# NEWSLETTER summer 2018

Above: Peter Willett (left) & Johann Gasteiger; see pages 1sbCI 1

## Table of Contents


*Chemical Information & Computer Applications Group Chair's Report* 3
*CICAG Planned and Proposed Future Meetings* 3
*The Formation of the CICAG* 4
*CICAG Committee News* 5
*Catalyst's new gallery - Harry Baker and 120 years of Chlorine!* 6
*The 8th Capps Green Zomaya Award* 7
*Tony Kent Strix Award and Annual Lecture 2018* 9
*Undergraduate Research Bursaries – CICAG Sponsorship in 2018* 11
*Chemical Information / Cheminformatics and Related Books* 12
*Meeting Report: Dial-a-Molecule Conference* 13
*Meeting Report: ICCS 2018* 14
*Meeting Report: Artificial Intelligence in Chemistry Meeting* 15
*Professor Peter Willett – a Celebration of his Research* 19
*InfoChem News* 20
*CAS / SciFinder / STN News* 21
*Other Chemical Information Related News* 23


Contributions to the CICAG Newsletter are welcome from all sources - please send to the Newsletter Editor:
**Stuart Newbold**, email: stuart@psandim.com

## Chemical Information & Computer Applications Group Chair's Report

This has been a busy few months for CICAG with a number of changes behind the scenes. Frank Longford, an EPSRC Doctoral Prize Research Fellow at the University of Southampton, has been co-opted onto the committee to co-ordinate the social media channels. Both the Twitter feed (https://twitter.com/RSC_CICAG) and LinkedIn page (https://www.linkedin.com/groups/1989945) now have more regular submissions and I encourage you to join in the convu tooedptot61(to)er4espB9(n)3(v)n yon co
members attending conferences to provide live Twitter updates, with Nthan Brown providing updates from the recent International Conference on Chemical Structures at Noordwijkerhout. This provided a con yuit for sciencosts not /S6c on m-9(n)-4(n)-4(ti)-11(n)3(g)1( to)-9( a)5(s)-11(k)1( q)-7(un)-4(s)1(t)--3(o)7(edp)50-0 from o

CICAG Committee os delioghted to an7(edp)5ounce toDr Samantha Hughes has agreed to join the Committee as a co-opted member-0 0 Samantha is currently an Associate Director at AstraZeneca in Cambridge and brings a wealth of experience in computational chemistry and cheminformatics. She is particularly keen to support

| Meeting | Date | Location | Further Information |
|---|---|---|---|
| Spectroscopic Data to Chemical Knowledge | 2018 | AstraZeneca, Macclesfield | To be organised jointly with the |

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

---

## CICAG Committee News

**New CICAG Committee Members**

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

Welcome, Samantha.

The Catalyst Science Discovery Centre and Museum in Widnes, Cheshire, the only UK museum celebrating the history of the chemical industry, has been awarded a grant of £36,159 to create a new permanent exhibition entitled ‘ ’ which is being funded by an AIM Biffa Award, through their History Makers Project. The new exhibition in a special Baker Gallery in Catalyst will celebrate the life and work of little-known local man but distinguished chemist, Dr Harry Baker (1859-1935), who was a student of two internationally famous chemists Professor Sir Henry Roscoe FRS in Manchester and Professor Robert Bunsen in Heidelberg. It will concentrate on his amazing work from 1897 onwards, (when he came with his family to Runcorn), on the development of chs39h10(me04 Tc(s)-7rTf7,7te)-7(2(R)1e)-7((i)-2(t)-3(i)t)-3(ohi

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

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

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

Full programme details:

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

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

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

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

---

## Undergraduate Research Bursaries – CICAG Sponsorship in 2018

A number of Undergraduate Research Bursaries, funded by the Royal Society of Chemistry are usually available to students in UK and the Republic of Ireland Chemistry and related departments from June-September each year. The purpose of the awards is to give experience of (S)-e

aYjEMC/P <MCID 30 BDC 0.003 Tc 0.1931 TD[(S1.241 TD[(S)-2(A)2(er)-1(n)7(t)4(s)5(s)5(t)7(n)-1p)2stne a-tonr1

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

Wiley Publishing
Editors: Thomas Engel, Johann Gasteiger

---

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

The Dial-a-Molecule conference on Predicting Reaction Outcomes was held at the Bradfield Centre on the Cambridge Science Park on Monday, 26 March, 2018. I

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

Robert Phipps (University of Cambridge) gave an account of how an experimentalist discovers new reactions, with examples from his work on non-covalent interaction controlled catalysis and enantioselective Minisci-type additions. Elizabeth Krenske (University of Queensland) then described how a superficially simple reaction, a base-catalysed silylation, required a worldwide consortium to investigate the mechanism, and then the mechanisms, of this highly complex process, and illustrated the care which is needed to draw conclusions about detailed reaction mechanisms. Ola Engkvist, from AstraZeneca, reported on the possibilities of automating reaction development and the opportunities that this gives for rapid and consistent data acquisition; these data, combined with data from publishers and open sources of information,

A highlight for me was the talk from Dr Greg Landrum (KNIME, but also the lead developer of RDKit), talking on QSAR modelling en-masse: “

.” [slides]. Greg used the KNIME Model Factory to test out many different combinations of modelling algorithms and molecular descriptors. In total, Greg generated 310,000 models in his workflow and reported key learnings from his work. Many of his models appeared very good in terms of model statistics, but on further introspection he noted that it is likely that these were either through overfitting or simply learning to discriminate the papers from which the data were derived. Greg’s mantras throughout the talk were:

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

A highlight from the conference for me was the touching celebration to Prof. Peter Willett of the University of Sheffield on reaching forty years of research in the field of Chemoinformatics from longstanding collaborators and friends Prof. Val Gillet and Dr Bob Clark. I may be biased as one of Peter’s many PhD students (current count is 75!), but I believe it is fair to say that if you do not know who Peter is, or have never read one of his papers, then you probably have to wonder what it is you actually do. Thankfully, Bob was able to stitch together his video tribute with Val’s live speech in Peter’s honour, and made the video available hereh0 0 9.96 118.2 5 refC2_0 1 Tf me wM001 -7(e w37(e)-7(s)-3o)-10-7(e)-7((e)-7(.003 Tc(wn-0.d)-5( )1(D)1e)-7(Sig

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

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

Presentation Summaries from Bursary Awardees

Speaker: **Marwin Segler, BenevolentAI**
Presentation: **Self-Driving Synthesis Planning**
Report by: **Po-Chang Shih, University College London**
The great talk given by Dr Marwin Segler clearly gives a broad view of current progress of AI technology in the field of retrosynthesis. In the beginning of the talk, a story about the challenge of developing AI in retrosynthesis was introduced – encoding ~20,000 synthetic chemistry rules took 10 years. This is absolutely startling, as there are apparently more than she(t)4(a)-3(rts)5(yi)5(n)-8-1(r)s4(s)5(r)-12rros0.186 Tw 10.193 0 Td[(3 Tw3

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

Speaker:

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

Fig. 2. How to Link from Compound to Literature

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

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

---

## CAS / SciFinder / STN News

## Other Chemical Information Related News

[All hyperlinks correct & working as of 9 July 2018]

Impact Factor Creator and Chemical Information Pioneer Eugene Garfield Honoured

Eugene Garfield was a pioneer, who in the 1950s started creating information tools such as the then ground-breaking Science Citation Index. Garfield, who died in 2017, founded the Institute for Scientific Information (ISI), and went on to develop a number of tools which helped transform the search and retrieval of scientific and chemical information.

https://cen.acs.org/articles/96/i15/Impact-factor-creator-chemical-information.html

Source: Chemical & Engineering News

Computational Chemistry: A Rising Tide of Women

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

https://pubs.acs.org/doi/10.1021/acs.jcim.8b00170

Source: Journal of Chemical Information and Modeling

ChemMaps Lets Researchers Navigate the Chemical Universe

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

https://news.ncsu.edu/2018/06/fourches-chemmaps/

Source: NC State University

ChemMaps: Tool to navigate chemical space aims to be 'Google Maps of drugs'

https://www.chemistryworld.com/news/tool-to-navigate-chemical-space-aims-to-be-google-maps-of-drugs/3009145.article

Source: Chemistry World

The Institutionalized Racism of Scholarly Publishing

Academic Journals, and the tag of 'predatory publishing'.

https://awayofhappening.wordpress.com/2018/06/09/the-institutionalized-racism-of-scholarly-publishing/

Source: Wordpress

The battle behind the periodic table's latest additions

Four new elements were added in 2015 with great fanfare — but some researchers complain the announcement was premature.

https://www.nature.com/articles/d41586-018-05371-y

Source: Nature

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

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

http://newsroom.wiley.com/press-release/all-corporate-news/wiley-4ut1-10id4(a)4(n)2(y.98 0 T26h)5

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

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

Info graphic Shows How Fair Use Promotes Creation of New Knowledge

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

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

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

Source: Cision PRWeb

---