RSC INTEREST GROUP
CHEMICAL INFORMATION AND
COMPUTER APPLICATIONS GROUP

NEWSLETTER summer 2019

Above: James Webster, the first recipient of the Peter Willett Award for Outstanding Poster Presentation (organised and sponsored by CICAG), receives his certificate from Professor Willett at the at the 8th Joint Sheffield Conference on Chemoinformatics (see page 7)

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

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

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, FRSC, email: stuart@psandim.com

Chemical Information & Computer Applications Group Chair's Report

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

Whilst RSC members can join up to 3 interest groups for free, in practice many members do not take up this opportunity. In an effort to highlight this CICAG ran an advertising campaign to coincide with annual subscription renewals at the start of the year. I'm delighted to report that these efforts were rewarded with an increase in CICAG membership from 388 to 440. I'd also like to thank all members of the committee for their Tweets, posts on LinkedIn and emails during the advertising campaign. Whilst the renewal letter is the traditional time for members to join interest groups you can make a request to join a group via email (membership@rsc.org) or telephone (01223 432141).

Social media is becoming an increasingly active way for communicating with members with both Twitter (https://twitter.com/RSC_CICAG) and LinkedIn (https://www.linkedin.com/groups/1989945/) gaining followers. Importantly the social media feeds provide an opportunity for communication with non-RSC members around the world. The CICAG website (http://www.rsccicag.org) is now active and we would be very interested to hear suggestions for additional content.

The April meeting "Workshop on Computational Tools for Drug Discovery" co-organised with the SCI Fine Chemicals Group, was a great success. The event ran at full capacity and the feedback from both delegates and software vendors has been excellent. We plan to repeat this event in 2 years' time.

The 8th Joint Sheffield Conference on Chemoinformatics is held every three years (https://cisrg.shef.ac.uk/shef2019/) and took place 17-19 June 2019 and CICAG are delighted to have sponsored this meeting; Nathan Brown is the CICAG representative on the organising committee and we hope for greater involvement in future meetings. CICAG are also delighted to announce the first winner of the Peter Willett Award for Outstanding Poster Presentation, James Webster.

After the outstanding success of the first meeting, the organisation of the 2nd RSC-CICAG/RSC-BMCS Artificial Intelligence in Chemistry Meeting is well underway, and will be held at Fitzwilliam College in Cambridge, 2-3 September 2019. We have a really outstanding line-up of speakers, and registration is now open.

Organisation of the joint meeting between RSC-CICAG and RSC-BMCS to acknowledge 20 years of the Rule of 5 is also now well underway. It will be held on Wednesday, 20 November 2019, at Sygnature Discovery, BioCity, Nottingham, UK. The meeting is generously sponsored by Sygnature and the line-up is almost complete, and it looks like it will be an entertaining day of science.

The last newsletter also included a contribution from Noel O'Boyle on a recent publication *Clarifying the Cahn-Ingold-Prelog rules* (https://doi.org/10.1021/acs.jcim.8b00324). We are delighted to include more contributions from external sources, these include Rich Apodaca (*WebAssembly*), Lucy White of CCDC (*Big Data; the 1 Millionth Structure*), Gabriele Cruciani (*Peter Goodford*), and Diana Leitch & Meryl Jameson (*Catalyst opens the new History Makers Gallery*), and Wendy A. Warr (*Meeting Report: AI3SD Workshop on AI in Drug Discovery and Drug Safety*). A big thank you to the external contributors to this issue; once again, I'd like to invite other contributions that would be of interest to the CICAG community.

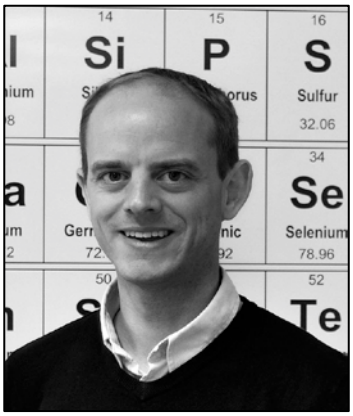
CICAG Planned and Proposed Future Meetings

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
Second AI in Chemistry Meeting	2-3 Sep 2019	Fitzwilliam Museum, Cambridge	2-Day meeting, registration now open
20 Years of the Rule of 5	20 Nov 2019	Sygnature Discovery, BioCity, Nottingham	Celebrating 20 years of Lipinski's rule, registration now open
Structure, Reaction and Patent Information for Small Organisations	TBD	TBD	Proposed joint meeting with RSC Consulting Group
Open Source Software 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)
A hundred days to Plan S	Sep 2020	TBD	Discussion on planned and possible changes in chemistry publishing

CICAG Committee News

CICAG Committee Member Dr Neil Berry – Professor of Chemistry and Department Head at the University of Liverpool



CICAG is delighted to hear that committee member Neil Berry has been promoted to Professor of Chemistry and Head of Department at the University of Liverpool. Professor Berry said: "*I am honoured to be Head of Department of Chemistry. Having been in the department for over 16 years I have seen it grow 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, and in the most recent 2014 Research Excellence Framework, was ranked 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 realise our vision of a Department delivering world-class research, an outstanding student experience and wide-reaching impact.*"

Peter Goodford: Pioneer of Structure-Based Drug Design & Molecular Interaction Fields

Contribution from Gabriele Cruciani, student, colleague, and friend of Peter

Peter Goodford was preparing a course of lectures on "Blood" for medical students at the Royal London Hospital in 1965, when he learned how Max Perutz in Cambridge was using X-ray crystallography to determine the structure of the haemoglobin tetramer (Hb4).

Peter thought that macromolecular structures like Perutz's Hb4 could possibly be used as scaffolds to design novel therapeutic agents, but he very soon realised that it might be difficult to examine such an unsubstantiated hypothesis in the clinical environment of a teaching hospital.

Goodford therefore moved to the pharmaceutical industry in 1967, and was appointed head of a new Department of Biophysics and Biochemistry at the Wellcome Research Laboratories near London. By the early 1970s he had convinced Wellcome's research director, John Vane, that the scaffold hypothesis should be tested, and that Wellcome should establish a laboratory for the X-ray crystallography of proteins. Max Perutz in Cambridge, David Phillips in Oxford, Herman Watson in Bristol, and Tom Blundell at the University of Sussex all provided active support for this project, and Goodford recruited crystallographers (John Champness, Christopher Beddell, and David Stammers) from these academic groups.

During the 1980s Goodford retired from Wellcome, and went to live with his family in a village outside Oxford and used his savings to buy a small VAX VMS DEC computer for his study at home. Goodford did not believe that it was possible to *design* therapeutic agents. He reserved this word for cars, railway tunnels, knives, forks and anything whose ideal properties could be precisely and completely specified on paper before it was actually made. In his view it was not normally possible to synthesise novel drug molecules with precisely defined ideal properties, and he would sometimes emphasise this opinion by asking colleagues how they would design and synthesise a flat methane molecule.

In Goodford's opinion suitable molecules for a specific therapeutic job could sometimes be *discovered*, and he thought that computational studies would be an aid in the discovery process. He believed that the findings could help with the generation of new concepts, and thus lead to the discovery of novel compounds for biological testing. He thought that computer programs should not simply deliver numbers (which might or might not be correct), but should promote new creative exploitable ideas in the mind of the user.

Goodford had this objective in mind when he devised his well-known GRID program as a tool for anyone studying a Target molecule of known structure (such as Hb4). GRID would draw the user's attention to places on the Target where a particular atom or chemical group (known as the "Probe") should interact favourably. Probes could be monatomic (e.g. carbonyl oxygen) or more complex (e.g. a complete amide group CO.NH2) or conceptual (e.g. the hydrophobic Probe), and the user had to decide which Probe or Probes he would like to study. He might for example select an aliphatic hydroxyl Probe if he wanted to synthesise sugar molecules, or a carboxyl Probe if he thought that acidic ligands would be appropriate for the chosen Target. However the use of GRID would more or less force him to think seriously about the structure and properties of both Target and Ligand.

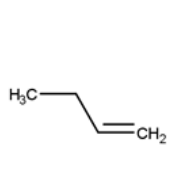
The American scientist Yvonne Martin heard about GRID during the 1980s, and invited Goodford to install a copy of the program in her laboratory. This was the start of a collaboration during which Peter came to realise that many people world-wide might be interested in the GRID method. Before long the program had been ported to other hardware platforms, and there were GRID users on every continent. Peter became a frequent visitor at academic departments from Barcelona to Sydney, from Helsinki to San Diego; and at pharmaceutical companies from Cincinnati to Osaka, and from Berlin to Melbourne. In 1989 an International School of Crystallography was organised by Peter at Erice, and a powerful Silicon Graphics computer was lugged up the second highest mountain in Sicily so that participants could try out the software for themselves. Installing that computer at that time, and getting it to work in such an isolated location, was a significant achievement for Silicon Graphics, and it is sobering to think that almost every participant could bring his own powerful notebook to the corresponding Erice school only twenty years later.

Goodford took the view that the success of a drug discovery project could only be judged by the number of new medicines which it spawned and by the therapeutic success of those novel compounds in helping sick people in the real world. He was very excited when Glaxo used GRID in the discovery of their anti-flu drug Relenza, the first of few real drugs designed with his method. Peter himself continued devising novel GRID algorithms for more complex Probes and for more flexible Targets until the year 2006, with the help of young scientists expert in modern methods of drug discovery and coming from all over the world.

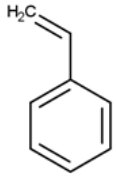
Peter passed away peacefully on 10th May 2019, aged 87.

New IUPAC Project to Develop SMILES+

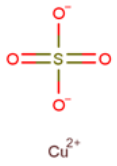
Contribution from CICAG Committee member Dr Helen Cooke, helen.cooke100@gmail.com, and the IUPAC SMILES+ Team



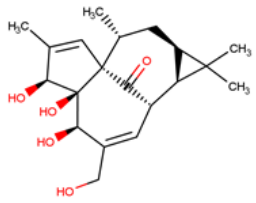
CCC=C



C=CC1=CC=CC=C1



[Cu++].[O-]S([O-])(=O)=O



C[C@@H]1C[C@@H]2[C@H]([C@@H]3C=C(CO)[C@@H](O)[C@]4(O)[C@@H](O)C(C)=C[C@@]14C3=O)C2(C)C

Most cheminformaticians will have used, taught or at least be familiar with the SMILES notation, developed by Daylight Chemical Information Systems in the late 1980s. Although SMILES is still widely utilised, unfortunately it is no longer centrally supported or developed to address new requirements, and there is no up-to-date documentation. The consequence is that non-standard SMILES dialects and extensions have been developed, usually for very good reasons, by individual organisations to address new use cases. But this evolution of SMILES is hampering interoperability and the sharing of information and data between systems and organisations. The '+' in SMILES+ is to indicate that it's intended that future approved extensions to a core IUPAC SMILES specification will be able to be accommodated.

This new four-year project, being undertaken by a global team led by Vincent Scalfani from the University of Alabama, was approved by IUPAC in April 2019 and reports in to its Committee on Publications and Chemical Standards.

It is sometimes asked why this project is necessary when InChI already exists. InChI and SMILES serve complementary chemical information use cases, and having an up-to-date, maintained, standard version of SMILES will be an important stepping stone to elaborate on these use cases. To ensure common objectives can be determined and achieved, the SMILES team is in collaboration with the InChI Trust.

Open SMILES, a community sponsored open-standards version of the SMILES language, will form the starting point for a working draft of the IUPAC SMILES+ specification which is being developed openly on GitHub.

Early goals for the SMILES+ project team include:

- Reviewing the existing Open SMILES specification, deciding which sections could be removed and identifying obvious gaps which should be addressed.
- Looking at existing documentation and manuals from software providers and building an annotated bibliography of current toolkit SMILES documentation.
- Reaching out to known providers of toolkits which make use of SMILES (CDK, ChemDoodle and OpenBabel are self-declared users of Open SMILES). This will assist development of a test suite of structures, including edge cases.
- Building a list of stakeholder contacts. We are seeking community input from the outset to help determine requirements. This stakeholder group will be engaged throughout the project - please contact Helen Cooke (helen.cooke100@gmail.com) if you are interested in joining this group.

Effective communication will be key to the success of this initiative. The project team has this in mind and will present regularly at appropriate events and meetings and communicate through a variety of channels. Look out for us at the InChI Symposium 23-24 August 2019 and the ACS Fall Meeting 25-29 August. Ultimately, the plan is for the specification to be published as an IUPAC recommendation.

Further information about the project is on the IUPAC website.

The 1st Peter Willett Award for Outstanding Poster Presentation

Contribution from CICAG Committee member Dr Nathan Brown, nathan.brown@benevolent.ai

The RSC-CICAG group is pleased to announce that the inaugural awardee of the Peter Willett Award has been given to James Webster from the University of Sheffield for his work on "*Reaction Vector Based Monte Carlo Tree Search for De Novo Design*". The award was given to James at the 8th Joint Sheffield Conference on Chemoinformatics on 19 June 2019. The judging panel consisted of Dr Nathan Brown (BenevolentAI), Dr Robert Clark (SimulationsPlus), and Dr Peter Ertl (Novartis Institutes for BioMedical Research). James received a certificate (photo), and £250 in vouchers.

In addition to the main award winner, honourable mentions were announced at the conference to: Christina Founti, Gian Marco Ghiandoni and Jess Stacey.

About the Award
The Peter Willett Award for Outstanding Poster Presentation is awarded triennially, coinciding with the Joint Sheffield Conference on Chemoinformatics, with this, the first award, being given last month. The award is named in honour of Professor Peter Willett for his outstanding contributions to the field of Chemoinformatics since the 1970s and recognises in particular his passion and commitment to educating the next generation of scientists.

Left: James Webster receives his certificate from Peter Willett at the Sheffield Conference (photo courtesy of Wendy A. Warr)

From the Award Winner
The Peter Willett award for outstanding poster presentation endorsed by RSC-CICAG is a great way to honour Professor Peter Willett's renowned contribution in the field of Chemoinformatics. When it was announced at the 8th Joint Sheffield Conference that I had been selected as the first recipient of this prestigious award, following my initial excitement and surprise my thoughts immediately turned to thank the amazing group of people who have provided me with unfailing encouragement and support.

Special thanks go to Professor Val Gillet, Professor Beining Chen, the entire Sheffield Chemoinformatics group and Mike Bodkin at Evotec, our industrial collaborator. Finally, I would like to thank the sponsors of the 8th Joint conference for making both this event and award possible.

My research interests stem from three areas: de novo design, reaction informatics and game playing. The title of the poster "Reaction Vector Based Monte Carlo Tree Search for De Novo Design" highlights the fusion of all three areas. Reaction based de novo design attempts to solve the problem of what to make next while simultaneously proposing a synthetic route to the designed compound. The fusion of Monte Carlo tree search, which has seen terrific success in the game playing literature, raises the performance and utility of the approach. By efficiently growing an asymmetric tree through the chemical space coupled with a simulation step to 'see' past local minima allows vastly improved performance over previous work. The poster is the culmination of 18 months of hard work for which the award is gratifying recognition.

James Webster

Abstract: De novo design [1] is an approach to rationally designing molecules which fit the desired property profile. De novo design is divided into three subfields; construction, scoring and search. Each component within a de novo design approach must be carefully balanced to chart an effective path through the vastness [2] of chemical space to areas of interest.

An active area of de novo design research is utilising reaction vectors for de novo design [3]. Reaction vector based de novo design utilises knowledge-based reaction transforms to build novel molecules systematically. Benefits of the reaction vector based approach include an increased likelihood the designed molecules are synthetically accessible and effective constraint on the size of chemical space to be explored. What is more reaction vectors also provide a prospective synthesis route to the designed molecules saving time on synthetic planning.

Previous work on search utilising reaction vectors has primarily explored the idea of greedy enumeration whereby all applicable transforms are applied in a breadth-first manner. This approach has a significant limitation in terms of depth due to the combinatorial explosion of products that are produced. Further refinement attempted to pick molecules to enumerate based on desirability criteria. This approach, however, led to an increased probability of being trapped in local optima: "the intermediate problem". Hence in this poster, we explore the use of the Monte Carlo tree search (MCTS) algorithm [4,5] as an alternative method of reaction vector based search.

Herein we describe our implementation of the reaction vector based Monte Carlo tree search (RV-MCTS) to search through chemical space effectively. To test our approach, we build a small, focused rediscovery benchmark tailored to reaction based de novo design. We demonstrate that the RV-MCTS approach can rediscover the majority of target compounds in the benchmark while simultaneously proposing a valid synthetic route.

1 M. Hartenfeller and G. Schneider, Wiley Interdiscip. Rev. Comput. Mol. Sci., 2011, 1, 742–759.
2 P. G. Polishchuk, T. I. Madzhidov and A. Varnek, J. Comput. Aided. Mol. Des., 2013, 27, 675–679.
3 H. Patel, M. J. Bodkin, B. Chen and V. J. Gillet, J. Chem. Inf. Model., 2009, 49, 1163–1184.
4 R. Coulom, in Computers and Games, 2006, pp. 72–83.
5 L. Kocsis and C. Szepesvári, in Machine Learning: ECML, 2006, pp. 282–293.

Big Data leads the way for Structural Chemistry

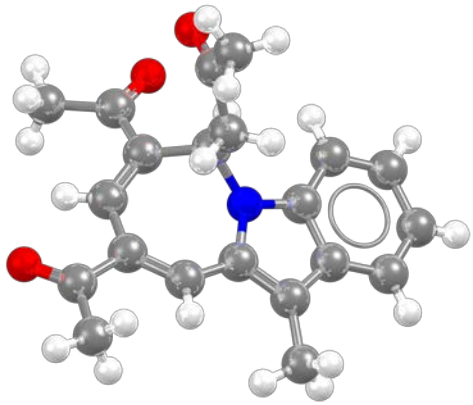
With thanks to Lucy White: lwhite@ccdc.cam.ac.uk and other members of the Cambridge Crystallographic Data Centre (CCDC) team

The Cambridge Structural Database has reached one million structures, leading the way in structural data to inform drug discovery and materials development.

CCDC (The Cambridge Crystallographic Data Centre), the world-leading experts in structural chemistry data, software and knowledge for materials and life science research and application, has recently announced a huge milestone for structural chemistry with the addition of the millionth structure into the Cambridge Structural Database (CSD).

The CSD is the world's repository of highly curated experimentally determined organic and metal-organic crystal structures. It is used globally by scientists in over 70 countries to understand how molecules behave and interact in three dimensions in the solid form and ultimately how this affects physical properties.

As the interest in 'Big Data' continues to grow in a time where machine learning and automation are changing the way pharmaceutical, agrochemical and many other industries work, reaching such a significant milestone is a huge achievement for the CCDC and the wider scientific community that contribute to and rely on this resource.



Left: CCDC's one millionth structure

Large volumes of data such as this enable scientists to generate more replete answers from a more complete and diverse volume of information, ensuring confidence in the insights being drawn from the data. Furthermore, CCDC's focus on ensuring the integrity of the data within the CSD through stringent quality assurance and control steps adds even more value and confidence that scientists are obtaining the highest quality information to inform their research.

This rich data resource, alongside advanced search, 3-D data mining, analysis and visualisation software from CCDC enables scientists from both industry and academia to further their research and predict new outcomes. In addition, knowledge derived from the CSD underpins computational chemistry and molecular modelling and is relied on by industry for the development and manufacturing of new drugs and within academia to teach chemistry.

Celebrating the achievement, Dr Jürgen Harter, CEO of CCDC has commented, "*This is truly an important milestone not only for CCDC but also for the wider scientific community. In addition to the value that lies in large sets of data like this to help scientists inform their research and decision making, we also pride ourselves on the high quality of the data, a result of the hard work of our expert in-house database team. Maintaining a policy of strict data interrogation ensures the value of the plentiful insights that can be drawn from the CSD, avoiding misinformation that can lead to wasted time, resources and ultimately cost.*"

CCDC announced the 1,000,000th structure to be a N-heterocycle produced by a chalcogen bonding catalyst activating multiple reactions steps sequentially. The structure was determined by Yao Wang and co-authors from Shandong University in China and published in the Journal of the American Chemical Society (JACS).

Dr Wang commented "*We are delighted to hear that our structure (1-(7,9-diacetyl-11-methyl-6H-azepino[1,2-a]indol-6-yl)propan-2-one; CSD Refcode XOPCAJ) is the millionth structure to enter the CSD! We have used the CSD for over ten years because it is an excellent platform to report new crystal structures and an outstanding database to find inspirable chemical structures. It is a valuable resource to us and to many other scientists around the world so we are very proud to be associated with this milestone for the community.*"

So, what's next for the CSD? Dr Harter believes that although the use of the CSD in the pharmaceutical and agrochemical industries is already well-established, it is now fast becoming a fundamental resource for research into new materials such as batteries, paints, pigments and dyes, and in particular the development of gas storage frameworks and tailored catalysts. As environmental contamination and sustainability become increasingly important there is considerable potential on a global scale. In addition, CCDC has noted a consistent rise in deposits from research taking place in China over recent years. "*It is an exciting time for life science and materials development research with markets such as China leading the way in scientific discovery and functional materials design. We are excited to see what insights we obtain from this market going forward*". CCDC also have plans to further draw on global insights and trends from the data to inform the direction of future research across different industries.

For more information visit: https://www.ccdc.cam.ac.uk/csd-1-million, or email Lucy White.

Catalyst opens new History Makers Gallery

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

On Friday 17 May, Catalyst was thrilled to officially open the Baker Gallery, a new exhibition which celebrates the work of local chemist Dr Harry Baker, a pioneer in the development of chlorine production at the Castner Kellner factory in Runcorn, on the site now owned by INOVYN.

The gallery has been funded by an Association of Independent Museums (AIM) Biffa Award, under a scheme called *History Makers*, which enables museums to recognise remarkable people that helped to shape our world.

Guests at the opening included members of the Baker Family, Chris Tane CEO of Inovyn, members of AIM, invited guests, and year 5 pupils from Weston Point Community Primary School who worked with the designer of the gallery to research the life and work of Harry Baker. They produced a newspaper which was given away at the event and created a song which they performed at the opening ceremony.

Dr Diana Leitch, Chair of Trustees and Acting Director of Catalyst commented "*It's almost 3 years ago that we had a dream to create a gallery dedicated to a local history maker and we hope that when you see the gallery you will understand its significance. Catalyst is very much about the past, present and the future, and the future is the young people that we inspire.*

We chose to work with Weston Point Community Primary School particularly because they are right next door to where Harry Baker worked from 1897 for the rest of his career, and where 122 years on, chlorine is still produced today, which is incredibly important."

Emma Chaplin, Director of AIM, the Association of Independent Museums, said "*It is fair to say that most of us had not heard of Harry Baker when we received the application from Catalyst but we were blown away by the passion and the huge connection to the site and the community. You've set the bar really high for other History Makers and from our point of view, it's great that we can showcase you to other museums across the country and that Biffa Award can show how landfill tax money is being invested, so, incredibly well done for making such a difference."*

More can be found out about Catalyst by visiting the website at www.catalyst.org.uk or contacting Meryl Jameson, Marketing Manager at Catalyst Science Discovery Centre, meryl@catalyst.org.uk.

The Association of Independent Museums (AIM) is a national charity which connects, supports and represents independent-thinking museums, galleries and heritage organisations, helping them to prosper. For more information contact Sassy Hicks sassy@aim-museums.co.uk.

Biffa Award**:** Since 1997, Biffa Award has awarded grants totalling more than £165 million to thousands of worthwhile community and environmental projects across the UK. The programme administers money donated by Biffa Group Ltd through the Landfill Communities Fund. For more information see www.biffa-award.org.

The AIM Biffa Award History Makers is a programme which enables museums to apply for funding to create new exhibitions featuring the lives and achievements of extraordinary, historical figures who have made a significant impact in the industrial, creative industries and arts, scientific or social history of the UK and in so doing, helping to shape the world we live in today.

WebAssembly

Contribution from Richard L. Apodaca, email: rich.apodaca@gmail.com

Richard L. Apodaca is a software developer and Ph.D. chemist living in La Jolla, California. His company, Metamolecular, creates innovative solutions to chemical information problems.

WebAssembly

Computational chemists and cheminformaticians looking for ways to deploy their software quicker and more easily may be interested in WebAssembly (aka "wasm").

The WebAssembly standard describes a platform-independent binary instruction format.[1] Its goal is to enable the production of sandboxed execution environments capable of running binary code at near-native speed. Since 2017 all major browsers (Chrome, Safari, Edge, Opera, and Firefox) have shipped with WebAssembly runtimes. Several standalone runtimes are also under development, including one from Intel.[2] A comprehensive and up-to-date collection of WebAssembly resources can be found on GitHub.[3]

WebAssembly traces its origins to the popular Emscripten compiler.[4] Emscripten's original purpose was to compile LLVM output to JavaScript for use in Web browsers. This worked far better than it should have, and before long very large codebases written in C and C++ were being compiled to JavaScript. Asm.js, a subset of JavaScript that can be executed efficiently, emerged as a way to make Emscripten's output run faster. Although Asm.js performed well during runtime, parsing became a bottleneck. WebAssembly's initial aim was to surmount this problem with a rapidly-parsable binary instruction format.

Any program that can be compiled to LLVM can potentially be further compiled to WebAssembly. This opens many opportunities for simplifying the distribution and use of chemistry software written in languages like C, C++, and even Fortran. Over at my blog, *Depth-First*, a multi-part series currently in progress details the compilation of InChI, a popular cheminformatics codebase written in C, to WebAssembly.[5]

WebAssembly's performance should make it attractive for computationally-demanding chemistry software. Two comparisons are worth considering: (1) WebAssembly vs. native; and (2) WebAssembly vs. JavaScript. A recent study compared WebAssembly/native performance in a suite of benchmarks.[6] As expected, the WebAssembly version ran consistently slower. However, the slowdown was remarkably small (average 1.43x on Firefox and 1.92x on Chrome; peak 2.6x on Firefox and 3.14x on Chrome). The authors pinpointed the source of these performance gaps, which may translate into even better and more consistent performance in the future. In another series of reports, the performance of WebAssembly and JavaScript implementations of a computationally intensive algorithm was compared.[7] Initially, the WebAssembly version ran faster, but after some superbly documented optimizations, the JavaScript version came out ahead. The speed of the WebAssembly version was then further optimised. There are two lessons here: (1) the modern JavaScript runtime is blazingly fast; and (2) extracting actionable lessons from benchmarks is tricky business.

Despite its good performance and impressive capabilities, the creators of WebAssembly consider the current version to be a proof-of-concept. As such, important features have been left out until clear demand emerges.[8] Some of the many proposed improvements include garbage collection, more efficient JavaScript bindings, and zero-cost exception handling. The rapid adoption of WebAssembly even at this early stage bodes well for future improvements in performance and functionality.

Language and performance no longer pose significant barriers to the in-browser deployment of vast swaths of chemistry software using WebAssembly. Software compiled in this way can access powerful browser APIs for GPU-accelerated graphics (WebGL), 2D vector and raster graphics (SVG and Canvas, respectively), and non-blocking computation (WebWorkers). Even if deployed outside the browser, WebAssembly offers an unprecedented cross-platform, embeddable, secure, and fast binary compilation target.

[1] https://webassembly.org
[2] https://github.com/intel/wasm-micro-runtime
[3] https://github.com/mbasso/awesome-wasm
[4] https://emscripten.org
[5] https://depth-first.com/articles/2019/05/15/compiling-inchi-to-webassembly-part-1/
[6] https://arxiv.org/abs/1901.09056
[7] http://fitzgeraldnick.com/2018/02/26/speed-without-wizardry.html
[8] https://webassembly.org/docs/future-features/

Tony Kent Strix Award and Annual Lecture 2019

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

1. UKeiG - 2019 Call for Tony Kent Strix Award Nominations

The UK electronic information Group (UKeiG) is now seeking nominations for 2019 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 28 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 2019 award should reach the judges by Friday 28 September 2019 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 2019 Strix Annual Lecture in London on the afternoon of Friday, 29 November.

From Nantwich to Oxygen: Joseph Priestley's Journey of Discovery

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

Since my article in CICAG's winter 2018/19 Newsletter, preparations for Nantwich Museum's exhibition to celebrate the International Year of the Periodic Table have been progressing well. At the time of writing, the information panels are almost complete, our artist has almost finished his paintings and plans are in place for a number of events.

At the exhibition, we will be introducing Nantwich Museum's visitors to Joseph Priestley, a non-conformist minister and school teacher in the town from 1758-61. At his school, unusually for the time, he taught science - his pupils, including some girls, were introduced to scientific equipment and experiments. Historians have suggested that Priestley's scientific teachings constituted the first ever science lessons. Throughout his life, as well as his discovery of oxygen and other gases, Priestley made significant contributions in other fields, such as religious and political philosophy, emerging as an important, though controversial, international figure and man of the Enlightenment.

Soon after Priestley's discovery of oxygen in 1774 the history of the Periodic Table begins, and his discovery will be used to introduce the Table to Museum visitors of all ages. The role and importance of oxygen is also a feature of the exhibition. We're proud of the embroidered Periodic Table which will be on display, made by the Museum's Craft Group, who have greatly enjoyed working on the project.

The exhibition runs from 14 August – 26 October 2019, and more information can be found in the RSC's Events Database and on the Nantwich Museum website.

Public events:

- Throughout the exhibition: Element Trail for children around the Museum
- 21 and 22 August: "Priestley's school in Nantwich" themed drop-in craft workshops for children
- 27 September: Periodic Table themed coffee morning
- 28 September: "Priestley's Element": drop-in workshop for children led by Heidi Dobbs, RSC
- 5 October: Half-day series of talks, featuring Gaye Blake-Roberts (Curator, Wedgwood Museum), Fabio Parmeggiani (School of Chemistry, University of Manchester), Helen Cooke (CICAG and N Staffs committees and Nantwich Museum).
  Booking required, email enquiries@nantwichmuseum.org.uk, tel. 01270 627104

Sponsors and supporters:
With thanks and gratitude to

- Nantwich Museum Trust Ltd
- RSC North Staffordshire Local Section
- RSC CICAG
- RSC Environmental Chemistry interest group
- RSC Historical Chemistry interest group
- RSC - IYPT

Meeting Report: Workshop on Computational Tools for Drug Discovery

Co-organised with the SCI Fine Chemicals Group, 10 April 2019

Contribution from CICAG Committee member Dr Neil Berry, email ngberry@liverpool.ac.uk

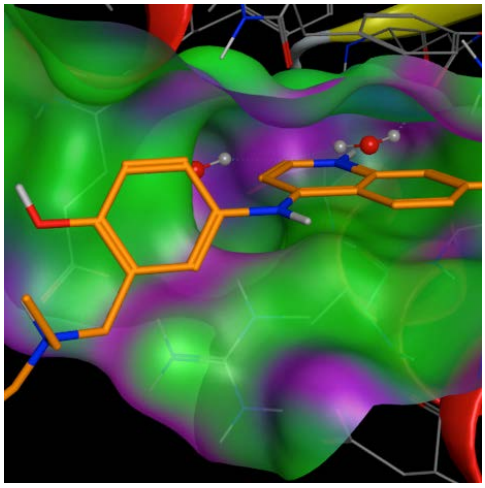
All scientists working in drug discovery need tools and techniques for handling chemical information. This workshop was intended to offer an opportunity to try out a range of software packages with expert tuition in different aspects of pre-clinical drug discovery. Attendees were able to choose from sessions covering data processing and visualisation, ligand and structure-based design, or ADMET prediction run by the software providers. All software and training materials required for the workshop were provided for attendees to install and run on their own laptops and use for a limited period afterwards.

Six software vendors took part: Optibrium, Cresset, Dotmatics, BioSolveIT, Knime and ChemAxon, with delegates able to choose to attend four of the six workshop sessions. All vendors provided the software for

download and installation prior to the event which meant the sessions could start without delay. Each session was 90 minutes, and most vendors used previously prepared training materials and had a helper on hand to circulate amongst the delegates to pick up any issues. The event was kindly sponsored by these software vendors along with Medchemica.

Interestingly, the majority of the delegates (60-70%) were from academia with the remainder from industry. The feedback was very positive with workshop content and venue scoring highly. There were a couple of request for certificate of attendance and suggestion that Continuing Professional Development (CPD) points could be awarded, something CICAG should think about for the next event. The feedback from the vendors was equally positive, several commented that it was a pleasure to get participants who had actively chosen to come rather than nominated by their boss. All vendors would like to be involved in future events and other vendors have requested that they be involved in future events.

The workshop was held at The Studio in Birmingham which proved to be a well-received location, being a minute's walk from the railway station and with ample parking nearby. All three rooms had good a/v facilities and ample power outlets to recharge laptops during the breaks. CICAG looks forward to organising a similar event in the future.



BioSolveIT ChemAxon cresset software dotmatics knowledge solutions Open for Innovation KNIME optibrium MedChemica

---

Meeting Report: AI3SD Workshop on AI in Drug Discovery and Drug Safety

Meeting report reproduced with the permission of Wendy A. Warr, email: wendy@warr.com

About the network

The Artificial Intelligence and Augmented Intelligence for Automated Investigations (AI^3) for Scientific Discovery Network+ (AI^3SD, http://www.ai3sd.org/) is funded by EPSRC, and hosted by the University of Southampton. The network 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 recent success of particular types of machine learning (e.g., deep neutral nets) has excited the interest of the scientific community in delivering insight into the complexity of the real world. This type of approach requires massive amounts of data to be trained. Traditional approaches to scientific discovery work with relatively small amounts of often uncertain data which are distilled by human insight to yield predictions and testable theories which may evolve as new data become available. The impact of "larger data" parallels the reality that almost all science now depends on computational assistance. Nevertheless the quantity of quality data needed to train the new AI systems is not directly available even with recent advances in automation. As a basis for the network the team at the University of Southampton proposes to use "amplification by simulation" as a key element of the cycle of automated experiments, simulation, AI

learning, prediction, comparison, design, and further experiments, to create the environment in which leading AI developments can be applied to chemical and materials discovery.

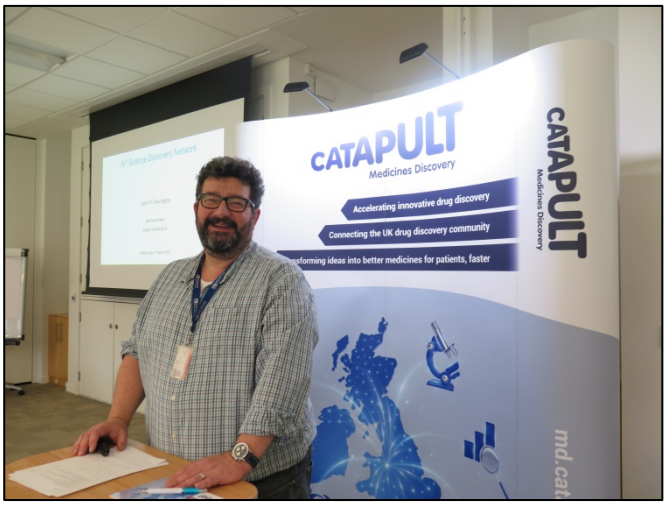
About Medicines Discovery Catapult

Catapults are not-for profit, independent centres which connect businesses with the U.K.'s research and academic communities. The Medicines Discovery Catapult (MDC, https://md.catapult.org.uk) is a national facility connecting the UK community to accelerate innovative drug discovery. It is an independent not-for-profit company, funded by Innovate UK, to bring together the large and diverse sector of industry, academia, charities, technologist, finance companies, small and medium enterprises (SMEs), and start-ups.

The Event

Drug discovery is a long and long-term scientific investigation involving interdisciplinary research methods coupled with large heterogeneous datasets. The research and data space in this area is vast, and AI^{3}SD and MDC believe that the use of AI and machine learning technologies can help spur on advances in this domain. The current workshop was designed to draw together those with a keen interest in using AI and machine learning technologies in the domain of drug discovery, both to aid future drug discovery, and to help improve drug safety. AI^{3}SD firmly believes that interdisciplinary collaboration is the key to many of these advances. At the workshop, keynote talks were interspersed with general group discussions and working groups around the key topics that arose.

Introduction to MDC Informatics



John P. Overington was the first speaker. He is the Chief Informatics Officer, Medicines Discovery Catapult. He leads the development and application of informatics approaches to promote and support innovative, fast-to-patient drug discovery in the United Kingdom through collaborative projects across the applied R&D community.

John described the informatics projects as "plumbing" and listed the current ones (not including two Innovate UK grants that have not yet been announced):

- theCollaboratory: inter-organisational data transfer from ELNs
- VESPA: a "multiscale" Bayesian network approach to variant effect prediction
- definitive a priori target validation using Mendelian randomisation
- AssayNet: directed graph of bioassays from gene to clinical trial and translational predictive models, annotated with supplier, academic lab, etc.
- CRISPY: MDC's collaborative intelligence platform, with AI-enabled search across the UK life science sector
- MPO-constrained optimisation using generative adversarial networks, computer vision, etc.
- Deep ADMET: SAR data "on demand" combining Optibrium's StarDrop and Intellegens' Alchemite (using deep learning),1 working with MDC under an Innovate UK grant
- drug combinations to improve efficacy, resistance, and safety
- addressing new target classes (the RNA-world, transporters, channels, etc.)
- cryptic pathogenic infectious agents

John gave more detail about CRISPY for collaborative intelligence.[2] It is difficult to identify people and organisations with specific skills and experience, and expert curation of resources is slow and expensive. CRISPY builds a live knowledge graph of UK drug discovery assets and capability, using data from Companies House, the Charities Commission, patents, published papers, grant applications, UK universities,

the Financial Control Authority, theses, the British Private Equity & Venture Capital Association, angel networks, professional societies, consultant networks, and so on. Natural language processing, named entity recognition, the software word2vec,[3] and faceted searching are used to provide a focused search engine for UK life science.

Search results can be visualised in various types of display. John showed some bar charts, and also hot spots for AI on a map of the United Kingdom. Most of the companies involved are around London and Cambridge but there is also one around Alderley Edge! Search terms are entered in a Google-like graphical user interface. John looked for "surface plasmon resonance". The system does not need an ontology: word2vec is able to find that "SPR", for example, is a synonym for "surface plasmon resonance". CRISPY returns a list of terms that can be selected (or deselected). John ticked five of the boxes for terms related to surface plasmon resonance, and obtained a table of the URLs and titles of 1,000 hits (from 40 different organisations) out of a total of 311,328 records. The search took only 1.49 seconds. More detail can be obtained for each hit. A preview column is also available in the table enabling a preview with marked-up text to be displayed on top of the table for a selected hit.

The system will make it easier to find collaborators, to "fill holes" in a project, and to study competitors. It is part of MDC's efforts to disrupt traditional skill sets, in a positive way, and, for example, to help chemists make the right compounds. Currently CRISPY is an internal tool for MDC staff, but John mentioned that he would be keen to find collaborators to apply and extend the system's content and uses.

Using Machine Learning to Drive Reaction-based *de Novo* Design

Val Gillet is a Professor of Chemoinformatics at the University of Sheffield where she heads the Chemoinformatics Research Group. She recommended a review[4] on *de novo* design. There was a flurry of attempts at *de novo* design in the 1990s. Three phases are involved in the design: making molecules, searching a vast chemical space, and scoring the molecules. In the 1990s restricted sampling of chemical space and scoring were difficult and most approaches were agnostic of synthesis.

Interest in *de novo* design went into abeyance but has re-emerged recently. Current approaches are reaction based, using rules, and generative models, using AI. AI methods typically use SMILES and no attention is paid to synthetic accessibility, except in that the system is trained on databases of molecules. Val's team uses reaction vectors (RVs) in an AI approach. RVs are counts of atom pairs removed from reactants and gained in products. Atom-pair descriptors of reaction components can be applied *in silico* to generate new products. In reality, a more complex descriptor is needed. A reaction centre one step out from the atom-pair is used in a sophisticated forward prediction algorithm, followed by fast structure generation. The reaction vectors are applied to previously unseen starting materials in order to suggest novel molecules for synthesis.[5] Each transformation that is applied has a precedent in the literature, and thus a high degree of confidence is established in the synthetic feasibility of the resulting molecules.

The approach has been implemented in KNIME and was validated by reproducing known reactions. For a wide range of reaction types 95% are well represented. Accuracy is not quite so good for more complex reaction types such as rearrangements: about 90% for 6000 reactions extracted from the *Journal of Medicinal Chemistry*. The literature source of the reactions is stored with the library of potential products.

The reaction based approach initially required datasets that were hand-crafted but a large collection of reactions has now been made publicly available[6,7] by NextMove Software. This database from US patents is referred to here as USPD. Schneider *et al.* have used text mining to extract 1.15 million unique, whole reaction schemes, including reaction roles and yields, from pharmaceutical patents. The reactions were assigned to well-known reaction types such as Wittig olefination or Buchwald-Hartwig amination using an expert system, and analysed to show the evolution of reaction types over time.

The Sheffield team have a process for cleaning the NextMove dataset prior to making reaction vectors. Stoichiometrically balanced reactions are needed, atom mapping is carried out, and catalysts etc. are removed. RVs are then calculated and validation carried out to see if the known product of a reaction can be generated. From the 1.8 million reactions,[6] 92,530 unique and validated RVs were obtained that also had classification labels. The distribution of the number of reactions per RV was highly skewed.

Val showed a flowchart for fully enumerated fragment expansion for a single step reaction from a starting set of 771 DSPL fragments, 24,000 Sigma Aldrich reagents and the USPD derived RVs, which produced 6.5 million virtual products. Multistep reactions cause a combinatorial explosion. The Sheffield team sought to mine USPD to find reaction classes more likely to be applied to a specific starting material. Their reaction recommender program does this, and reduces the number of predictions to synthetically accessible products. For a given starting material, the applicable RVs are limited to those that belong to reaction classes suggested by the recommender. The recommender therefore requires that reactions are grouped by class.

A classification procedure was first developed using reactions extracted from the USPD which were represented by RVs encoded as dynamic fingerprints. This approach is similar to that reported[8] by NextMove and Novartis, but Val's team modified the classification scheme to be compatible with the RVs. Reactions were grouped into a four-layer hierarchy. Validation involved an external dataset of 25,000 unseen RVs from USPD. Performance was similar to that reported by the Novartis team but there were 336 classes not 50 and these were combined with a confidence predictor.

The reaction recommender starts with starting materials and reaction class and aims to take account of features outside the reaction centre. The hope is to reduce the size of the chemical space while increasing the number of synthetically accessible molecules. This is now a multi-label classification problem. A starting set of classified reactions is used. Each reaction is represented by only a starting material. Starting materials represented by identical descriptors are merged and reaction labels are appended.

The recommender is trained on 1.1 million cleaned reactions from USPD. Starting materials and reaction class are extracted, duplicates are removed, descriptors are generated, and training and validation are carried out. Extensive experimentation is necessary to make the best decision on reaction class layers, learning methods etc. Two layers appear to be better than three. Val presented a table of the different types of descriptor used.

Retrospective validation was carried out with 26,000 reactions from the *Journal of Medicinal Chemistry* not seen by the model. The reactions were cleaned and classified, and then starting materials were extracted and duplicates removed. The trained recommender was used for each starting material and the reaction suggested was compared with the true class. Val presented some results. In some cases there was no recommendation or a wrong recommendation. The results could be improved if sufficient training data were used. Some wrong recommendations were due to a starting material having more than one reaction centre and the "incorrect" one being suggested. These could be corrected by applying the recommended reaction and making a new recommendation for that product.

In a second validation, behaviours with and without the recommender were compared, starting with fragments from DPSL and reagents from Sigma Aldrich. The results showed that the recommender is successful in reducing the number of products, and in saving time. The synthetic accessibility was measured using the MOE rsynth descriptor and mean SA score.[9] The number of products decreases and the synthetic accessibility scores improve as the classification labelling becomes more fine-grained.
This work is being carried in collaboration with Mike Bodkin and others at Evotec. The team is currently exploring the use of the recommender in augmented *de novo* design and in fully automated design.

Re-energising Small Molecule Drug Discovery

Willem van Hoorn (pictured left with the logo modified from Darwin's tree of life) is Chief Decision Scientist at Exscientia. He said that although 90% of drugs are small molecules, and 50% of clinical trials are for small molecules, small molecule drug discovery remains inefficient: it takes five years to get 2500 compounds from idea to drug candidate.[10] Hit to candidate is the most expensive part of drug discovery in terms of cost of capital. Exscientia's vision is to produce a candidate from 500 compounds in just 1.5 years.

Pioneering automated drug design methodologies[11] developed by Professor Andrew Hopkins and other researchers at the University of Dundee led to the spin out of Exscientia in 2012. The company now has an office in Oxford as well as the original one in Dundee. It has become a scale-up, not a start-up company, with clients' molecules heading towards the clinic. The company's AI-driven systems actively learn best practice from vast repositories of discovery data and are further enhanced with knowledge acquired from seasoned drug hunters. The proprietary AI design module uses a genetic algorithm to predict new structures; machine learning did not work as well or as quickly.

Novel compounds prioritised for synthesis by the AI systems simultaneously balance potency, selectivity and pharmacokinetic criteria in order to deliver successful experimental outcomes. By applying a rapid design-make-test cycle, the Exscientia AI platform actively learns from the preceding experimental results and rapidly evolves compounds towards the desired candidate criteria. Exscientia systems learn from both existing data resources and experimental data from each design cycle. The company is developing single-target small molecules as well as compounds with more challenging target product profiles, through a novel bispecific small molecule strategy (compounds with dual pharmacology in an integrated pharmacophore) and phenotypic-driven drug design.

In 2015, the pharmaceutical firm Sunovion asked 10 of its chemists to play a game[12] to see who could discover the best leads for new drugs. The chemists were presented with hundreds of chemical structures, just 10 of which were labelled with information on their biological effects. The experts had to select other molecules that could turn out to be drug candidates. The 11th player was an Exscientia computer algorithm. The chemists took several hours, the computer only milliseconds. Only one chemist out of the 10 beat the machine. By 2017 the machine was as good as the best chemist.

Nevertheless the human is not redundant. In 1996, the first chess match between world chess champion Garry Kasparov and the IBM supercomputer Deep Blue was won by Kasparov. The second, in 1997, was won by Deep Blue. The 1997 match was the first defeat of a reigning world chess champion by a computer under tournament conditions. In 2016, in a five-game Go match between 18-time world champion Lee Sedol and AlphaGo, a computer Go program developed by Google DeepMind, AlphaGo won all but the fourth game. Advanced Chess, sometimes called centaur chess, was introduced by Garry Kasparov in 1998. A centaur chess player is one who plays the game by marrying human intuition, creativity and empathy with a computer's brute-force ability to remember and calculate a staggering number of chess moves. By 2008 it had been shown that, not surprisingly, a centaur beats the solo human, but less surprisingly, a centaur beats the solo computer (under certain time constraints).

In Exscientia's Centaur Chemist technology the machine and the chemist work together. The principle is similar to how a human would learn, but the AI process is far more effective at identifying and assimilating

multiple subtle and complex trends to balance potency, selectivity and pharmacokinetic criteria. For example, the decision to focus on, say, hERG demands human strategic thinking, but finding compounds that fulfil that brief is better done by the AI. Willem presented some case studies.

In a collaborative psychiatric drug design project between Exscientia and Sunovion, the research focused on developing novel approaches to compound design by analysing data arising from phenotypic drug discovery. This collaboration builds upon Exscientia's delivery of novel bispecific compounds that combine activities at the GPCR and ion channel target families. In the work reported by Willem, one target had literature and patent SAR, but the other had limited literature SAR. Exscientia's clients had worked previously on the second target (in a single-target approach) and had run a high throughput screen. They shared the screening data so that Exscientia could build a model. In the lead identification stage 20 compounds, on average, were made per two-week cycle. Willem presented a table showing which compounds were taken into lead optimisation. Just over 100 compounds were made in two series. There were 80 further compounds for each prioritised scaffold.

Exscientia and Sumitomo Dainippon researchers are working in partnership to deliver novel multi-target small molecules that have potential to deliver enhanced therapeutic performance in the treatment of psychiatric diseases. The first compound passed over to Sumitomo Dainippon for further internal development is a bispecific, dual-agonist compound that selectively activates two GPCR receptors from two distinct families. The accelerated project delivered the molecule that fulfilled development quality criteria in only 12 months. Starting from a product concept, fewer than 400 compounds needed to be synthesized[13] to identify molecules that matched the required development criteria. Clinical trials will begin for one compound in 2019.

It has been said that AI will not replace chemists but chemists who do not use AI will be replaced by those who do. The same maxim can be applied to companies: those that use AI will out-compete the ones who do not. The positive message is that chemists and companies who do not use AI can become chemists and companies who do.

Understanding the holes in the metabolome

Nicola Richmond is a director in the Artificial Intelligence and Machine Learning team at GlaxoSmithKline (GSK). She described some work carried out by Casey Wojcik, a postgraduate researcher at Stanford University who worked at GSK under the Cambridge Mathematics Placement scheme after finishing Part III of the Mathematical Tripos at the University of Cambridge.

GSK's motivation for the work was to develop objective approaches to analysing metabolomics data. Metabolomics is the scientific study of chemical processes involving metabolites, the small molecule intermediates and products of metabolism. GSK wanted to find out if they could add more insights with topological data analysis (TDA) in an objective way. Nicola showed a typical "hairball" network: the simplicity and beauty of node-link diagrams turns into clutter and confusion when the number of nodes and links gets too high.

TDA is a mathematical approach to the analysis of datasets using techniques from topology. TDA provides a general framework to analyse datasets that are high-dimensional, incomplete and noisy, in a manner that is robust to noise. TDA has combined algebraic topology and other tools from pure mathematics to allow mathematically rigorous study of shape.

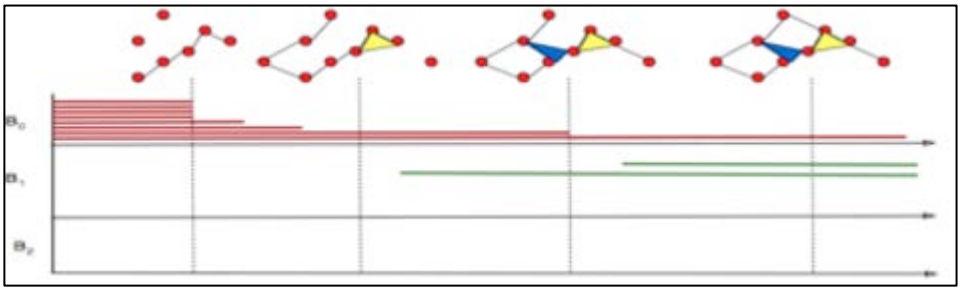
Topology classifies spaces based on their invariant properties. A well-known example is the similarity between a doughnut and a coffee mug: you can deform a doughnut into a coffee mug. The shapes can

stretch or deform but not break. Data also have structure and shape. TDA can be used to study the shape of data, in particular, connectedness and voids.

A simplicial complex is a geometrical representation of a topological space which is realised as a union of simplices, such as points (0-simplex), line segments (1-simplex), triangles (2-simplex), tetrahedrons (3-simplex) and other higher dimensional cousins. Simplicial complexes provide a simple combinatorial way to describe certain topological spaces.

You can build a family of simplicial complexes from a point cloud and study the complexes in order to study the point cloud. Persistent homology is a method for computing topological features of a space at different spatial resolutions. More persistent features are detected over a wide range of spatial scales and are deemed more likely to represent true features of the underlying space rather than artefacts of sampling, noise, or particular choice of parameters.

Topological spaces can be characterised by using topological invariants: algebraic objects which are invariant under homeomorphisms. Homology, one of the topological invariants, is a mechanism for counting the number of *n*-dimensional holes. Persistent homology is the computational implementation of homology, allowing you to describe a simplicial complex in terms of *n*-dimensional holes. These holes are expressed by Betti numbers, *Bi*, where B0 corresponds to the number of connected components, B1 corresponds to the number of planar holes, B2 corresponds to the number of voids in solid objects (2-dimensional holes), and so on. Persistent homology is an incremental construction of the final filtered simplicial complex. In the figure below, the Betti numbers are visualised through barcodes. Barcodes are a graphical representation of Betti generators whose horizontal axis corresponds to the filtration parameter and whose vertical axis represents ordering of homology generators. Barcodes could be useful descriptors for input to machine learning approaches.



A weighted network is a network where the ties among nodes have weights assigned to them. In a social network, the weights represent how well people know each other. In a metabolic network the weights relate to reaction rates. Topological features are important. While the null hypothesis (H0) in any experiment or research project is that the connection or conclusion suggested by the experiment is false, the alternative hypothesis (H1) is always the assertion that there is a meaningful connection to be investigated. Open source software called jHOLES is available for computing persistent homology.[14]
Nicola described Casey's project using TDA and metabolic networks using metabolomics data on Panobinostat, a drug by Novartis for the treatment of multiple myeloma and other cancers. It is a histone deacetylase inhibitor, which promotes cell-cycle arrest and apoptosis of tumour cells *via* multiple pathways.

Casey used the Kyoto Encyclopaedia of Genes and Genomes (KEGG) to make a metabolic network and computed correlation coefficients for each metabolite with panobinostat and constructed a network with edges weighted by the average correlation coefficients of the end vertices. Next jHOLES was used to study cycles, with visualisation in Cytoscape, an open source software tool for visualising complex networks and integrating these with any type of attribute data. The results were grouped by cycle and coloured by pathway.

It could be seen that the first few cycles most strongly correlate with treatment. One pathway was singled out: a tricarboxylic acid (TCA) cycle for cysteine and methionine metabolism. That cycle is known to interact strongly with the drug. The practitioner can understand this. Note that this was a purely data-driven identification.

Nicola outlined the benefits of the methodology. It singles out strongly correlating subnetworks which are small enough to inspect manually. It can detect affected pathways, key metabolites, and cross-talk between pathways. Data imputation through diffusion allowed for unmeasured metabolites.

In summary, persistent homology has a strong theoretical foundation and is useful for understanding changing network structure, especially in metabolism. It is easy to compute and simplifies the task of analysing complex networks. GSK wants to understand systems biology. A lot of the data are present as omics datasets, and images from phenotypic screening could be used to advantage. Pure mathematics has a role to play in systems biology.

Discussion

Interspersed with the talks were discussions among six groups of attendees. Professor Jeremey Frey of the University of Southampton and AI^{3}SD introduced the discussions; and Samantha Kanza, the AI^{3}SD Network Coordinator, ably documented the feedback gathered on all the flipcharts. In Jeremy's introduction he said that we have some (mediocre) property prediction algorithms and some budding *de novo* design algorithms, but the number of applications that combine the two is limited. We need to make molecules with the right properties.

As a discipline, chemistry is splitting into biochemistry, molecular biology, nanotechnology, and so on. These sub-disciplines need to be linked so that data can be transferred across them. The subject of the discussions was the way that AI can transform science, and, in particular, applications of AI in drug discovery and use. By the end of the day Jeremy hoped to have a summary of the issues, and an agenda for making progress in addressing them.

Samantha analysed the feedback from each of the six groups and the ideas summarised on various flipcharts. She drew them together into seven themes: skills and training; data quality, access, collation, and reproducibility; interdisciplinary data sharing and interoperability; optimisation of the drug discovery process (reducing attrition, failure and cost); explainable AI and models; making decisions, and trustworthiness of machine learning; and miscellaneous other areas such as automated synthesis and candidate selection.

Conclusions

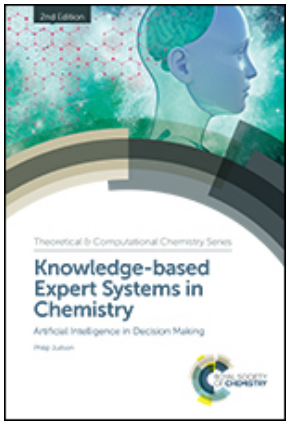
We, the attendees wanted to consolidate a view on the impact that AI will have over the next 5-10 years on the drug discovery process. We need tips for embarking on this journey. What does a good, AI-driven project look like? What are the outputs, constraints and limitations? Are the data reproducibile? Different communities have developed methods for handling "big data" but in drug discovery we do not always have big data. Instead we have complicated data on a few things. There are gaps in the suite of tools we can use. Lists of standards, resources and ontologies are needed. We must focus on the biology as well as the chemistry. AI is not really new, but it now has a lot of momentum, and it has caused some damage. We need a roadmap with realistic timescales. Since we are "punching above our weight" for AI, and we have a great range of talent, why do we not make greater demands on the government to support and retain this capability?

References
(1)Whitehead, T. M.; Irwin, B. W. J.; Hunt, P.; Segall, M. D.; Conduit, G. J. Imputation of Assay Bioactivity Data Using Deep Learning. *J. Chem. Inf. Model.* **2019,** *59* (3), 1197-1204.
(2)Markova, D.; McArthur, A. *Collaborative Intelligence. Thinking with People Who Think Differently*; Penguin Random House: New York, NY, 2015.
(3)Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. 2013, arXiv.org e-Print archive. https://arxiv.org/abs/1301.3781 (accessed March 18, 2019).
(4)Hartenfeller, M.; Schneider, G. Enabling future drug discovery by de novo design. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011,** *1* (5), 742-759.
(5)Patel, H.; Bodkin, M. J.; Chen, B.; Gillet, V. J. Knowledge-Based Approach to de Novo Design Using Reaction Vectors. *J. Chem. Inf. Model.* **2009,** *49* (5), 1163-1184.
(6)Lowe, D. M. Chemical reactions from US patents (1976-Sep2016). https://figshare.com/articles/Chemical_reactions_from_US_patents_1976-Sep2016_/5104873 (accessed March 20, 2019).
(7)Lowe, D. M. Extraction of chemical structures and reactions from the literature, Ph. D. Thesis, University of Cambridge. https://www.repository.cam.ac.uk/bitstream/handle/1810/244727/lowethesis.pdf?sequence=1&isAllowed=y (accessed March 20, 2019).
(8)Schneider, N.; Lowe, D. M.; Sayle, R. A.; Tarselli, M. A.; Landrum, G. A. Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter. *J. Med. Chem.* **2016,** *59* (9), 4385-4402.
(9)Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminf.* **2009,** *1*, 8.
(10)Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010,** *9* (3), 203-214.
(11)Besnard, J.; Ruda, G. F.; Setola, V.; Abecassis, K.; Rodriguiz, R. M.; Huang, X.-P.; Norval, S.; Sassano, M. F.; Shin, A. I.; Webster, L. A.; Simeons, F. R. C.; Stojanovski, L.; Prat, A.; Seidah, N. G.; Constam, D. B.; Bickerton, G. R.; Read, K. D.; Wetsel, W. C.; Gilbert, I. H.; Roth, B. L.; Hopkins, A. L. Automated design of ligands to polypharmacological profiles. *Nature* **2012,** *492* (7428), 215-220.
(12)Mullard, A. The drug-maker's guide to the galaxy. *Nature* **2017,** *549* (7673), 445-447.
(13)Yoshinaga, H.; Uemachi, H.; Ohno, T.; Besnard, J. Preparation of 2,6-disubstituted pyridine compounds for the treatment of symptoms associated with anxiety disorders. WO2018168738A1, 2018.
(14)Binchi, J.; Merelli, E.; Rucco, M.; Petri, G.; Vaccarino, F. jHoles: A Tool for Understanding Biological Complex Networks via Clique Weight Rank Persistent Homology. *Electronic Notes in Theoretical Computer Science* **2014,** *306*, 5-18.

Chemical Information / Cheminformatics and Related Books

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

Knowledge-based Expert Systems in Chemistry

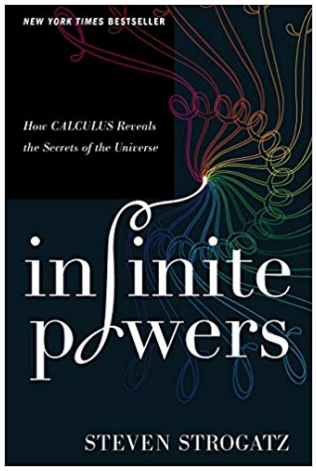


There have been significant developments in the use of knowledge-based expert systems in chemistry since the first edition of this book was published in 2009. This new edition has been thoroughly revised and updated to reflect these advances.

The underlying theme of the book is still the need for computer systems that work with uncertain or qualitative data to support decision-making based on reasoned judgements. With the continuing evolution of regulations for the assessment of chemical hazards, and changes in thinking about how scientific decisions should be made, that need is ever greater.
Written by a pioneer in the field, this book provides an essential

reference for anyone interested in the uses of artificial intelligence for decision making in chemistry.
RSC Publishing
Author: Philip Judson

Infinite Powers: How Calculus Reveals the Secrets of the Universe



Midway through Infinite Powers, Steven Strogatz writes that Isaac Newton and Gottfried Wilhelm Leibniz both "died in excruciating pain while suffering from calculi — a bladder stone for Newton, a kidney stone for Leibniz". It was a cruelly ironic end for the scientists who independently invented calculus: the word comes from the Latin for 'small stone', in reference to pebbles once used for counting.

Such fascinating anecdotes abound in Infinite Powers. Strogatz, a mathematician working in nonlinear dynamics and complex systems, has written a romp through the history of calculus — the study of how things change. Starting with the ancient Greeks, the book ends with connections between the field and artificial intelligence and machine learning. Calculus was key to working with Newton's laws of motion, which stimulated the Industrial Revolution. It is also central to quantum mechanics, which underpins the modern revolution in computers and communications. The book is a roll call of luminaries, including Galileo Galilei, Johannes Kepler, René Descartes and Pierre de Fermat.
Review extract with thanks to Nature.
Houghton Mifflin Harcourt
Author: Steven Strogatz

News from CAS

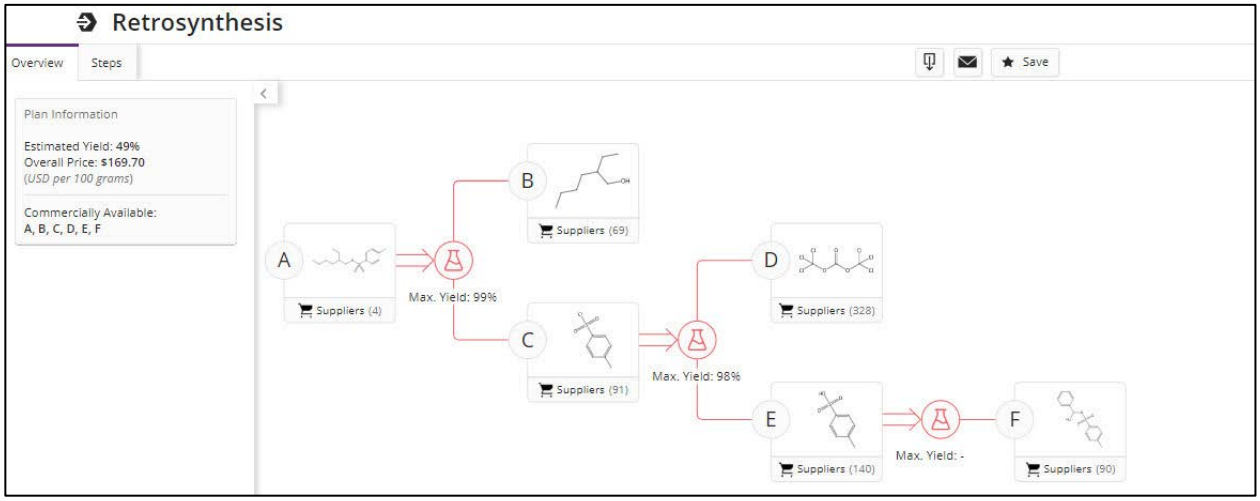
Contribution from Dr Anne Jones, email: ajones2@acs-i.org

SciFinder^n
We are delighted to announce the launch of Phase 1 of our new Retrosynthesis tool in SciFinder^n!
Retrosynthesis planning is the most significant CAS development in SciFinder^n history and is the best step towards computer aided synthetic design (CASD).

The NEW retrosynthesis planning capability in SciFinder^n gives organic chemists the power to quickly:

- Generate retrosynthetic plans for known molecules or intermediates
- Organize synthetic routes in a convenient, easy to navigate interface
- Identify material suppliers quickly and easily
- Save, download and share selected synthetic plans with key stakeholders

For any known molecule, SciFinder^n will perform a full retrosynthetic analysis utilizing the renowned CAS collection of reactions, presenting results in a highly intuitive retrosynthesis plan.



Alternative Route Selector

The route recommendation is offered based on a number of criteria including:

- simplicity of synthesis
- expected yield
- availability and cost of needed materials

Additionally SciFinder^{n} allows you to navigate available alternative routes and upon selection will regenerate the entire plan in seconds. SciFinder^{n} offers a recommended retrosynthetic route from a full analysis of the authoritative collection of reactions from CAS. The synthetic plan provides an estimate of the yield, cost and a summary of the available synthetic alternatives at each step along the way.

STN

New Web Address for New STN

Effective July 1, 2019, the web address for new STN will change from www.stn.org to nstn.stn.org. You can continue to use your existing loginID and password to sign in to new STN. If you are a new STN user and you have not yet tried STNext, we encourage you to try it instead, at next.stn.org. STNext is the preferred STN platform and is the only platform that is being actively developed with new features, functionality and content, offering you a superior searching experience.

Updates to World Surface Coatings Abstracts (file WSCA) Have Resumed

The database WSCA (World Surface Coatings Abstracts) is up-to-date again including new data of Coatings Online produced by PRA World Ltd. Previously, the database had been static on STN since 2013. For the years 2013 to 2018, more than 200,000 records have been added. The database provides scientific and business information on paints, coatings and related raw materials. WSCA will be updated monthly, and SDIs are again available. The Database Summary Sheet for WSCA has been revised to reflect the updated information.

PatentPak® Further Increases Your Efficiency by Adding 14 Additional Authorities to Coverage – available on SciFinder and STN

PatentPak® is an integrated workflow solution designed to radically reduce time spent acquiring and searching through full-text patents to find vital chemistry insights. PatentPak connects you to over 17 million searchable, full-text patents covered in CAplus^{SM}. The number of authorities PatentPak covers has just increased from 32 to 46, with the addition of the following: Argentina, Bulgaria. Czech Republic, Denmark, Finland, Hungary, Israel, Italy, Malaysia, Moldova, Netherlands, Phillipines, Poland and South Africa. Coverage for these new authorities is for basic patents, back to 2016. Organic chemistry patents are now complete for these authorities, and biological and agricultural patents are now underway.

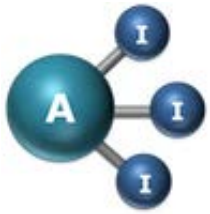
Aptean and CAS collaborate to offer CAS Biosequences content in GenomeQuest
Combining the CAS Biosequences™ Module and the GQ-Pat database on a single platform delivers a unique, comprehensive "one-stop" search solution for biotechnology intellectual property. Aptean has partnered with CAS, a division of the American Chemical Society, to make CAS Biosequences content available on the GenomeQuest sequence search platform. Aptean GenomeQuest is an intellectual property (IP) sequence search tool that integrates the world's largest, most up-to-date IP sequence database with state-of-the-art sequence comparison algorithms for powerful patent sequence search and analysis. Adding the CAS Biosequences Module to the Aptean GenomeQuest platform provides an additional source of human-curated biosequence content disclosed in patents and journals. This leads to an extension of the comprehensiveness of biotechnology IP searches and provides efficiency for users by making both of these critical resources accessible through GenomeQuest's powerful, user-friendly interface.

CAS Case Studies
CAS has been involved in the production of a number of case studies, available via our website. The most recent looks as the impact of data quality on machine learning results.
The case study can be downloaded at
https://www.cas.org/resources/case-studies/data-quality-impacts-machine-learning

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.

News from AI3SD

Contribution from AI3SD Network+ Coordinator Samantha Kanza, email: s.kanza@ai3sd.org



AI3SD Launched at the SCI in December 2018, and has been active for just over half a year. Since December we have hosted five workshops and put out our first funding call. We have another five meetings planned for 2019 and already have one formally organised for 2020. This piece will provide a summary of our activities to date, and detail what is coming next for 2019/2020.

18-19 November 2019, Winchester Science Centre and Holiday Inn
Registration for our AI3SD Network Conference 2019 is now open. Please register and share this event with colleagues who you think may be interested. As some of you will know, we launched in December 2018, and this conference marks the end of our first year. This is a two day event with a mixture of keynote talks from experts in different areas of AI for Scientific Discovery, and discussions around different research areas. There will be dedicated time for networking and we will be implementing a smart badge system whereby attendees can mark their badges according to whether they are looking for a collaborator, employment, job candidates, PhD students etc. We will report on the activities of AI3SD over the last year, including the workshops and hackathons we have run and attended, and there will be an opportunity to hear from the projects we funded in AI3SD-FundingCall1. On the evening of the 18th there will be drinks and networking in the Winchester Science Centre followed by a pre-dinner talk by famous Science Communicator Steve Mould. This will be followed by a formal conference dinner at the Holiday Inn.

The conference includes sessions for contributed talks and the presentation of virtual posters through flash talks. Anyone wishing to submit a poster or an abstract for a talk should complete the AI3SD 2019 Poster Submission Form or the AI3SD 2019 Abstract Submission Form. Please don't hesitate to get in touch if you have any queries about the conference or AI3SD.

AI3SD Other Upcoming Events for 2019/2020

AI/ML in Chemical Discovery & Development - 18-19 July 2019
Weetwood Hall, Otley Rd, Leeds LS16 5PS
This will be a two day residential event discuss to bring together different stakeholders from industry and academia to discuss applications of AI and Machine Learning in Chemical Discovery and Development. It will consist of a series of structured discussions and networking opportunities.

AI3SD Town Meeting & IP for Scientific Discovery Workshop - 11th September 2019
Wide Lane, University of Southampton Sports Ground, Wide Lane, Eastleigh SO50 5PE
The IP Workshop will be co-run with ClearViewIP who are funded by the European Commission for this work.

AI3SD & Pistoia Hackathon - 12-13th September 2019
University of Southampton Sports Ground, Wide Lane, Eastleigh SO50 5PE
This will be a two day hackathon run with the Pistoia Alliance.

AI3SD & Internet of Food Things Network+ AI and Allergen Detection Workshop - 17 October 2019
Society of Chemical Industry, 15 Belgrave Square, Belgravia, London SW1X 8PS
This will be a one day event with some keynotes and expert talks followed by a series of structured discussions around five different aspects of using AI technologies in the Food Industry and Allergen Detection.

AI3SD, Dial-a-Molecule & Directed Assembly Computationally Predicting Reaction Outcomes and Optimum Synthetic Routes – 9-11th March 2020
DeVere Totworth Court Hotel, Tortworth, Wotton-under-Edge GL12 8HH
This will be a three day residential event to examine the state of the art and future opportunities for the key problem of being able to predict the outcome of unknown chemical reactions, and consequently the ability to design optimum synthetic routes to desired molecules. The event will be a mixture of keynote talks, contributed talks and discussion sessions, and there will be an opportunity to submit posters and abstracts for short talks. Bartosz Grzybowski and Pierre Baldi have agreed to contribute plenary talks, and registration and submission opening details will be appearing soon.

****

AI3SD Events Hosted in 2019
AI3SD has hosted six events so far in 2019. More information on these events can be found at http://www.ai3sd.org/news/eventcoverage where blog posts and links to reports will be added in due course.

Network+ Town Meeting - 22 January 2019
AI3SD Hosted a Network+ Town Meeting at the SCI in London. This purpose of this meeting was to give prospective applicants the opportunity to ask questions about our first funding call, and also to find other organisations to collaborate with. There was time set aside for specific networking to find collaborators, and upon request from our attendees we created a Collaborators Page on this website for members who are actively seeking collaborators in research areas linked to AI3SD (both for this funding call and in general). All questions and answers were written up and added to our funding call page as an FAQ.

Molecules, Graphs & AI Workshop - 6 February 2019

On the 6 February AI3SD hosted a Molecules, Graphs & AI Workshop at the picturesque Ageas Bowl in Botley. This was a lively workshop with a strong academic presence from multiple universities, and some industry representatives also. The day was made up of three keynote talks and working group discussions on topics that were elicited from initial open discussions after the presentations. The full event page can be found here.

AI3SD & MDC AI in Drug Discovery & Drug Safety Workshop - 6 March 2019

On the 6 March AI3SD teamed up with the Medicines Discovery catapult to run a workshop at Alderley Park Conference Centre on AI in Drug Discovery & Drug Safety. This was a very well received event; there was a strong industry presence among the attendees and several representatives from different Universities. There were four keynote talks and working group discussions on topics that were elicited from initial open discussions after the presentations. The full report written by Dr Wendy Warr is on page 15.

AI for Materials Discovery Workshop - 19 March 2019

On the 19 March AI3SD hosted an AI for Materials Discovery Workshop at the University of Southampton. This was a very well attended workshop, with over 60 participants attending across the University of Southampton and representatives from many other Universities and Research Institutes, with some industry attendees also. This workshop ran across an afternoon and was made up of five keynote talks which generated a lot of questions and discussions. This was particularly useful for AI3SD, as it brought in many new members and raised awareness of the Network+ across different research groups. The event page complete with report written by Dr Nicola Knight and Dr Colin Bird is available here.

AI3SD Semantics & Knowledge Learning for Chemical Design Workshop - 1 May 2019

This workshop consisted of a small group of experts who are interested in using semantic web technologies to improve and progress scientific discovery. There were five keynote talks followed by some group discussions. The full event page can be found here.

AI3SD Funding Call 1

In January 2019 we put out our first funding call, and the results were announced in March. We have funded three projects for AI3SD-FundingCall1 which are –

- Project Title: Predicting the Activity of Drug Candidates where there is No Target
- Principal Investigator: Professor Matthew Todd
- Lead Institution: UCL
- Project Partners: Auromind, ExScientia & Intellegens

- Project Title: 'Next-next' Generation Quantum DNA Sequencing with Chemical Surface Design and Capsule Nets
- Principal Investigator: Professor Tim Albrecht
- Lead Institution: University of Birmingham
- Project Partner: City, University of London

- Project Title: Deep learning enhanced quantum chemistry: Pushing the limits of Materials Discovery
- Principal Investigator: Dr Reinhard J. Maurer
- Lead Institution: University of Warwick
- Project Partners: University of Strathclyde & Carnegie Mellon University

AI3SD Funding Call 2

Our second funding call will be announced on the 2 August 2019.

As before there will be a Network+ Town Meeting on the 11th September for prospective applicants to ask questions and seek collaborators. We will be making time available at each AI3SD Meeting for project networking, and there will an opportunity to present flash talks at our AI3SD Conference 2019 to seek additional collaborators.

The Key dates are as follows:

- 2 August 2019 – Announce Call on Website
- 19 December 2019 – Application Deadline
- 5 February 2020 – Formal Call Announce
- More information will be posted here in due course: AI3SD-FundingCall2 Page
- AI3SD Publication

Recent Publication

Dr Samantha Kanza & Professor Jeremy Frey had an article accepted into Expert Opinion in Drug Discovery entitled: "A new wave of innovation in Semantic Web Tools for Drug Discovery".

If you have any relevant publications you want highlighted please let us know.

AI3SD Members Events of Interest

We have added a Meetings of Interest page on our website where we have a comprehensive list of meetings that AI3SD Members might be interested in attending.

AI3SD Membership

We now have 328 members on the mailing list which covers individuals from over 30 different academic institutions and over 35 different companies. Membership of this list defines membership of the AI3SD Network. To join the network, please send an email to listserv@jiscmail.ac.uk with the following details:

- Subject: Subscribe
- Message: SUBSCRIBE AI3SD Firstname Lastname

Or go to: https://www.jiscmail.ac.uk/AI3SD

Events and Positions – Cambridge Cheminformatics Network

Contribution from Dr Andreas Bender, email: ab454@cam.ac.uk

CICAG members may be interested in this series of events and vacancies. In the meantime, the next Cambridge Cheminformatics Network Meeting will then be held on 4 September 2019, starting at 4.00 pm in the University's Chemistry Department on Lensfield Road - if you would like to present at this or a future occasion please just let me know.

1. **Events**

2-3 September 2019
2nd RSC-BMCS / RSC-CICAG Artificial Intelligence in Chemistry
Poster applications open until 5 July, Bursary applications open until 15 July
Fitzwilliam College, Cambridge

8 September 2019
"**Solubility Challenge Revisited**" **- solubility datasets are provided, submissions due 8 September 2019**

8 October 2019
ChEMBL@10 one-day symposium

21 October 2019
"**The Application of in silico models to support in vitro studies**"
IVTS One-day Workshop, Lhasa, Leeds

3-5 November 2019
15th German Conference on Chemoinformatics

20 November 2019
20 years of Rule of 5
Sygnature Discovery, BioCity, Nottingham

2. Vacancies in the Cheminformatics/CADD field and beyond

UK
Computational Chemist, Drug Discovery
Nanna Therapeutics, Cambridge

Postdoc Fellow - Deep Learning in Toxicity Prediction
AstraZeneca, Cambridge

Postdoc Fellow - AI to predict drug toxicity from multiomics data
AstraZeneca, Cambridge

Post-doctoral Researcher - Cheminformatics
Optibrium, Cambridge

Scientific Programmer
Healx, Cambridge

Computational Drug Designer
Exscientia, Oxford

Postdoc Position "Characterization of phenotype-genotype relationships in development and cancer at single-cell resolution"
European Bioinformatics Institute/Sanger, Hinxton

Postdoc Position "Analysis of Cancer Genomics Data"
European Bioinformatics Institute, Hinxton

Structure-Based Drug Designer, Fixed-Term Position (with option to extend), starting 9/2019
Cambridge, UK
- Please contact Andreas directly in case of interest

PB/PK Modeller, Fixed-Term Position
Yorkshire, UK
- Please contact Andreas directly in case of interest

Europe
PhD Position "In-silico studies of ion channels and their modulators using computational molecular design"
WWU Muenster, Muenster/Germany

Postdoc Fellow - Use AI to discover novel chemical reactions
Events and Positions - Cambridge Cheminformatics Network

Postdoc Fellow - Use AI to discover novel chemical reactions
AstraZeneca, Gothenburg/Sweden

Principal Scientist Computational Biology
Boehringer Ingelheim, Biberach/Germany

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 3 July 2019*]

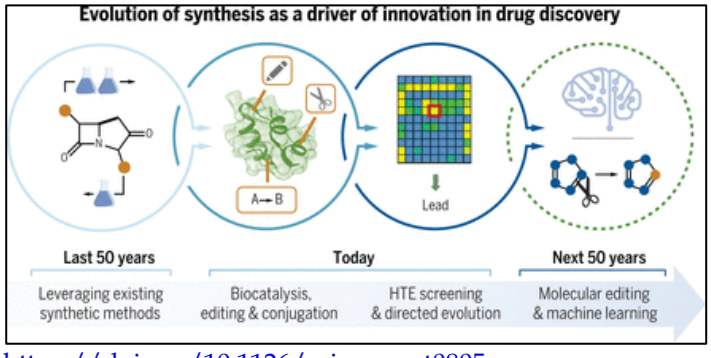
The Research Ecosystem is Changing Rapidly and Profoundly, Elsevier/Ipsos MORI Study Shows
In 10 years' time, the academic world will see new funding models, methods of collaboration, and ways of conceptualizing research and measuring its impact – all driven by advances in technology. A new report, "*Research Futures - drivers and scenarios for the next decade*", lays out three plausible future scenarios showing how the world of research could transform over the next ten years. These are:

- "Brave open world" – considers the rise of open science;
- "Tech titans" – examines the growing influence and dominance of technology and technology companies; and
- "Eastern ascendance" – considers a fragmented world in which China plays a key role.

https://www.prnewswire.com/news-releases/the-research-ecosystem-is-changing-rapidly-and-profoundly-elsevieripsos-mori-study-shows-300795773.html
Source: PR Newswire

Opinion: Martyn Poliakoff, David Cole-Hamilton: IYPT is Good News for Chemistry
Brimming with opportunity, IYPT is putting chemistry and chemists centre stage.
https://www.chemistryworld.com/opinion/iypt-is-good-news-for-chemistry/3010109.article
Martyn Poliakoff - The University of Nottingham
David Cole-Hamilton - University of St Andrews
Source: Chemistry World

The Importance of Synthetic Chemistry in the Pharmaceutical Industry
Merck & Co., Inc. and Janssen Research & Development LLC highlight small-molecule catalysts stimulated by visible light, enzymes engineered for versatility beyond their intrinsic function, and bio-orthogonal reactions to selectively modify proteins for conjugation. High-throughput techniques are also poised to accelerate methods optimization from small-scale discovery to large-scale production, and complementary machine-learning approaches are just coming into focus.

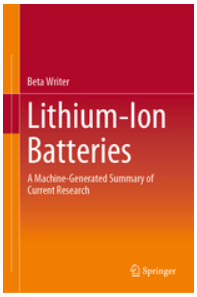


https://doi.org/10.1126/science.aat0805
Source: Science, AAAS

PDFs Dragged into Digital age with 3D Interactive Chemical Structures
Fed up with the limited information 2D figures can convey about complex molecular structures, a team led by Kerry Gilmore at the Max Planck Institute of Colloids and Interfaces in Germany and Nabyl Merbouh at

Simon Fraser University in Canada has developed interactive 3D images that can be easily embedded into PDFs.
https://www.chemistryworld.com/news/pdfs-dragged-into-digital-age-with-3d-interactive-chemical-structures/3010126.article
Source: Chemistry World

Springer Nature Publishes its First Machine-Generated Book



Innovative book prototype provides a compelling machine-generated overview about the latest research on lithium-ion batteries, automatically compiled by an algorithm developed in collaboration with the Applied Computational Linguistics lab of Goethe University Frankfurt/Main.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-machine-generated-book/16590134
Source: Springer Nature

Death of the author
The first machine-generated book in chemistry is here. But is it any good? And what future does it herald?
https://www.chemistryworld.com/opinion/death-of-the-author/3010412.article
Source: Chemistry World

Info Pros and Artificial Intelligence: New White Paper from Springer-Nature
Librarians and information professionals have always been on the cutting edge of information retrieval, management and transformation, and "insight as a service" has always been a staple in their repertoire. Info pros can take an even more prominent role in the coming years in supporting artificial intelligence initiatives within their organisations. In this white paper, Mary Ellen Bates looks at how info pros can help lead the AI discussion and contribute their unique skill set to this transformational technology.
White Paper
Source: Springer-Nature

INPI and CAS Collaborate to Improve Patent Application Examination Efficiency with new AI approach
The National Institute of Industrial Property (INPI) of Brazil and CAS, a division of the American Chemical Society, have initiated a collaboration to speed the processing of patent applications at INPI by improving the efficiency of the patent examination process. CAS will apply data from its human-curated content collection in a new artificial intelligence (AI) solution. The technology combines scoring similarity, workflow classification and knowledge-space exploration capabilities to uncover a reliable set of prior art, or previously published information relevant to the subject innovation, which expedites patent examiners' work.
https://www.cas.org/resources/press-releases/cas-inpi-collaboration
Source: CAS

How Researchers Discover, Read and use Scholarly Literature: International Trends
A new international study explores how researchers find and use scholarly literature for their work and finds differences by age, discipline and geography.
https://www.infotoday.eu/Articles/Editorial/Featured-Articles/How-researchers-discover-read-and-use-scholarly-literature-international-trends-131404.aspx
Source: Information Today

Computational Chemists Welcome 'Living' Journal
Journal will publish educational papers and allow continual updates.
https://doi.org/10.1021/cen-09701-scicon3
Source: CEN

Google Assistant now comes with a real-time translator for 27 languages
https://www.technologyreview.com/the-download/612730/google-assistant-now-comes-with-a-real-time-translator-for-27-languages/
Source: MIT Technology Review

Clarivate Analytics Joins Accenture's Life Sciences Ecosystem to Help Accelerate Innovation in Drug Discovery and Scientific Research

Clarivate Analytics, the global leader in providing trusted insights and analytics to accelerate the pace of innovation, has joined Accenture's open partner ecosystem, which is designed to help independent software vendors (ISVs), specialized content providers, and life sciences companies team more effectively to advance drug discovery efforts and improve patient outcomes.

https://www.prnewswire.com/news-releases/clarivate-analytics-joins-accentures-life-sciences-ecosystem-to-help-accelerate-innovation-in-drug-discovery-and-scientific-research-300775817.html

Source: PR Newswire

Google Scientist: AI may not be 'cure-all for Discovering New Drugs

Promises that Artificial Intelligence will create life-saving drugs faster than humans and healthcare as may be overblown, one of Google's top computer scientists has suggested.

https://www.telegraph.co.uk/technology/2019/01/13/google-scientist-ai-may-not-cure-all-discovering-new-drugs/

Source: Daily Telegraph

Not just for bots: The Changing Career Landscape in AI

Artificial intelligence (AI) is impacting science in new and exciting ways as researchers are- using it to better understand society and find solutions to problems across diverse disciplines; and as the application of AI expands, so too do the career opportunities.

https://www.sciencemag.org/features/2018/11/not-just-bots-changing-career-landscape-ai

Source: Science, AAAS

Moving Beyond Common Myths to Deliver Real Value with Artificial Intelligence

- Myth 1: Artificial intelligence can make business decisions for you
- Myth 2: Every business challenge can be helped by artificial intelligence
- Myth 3: Artificial intelligence can process any and all available data

https://www.cas.org/blog/moving-beyond-common-myths-deliver-real-value-artificial-intelligence

Source: CAS

Wiley Strikes Open Access Deal With German Universities and Libraries

German academic institutions have reached a 'ground breaking' nationwide deal to allow their researchers to make their work freely available around the world in journals published by Wiley – at no extra cost. For an agreed annual fee they'll also have access to all Wiley's content back to 1997.

https://www.chemistryworld.com/news/wiley-strikes-open-access-deal-with-german-universities-and-libraries/3010008.article

Source: Chemistry World

Novartis and the University of Oxford's Big Data Institute to establish world-leading research alliance using artificial intelligence to understand complex diseases and improve drug development

- Unique five-year alliance established between industry and academia to advance the analysis and interpretation of ultra large and multiple datasets, using cutting edge technology including artificial intelligence machine learning
- These technologies have the ability to spot disease patterns and signals much earlier in their manifestation, translating more quickly into methods that will support clinical decision making
- The alliance will be transformative for how Novartis adopts data science at scale from multiple data sources, beginning with flagship programs in multiple sclerosis, dermatology and rheumatology

https://www.novartis.co.uk/news/media-releases/novartis-and-university-oxford%27s-big-data-institute-establish-world-leading

Source: Novartis UK

Leading Global CRO/CDMO AMRI Commits to 5-year Agreement for SciFinder from CAS to Enable Efficient Research and Client Success

As pharmaceutical companies lean more heavily on outsourcing to advance their pipeline, leading contract research, development and manufacturing organizations need the best-in-class scientific information solution to stay ahead of the curve.

https://www.prnewswire.co.uk/news-releases/leading-global-cro-cdmo-amri-commits-to-5-year-agreement-for-scifinder-from-cas-to-enable-efficient-research-and-client-success-855888324.html

Source: PR Newswire

67 Bricks help the Royal Society of Chemistry streamline its Open Access system
The open access (OA) landscape is changing rapidly and 67 Bricks' client the Royal Society of Chemistry wanted to improve and speed up the way it managed its OA processes and customer experience. The Royal Society of Chemistry brought in 67 Bricks to help buiild a new, automated OA system that would be flexible and robust enough to grow alongside their Open Access programme, evolve with the changing landscape and support staff and external users simultaneously.
http://www.67bricks.com/index.php/all-news/171-rsc-open-access-system
Source: 67 Bricks

Berlin-based Market Intelligence Platform IPlytics Raises Millions to Disrupt Patent Analytics
With over 500,000 new patent applications every month, companies have their work cut out for them when it comes to keeping track of new technologies and their competition – and patent disputes between companies can put entire industries in gridlock.
Berlin-based startup IPlytics attempts to solve this problem with its market intelligence platform, which maps multiple data sources to allow companies to search and analyse technology, their competition, market landscapes, and patents.
https://www.eu-startups.com/2019/03/berlin-based-market-intelligence-platform-iplytics-raises-millions-to-disrupt-patent-analytics/
Source: EU-Startups

BINOPtimal: a web tool for optimal chiral phosphoric acid catalyst selection
A catalyst selection program, BINOPtimal, has been developed. This interactive web tool selects the best performing chiral phosphoric acid catalysts from analysis of the starting materials, imine and nucleophile, on the basis of rules derived from the transformations within its database. This procedure has been applied to an example transformation demonstrating the potential to assist reaction design.
https://pubs.rsc.org/en/content/articlelanding/2019/cc/c8cc09344j#!divAbstract
Source: Chemical Communications

EduDataHub Launches Market Intelligence Solution for Higher Education Academic Publishing
EduDataHub, a data-centric market research firm, announces the release of Academic Publishing Dashboard, the only comprehensive market intelligence solution for an estimated $3.4 billion U.S. higher education publishing industry. Leveraging EduDataHub's database of more than two decades of textbook adoptions and sales data, the solution is designed to help publishers, Ed-Tech and other companies operating within the space to make informed decisions in matching their products to courses offered in colleges and universities.
https://www.prweb.com/releases/edudatahub_launches_market_intelligence_solution_for_higher_education_academic_publishing/prweb16033846.htm
Source: PR Web

BASF is Combining Digitalization and IT Expertise in a Single Division
The "Digitalization & Information Services" division aims to accelerate the digital transformation at BASF.
https://www.basf.com/global/en/media/news-releases/2019/01/p-19-126.html
Source: BASF

App Predicts Pharma Process Waste
To help pharma chemists find less wasteful routes to their products, members of the American Chemical Society's Green Chemistry Institute have developed a web-based tool to predict the efficiencies of syntheses.
https://pubs.acs.org/doi/10.1021/cen-09704-scicon5
Source: Chemical & Engineering News

Copyright Clearance Center Announces RightFind® Suite for Emerging Life Science Companies
This content workflow solution provides fast, easy access to a full range of published content – anytime, anywhere – providing a more efficient research process and competitive advantage.
https://www.businesswire.com/news/home/20190129005193/en/
Source: Business Wire

Institute for Scientific Information Launches Global Research Report – Profiles not Metrics
The Global Research Reports are a new series of publications aimed at those who deal with research in academia, corporations, publishers and governments to inform, to stimulate debate and to demonstrate the rich information potential of research data.
https://clarivate.com/blog/news/institute-for-scientific-information-launches-global-research-report-profiles-not-metrics/
Source: Clarivate

Kumsal Bayazit appointed Chief Executive Officer of Elsevier
https://www.elsevier.com/about/press-releases/corporate/kumsal-bayazit-appointed-chief-executive-officer-of-elsevier
Source: Elsevier

How to get your work on the Cover of a Journal
First, do some great science, obviously. Next, take some advice from scientific photographer Felice Frankel and learn to take an exceptional photo with only a mobile phone or a flatbed scanner. Kelly Krause, creative director of Nature and the Nature journals, walks you through some of the stunning images and top tips from Frankel's new book.
https://www.nature.com/articles/d41586-019-00349-w
Source: Nature

ARL Initiative to Recruit a Diverse Workforce Renamed Kaleidoscope Program
The name Kaleidoscope signifies responsiveness to fluid landscapes, needs, and cultures, which is an important element of diversity, equity, and inclusion in research libraries and in society more broadly. Much like a kaleidoscope changes as it is moved, reflecting diverse colors, shapes, and perspectives, the Kaleidoscope Program ensures the evolution of the library and archives profession to reflect and embrace the diverse reality of the communities we serve.
https://www.arl.org/news/arl-news/4716-arl-initiative-to-recruit-a-diverse-workforce-renamed-kaleidoscope-program#.XFLYfFUzbIV
Source: Association of Research Libraries

Solving global challenges starts with open data -- new report launched today
Better incentives for researchers and fewer barriers between technological systems are key to kickstarting a revolution in open data, according to 'Realising Potential,' a report released today by the Open Research Data Task Force.
https://www.eurekalert.org/pub_releases/2019-02/uow-sgc020419.php
Source: EurekAlert! AAAS

Thousands of scientists run up against Elsevier's paywall
Researchers have been left without access to new papers as libraries and the major publisher fail to agree on subscription deals.
https://www.nature.com/articles/d41586-019-00492-4
Source: Nature

How Synthetic Chemistry can Inspire Drug Design
Within Merck & Co., Inc. and in contrast to the actions of many other pharma companies, we promote innovation in synthetic chemistry as being critical to success in drug discovery and development. Furthermore, new developments in synthetic methods, biocatalysis, chemoinformatics, and reaction miniaturization empower the speed and quality of discovery...
https://www.cas.org/blog/how-synthetic-chemistry-can-inspire-drug-design
Source: CAS

AI is Reinventing the way we Invent
The biggest impact of artificial intelligence will be to help humans make discoveries we couldn't make on our own.
https://www.technologyreview.com/s/612898/ai-is-reinventing-the-way-we-invent/
Source: MIT Technology Review

Rare Trial of Open Peer Review Allays Common Concerns
Study suggests that making reviewers' reports freely readable doesn't compromise peer-review process.
https://www.nature.com/articles/d41586-019-00500-7
Source: Nature

Physics societies warn of 'irrecoverable damage' from European open-access plan
The European Physical Society (EPS) has warned that a major open-access initiative in Europe could cause "irrecoverable damage" if it is implemented too quickly. In a statement, the EPS says that while it welcomes the proposal – known as Plan S — as a "medium to long-term vision", its proponents must get more support by engaging further with the scientific community.
https://physicsworld.com/a/physics-societies-warn-of-irrecoverable-damage-from-european-open-access-plan/
Source: PhysicsWorld

Indian payment-for-papers proposal rattles scientists
Researchers say the policy could intensify existing issues with research quality and misconduct.
https://www.nature.com/articles/d41586-019-00514-1
Source: Nature

BASF's investment into Alchemist Accelerator promotes digital innovations in the chemical industry
- Alchemist will fund digital startups driving growth of the chemical industry
- BASF accelerates its outreach towards technologies like artificial intelligence, internet of things and robotics

https://www.basf.com/global/en/media/news-releases/2019/02/p-19-139.html
Source: BASF

EPA Releases First Major Update to Chemicals List in 40 Years
A key result of the update is that less than half the total number of chemicals on the current TSCA Inventory are currently in commercial use. As the result of a tremendous effort on behalf of thousands of stakeholders and manufacturers from across the country, this information will help EPA focus risk evaluation efforts on chemicals that are still on the market.
https://www.epa.gov/newsreleases/epa-releases-first-major-update-chemicals-list-40-years
Source: EPA News Releases

Big pharma is embracing open-access publishing like never before
Proportion of open-access publications with authors from the pharmaceutical industry doubled between 2009 and 2016.
https://www.nature.com/articles/d41586-019-00610-2
Source: Nature

Huge US University Cancels Subscription with Elsevier
University of California system and Dutch publisher fail to strike deal that would allow researchers to publish under open-access terms.
https://www.nature.com/articles/d41586-019-00758-x
Source: Nature

ResearchGate and Springer Nature embark on pilot to deliver seamless discovery and an enhanced the reading experience
Full-text articles published in select Nature journals since November 2017 will be rolled out to researchers' ResearchGate profiles, making it easier to read or download research on or off campus.
https://group.springernature.com/gp/group/media/press-releases/researchgate-and-springer-nature-embark-on-pilot/16518868
Source: Springer Nature

Twitter chemistry poster conference continues to grow as it celebrates its fifth anniversary
Run entirely on social media, researchers can present virtual posters in 12 different categories, ask questions and network – all without having to leave the comfort of their office or home.

https://www.chemistryworld.com/news/twitter-chemistry-poster-conference-continues-to-grow-as-it-celebrates-its-fifth-anniversary/3010177.article
Source: Chemistry World

Peer-review Experiments Tracked in Online Repository
ReimagineReview records trials that are probing the pros and cons of different approaches to review.
https://www.nature.com/articles/d41586-019-00777-8
Source: Nature

Data Quality: The not-so Secret Sauce for AI and Machine Learning
If AI or machine learning algorithms aren't living up to your expectations, could data be the culprit?
https://www.cas.org/blog/data-quality-not-so-secret-sauce-ai-and-machine-learning
Source: CAS

Global Chemoinformatics Market to Grow at Approx 18.7% During 2019-2024
Growth Propelled by Innovations & Advancements in the Drug Development Process and Increased Demand of Personalised Medicine.
https://www.businesswire.com/news/home/20190312005679/en/Global-Chemoinformatics-Market-Grow-Approx-18.7-2019-2024
Source: Business Wire

£79m Supercomputer set to Boost UK Capability
The supercomputer is hundreds of thousands of times more powerful than a traditional desktop computer and will be used to run massive research simulations.
https://www.ed.ac.uk/news/2019/ps79m-supercomputer-set-to-boost-uk-capability
Source: University of Edinburgh

New investment in European Bioinformatics Institute to tackle life threatening diseases
European Bioinformatics Institute in Cambridge receives £45 million government investment to increase centre's computing, storage and building capacity.
https://www.gov.uk/government/news/new-investment-in-european-bioinformatics-institute-to-tackle-life-threatening-diseases
Source: Gov.uk

Reprints Desk Announces Launch of Article Galaxy Gadget Store
Powered by Reprints Desk's Award-Winning Research Platform, the Gadget Store Introduces a Novel Ecosystem of Science Apps Designed for Individual Researchers.
https://info.reprintsdesk.com/about/newsroom/reprints-desk-announces-launch-of-article-galaxy-gadget-store
Source: Reprints Desk

The Marriage of Big Pharma and Biotech
https://www.rdmag.com/article/2019/03/marriage-big-pharma-and-biotech
Source: R&D Magazine

Drawing on experience to help scientists tell their stories
Well-designed illustrations can help to clarify research in publications, grant applications and presentations. But reducing research to a simple visual representation is a skill few scientists are taught.
Popular tools for creating images, such as Microsoft PowerPoint and Adobe Illustrator, were not designed with scientific research in mind. As a result, even seemingly simple tasks — drawing a biochemical pathway, for instance — can take hours of painstaking work. Enter BioRender, a browser-based illustration programme designed specifically for scientific illustration.
https://www.natureindex.com/news-blog/drawing-on-experience-to-help-scientists-tell-their-stories
Source: Nature Index

Altmetric and Nature awarded funding from the Google Digital News Innovation Fund
Data science company Altmetric and leading multidisciplinary science journal Nature will receive funding from the Google Digital News Innovation Fund to build a novel tool for measuring the impact of journalism.
https://www.digital-science.com/press-releases/altmetric-and-nature-awarded-funding-from-the-google-digital-news-innovation-fund/

Source: Digital Science

Interface Reactions on the Materials Project
The Interface Reactions app, is an application that calculates possible interface reactions between solids using thermodynamic driving forces.
https://materialsproject.org/static/newsletter/announce-interfacerxns.html
Source:

It's back: Chemical e-commerce tries for an act 2
Almost 2 decades after a series of flops, new web portals are emerging for online buying and selling.
https://cen.acs.org/business/informatics/s-back-Chemical-e-commerce/97/i12
Source: Chemical & Engineering News

Machine Learning Reveals Rapid Material Classification
A research team at The University of Tokyo has developed a powerful machine learning algorithm that predicts the properties and structures of unknown samples from an electron spectrum.
https://www.eurekalert.org/pub_releases/2019-03/iois-mlr032519.php
Source: EurekAlert! Science News

Top UK funders don't support most highly cited health researchers
https://www.chemistryworld.com/news/top-uk-funders-dont-support-most-highly-cited-health-researchers/3010293.article
Source: Chemistry World

Plagiarism detectors are a crutch, and a problem
Academics and editors need to stop pretending that software always catches recycled text and start reading more carefully.
https://www.nature.com/articles/d41586-019-00893-5
Source: Nature

American Chemical Society and Max Planck institutes partner on transformative open access plan
The four-year transformative agreement provides researchers affiliated with Max Planck institutes the opportunity to disseminate immediately, under an open access license, 100 percent of their research articles upon acceptance and publication by a peer-reviewed ACS journal. Max Planck researchers also benefit from full reader access to all ACS Publications journals and Chemical & Engineering News.
https://www.acs.org/content/acs/en/pressroom/newsreleases/2019/march/acs-and-max-planck-institutes-partner-on-transformative-open-access-plan.html
Source: American Chemical Society

Predatory publisher OMICS fined $50 million by US court for defrauding scientists
The publisher and conference organiser OMICS Group has been ordered by a federal court in the US to pay $50.1 million (£38.3 million) to the Federal Trade Commission (FTC) for deceptive and predatory publishing activities.
https://www.chemistryworld.com/news/predatory-publisher-omics-fined-50-million-by-us-court-for-defrauding-scientists/3010348.article
Source: Chemistry World

Future Science Group launches new Initiative, PharmaPlatforms
Taking published content from FSG's journals, knowledge networks and content hubs, PharmaPlatforms will provide a range of solutions that enrich the information presented, engage the reader, and reach a broad scientific and medical professional audience.
https://www.future-science-group.com/new-initiative-pharmaplatforms
Source: Future Science

Digital Science welcomes Gigantum and Ripeta to the family to help increase reproducibility in research
Both companies are playing a key role in making scientific research reproducible and more transparent. Ripeta is developing a "credit report" for scientific publications to assess and help improve the transparency needed to effectively communicate research, while Gigantum is an emerging innovator in the area of data science platforms, supporting large-scale, data-oriented scientific research.

https://www.digital-science.com/press-releases/digital-science-welcomes-gigantum-and-ripeta-to-the-family-to-help-increase-reproducibility-in-research/
Source: Digital Science

Springer Nature and DrugPatentWatch partner to provide patent information for drug database
A new module focusing on patents has been added to the drug pipeline tool AdisInsight. The module allows users to find out which drugs have been patented and when regulatory protection may expire.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-and-drugpatentwatch-partner/16648806
Source: Springer Nature

PerkinElmer Joins Accenture's Life Sciences Ecosystem to Help Drive Innovation in Drug Discovery and Scientific Research
The ecosystem is an integral part of Accenture's cloud-based informatics research platform focused on improving productivity, efficiency and innovation in the drug discovery process.
https://www.businesswire.com/news/home/20190416005541/en/
Source: Business Wire

Three-year trial shows support for recognizing peer reviewers
Thousands of Nature referees have chosen to be publicly acknowledged.
https://www.nature.com/articles/d41586-019-01162-1
Source: Nature

Finding cures faster: AI charting the course into new regions of chemical space
An analysis of CAS REGISTRYSM shows that the pace of innovation is accelerating as pharmaceutical companies race to discover new drugs in the uncharted regions of a dauntingly large chemical cosmos.
https://www.cas.org/blog/finding-cures-faster-ai-charting-course-new-regions-chemical-space
Source: CAS

BASF invests in Quantum Computing Startup Zapata Computing
Zapata's technology platform is an innovative approach to leverage the potential of this novel technology Applications of quantum computing offer opportunities to meet the challenges of the chemical industry.
https://www.basf.com/global/en/media/news-releases/2019/04/p-19-178.html
Source: BASF

Mathematica 12 now has chemistry support
The new release of Mathematica version 12 includes many additional features, including some Chemistry developments.
https://www.macinchem.org/blog/files/b58d7e4e9604f31d2e4496210f4a9f52-2454.php
Source: Macs in Chemistry

Healthcare Blockchain Technology Market to hit $1.6bn by 2025
https://pharmaphorum.com/partner-content/healthcare-blockchain-technology-market-will-reach-usd-1-6-bn-by-2025/?utm_sq=g275tqufdz
Source: Pharmaphorum

Farewell to "Watson For Drug Discovery"
STAT is reporting that IBM has stopped trying to sell their "Watson for Drug Discovery" machine learning/AI tool, according to sources within the company.
https://blogs.sciencemag.org/pipeline/archives/2019/04/18/farewell-to-watson-for-drug-discovery
Source: In the Pipeline, AAAS

Elsevier agrees to first read-and-publish deal
A Norwegian consortium has signed a new kind of subscription deal with Elsevier that includes open-access publishing -- a first for the publisher. But the new rights come at a cost.
https://www.insidehighered.com/news/2019/04/24/elsevier-agrees-first-read-and-publish-deal
Source: Inside Higher Ed

Automation: Chemistry shoots for the Moon
A new class of chemical instrumentation seeks to alleviate the tedium and complexity of organic syntheses.
https://www.nature.com/articles/d41586-019-01246-y
Source: Nature

AI for Molecular Design
Artificial intelligence is set to speed up the development process for new pharmaceuticals.
https://www.wsj.com/articles/ai-for-molecular-design-11556639110
Source: Wall Street Journal

Dundee professor in line for £400m payday from Exscientia flotation
Dundee University spin-off Exscientia, which uses AI to help find blockbuster pharma drugs, is considering an initial public offering, and will be ready to float by next May.
https://www.telegraph.co.uk/technology/2019/05/05/dundee-professors-could-line-500m-future-exscientia-debut/
Source: Daily Telegraph

Springer Nature seeks views on new approach aimed at accelerating the transition to open access
Springer Nature sets out a new approach to research publishing with the goal of increasing the demand from authors to immediately publish their research open access (OA) and growing the supply of journals able to publish OA, potentially enabling even highly selective journals such as Nature to transition to OA.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-seeks-views-on-new-approach/16707864
Source: Springer Nature

LANXESS to deploy artificial intelligence in product development
Much faster development of high-performance plastics thanks to partnership with materials AI leader Citrine Informatics.
https://lanxess.com/en/corporate/media/press-releases/2019-00029e/
Source: LANXESS

Everything you ever wanted to know about global AI patent activity
WIPO released its first of a new series, Technology Trends, earlier this year, taking a deep-dive into the artificial intelligence patent landscape.
https://www.iam-media.com/litigation/everything-you-ever-wanted-know-about-global-ai-patent-activity
Source: IAM

Artificial Intelligence and Drug Discovery
Artificial Intelligence (AI) may have some interesting consequences for the patentability of new drug substances.
https://www.vennershipley.co.uk/resources/publications/2019/04/09/artificial-intelligence-and-drug-discovery
Source: Venner Shipley

MarkLogic Launches Pharma Research Hub to Accelerate Drug Research and Results
A fully managed cloud data hub that speeds up the drug discovery process and fuels the use of artificial intelligence.
https://www.businesswire.com/news/home/20190514005104/en/
Source: Business Wire

American Chemical Society introduces reimagined publications web experience, powered by Atypon
The new web platform delivers global access to all ACS journals and books, as well as to the C&ENGlobal Enterprise edition. The website supports over 6,000 universities, corporations and government research centers worldwide, hosts over 1.4 million peer-reviewed articles and delivers over 130 million article requests per year.
https://www.atypon.com/news/american-chemical-society-introduces-reimagined-publications-web-experience-powered-by-atypon/
Source: ACS

Cyclica Launches Ligand DesignTM, a Powerful Multi-Targeted Drug Design Technology at Collision Conference Toronto
Cyclica, a leading biotechnology company that commercialized Ligand Express® announced its novel, first-in-class drug design technology, Ligand DesignTM on stage at the Collision Conference in Toronto. This revolutionary platform empowers scientists in the pharmaceutical industry to generate novel drug-like compounds by combining powerful methods for molecule generation, ADMET prediction, and high-throughput ligand-protein interaction prediction.
https://www.businesswire.com/news/home/20190522005672/en/Cyclica-Launches-Ligand-DesignTM-Powerful-Multi-Targeted-Drug
Source: Business Wire

CAS reaches 150 millionth substance
Carbonitrile derivative is a promising compound for the treatment of cancer and immunological diseases.
https://cen.acs.org/acs-news/programs/CAS-reaches-150-millionth-substance/97/web/2019/05
Source: Chemical & Engineering News

From cat photos to self-driving cars: Imperial's power boost for clever machines
Imperial researchers will lead a new international EPSRC Centre to develop pioneering approaches to machine learning.
https://www.imperial.ac.uk/news/191368/from-photos-selfdriving-cars-imperials-power/
Source: Imperial College London

McGraw-Hill and Vantage Learning Collaborating to Create Artificially Intelligent Automated Writing Feedback Tool
New Technology Adds Another Powerful Capability to McGraw-Hill Digital Course Materials for College Students.
https://www.mheducation.com/news-media/press-releases/mcgraw-hill-vantage-learning-create-ai-writing-tool.html
Source: McGraw-Hill

SOCMA Publishes Updated Handbook for Toxic Substances Control Act (TSCA) Compliance
First published in 1995, the revised handbook has been used by facility, operations, regulatory and environmental management teams as the definitive guide for not only regulatory implementation, but also internal reviews and audit preparation.
https://www.prnewswire.com/news-releases/socma-publishes-updated-handbook-for-toxic-substances-control-act-tsca-compliance-300860659.html
Source: Cision PR Newswire

Researchers warn open access Plan S may still be too rushed, despite one-year delay
Updated guidelines from cOAlition S funders draw mixed reactions from scientists and publishers.
https://www.chemistryworld.com/news/researchers-warn-open-access-plan-s-may-still-be-too-rushed-despite-one-year-delay/3010576.article
Source: Chemistry World

The Web of Science Group Launches New Publisher Analytics Reports
The ready-to-use reports provide a snapshot of journals' performance and inform competitor analysis.
https://clarivate.com/blog/science-research-connect/the-web-of-science-group-launches-new-publisher-analytics-reports/
Source: Clarivate

Craft beautiful equations in Word with LaTeX
Manufacturers are ditching equation editors in word-processing software in favour of the LaTeX typesetting language. Here's how to get started.
https://www.nature.com/articles/d41586-019-01796-1
Source: Nature

'Broken access' publishing corrodes quality
Funders should award competitive grants directly to journals to underwrite the costs of open access, urges Adriano Aguzzi.
https://www.nature.com/articles/d41586-019-01787-2
Source: Nature

Aurora Supercomputer to Empower Advanced Chemistry Research
The Aurora Supercomputer will arrive at the Argonne Leadership Computing Facility (ALCF) in 2021.
https://www.technologynetworks.com/informatics/articles/aurora-supercomputer-to-empower-advanced-chemistry-research-320550
Source: Technology Networks

Group targets a platform for collaborative data management in drug discovery
A consortium of pharmaceutical companies, academic researchers, and technology suppliers has announced a 3-year, $20 million project to apply machine learning and blockchain security to data management in drug development.
https://cen.acs.org/business/informatics/Group-targets-platform-collaborative-data/97/i23?PageSpeed=noscript
Source: Chemical & Engineering News

Elsevier Announces the International Center for the Study of Research
A groundbreaking new Center tasked with examining and advancing the evaluation of research across all fields of knowledge production has been launched by Elsevier, the information analytics business specializing in science and health. The International Center for the Study of Research (ICSR) will work closely with the research community to review and develop the use of qualitative and quantitative metrics of research evaluation to create a more transparent and robust approach to research assessment.
https://www.prnewswire.com/news-releases/elsevier-announces-the-international-center-for-the-study-of-research-300871135.html
Source: Cision PR Newswire

The Web of Science Group Releases 2019 Journal Citation Reports, Revealing the World's Most Influential Journals
The 2019 release aggregates the meaningful connections of citations created by the research community through the delivery of a rich array of publisher-independent data, metrics and analysis of the world's high-quality academic journals.
https://www.prnewswire.com/news-releases/the-web-of-science-group-releases-2019-journal-citation-reports-revealing-the-worlds-most-influential-journals-300871565.html
Source: Cision PR Newswire

Springer Nature launches new video portfolio
Springer Nature Video offers insights, research and training material from experts/ Medicine and Professional and Applied Computing will be the first video subject collections.
https://group.springernature.com/gp/group/media/press-releases/springer-nature-launches-new-video-portfolio/16825682
Source: Springer Nature