parameters. The technique used to tackle the problem is a derivative of the well-known k-means++ method, and det-K, a way for determining the best value of clusters. In particular, they use a hierarchical approach where clusters are recursively found until eventually they are pure. She found that her data was best described by 2 clusters according to the det-K method, and identified items in the first cluster such as alkanes and nitriles, while in the second benzene and molecules with double rings were found. The physical properties that cause most of the separation were molecular weight, melting point, boiling point and molar volume. Another clustering with k=2 was made and found the most important properties to be dipole moment, dielectric constant and molar volume. Hierarchical k-cluster, on the other hand, was shown to separate solvent data into functional groups, allowing for novel re-arrangement. Further work encompasses repeating this analysis in order to compare results with the previous PCA methods, and with solvents that feature incomplete data.

Speaker: **Colin Batchelor, Royal Society of Chemistry**
Presentation: **Deep learning and Chemical Data**
Written by: Caroline Bushdid, Institut de Chimie de Nice
Colin Batchelor's talk presented the results of using deep artificial neural networks (DNN) to analyse spectra and unstructured text related to chemistry. In particular, he presented this technique as applied to three topics: NMR spectra analysis, Chemical named-entity recognition and protein/molecule relationship extraction.
In the case of NMR, the spectra were given as an input in the form of a vector, and the output was the presence of certain functional groups. He found that DNN outperformed classical Random Forest models and the model was capable of predicting accurately functional groups such as iodine, nitro groups and even aliphatic groups, but natural products identification remained challenging. Overall, deep learning did not perform as well as humans in this task, probably due to the need for more data.
In contrast, when DL was used to recognise chemical names and patterns, the numerical model was found to perform as well as humans. The nature of the input data (i.e. unstructured text) made the task more difficult than expected. The limit in the performance of the model was found to be mainly due to humans disagreeing in the annotations. He explored using a word-base system as compared to a character-based system and found that the best results were obtained by using an ensemble of these techniques.
Finally Colin used DL to extract relationships on how molecules affect proteins. The input consisted of text (for example from CHEMPROT) which resulted in the output of a vector which held information such as "no relation", "up regulation", "down regulation", "agonist", "antagonist", etc., etc. In this case the model performed less well than humans, underlying that, similarly to the case of NMR spectra, the task is much harder for DL.

Speaker: **Darren Green, GlaxoSmithKline**
Presentation: **Automation, Analytics and AI**
Written by: **Sam Munday, University of Southampton**
Dr Darren Green's talk on automation, analytics and AI highlighted the role that humans will continue to have alongside an increasingly powerful AI in the field of drug discovery. In comparison to other speakers, Darren said his belief was that that the chemical community has yet to engage with true AI systems and that we are currently still applying automation techniques, rather than intelligent algorithms.
Darren started with an overview of his experience within the field of molecular design at GSK UK. He described the changing role of the medicinal chemist, with intuition borne from experience giving way to a more data-driven systematic approach to identifying possible lead compounds. He highlighted how this will lead to increasing productivity, as the chemist uses models built from GSK's vast database to refine and direct their search, leading to less unsuccessful paths being followed.
Dr Green's view on the future of AI in chemistry drew from his experiences over the past 10 years, including the rise of "evolutionary algorithms" up to deep neural network techniques. His talk was funny and engaging and a great way to round off what was an already fantastic event.

Professor Peter Willett – a Celebration of his Research

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

Peter Willett (left) with Johnny Gasteiger at the Strasbourg Summer School, 2014
Photo courtesy of Nathan Brown

This year we celebrate over 40 years of Peter Willett's contributions to cheminformatics, truly a significant milestone in an illustrious career.

As most readers will know, Peter is Professor of Information Science in the Information School at the University of Sheffield. His achievements were recognised recently at the 11th International Conference on Chemical Structures, held in Noordwijkerhout, Netherlands, where there was a session dedicated to him, led by Professor Val Gillet; a recording of Val's speech is available on YouTube, which also contains a tribute from Bob Clark, formerly of Tripos.

To summarise Peter's career so far (and Val says that his c.v. stretches to 53 pages!), he obtained an MA in Natural Sciences (Chemistry) from Oxford University, after which he obtained an MSc in Information Science in the School of Librarianship and Information Science (as it was then called) at the University of Sheffield, followed by his PhD on the indexing of chemical reactions and post-doctoral work on the automatic classification of document databases. He was then appointed to the School's academic staff, where his research has focused on cheminformatics, information retrieval and bibliometrics. Always keen to ensure there are practical applications for his research, Peter's work is supported by many industrial organisations, especially pharmaceutical, biotech and software companies. Algorithms developed through his group's research have been deployed widely in IT solutions.

Peter's career so far can be illustrated with some data:

- 578 publications including 16 books
- Over 22,000 citations in Google Scholar
- Supervised 72 PhD students
- Editorial roles on 18 journals and reviewer for 126
- Recipient of the Tony Kent Strix award
- Youngest recipient of the Herman Skolnik award
- First recipient of the Mike Lynch award
- Only the second non-US recipient of the ACS's Computers in Chemistry award

But the data and awards don't tell the full story. Those of us who have had the pleasure of working with Peter know that in addition to being a true expert in his field, he is generous with his time and supportive of colleagues and collaborators. Personally, I had the pleasure of working with him during the late 1990s and early 2000s when both the University of Sheffield's Department of Info Studies and UMIST's (University of Manchester Institute of Science and Technology) Chemistry Department (where I worked) had both been awarded funding from EPSRC to support MSc courses in Cheminformatics. Far from being rivals, we were able to help each other out and share experiences. Peter delivered lectures for one of our units and was the external examiner for our course, and I was able to support his course by delivering tuition on the use of Beilstein CrossFire.

Peter has also contributed to CICAG's scientific meetings, including the popular "Celebrating the History of Chemical Information" meeting in 2010. His talk *Chemoinformatics: historical development of database methods* explored the emergence of the discipline in the late 1990s, a consequence of the data explosion resulting from combinatorial chemistry, and its foundations which go back much further.

The story of Peter's research, teaching and contributions to our field is not over. We look forward to the next chapter in his success.

InfoChem News

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

PubChem is, arguably, the largest corpus of publicly available chemical information. It collects chemical information from hundreds of data sources, and disseminates it to the public free of charge. Yet, although it contains 94.6 million unique chemical structures, very few of them (around 470,000) had direct links back into the primary literature. Enter Springer Nature, delivering on its commitment to Linked Open Data. In October 2017 PubChem added more than 26 million links to scientific articles, thanks to contributions from the publisher. Of these, 1.6 million links point to open access or free-to-read documents.

The new data are created using InfoChem's Chemical Named Entity Recognition engine ICANNOTATOR. It is a smart ranking algorithm, ensuring that only relevant citations and molecules are deposited (who needs articles where, for example, an uninteresting solvent was mentioned just once?). Now Springer Nature and InfoChem have gone one step further by implementing a fully automated workflow for the deposition of chemical compounds to PubChem, including links to scientific documents from Springer Nature.

While in the first phase of the project many parts of the processing had to be done manually, the great advance now is having a workflow in place that no longer requires human interaction. All the post-publication processing steps are fully automated, from chemical annotation and relevancy calculation, to finally uploading and ingesting the literature data into the PubChem Compound database, which makes the process smart, rapid and cost effective.

InfoChem has also added new Nature.com titles like Nature Chemical Biology, Nature Chemistry or Nature Catalysis to the workflow, so that links into these journals from compounds extracted from full-text are also visible on PubChem. All extracted compounds are linked to at least one Springer Nature document; from February 2018 until now more than nine hundred thousand links back into the primary literature have been added (Fig. 1). Over one thousand new compounds per month on average were delivered in the period from February to May 2018.

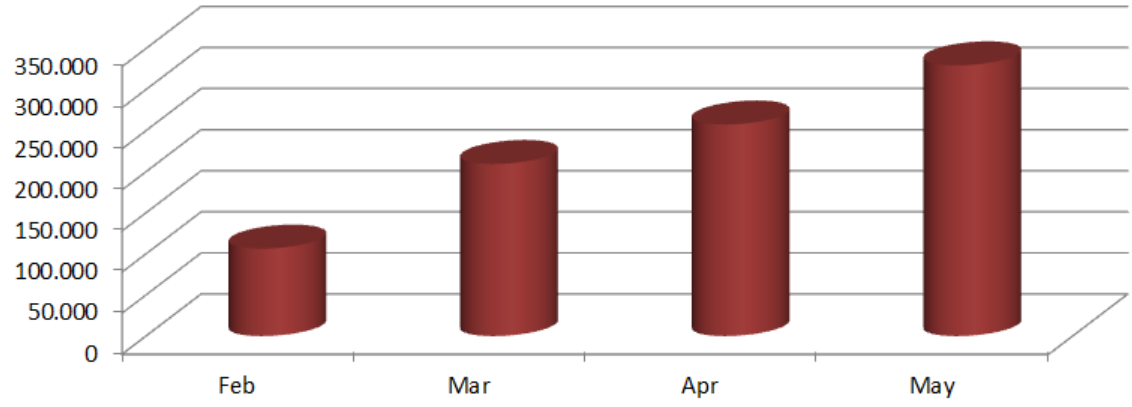


Fig. 1: New Citation Links from PubChem Compounds to Springer Nature Documents

Chemical compounds and their source documents in Springer Nature can be searched by substructure through PubChem, improving the accessibility and discoverability of information about chemical compounds. Each chemical record in PubChem with a "Springer Nature References" literature section includes a table with links to Springer Nature documents (Figs. 2-3).

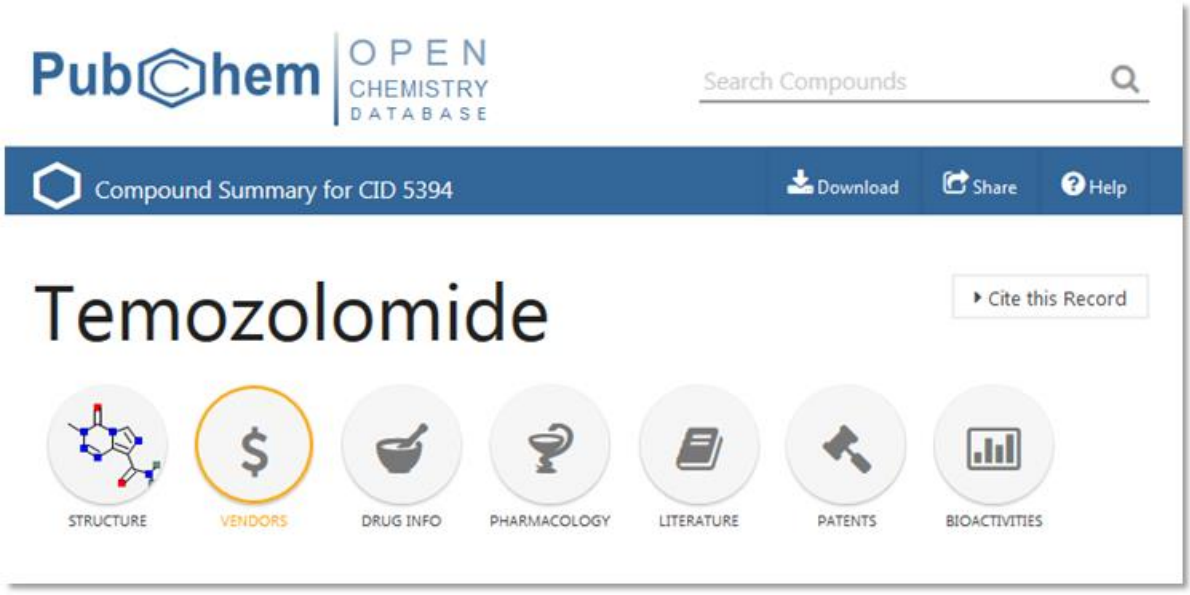


Fig. 2. How to Link from Compound to Literature

Download

1 to 5 of 7,080 1 2 3 ... 1,416 Relevance

Title	Journal or Book	Year	PMID
Temozolomide/PLGA microparticles: a new protocol for treatment of glioma in rats	Medical Oncology	2010	20396980
Role of temozolomide in the treatment of newly diagnosed diffuse brainstem glioma in children: experience at a single institution	Child's Nervous System	2010	20217098
Population pharmacokinetics of temozolomide and metabolites in infants and children with primary central nervous system tumors	Cancer Chemotherapy and Pharmacology	2003	13680158
ATM inhibitor KU-55933 increases the TMZ responsiveness of only inherently TMZ sensitive GBM cells	Journal of Neuro-Oncology	2012	23054561
JCOG0911 INTEGRA study: a randomized screening phase II trial of interferonβ plus temozolomide in comparison with temozolomide alone for newly diagnosed glioblastoma	Journal of Neuro-Oncology	2018	29557060

from Springer Nature

Fig. 3: Links from the Compound "Temozolomide" to the Primary Literature

Springer Nature's initial contribution of links had the effect of doubling, from roughly 470,000 to more than one million, the number of chemical structures in PubChem having links to scientific documents. Moreover some 40% of the compounds added were novel to PubChem. Now in 2018, yet more links have been added, increasing yet more the value of PubChem for researchers worldwide. And as the deposition is continuous, the number of links and structures will continue to increase. Enthusiastic feedback from users has made it clear that there is a demand for more literature information in PubChem: hopefully other scientific publishers will follow the example of Springer Nature.

CAS / SciFinder / STN News

Contributed by CAS Applications Specialist in the UK & Ireland, Dr Anne Jones, email annejones@acs-i.org.

SciFinder^n with ChemPlanner

SciFinder^n allows scientists to do better research in less time. SciFinder^n accelerates research by providing actionable results such as step-by-step synthetic methods and hard to find chemistry in patents.

- Access a comprehensive collection of content covering chemistry and related sciences in multiple languages from around the globe
- Pinpoint the most valuable information with chemical search relevance that helps you know where to start
- Accelerate your research with a streamlined, frictionless interface
- Save time with direct access to patent documents and step-wise synthetic procedures and methods

ChemPlanner® is a state-of-the-art workflow tool that uses computer-aided synthesis design (CASD) to help chemists efficiently identify the best synthetic routes from commercially available building blocks to their target molecules.

The integration of ChemPlanner with SciFinder^n is making excellent progress and we will announce further details and timescales when they are available.

STNext

STNext™ is a modern new gateway to the comprehensive content and powerful features of STN® that reflects the needs and priorities of today's patent experts.

- Freedom to securely use anytime anywhere
- IT-free automatic and ongoing improvements
- Familiar tools and capabilities from other STN products making transition a snap

If you have not yet tried *STNext* or want to know more, then please feel free to access the website or contact us directly https://next.stn.org/. The site is accessible with your regular STN Login and password.

CAS Blog

CAS is now producing a weekly blog with posts covering a large range of interests and technologies relating to today's scientific industries.

Our new blog started back in February this year and can be found at https://www.cas.org/blog. Subscribing is easy, just click the yellow subscribe button at the top of the page.

CAS Services

CAS is offering custom services based on our content collection, subject matter expertise and specialised technologies. We have grouped these into a few different categories which are Technology, Content, Knowledge and Professional Services.

If your organisation needs any assistance in these areas then CAS can help. Please refer to https://www.cas.org/services for further information.

CAS UK Training in 2018

In addition to the e-learning materials, CAS continues to offer instructor-led training for STN, SciFinder, NCI Global, ChemZent™, PatentPak™ and MethodsNow in the UK.

We conduct 'in-house' or WebEx training sessions on all aspects of searching CAS solutions.

I have also published a YouTube channel that is being populated with short search tips across all our CAS products. You are welcome to subscribe and I am also happy to take suggestions for training ideas or topics.

https://www.youtube.com/channel/UCAJK7P6ziR-WA717euVP9ig

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 9 July 2018*]

Impact Factor Creator and Chemical Information Pioneer Eugene Garfield Honoured
Eugene Garfield was a pioneer, who in the 1950s started creating information tools such as the then ground-breaking Science Citation Index. Garfield, who died in 2017, founded the Institute for Scientific Information (ISI), and went on to develop a number of tools which helped transform the search and retrieval of scientific and chemical information.
https://cen.acs.org/articles/96/i15/Impact-factor-creator-chemical-information.html
Source: Chemical & Engineering News

Computational Chemistry: A Rising Tide of Women
The authors of this brief article (Katharine Holloway and Georgia McGaughey of Vertex) investigated the topic of gender diversity in computational chemistry on the basis of similar recent publications in the related fields of medicinal chemistry and computational biology. They also examined historical demographics in the attendance at the Gordon Research Conferences on Computer-Aided Drug Design and Computational Chemistry, and membership in the Computers in Chemistry Division of the ACS. This analysis suggests that female representation in computational chemistry has risen steadily over the last 40 years, and likely stands at around 25%, which appears to slightly exceed that of the neighbouring fields of computer science and medicinal chemistry.
https://pubs.acs.org/doi/10.1021/acs.jcim.8b00170
Source: Journal of Chemical Information and Modeling

ChemMaps Lets Researchers Navigate the Chemical Universe
Researchers from North Carolina State University have created a new online service – ChemMaps – that allows users to interactively navigate the chemical space of over 8,000 drugs and 47,000 environmental compounds in 3D and real time. ChemMaps is designed to be a central resource for students and researchers who want to easily visualize and study complicated sets of chemical structures. The first release of the free-to-use website is available at www.chemmaps.com.
https://news.ncsu.edu/2018/06/fourches-chemmaps/
Source: NC State University

ChemMaps: Tool to navigate chemical space aims to be 'Google Maps of drugs'
https://www.chemistryworld.com/news/tool-to-navigate-chemical-space-aims-to-be-google-maps-of-drugs/3009145.article
Source: Chemistry World

The Institutionalized Racism of Scholarly Publishing
Academic Journals, and the tag of 'predatory publishing'.
https://awayofhappening.wordpress.com/2018/06/09/the-institutionalized-racism-of-scholarly-publishing/
Source: Wordpress

The battle behind the periodic table's latest additions
Four new elements were added in 2015 with great fanfare — but some researchers complain the announcement was premature.
https://www.nature.com/articles/d41586-018-05371-y
Source: Nature

Indian institute investigates nanoscientists for indiscriminate image manipulation
Four new retractions bring the total for Rashmi Madhuri and Prashant Sharma up to 14, with at least another 50 papers potentially affected.
https://www.chemistryworld.com/news/indian-institute-investigates-nanoscientists-for-indiscriminate-image-manipulation-/3009159.article
Source: Chemistry World

ProQuest Expands Remote and Mobile Access with Google Scholar
https://www.proquest.com/about/news/2018/ProQuest-Expands-Remote-and-Mobile-Access-with-Google-Scholar.html
Source: ProQuest

Springer Nature works with tech industry to find innovative solutions for researchers
The company has partnered with leading global technology forum The Next Web (TNW) Conference to explore how technology can help improve the sharing, discoverability and use of research data.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-innovation-launchpad-meetups/15789188
Source: Sringer Nature

Why thousands of AI researchers are boycotting the new Nature journal
"Many in our research community see the Nature brand as a poor proxy for academic quality. We resist the intrusion of for-profit publishing into our field. As a result, at the time of writing, more than 3,000 researchers, including many leading names in the field from both industry and academia, have signed a statement refusing to submit, review or edit for this new journal. We see no role for closed access or author-fee publication in the future of machine-learning research. We believe the adoption of this new journal as an outlet of record for the machine-learning community would be a retrograde step."
https://www.theguardian.com/science/blog/2018/may/29/why-thousands-of-ai-researchers-are-boycotting-the-new-nature-journal
Source: The Guardian

Clarivate Analytics Invites Applications for 2018 Eugene Garfield Award for Innovation in Citation Analysis
Launched in 2017 in memory of Dr Garfield and his pioneering work in scientometrics, this international award recognises researchers at any stage of their career developing an innovative approach in citation analysis that improves the impact of how scientific research is evaluated.
https://clarivate.com/blog/news/clarivate-analytics-invites-applications-2018-eugene-garfield-award-innovation-citation-analysis/
Source: Clarivate Analytics

Nanoscale Advances: a new open access journal for nanoscience
A new journal from the RSC and launched in partnership with leading nanoscience research centre the National Center for Nanoscience and Technology (NCNST), of Beijing, China. Nanoscale Advances will showcase interdisciplinary advances across the breadth of nanoscience and nanotechnology, and offers a home for research from all over the world.
http://www.rsc.org/news-events/articles/2018/jun/nanoscale-advances-launch/
Source: RSC Publishing

Clarivate Analytics Releases Enhanced 2018 Journal Citation Reports Highlighting the World's Most Influential Journals
https://clarivate.com/blog/news/clarivate-analytics-releases-JCR-2018
Source: Clarivate Analytics

Pittsburgh Life Sciences Greenhouse (PLSG) Introduces HealthcareData.Center for Life Sciences Innovators

HealthcareData.Center is designed to share knowledge on the U.S. healthcare system with students, researchers, academics, entrepreneurs, and investor communities.

https://www.prnewswire.com/news-releases/plsg-introduces-healthcaredatacenter-for-life-sciences-innovators-300672254.html

Source: PR Newswire

Elsevier are corrupting open science in Europe

Elsevier - one of the largest and most notorious scholarly publishers - are monitoring Open Science in the EU on behalf of the European Commission. Jon Tennant argues that they cannot be trusted.

https://www.theguardian.com/science/political-science/2018/jun/29/elsevier-are-corrupting-open-science-in-europe

Source: The Guardian

Report shows value of hybrid journals to research community and beyond

The report from Springer Nature shows that OA articles published in hybrid journals attract more downloads, citations, and attention compared with non-OA articles in hybrid journals.

https://group.springernature.com/gp/group/media/press-releases/report-shows-value-of-hybrid-journals-to-research-community/15894508

Source: Springer Nature

Unicode Consortium releases new science emoji

The latest Unicode update, Unicode 11.0, includes the following science-themed emoji: lab coat, goggles, test tube, petri dish, DNA, microbe, compass, abacus, and fire extinguisher.

https://cen.acs.org/articles/96/web/2018/06/Unicode-Consortium-releases-new-science.html

Source: Chemical & Engineering News

Reed Tech® Acquires PatentSight - a Provider of Analytics Solutions that Help Companies Monetise and Value IP Assets

PatentSight provides technology-driven organizations access to unique data analytics to gain valuable insights for strategic IP management.

https://www.prnewswire.com/news-releases/reed-tech-acquires-patentsight-a-provider-of-analytics-solutions-that-help-companies-monetize-and-value-ip-assets-300649386.html

Source: PR Newswire

India culls 4,305 dubious journals from approved list

The higher-education advisory agency is cleaning up its registry of approved journals, but academics say the list should be abolished.

https://www.natureindex.com/news-blog/india-culls-four-thousand-three-hundred-dubious-journals-from-approved-list

Source: Nature Index

Elsevier Teams up with PerkinElmer to enable faster, more intuitive Chemistry Research

The first phase of the collaboration will see the integration of Elsevier's flagship chemistry database, Reaxys, with PerkinElmer's ChemDraw® software.

https://www.prnewswire.com/news-releases/elsevier-teams-up-with-perkinelmer-to-enable-faster-more-intuitive-chemistry-research-682028961.html

Source: PR Newswire

EU: Recommendation on access to and preservation of Scientific Information

The European Commission has released a new set of recommendations to the Member States that offer guidance and propose the best way to implement and support open science practices. The new

recommendations include sections on incentives, rewards and require action plans from member states with concrete and measurable objectives. They are to replace the recommendations of 2012 which set out clear guidelines as to how publically funded work should be made openly and freely available.
https://www.openaire.eu/recommendation-to-the-ms
Source: OpenAIRE

Sweden stands up for open access - cancels agreement with Elsevier
Large science publisher Elsevier does not meet the requirements of Swedish universities and research institutes. Bibsam Consortium has after 20 years decided not to renew the agreement with the scientific publisher.
http://openaccess.blogg.kb.se/2018/05/16/sweden-stands-up-for-open-access-cancels-agreement-with-elsevier/
Source: OpenAccess.se

Springer Nature and ResearchGate announce new cooperation to make it easier to navigate the sharing of academic journal articles
In a cooperation agreement, Springer Nature and ResearchGate, along with Cambridge University Press and Thieme, will work together on the sharing of articles on the scholarly collaboration platform.
https://www.springernature.com/gp/group/media/press-releases/springer-nature-and-researchgate-announce-new-cooperation/15705990
Source: Springer Nature

Web of Science owner buys tool that offers one-click access to journal articles
Clarivate Analytics — has bought Kopernio, whose tool gives researchers one-click, legal access to journal articles even when off campus.
https://www.nature.com/articles/d41586-018-04414-8
Source: Nature.com

New integration of over 8 million patent records enables users to explore potential economic impacts of scholarly work
Data science company Altmetric announced the addition of worldwide patent data from IFI CLAIMS®, making it possible for users to more easily track the commercialization and potential economic impacts of their organizations' research.
https://www.altmetric.com/press/press-releases/patent-data-in-altmetric-highlights-the-commercialization-of-research/
Source: Altmetric

Research England invests £67 million in collaborative projects to drive university commercialisation
https://re.ukri.org/news-events-publications/news/research-england-invests-67-million-in-collaborative-projects-to-drive-university-commercialisation/
Source: Research England

New software could run the lab of the future
ChemOS helps chemists by letting robots perform experiments in their labs
https://cen.acs.org/articles/96/web/2018/03/New-software-run-lab-future.html
Source: Chemical & Engineering News

Chemistry's Image Problem
Ever wondered why the media's representation of chemistry and chemists never seems to hit the mark? The limited choice of stock imagery available might hold the answer.
https://www.chemistryworld.com/opinion/chemistrys-image-problem/3008806.article
Source: Opinion, Chemistry World

Wiley and Lumina Datamatics increase the discoverability of millions of images from Wiley Online Library

The new partnership will enable the discovery and legal use of Wiley's collection of images and figures on Lumina's Rights Platform. Lumina Datamatics will increase the accessibility of critical figures and images found in Wiley's publications, while empowering users to search, identify and clear permission to reuse the most appropriate images for their needs.

http://newsroom.wiley.com/press-release/all-corporate-news/wiley-and-lumina-datamatics-increase-discoverability-millions-image

Source: Wiley

Need to make a molecule? Ask this AI for instructions

An AI tool that has digested nearly every reaction ever performed could transform chemistry.

https://www.nature.com/articles/d41586-018-03977-w

Source: Nature Comment

Scientific journal officially retracts controversial CRISPR study

Last year, a study published in the journal Nature Methods caused controversy in the scientific community after claiming to find over 100 unintended large genetic deletions or insertions related to the CRISPR gene-editing process. Now the journal has officially retracted the paper after a thorough review found the main claims in the study were not sufficiently backed up by data.

https://newatlas.com/controversial-crispr-study-retracted-journal/54046/

Source: New Atlas

Partnership with Leading Societies to Support ChemRxiv™

The RSC, ACS and the German Chemical Society (GDCh) have entered into a partnership to support the financial and strategic development of ChemRxiv™, the premier preprint server for the global chemistry community.

http://www.rsc.org/news-events/articles/2018/mar/partnership-supporting-chemrxiv/

Source: RSC

https://www.acs.org/content/acs/en/pressroom/newsreleases/2018/march/acs-partners-with-leading-societies-to-support-chemrxiv.html

Source: ACS

https://www.gdch.de/service-information/oeffentlichkeitsarbeit/pressedienst-chemie.html

Source: GDCh

Data Management made Simple

Keeping your research data freely available is crucial for open science — and your funding could depend on it.

https://www.nature.com/articles/d41586-018-03071-1

Source: Nature Comment

The undercover academic keeping tabs on 'predatory' publishing

Following the shutdown of Beall's list, blacklists that warn against questionable publishers are in demand.

https://www.nature.com/articles/d41586-018-02921-2

Source: Nature

AIP Publishing Introduces Chinese Language Author Services

AIP Publishing has launched a suite of Mandarin-language author services to facilitate submissions from one of the world's fastest growing physical science research communities. By providing Chinese authors with resources to make the publishing process easier and faster, AIP Author Services is leading the effort to develop stronger connections between researchers across the globe. The Mandarin-language version of the AIP Author Services site is now available at http://aipauthorservices.cn.

https://publishing.aip.org/publishing/news/aip-publishing-introduces-chinese-language-author-services

Source: AIP

Researchers' challenges in sharing data cross geographic borders and disciplines
A report from Springer Nature reveals strong support for data sharing globally, yet identifies a number of common problems amongst researchers when they endeavour to share their data. *Practical Challenges for Researchers in Data Sharing*, based on one of the largest surveys on the subject of research data, reinforces previous findings on the challenges faced by researchers in sharing their data.
https://www.springernature.com/gp/group/media/press-releases/researchers-challenges-in-sharing-data-cross-geographic-borders-/15545272
Source: Springer Nature

Digital Science and Katalysis Lead Initiative to Explore Blockchain Technologies for Peer Review
The initiative is an important step towards a fairer and more transparent ecosystem for peer review and explores the utility of decentralized data stores in supporting trusted assertions that connect researchers to their activities.
https://www.digital-science.com/press-releases/digital-science-and-katalysis-lead-initiative-to-explore-blockchain-technologies-for-peer-review/
Source: Digital Science

Chemists test computer-planned syntheses for the first time
Planning efficient synthetic routes can seem like a dark art or feel like a Herculean labour of literature review. Chemists, for the first time, have tested a computer program's ability to plan complete syntheses without human help, following the proposed routes in the lab. Can Chematica plan better synthetic routes than people can? Some chemists are sceptical.
https://cen.acs.org/articles/96/i10/Chemists-test-computer-planned-syntheses.html?utm_source=NonMember&utm_medium=Newsletter&utm_campaign=CEN
Source: Chemical & Engineering News

Computer Literate?
Chemists are probably more adept with machines than they realise. Many of them program every day – plugging molecular inputs into experimental instructions – and there's a far more complex set of commands involved in programming matter than you'll find in any coding language.
https://www.chemistryworld.com/opinion/computer-literate/3008698.article
Source: Chemistry World

The Race to Knowledge: How Researchers Benefit from New Scientific Literature Access, Acquisition and Analysis Technologies
For most biotech and medical research organizations, accessing exactly the right information in published journal articles is crucial.
https://www.rdmag.com/article/2018/03/race-knowledge-how-researchers-benefit-new-scientific-literature-access-acquisition-and-analysis
Source: RD Management

Twitter launches Bookmarks, a private way to save tweets
https://techcrunch.com/2018/02/28/twitter-launches-bookmarks-a-private-way-to-save-tweets/
Source: TechCrunch

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry
https://jcheminf.springeropen.com/articles/10.1186/s13321-017-0247-6
Source: SpringerOpen

Open access is the future – Springer Nature survey

The overwhelming majority of professional staff in research institutions and libraries across the globe view open access as the future of academic and scientific publishing. However, many are not satisfied with the speed of transition.
https://www.researchinformation.info/news/open-access-future-%E2%80%93-springer-nature-survey
Source: Research Information

Computers outperform lab rats in detecting toxic chemicals
https://www.eurekalert.org/pub_releases/2018-02/s-ms-col021618.php
Source: EurekAlert!

Infographic Shows How Fair Use Promotes Creation of New Knowledge
http://www.arl.org/news/arl-news/4474-infographic-shows-how-fair-use-promotes-creation-of-new-knowledge
Source: Association of Research Libraries®

Researchers have finally created a tool to spot duplicated images across thousands of papers
https://www.nature.com/articles/d41586-018-02421-3
Source: Nature

Reshaping the IUPAC Gold Book for a Digital World
On January 24, ACS hosted a webinar featuring Dr Stuart Chalk discussing the ongoing work to update the Gold Book's website and repurpose it for modern research. Chalk, an Associate Professor of Chemistry at the University of North Florida, is leading the redevelopment of the IUPAC Gold Book website and is a titular member of IUPAC Committee on Publications and Chemical Data Standards.
http://axial.acs.org/2018/02/05/iupac-gold-book-digital/
Source: ACS Axial

Chemists harness artificial intelligence to predict the future (of chemical reactions)
Using machine learning, to predict multi-dimensional reaction yields.
https://www.sciencedaily.com/releases/2018/02/180215141740.htm
Source: ScienceDaily

New ACS journal dedicated to pharmacology, biochemistry and drug discovery research
ACS Pharmacology & Translational Science presents a new opportunity to enrich and expand the conversation featuring fundamental and applied research in pharmacology and translational science. The journal will provide a venue for leading pharmacology research, a field that encompasses the study of the science of drug action on biological systems.
https://pubs.acs.org/journal/aptsfn
Source: ACS

A picture speaks a thousand words in new scientific journal
The Journal of Sketching Science is a new scholarly publishing platform that summarises complex scientific research as illustrations. The new journal is envisaged as a companion to traditional scientific journals.
https://www.chemistryworld.com/news/a-picture-speaks-a-thousand-words-in-new-scientific-journal/3008595.article
Source: Chemistry World

Screen reader plus keyboard helps blind, low-vision users browse modern webpages
A new approach developed by engineers at the University of Washington and Carnegie Mellon University uses the keyboard as a two-dimensional way to access tables, maps and nested lists.
https://www.washington.edu/news/2018/04/18/screen-reader-plus-keyboard-helps-blind-low-vision-users-browse-modern-webpages/
Source: University of Washington

Academic Software "Kopernio" Receives Funding from Innovate UK to Accelerate One-Click Journal Article Access for Academics and Researchers Worldwide

Kopernio is a free web-based software that enables one-click access to academic journal articles for more than 9m researchers and scientists worldwide and across a wide range of academic and researcher workflows. It intelligently activates on over 20,000 journal websites, databases, and search platforms such as PubMed and Google Scholar.

http://www.prweb.com/releases/2018/02/prweb15148605.htm

Source: Cision PRWeb
