ROYAL SOCIETY OF CHEMISTRY | CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP

NEWSLETTER Winter 2015

Doug Veal with his RSC commemorative plate: see page 4

.

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: http://www.rsc.org/CICAG

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

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

https://twitter.com/RSC_CICAG

QR Code

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

Table of Contents

Chemical Information & Computer Applications Group Chair's Report

Contributed by RSC CICAG Chair Dr Helen Cooke, email: helen.cooke100@gmail.com

In our Summer 2015 Newsletter, I mentioned that our Committee had been enhanced by recruitment of a number of new members. I'm pleased to report that since then the new, enlarged team has had two stimulating and productive meetings, with plenty of exciting ideas emerging for how we can advance excellence in the chemical sciences through our activities.

We are planning an exciting programme of events for 2016 (see separate article). We will also be considering the possibility of hosting virtual meetings (by teleconference), holding more meetings outside London, and working with industrial organisations to explore opportunities for hosting meetings at their sites. We will also be updating the content of the CICAG's web pages to coincide with the next phase of the RSC's web site refresh. Please watch out for announcements via emails circulated to CICAG members from the RSC, Twitter and LinkedIn.

We are continuing to partner with other organisations and interest groups, and held a successful scientific meeting in October which was organised jointly with the AAMG (Automation and Analytical Management Group), entitled "Measurement, Information and Innovation: Digital Disruption in the Chemical Sciences". In 2016, we will be working with two external organisations, the Chemical Structure Association Trust and the Society of Chemical Industry, to organise scientific meetings.

After 6 years, Lindsay Battle, editor of the CICAG Newsletter, has decided to resign from this role. Lindsay has done a superb job, ensuring that the content has been consistent in style, that Newsletters are published in a timely fashion, and that everyone has had the opportunity to contribute. Thank you Lindsay for all your hard work for CICAG. Although Lindsay's contributions will be greatly missed, I am pleased to announce that Stuart Newbold will be taking over the role of editor. If there are topics you would like to see covered in the Newsletter, please get in touch with Stuart at stuart@psandim.com or via 01223 709136/07904 688799.

To keep in touch with the CICAG, please follow us on Twitter @RSC_CICAG and LinkedIn, or join our MyRSC group (links on the front page of this Newsletter). Our Social Media Editor, Keith White, makes

frequent posts and looks forward to your comments and interactions. Please also feel free to contact me directly with any questions or suggestions as to how your Committee can support you.

Recognition for Doug Veal's long service

In our Summer 2015 Newsletter it was reported that Doug Veal had decided to retire from the Committee after 25 years' service. In recognition of Doug's long service and many contributions to the CICAG and its predecessor, the Chemical Information Group, in the autumn Doug was presented with a commemorative plate and certificate. Although no longer a Committee member, Doug will continue to be the CICAG's point of contact for the Tony Kent Strix award (http://www.cilip.org.uk/uk-einformation-group/awards-and-bursaries/tony-kent-strix-award), which the CICAG sponsors.

Proposed CICAG Meetings for 2016

Our Committee is working on a broad range of scientific meetings for 2016, which will be held at the RSC's London headquarters at Burlington House and elsewhere in the UK. Working titles for the proposed meetings are:

- "**Mobile chemistry: content, compute, collaborate**". Planned for summer 2016. Speakers currently being recruited.
- "**Chemical structure representation: what would Dalton do now?**". To coincide with the 250th anniversary of John Dalton's birth. Meeting to be co-sponsored by the Chemical Structure Association Trust. Planned for autumn 2016. Speakers currently being recruited.
- **"Chemoinformatics for Drug Design"**. To be organised jointly with the Society of Chemical Industry. Planned for autumn 2016.

More details will be available from the RSC's Events listing and on the CICAG website http://www.rsc.org/CICAG nearer the time.

Meeting Report: Measurement, Information and Innovation
Digital Disruption in the Chemical Laboratory

Contributed by RSC CICAG Chair Dr Helen Cooke

This one-day scientific meeting, held on 16 October 2015 at Burlington House, was organised in partnership with the RSC's Automation and Analytical Management Group (AAMG). The meeting addressed the challenges, opportunities and issues that arise through the digital transformation of laboratory work, in particular concerning the long-term processes for storing, maintaining, finding, using, interpreting and understanding scientific data and information. Automation has significant implications for chemists and the skill set and expertise needed to optimise laboratory processes while ensuring opportunities for innovation are not reduced. Although the 'scientific method' continues to define the basic principles of laboratory work, digital technologies raise new questions about:

- The authenticity and integrity of data acquisition
- The collation and management of data and information
- The interpretation and derivation of meaning from data

In chemistry, we have the opportunity to learn from other's mistakes, and to manage the digital transformation in a controlled way. The meeting examined challenges, case studies, and there was plenty of time for lively discussions involving the audience as well as speakers.

Ten expert speakers with a diversity of experience from a broad range of industrial and academic organisations presented at the meeting. The topics and speakers were:

- *The purpose, practicalities, pitfalls and policies of managing and sharing data in the UK.* Dr Danny Kingsley, Head of Scholarly Communication, Cambridge University.
- *Allotrope Foundation, how a data standard and taxonomy for analytical data will support data integration and knowledge generation.* Dr Gerhard Noelken, Technology and Innovation Group, Pfizer Pharmaceutical Sciences.
- *Why we should embrace learnings from other industries.* Peter Boogaard, Industrial Lab Automation.
- *HELM (Hierarchical Editing Language for Macromolecules): A Pistoia Alliance project providing a standard for the storage and exchange of macromolecule structural information in the information ecosystem.* John Wise, Executive Director Business Development - Europe, Pistoia Alliance.
- *Ensuring the machine learner gets the right tuition and AI gets its nose; standardisation and benchmarking in diagnostic marker discovery in exhaled breath.* Professor C.L. Paul Thomas, Centre for Analytical Science, Chemistry Department, Loughborough University.
- *Building the utopian dream – toward the integrated, paperless drug discovery lab.* Dr Allan Jordan, Head of Chemistry, Drug Discovery Unit, Cancer Research UK Manchester Institute.
- *Using machine learning and cloud technologies for point-of-care in vitro diagnostics.* Dr Omer Casher, Chief Data Officer, SIME Diagnostics, Stevenage Bioscience Catalyst.
- *Enabling creative scientists to focus on science: automation, workflows, dashboards, insight, aha!!!* Martin Owen, Data to Insight Activist, GlaxoSmithKline.
- *Building the links*. Mr Richard Kidd, Royal Society of Chemistry.

Some speakers emphasised the difficulty of acquiring data, others the challenges of structuring data sufficiently for it to be useful, and others the problems with doing anything with the data that is available. A huge amount of data is available, more is continuously being generated, and ideas and methods for structuring data are developing and being used. Whilst the best paths forward may not be obvious, there are extraordinary opportunities ahead.

It was suggested during one of the discussions that this meeting could be repeated outside London, possibly at local industrial organisations, where company employees and others could attend. This is something CICAG and AAMG will be exploring in 2016.

The presentation slides and abstracts are available here:
http://www.rsc.org/events/detail/18885/measurement-information-and-innovation-digital-disruption-inthe-chemical-sciences

---

Tony Kent Strix Award

The UK e-Information Group, Geological Society, Burlington House, 6th.November 2015

The Tony Kent Strix Award is presented at the Strix Annual Lecture which is organised by the UK e-Information Group (UKeiG) in conjunction with the RSC CICAG, the British Computer Society Information Retrieval Specialist Group (BCS IRSG) and the UK Chapter of the International Society for Knowledge Organisation (ISKO UK), in recognition of outstanding contributions to the field of information retrieval and has been sponsored in part by CICAG for a number of years. The presentations reviewed here were the first in the UKeiG Annual Lecture Series, which have been sponsored by Google. UKeiG is a Special Interest Group of CILIP – the Chartered Institute of Library and Information Professionals. The 2014 Award winner, Dr Susan T. Dumais of Microsoft Research, presented this first Annual Lecture.

The meeting began with a brief biographical introduction of Tony Kent by Doug Veal, a former CICAG chairperson who worked with Tony Kent for 13 years. Tony Kent was a naturalist who started his career as a lecturer at the University of Nottingham in the 1960s. At about the same time the then Chemical Society had set up a research unit at Nottingham and he became its leader, writing software to search and retrieve information from Chemical Abstracts tapes. In subsequent collaboration with Jan Wyllie they later

developed the STRIX software and database searching package – named after a genus of owls who have, of course, the ability to see in the dark.

Speakers at the 2015 Tony Kent Strix Annual Lecture
Susan Dumais, Doug Veal and Jan Wyllie

Remembering Futures Past. Tony Kent and the Lost Vision of a Boolean Universe (Jan Wyllie)

Jan Wyllie began his talk by discussing how he and Tony Kent had in the 1980s had developed STRIX, a structured hierarchical database for information management which allowed full text retrieval and used, in the days before graphical user interfaces, a DOS command line. Working from home, their company was modestly successful, selling 450 copies of the software to a wide variety of customers. The code for the database is still being developed and has now been made Open Source.

As a postscript, he added a quote from Tony Kent (*The Intelligent Enterprise*, June 1991) about any future reliance on Artificial Intelligence (AI) to take away human responsibility for decision making:

> *Finding useful information is an intelligent process requiring intelligent people. At the end of the day, only the intelligent can recognise what is useful.*

Understanding and Improving Search using Large-Scale Behavioural Data (Dr Susan T. Dumois, Microsoft Research)

A presentation about how the study of large-scale behavioural data logs for web search engines has changed how web systems have been designed.

Although it is only 20 years since internet search engines were first introduced, they have changed the way we behave in ways that nobody could have anticipated – e.g. the way we buy things, plan holidays, search medical conditions.

Analysis of server-side behavioural logs maintained by search engine providers often provide surprising insights into how people interact with search engines. These logs record the queries and results of online searches. Often people are not trying to find "things" (i.e. retrieve information), but to find websites (i.e. to navigate). Further analysis of the many millions of queries show that the top 250 queries account for 10% of all queries (e.g. weather, horoscopes, sex, ebay), but that 40% of queries occur only once. There is also a periodicity for daytime variation (i.e. peaks for morning, afternoon, night browsing): finance -> food &

events -> porn. For one individual, on average 30% of queries are actually repeats of ones made previously – this opens the possibility of personalised service.

Although these logs show what people are doing, deducing as to why is another matter. Additionally providers would like to know the answers to questions such as: *was the result successful? Were people happy with the results?* So in addition to studying general behavioural (i.e. observational) logs, controlled experimental methods have been developed – laboratory studies, panel studies and large-scale log studies.

A most stimulating talk, revealing how search engine providers are trying to better understand and improve their services. For brevity's sake these notes are only a selection of the topics covered. For further information see: *http://research.microsoft.com/~sdumais*

Alan Tonge, November 2015

Other Awards - Calls for nominations

CSA Trust Grants: Applications Invited: The CSA Grant Program has been created to provide funding for the career development of young researchers who have demonstrated excellence in their education, research or development activities that are related to the systems and methods used to store, process and retrieve information about chemical

2017 Herman Skolnik Award: Call for Nominations: Awarded by the ACS Division of Chemical Information to recognize outstanding contributions to and achievements in the theory and practice of chemical information science. **Closing date: 1st June 2016**.

2016 Lucille M. Wert Scholarship: Call for Applications: Designed to help persons with an interest in the fields of chemistry and information to pursue graduate study in library, information, or computer science. **Closing date: 1st February 2016**

2016 CINF Scholarship for Scientific Excellence: Call for Applications. Awarded by the ACS Division of Chemical Information, co-sponsored by InfoChem and Springer, and designed to reward students in chemical information and related sciences for scientific excellence and to foster their involvement in CINF.

Other Awards - Recent Recipients

2016 Herman Skolnik Award - recipients are **Stephen Bryant** and **Evan Bolton** for their work on developing, maintaining, and expanding the web-based NIH/NLM/NCBI PubChem database and related software capabilities and analytic tools to enhance the scientific discovery process.

Dr Stephen Heller Receives the 2015 Patterson-Crane Award for his work on the development of the IUPAC International Chemical Identifier (InChI).

Jason Farradane Award: UKeiG is pleased to announce that the winner in 2015 is **Sheila Webber**, nominated for excellence in education and teaching in information science and for raising the profile of the information profession.

Student Bursaries

A number of Undergraduate Research Bursaries, funded by the Royal Society of Chemistry, will be available to students in UK and the Republic of Ireland Chemistry and related departments from June-September 2016. The purpose of the awards is to give experience of research to undergraduates with research potential in the

middle years (i.e. 2/3, 2/4 or 3/4) of their degree and to encourage them to consider a career in scientific research. Applications relevant to the interests of the CICAG will be welcomed in the areas of cheminformatics, chemical information, chemical data management, chemistry data analytics, chemistry IT solutions and applications. Further information can be found here:
http://www.rsc.org/Education/HEstudents/undergraduate-bursary.asp.

Chemical Information / Cheminformatics and related Books

1. **In Silico Medicinal Chemistry: Computational Methods to Support Drug Design**, Nathan Brown, RSC Nov 2015

This truly is a golden age of *in silico* medicinal chemistry as a data science, which is essentially what it is, using huge quantities of heterogeneous data (so-called big data) and various modelling techniques from statistical learning methods to do structure-based modelling. All of these and more are covered in this book, written by Nathan Brown at The Institute of Cancer Research, London and recently published by The RSC.

The title of this book, *In Silico* Medicinal Chemistry, is intended as an umbrella term for all approaches that use computers in chemistry to benefit medicinal chemistry and drug design and discovery. In this way, one can see *In Silico* Medicinal Chemistry as covering aspects of: chemoinformatics (also called cheminformatics), molecular modelling and computational chemistry. This book is not intended to be all-inclusive and exhaustive, but rather to make a solid foundation from which the reader can pursue aspects that most interest them or are relevant to a particular scientific challenge. Each chapter concludes with a selected list of key references to which the interested reader is directed for more in-depth information around specific subject areas from leading monographs in those areas.

The book covers the fundamentals of the field first: how we represent and visualise molecules in the computer, and how they may be compared. The first section begins with a brief history and introduction to mathematical graph theory and its close links with chemistry and molecular representations harking back to the advent of atomistic theory and even earlier to the foundations of graph theory with Euler. Representing molecules in the computer is essential for whatever subsequently needs to be achieved in the computer. For some applications it may be possible to have more complex representations, but more complex representations will typically require more complex calculations to analyse and make best use of the data. The methods by which we compare molecules also lie at the heart of computational chemistry. Similarity is a philosophical concept, but it is essential to consider the different types of similarities that may be measured and how they may be applied. All of these topics are covered here.

The second section considers the many different ways we can describe molecules in the computer. The old parable of the 'Six Blind Men and the Elephant' written by John Godfrey Saxe, from ancient tales, highlights challenges in measuring similarity and understanding differences. In the parable, six blind men were each asked to describe an elephant. The first blind man suggested that the elephant was like a wall because he felt its body. The second thought it like a snake, having touched its trunk. The third identified it as like a spear when feeling its tusk, and so on. This parable highlights the importance of recognising and understanding the concept of similarity and why it is important. The section begins with physicochemical descriptors, from which it possible to calculate properties that are measurable, with a high degree of accuracy. The second chapter moves onto topological descriptors that encode aspects of molecular graph representation, whether through the calculation of a single value that encapsulates an aspect of the molecular graph but is interpretable, or large quantities of complex descriptors that do not tend to be so interpretable, but are highly efficient and effective. The third class of molecular descriptor is the topographical or geometric descriptor that encodes information about the shapes and geometries of molecules, since clearly they are typically not flat, or static, entities.

The third section considers statistical learning methods, an integral aspect of computational drug discovery, and some of the best methods we have to investigate different properties. An introduction to statistical learning is presented, prior to splitting out two different aspects of statistical learning: unsupervised and supervised learning. Unsupervised learning uses statistical methods to understand the structure of data and how different objects, described by variables, relate to each other. This is important in understanding the proximity or otherwise of our data points, in our case molecules, and is integral to the concepts of molecular similarity and diversity in chemical databases and techniques used in many methods. Supervised learning still uses the descriptions of our objects - molecules, but the objective here is to relate these to another variable or variables. In chemistry, supervised learning can be used to make predictions about molecules before they are synthesised. This predictive learning can be very powerful in computational chemistry since we can explore that vast space of possible small molecules discussed earlier in a much more effective and rapid way. Lastly, a discussion and some advice on best practices in statistical learning are given to assist the practicing scientist using computers to make statistical analyses or summaries.

The next section moves on to explicit applications of computational methods in drug discovery. These methods are well known in the field and use aspects of all of the previously discussed concepts and methods. Similarity searching is up first, which is focussed on the identification of molecules that are similar to those that are already known, but also comparing large numbers of molecules for similarity and diversity. One of the most important aspects of similarity searching is the introduction of the concept of virtual screening, where new and interesting molecules can be identified by using ones that are already known, but with a similarity measure that is relevant to the challenge being addressed.

The second chapter in this section covers the twin concepts of bioisosteric replacements and scaffold hopping. These two concepts are related to similarity searching, which was mentioned previously, but instead of trying to identify molecules that tend to have structural similarities, this approach looks for functional similarity, with little regard for the underlying structure. This is becoming increasingly important in drug discovery as it allows projects to move away from troublesome regions of chemistry space that, although important for potency, may exhibit other issues that are undesirable in drugs.

The third chapter covers clustering and diversity analysis, which are essentially two sides of the same coin. Cluster analysis permits the identification of natural groupings of objects, molecules, based on molecular descriptors and example of the application of unsupervised learning. Using cluster analysis it is possible to select clusters of interesting molecules for follow-up or, using molecular diversity to select a subset of molecules that are different to each other. Whereas cluster analysis is an example of unsupervised learning, Quantitative Structure–Activity Relationships (QSARs) are an example of supervised statistical learning methods. Here, the objective is to correlate molecular structure with known biological endpoints, such as enzyme potency, and build a statistical model. The benefit of such a model, a QSAR, is that it may, with care and caution, be applied to predict for molecules that have not been tested, and have not even been synthesised. This allows vast virtual libraries to be analysed and prioritised to allow the focus to rest on those molecules that are most likely to succeed.

Since proteins began being crystallised and their structures identified through X-ray crystallography, the structures have held the promise of allowing the optimisation of new molecular entities *in silico* that are predicted to be enhanced in potency against the enzyme-binding site of interest. Protein-ligand docking methods have been developed for more than 30 years to model virtual molecules that are more optimal in interactions and potency than their predecessors. Many new methods and developments have been made and the predictive abilities of docking have improved greatly over the years. Still, however, challenges remain. The above chapter considers the methods that have been developed, an understanding of how to validate docking models, and finally how best to use the methods.

The last chapter in this section covers *de novo* design, arguably the pinnacle of computational drug discovery. The grand objective in *de novo* design is to design molecules in the computer that are entirely optimised for each of the objectives of interest. Clearly, the discipline is not that close to being able to achieve such a grand

challenge, but much headway has been made, particularly in recent years, utilising all of the methods that go before in this book. A brief history of *de novo* design is given in structure- and ligand-based methods, with a final view towards the future and the incorporation of multiple objectives in *de novo* design workflows.

The penultimate section of the book looks at a few successful case studies and methods that have been applied in every stage of drug discovery, from aspects of target validation in terms of druggability analyses and hit discovery, through to moving from hit compounds to leads and the optimisation of those leads. Some examples of methods that have or can be used in these in these are covered to set the context of the field and its level of importance through the drug discovery pipeline.

Lastly, the book concludes with the 'Ghosts of Christmases Past, Present and Yet to Come' after Charles Dickens. This chapter represents the importance of remembering where we came from and respecting the contributions of the giants that came before us; it reflects on where we are, how we got here and what has been achieved in recent years; and lastly, the chapter discusses what needs to be addressed in future, how can we achieve this and what we all need to do to prepare for the future.

This book is intended as an overview of a vast field, with thousands of scientists working in it worldwide. Each chapter has a set of key, yet not extensive, references as guides to where the interested reader may go next in the development of their skills and expertise in this complex field, no matter what it may be called. The book is ideal for undergraduate, graduate, and doctoral students relatively new to the field and requiring an overview of the key methods. The book is also useful as a ready reckoner for the seasoned professional in modelling and to the medicinal chemist looking for a deeper understanding of the tools and techniques that could be employed to further their drug design work.

Author: Nathan Brown; Print publication date: 02 Nov 2015; Copyright: Nathan Brown, 2016
Print ISBN: 978-1-78262-163-8; PDF eISBN: 978-1-78262-260-4; EPUB eISBN: 978-1-78262-749-4
DOI: http://dx.doi.org/10.1039/9781782622604

2. **The Handbook of Medicinal Chemistry**: is a new ibook providing insight and advice for medicinal chemists.
Drug discovery is a constantly developing and expanding area of research. Developed to provide a comprehensive guide, the Handbook of Medicinal Chemistry covers the past, present and future of the entire drug development process. Highlighting the recent successes and failures in drug discovery, the book helps readers to understand the factors governing modern drug discovery from the initial concept through to a marketed medicine. With chapters covering a wide range of topics from drug discovery processes and optimization, development of synthetic routes, pharmaceutical properties and computational biology, the handbook aims to enable medicinal chemists to apply their academic understanding to every aspect of drug discovery. Each chapter includes expert advice to not only provide a rigorous understanding of the principles being discussed, but to provide useful hints and tips gained from within the pharmaceutical industry. This expertise, combined with project case studies, highlighting and discussing all areas of successful projects, make this an essential handbook for all those involved in pharmaceutical development.
A free app has also been created in collaboration with the editors of the book. The Medicinal Chemistry Toolkit provides a suite of resources to support the day to day work of a medicinal chemist.

3. **Managing Scientific Information and Research Data**, **S. Baykoucheva,** Chandos Publishing, 2015, ISBN 9780081001950
See review in Chemical Information Bulletin Fall 2015

4. **Exploring Materials through Patent Information**, RSC, Oct 2014, Print ISBN: 978-1-78262-112-6
See review in Johnson Matthey Technol. Rev. 2015

National Chemical Database Service News

National Chemical Database Service (NCDS) now includes **CrystalWorks**, allowing you to do federated searches across CSD, ICSD and CrystMet.

British Patent Information Professionals Presentations: "A Day in the Life of an Independent Patent Searcher"

The British Patent Information Professionals (BPIP) group is an informal group of UK-based representatives of organisations with an interest in patent information. The group can trace its history back to 1990 when it started off as the Derwent UK User Group. Today the users enjoy the choice of a much wider range of patent information products far beyond those on offer in the 1990s and with much greater sophistication and content. BPIP meets once a year to discuss various issues, share searching tips and give feedback from patent information meetings. Meetings usually include presentations from both vendors and BPIP members. (BPIP Chairperson – Susan.Bates@shell.com).

BPIP's 44th meeting took place on 21 October 2015 at UCB's headquarters in Slough. There were presentations from two BPIP members, Pam Toplis of Serengeti Information Services Ltd and Stuart Newbold of Patent Search and Information Management Ltd, both chemistry information specialists, who shared their experiences of what life has been like outside the corporate information environment with "*A Day in the Life of an Independent Patent Searcher*". Pam established Serengeti in 2006 after a career at ICI/Zeneca/Avecia from 1985, whereas Stuart had a mixed career on both the information provider side – various roles at Derwent 1986-1999 and as part of the STN marketing team at the RSC 1999-2003 before spending 10 years at Astex Pharmaceuticals (now part of Otsuka) before starting PSIM. There was a surprising degree of overlap in the BPIP presentations – from resources used (STN – perhaps not surprising for chemistry/sequence work, and PatBase), the client bases (both industry and law practice), search types (FTO, validity, patentability, landscape, current awareness, etc.), and degree of complexity of searching and reporting requirements. Both Pam and Stuart highlighted a number of advantages and disadvantages associated with self-employment; they both cited the flexibility, work variety, control and esteem that comes with running their own companies. If given the choice, both said they now prefer being independent to being back in industry as employees.

Serengeti Information Services Ltd
Email: mail@searchserengeti.com
Telephone: +44(0)1257 274638
Mobile: +44(0)7776 180585
Website: http://www.searchserengeti.com

Patent Search and Information Management Ltd
Email: stuart@psandim.com
Telephone: +44(0)1223 709136
Mobile: +44(0)7904 688799
Website: http://www.psandim.com

CAS / SciFinder / STN News

Contributed by Dr Anne Jones, CAS Applications Specialist UK & Ireland

CAS News

CAS Assigns the 100 Millionth CAS Registry Number® to a Substance Designed to Treat Acute Myeloid Leukemia

On June 29, 2015, CAS registered the 100 millionth chemical substance in CAS REGISTRY℠ and celebrated its 50th anniversary as the world's largest database of unique chemical substances.

With a steady increase in patent activity around the globe, it is not surprising the 100 millionth small molecule registration comes from a patent. In this case, the substance was reported in a World Intellectual Property Organization (WIPO) patent (WO 2015081280) from Coferon, Inc. (Coferon) in Stony Brook, NY. The inventors claim the molecule, CAS RN® 1786400-23-4, available in both SciFinder® and STN®, is a novel therapeutic designed to treat acute myeloid leukemia.

SciFinder News

SciFinder® Release Introduces New PatentPak™ Interactive Chemistry Viewer
A recent release of SciFinder introduced the PatentPak Interactive Chemistry Viewer.

PatentPak is a robust, patent workflow solution available in SciFinder®. Designed to radically reduce time spent acquiring and searching through patents to find vital chemistry, PatentPak saves users up to half of their time by providing instant access to hard-to-find chemistry and patent families in languages users know.

Customers have already told us that PatentPak has improved their workflows, and now the interactive viewer adds functionality that takes the value of PatentPak to a new level. There is no other tool on the market that offers these benefits.

Key Features:

- Full-text patent documents from 11 major patent offices (an additional 20 offices to be added very soon)
- Patent family coverage in multiple languages, including English, German, Chinese, Japanese, French, Korean and Russian
- Patent page numbers for indexed key substances
- Unique, interactive patent document viewer
- Exportable patent PDFs with or without CAS analyst mark-up for sharing and printing

The benefits of these new features include –

1. **For the Interactive Document Viewer**

- Presents searchable, full-text of patents with interactive indexing and annotations that inform the user's research
- Offers direct links to and from SciFinder that promote serendipitous searching and exploration

2. **For the Key Substances Sidebar**

- CAS scientists compile a list of key substances in each patent so researchers can quickly identify and navigate to substances and chemistry of interest
- Direct page number links take researchers to the precise location of the substance in the patent, eliminating the need to read the patent page by page
- Hard-to-find substances embedded in text and R group tables link directly to a visual representation of the structure wherever possible
- Each substance in the sidebar includes the authoritative CAS Registry Number® and a link to the substance's SciFinder record in CAS REGISTRY℠, making interpretation of patent information easier than ever

STN News

PatentPak™ is also Coming Soon to Classic STN®!

In response to strong customer demand, PatentPak will be made available in classic STN in January 2016.

- PatentPak will be available in CA℠/CAplus℠ via STN Express® and STN on the Web℠
- PatentPak will be visible only to customers who subscribe to PatentPak on STN

- PatentPak will be included in all GVP offers and renewal quotes for contracts beginning January 2016 or later

PatentPak Key Features

- Full-text patent PDFs from 11 major patent offices (soon to increase with an additional 20 offices)
- Patent family coverage in multiple languages, including English, German, Chinese, Japanese, French, Korean and Russian
- Page number links for indexed substances in CA/CAplus records directly link to interactive patent viewer
- Exportable patent PDFs with or without CAS analyst mark-up for sharing and printing

Trials for PatentPak on STN are now available.

Derwent Markush Resource (file name DWPIM) on STN® - The unified solution for generic chemical structure searching comes true

The latest STN release, on November 19, 2015, delivers the Derwent Markush Resource (file name DWPIM) from Thomson Reuters to join MARPAT® from CAS, making new STN the only unified Markush solution, offering chemical patent searchers unsurpassed coverage and highly efficient access to the rapidly growing number of chemical structures disclosed in patents.

DWPIM seamlessly integrates into the Thomson Reuters / Derwent chemistry content on new STN.

For the first time, customers are able to conduct all of their structure and Markush searches on a single platform using premier chemical structure databases, including CAS REGISTRY℠, Derwent Chemical Resource (DCR), MARPAT, DWPIM and ReaxysFile™ .

CAS Training in the UK 2016

CAS Training in the UK 2016

In addition to e-learning materials, CAS continues to offer instructor-led training for both STN® and SciFinder® in the UK.

We conduct 'in-house' WebEx training sessions on all aspects of STN or SciFinder searching.If you wish to know more about any CAS products, or would like further information or help with STN or SciFinder, then please contact annejones@acsi.info.

Cambridge Cheminformatics Network Meetings 2016

Dates have now been agreed for next year's meetings of the Cambridge Cheminformatics Network. Meetings take place on Wednesdays at 4.00 p.m.

- 17 February at the CCDC, Cambridge (time fixed, final location TBC)
- 25 May at the Centre for Molecular Informatics (former Unilever Centre), Cambridge
- 31 August at the Centre for Molecular Informatics (former Unilever Centre), Cambridge
- 7 December at the European Bioinformatics Institute, Hinxton (time fixed, final location TBC)

If you would like to present at one of the above meetings please just let Andreas Bender know (ab454@cam.ac.uk). See also http://c-inf.net/.

Other Meetings/Conferences in 2016

Jan 12-13: **International Conference on Data and Information Management**, Loughborough

Jan 20: **Discovery and Discoverability**, London

Mar 20: **3rd international workshop on Bibliometric-enhanced Information Retrieval** (BIR 2016), to be held as part of the **38th European Conference on Information Retrieval** (ECIR), Padova, Italy

Mar 2: **http://www.gesis.org/en/events/events-archive/conferences/ecirworkshop2016/** the missing link between Information Retrieval and bibliometrics/scientometrics, and how it might be used to enhance retrieval processes in digital libraries both large and small

Mar 20: **LILAC: Librarians' Information Literacy Annual Conference**, Dublin

Apr 11-13: **UKSG Annual Conference and Exhibition**, Bournemouth

May 16-19: **6th International m-libraries conference**: Smart Libraries: Re-inventing libraries for a changing world, Milton Keynes

May 23: **Research Data Management for Information Professionals**, UKeiG Training, London

May 26-27: **18th International Conference on Chemical Structures**, Tokyo, Japan

Jun 27-29: **ARLG Conference 2016**: Are you being served? Serving our learners in a changing climate, Birmingham

Jul 4-6: **Seventh Joint Sheffield Conference on Chemoinformatics**, Molecular Graphics and Modelling Society and the Chemical Structure Association Trust, Sheffield

---

Recent meetings

which you may have missed, but can follow up online

Jul 2015 **Subject librarians: time for a fresh look?** University of Hertfordshire

Oct 2015 **The RSC Patent Debate – Does chemistry benefit from the patent system?** See report in IP Information & News Nov 2015

Nov 27: **Technology to support teaching**, University Science & Technology Librarians' Group (USTLG) meeting, London

---

People News

John Buckingham

Contributed by Dr Fiona Macdonald, Taylor & Francis Group

It is with great sadness that we report the untimely death of John Buckingham in a car accident on 28 August 2015. Born in 1943, John was an organic chemist who obtained his DPhil at Sussex University, studied at ETH with Albert Eschenmoser and worked with Bill Klyne at Westfield College, University of London. In the late 1970's he was hired by Chapman & Hall, the publishers of Heilbron's Dictionary of Organic Compounds, to publish a new edition of Heilbron and to convert the information into a database. Several editions later, the DOC, as it is now known, is published online only and has spawned several daughter

products, most notably Dictionary of Natural Products (DNP). John's passion for natural product chemistry and his expertise in chemical nomenclature made him the driving force of DNP and he remained fully engaged to the end, updating, editing and proof reading entries for the next edition. He was also the co-author of Natural Products Desk Reference, which will be published posthumously this year. He also possessed a deep knowledge of the history of chemistry, and wrote two popular science books Chasing the Molecule and Bitter Nemesis: The Intimate History of Strychnine. A talented writer, engaging colleague, and dear friend, he will be sorely missed by everyone who was fortunate enough to know him.

Other News Items

Understanding Impact Factors, Wiley ChemViews, Jul 2015

Dept. for Business, Innovation and Skills, metrics and non-selective QR allocation, Martin Eve blog Nov 2015

Many InChIs and quite some feat, Wendy Warr, J Comput Aided Mol Des Aug 2015

Digital Instructional Materials Gaining Ground in Science/STEM Education, Simba Information Jun 2015

DCL Announces Technology Breakthrough at USPTO: Over 1 Million Pages Each Month Automatically Transformed Into Searchable, Minable Content, Jun 2015

Elsevier Solutions Address the Challenges of Lead Optimization in Early Drug Discovery, PR Newswire Jun 2015

Long wait for publication plagues many journals: Scientific publishing requires patience, even after a paper has been formally accepted, Nature Jul 2015

A numbers game: Institutions must be plain about research metrics if academics are to engage with them, Nature Jul 2015

We need a measured approach to metrics: Quantitative indicators of research output can inform decisions but must be supported by robust analysis, Nature Jul 2015

Several Centuries of Centrality - chemistry as the 'central science', Dana Roth, ACS Central Science Jun 2015

British Library and Publishing Technology collaborate to make Document Supply Service content available via ingentaconnect, Jul 2015

Building the 21st century scientist: Nature in collaboration with Scientific American takes a look at the promise and challenges of bringing STEM education in line with decades of education research, Nature Jul 2015

The Problem With Patents - commentary on drug discovery and the pharma industry, blog post by Derek Lowe, Aug 2015

Guidelines for Preparing Patent Landscape Reports, WIPO

How The Internet Changed Chemistry - reporters and outside contributors explore the impact of taking the central science online, C&ENews Aug 2015

European consortium develops new approaches for dealing with Big Data, EurekAlert Aug 2015

Royal Society of Chemistry and Royal Society launch journal collaboration

RSC Periodic Table Free app for mobile phones and tablets; see also YouTube video

Wikipedians reach out to academics, Nature News & Comment Sep 2105

OCLC prints last library catalogue cards

Retracted papers get hooked up to linguistic lie-detector, Chemistry World Nov 2015; Original paper in J. Lang. Soc. Psychol.

Chemical Information and Computation 2015, Number One. 249th ACS National Meeting and Exposition, Denver, March 2015
Wendy Warr & Associates' 45th ACS report covers papers on Drug Discovery; Molecular and Structural 2D and 3D Chemical Fingerprinting; Getting to the Best Reaction: Tools for Finding a Needle in a Haystack; Research Results: Reproducibility, Reporting, Sharing and Plagiarism; Data Format Standards for Computational Chemistry and Defining "Value" in Scholarly Communications, plus a news section covering people, awards, and 65 organisations. See Contents list and order forms.

UKeiG Position Paper on Information Management - a short position paper on the recently released ISO9001:2015 for quality management systems. The paper briefly notes the eight quality management system principles and - importantly - highlights its relevance to the library, knowledge and information management sector. UKeiG Nov 2015

New RSC journals launching in 2016:

- Reaction Chemistry & Engineering - will publish high-impact research at the interface of chemical engineering and chemistry
- Nanoscale Horizons - the home for rapid reports of exceptional significance in nanoscience and nanotechnology
- Molecular Systems Design & Engineering - for cutting-edge research into how understanding of molecular properties, behaviour and interactions can be used to design and assemble better materials, systems, and processes to achieve specific functions. Crossing the boundary between fundamental and applied science, this unique new journal is a joint venture between the RSC and the Institution of Chemical Engineers

---

Note from the editor: This will be the last edition edited by Lindsay Battle; please send all items for future CICAG Newsletters to the new newsletter editor: **Stuart Newbold**, email: stuart@psandim.com