Chemical Structure Association Trust

ISSUE 4

NEWSLETTER

SUMMER 2003

Personalia

Helen Schofield and Frank Cooke got married on 21st December 2002. They're clocking up lots of air miles with USAir on the Manchester to Philadelphia flights while trying to plan the future.

After 18 years at MDL and 22 years at Roche, Guenter Grethe is now going to find more time for his family, particularly his granddaughter, and for his hobbies, including gardening and oenology. But he will not completely disappear from the scene. He will do some consulting and will continue to be involved in ACS CINF and CSA Trust activities.

Mark Canales and Rudy Potenzone have founded ChemTelligence Partners Inc. (http://www.chemtelligence.com), a company supplying consulting services in the strategic deployment and use of information-based tools in support of life sciences research.

Peter Gund and Herschel Weintraub are now working for IBM Life Sciences, Worldwide Pharma R&D Solutions.

Dr Ursula Schoch-Gruebler, Senior Vice-President, BASF Group Information Center at BASF in Germany, is being honoured as the recipient of the International Patent Information (IPI) Award 2003 (http://www.IPI-Award.com).

Pamela Scott has moved into the Legal group at Pfizer, Groton, doing patent and chemistry searching.

Yvonne McCormick has recently joined CAS in Europe as Regional Marketing Manager. She will be working with Barry Dunne supporting SciFinder and other CAS products in the UK and Ireland.

The 2003 MDL Excellence in Informatics Fellowship has been awarded to Manojkumar Jain, who will study chemical informatics at the School of Informatics, Indiana University.

The 2003 Daylight Innovation in Chemical Informatics Fellowship has been awarded to Jianyong Zhu who is to study chemical informatics at the School of Informatics, Indiana University.

Sponsorship Secretary John Holliday will maintain the register of donors. Please send donations to him: Dr John Holliday, Department Of Information Studies, University of Sheffield, Sheffield, S10 2TN, UK. Tel: +44-(0)114-222-2685; e-mail: J.D.Holliday@sheffield.ac.uk

---

Visit the CSA Trust website at http://www.csa-trust.org

Inside...

Forty years on: the Department of Information Studies at the University of Sheffield

Peter Willett

2003 marks the 40th anniversary of the establishment of the Department of Information Studies at the University of Sheffield [1]. The Department was originally the Postgraduate School of Librarianship, changing in 1967 to the Postgraduate School of Librarianship and Information Science, when the Department introduced an MSc in Information Science, and then to its current title in 1981. Chemoinformatics has formed one of the principal foci of research activity in the Department for many years, and this short paper summarises the major studies that have been carried out in this area. More detailed accounts have been published elsewhere [2,3].

How could an academic department set up to study librarianship have become involved in chemoinformatics research? This arose from the arrival in 1965 of Mike Lynch. He had previously been the Head of Basic Research at CAS, when it was carrying out some of the first experiments anywhere in the world on the use of computers for the processing of both textual and chemical structural information. On arrival in Sheffield, Mike rapidly established research programmes in both of these areas, and this dual focus of computational activity has continued with great success right up to the present day, with research in one area often providing the spur for subsequent developments in the other [4]. The chemoinformatics research has been in six major areas: reaction indexing; substructure searching, first in 2D and then in 3D databases; similarity searching and clustering; chemical patents; molecular diversity; and structural bioinformatics.

The first major area to be tackled was the development of automatic methods for the indexing, storage and retrieval of chemical reactions. Particular attention was given to the development of algorithms that could compare sets of reactant and product molecules to identify the reaction centres (i.e. the substructural moieties that had been changed during a reaction). Work continued on this for over a decade before an efficient and effective graph matching procedure was identified which has formed, with further development, the basis for the many public and in-house reaction retrieval systems.

The late 1960s saw the start of research into the selection of fragment screens for chemical substructure searching. The work involved a detailed analysis of the frequencies of occurrence of various types of algorithmic fragment. This statistical information then formed the input to a screen selection procedure. This resulted in the selection of a set of screens that occurred approximately equifrequently in the file and that could be used to reduce the computational requirements of the time-consuming atom-by-atom stage of a substructure search. Further studies evaluated the effectiveness of substructure search systems based upon the resulting screen sets, the statistical independence of screen assignments, and the relationship between query and structure characteristics, among other things. These frequency analysis and algorithmic fragment generation procedures formed the basis for many subsequent systems for 2D substructure searching.

The early 1980s saw the start of studies into methods for the representation and searching of the generic, Markush structures in chemical patents. This work continued for some 15 years, and involved very substantial modifications and enhancements to the connection-table, screening and atom-by-atom procedures of conventional substructure searching systems of the time. The studies were highly successful, resulting in a body of algorithms and data structures that paved the way for the sophisticated MARPAT and Markush DARC systems that are the current methods of choice for structure-based access to generic chemical structures.

This period also saw the start of the Department's work into chemical similarity measures, an area of work is still on-going. It was realised that the fragment substructures that had been developed for substructure searching could also be used to quantify the degree of structural similarity between pairs of molecules. This led to studies into the effectiveness of fragment-based similarity measures for both searching and clustering 2D chemical databases. This work highlighted the general applicability of the Tanimoto coefficient and the Jarvis–Patrick clustering method for chemical applications, approaches that are still used. Originally undertaken mainly to provide an alternative to conventional substructure searching, the principal value of this work now is probably as a simple, but effective, method for virtual screening and compound selection.

Thus far, all of the work had considered only 2D chemical structures. The introduction of structure-generation programs like CONCORD stimulated interest in how to extend the existing systems for substructure searching to encompass the availability of atomic co-ordinate data. The initial studies focused on rigid (i.e. single conformation) 3D structures, and simple modifications of conventional screening and atom-by-atom searching routines were soon developed that provided an effective means of carrying pharmacophore searches. Later work considered the extension of these

ideas to the inherently more difficult problem of flexible 3D substructure searching.

The work on 3D substructure searching, and related studies of methods for 3D similarity searching, was a direct precursor of our work in bioinformatics, and took place in collaboration with Professor Peter Artymiuk and his colleagues in the Department of Molecular Biology and Biotechnology at Sheffield over some 15 years. We sought to apply the graph-matching and genetic algorithm methods developed for processing small 3D molecules to databases of biological macromolecules, mainly 3D protein structures but also carbohydrate and, currently, nucleic acid structures. The principal contribution of this work has been its focus on the comparison and searching of biological structures, rather than the biological sequences.

The Department's involvement in 3D substructure searching aroused considerable industrial interest, which resulted in a near-doubling in the size of the research group in just a few years. It also led to an increasing focus on drug discovery applications in general, rather than on the database searching applications of much of our previous research, with projects on topics such as ligand docking, electrostatic field fitting, 3D QSAR and pharmacophore mapping. Drug discovery also provided the impetus for our involvement, from the mid-1990s onwards, with methods for molecular diversity analysis. Our initial work here, on strategies for selecting diverse database subsets, was based on our existing expertise in the calculation of molecular similarity and dissimilarity, but this soon developed into a whole range of related areas, such as diversity indices, novel types of structure representation for compound selection, and algorithms for the design of drug-like combinatorial libraries.

It is easy to adopt too self-congratulatory a tone in a review such as this, and we have hence recently carried out a citation analysis to establish how much our research has been used (or at least cited) by others [3]. Specifically, we considered citations relating to the five members of the full-time academic staff at Sheffield who have carried out chemoinformatics research: George Adamson, Val Gillet, John Holliday, Mike Lynch and the present author. Searches of the ISI Web of Science identified a total of 4845 unique citations to 321 publications (probably an underestimate as only post-1980 citation data are currently available on the Web). These citations appeared in 411 different journals, including not only chemical, chemoinformatics, biological, bioinformatics, computing and information science journals, but also titles across the physical, life and medical sciences. The citations came from 910 different organisations in 54 different countries, with many of the citations coming from chemical database and software companies and from the majority of the major multinational pharmaceutical companies. These figures demonstrate clearly the breadth of influence of the Sheffield research, as does the number of members of the chemoinformatics industry who have been associated with the Department at some time as research collaborators, post-doctoral researchers, or MSc and PhD students. It is also very pleasing to note that many of these people have additionally been involved with, or acted as an officer of, the CSA Trust or its predecessors (the Chemical Notation Association and the Chemical Structure Association).

Chemoinformatics at Sheffield continues to flourish. In 2000, we introduced the first specialist MSc chemoinformatics programme, with funding from the UK Engineering and Physical Sciences Research Council and a consortium of agrochemical, database, pharmaceutical and software companies. The programme takes about a dozen students a year, with the aim of alleviating the current severe shortage of graduates with specialist chemoinformatics skills [5]. Also in 2000, we ran what is proving to be a biennial international chemoinformatics course, and in 2001 the first of our annual 3-day training courses. Both series of meetings are run on behalf of the Chemical Structure Association Trust and the Molecular Graphics and Modelling Society. We have maintained the level of our research, with six post-doctoral research staff and 13 PhD students currently in the group, many of whom are funded in whole or in part by industry. Indeed, much of our success over the years has been due to the help from our many industrial collaborators. They have been generous indeed in providing research funding, data, software and valuable insights that have helped to ensure the relevance of the problems that we have addressed.

1. Saunders, W.L. Challenge and opportunity: Sheffield's new postgraduate library school. *Aslib Proceedings*, 16 (1964) 105-115.
2. Lynch, M.F. and Willett, P. Information retrieval research in the Department of Information Studies, University of Sheffield: 1965-1985. *Journal of Information Science*, 13 (1987) 221-234.
3. Bishop, N., Gillet, V.J., Holliday, J.D. and Willett, P. Chemoinformatics research at the University of Sheffield: a history and citation analysis. *Journal of Information Science*, in press.
4. Willett, P. Textual and chemical information retrieval: different applications but similar algorithms. *Information Research*, 5(2) (2000) at http://informationr.net/ir5-2/paper69.html
5. Schofield, H., Wiggins, G. and Willett, P. Recent developments in chemoinformatics education. *Drug Discovery Today*, 6 (2001) 931-934.

Blogging, RSS and Metadata

Henry Rzepa challenges the chemistry community to get enthusiastic!

It is interesting how developments by one group of people are regarded by other communities. The chemistry community does not appear to have adopted the use of metadata, RSS and Blogging with any enthusiasm. Metadata may be considered worthy, but boring, RSS is yet another acronym, and hence nerdish, and Blogging is a turn-off for purists, but a revolution for others. I should like to explain why I think enthusiasm is deserved!

Originally, websites were mainly controlled by webmasters. A small number of these sites became "must visit" sites, because of their content. Then static content started to be replaced by dynamic content generated from databases etc; (e.g. using PHP/MySQL). The poor user had to visit almost on a daily basis to get the most out of a site, and of course no-one has to the time to visit even a few sites on a regular basis. How could you find out if the site (in its daily reincarnation) might be of interest?

In about 1999, Netscape, followed by others, introduced "Web Logging", pronounced "Blogging". This is essentially a regularly updated (probably daily) diary of items and events, probably written by an individual, and disseminated via a personal web server. Suddenly faced with perhaps having to visit thousands or more websites, the "Blogging" community woke up to "metadata". A standard, based on XML, for specifying this metadata was invented, called RSS. Being more a cultural phenomenon (and not invented by computer scientists, who looked down on much of this), it almost inevitably bifurcated into RSS 1.0 (the purist camp, who based it on a standard known as RDF) and RSS 2.0 (yes, it's confusing, based on an "easy-to-use" version proposed by Dave Winer).

There are many ways of thinking about RSS. The one I find most rewarding is to think of it doing for hyperlinks what HTML did for content; that is, it provides a formal link management format which separates most importantly the content from the manner in which it is presented to the user. So a Blog can now promote its presence by offering what is called an RSS feed (or channel) with carefully crafted metadata indicating what items there are, when/where they were posted, who wrote them, and (perhaps most importantly) what can be done with them and how they relate to each other.

RSS channels can now be aggregated, or organised into subjects, themes, communities etc, using RSS aggregators (of which there are around 10 now for various OS platforms). You can search for them using "syndicators" and also perform Google-like searches on just them (and not the web in general).

So why should the chemistry community generate enthusiasm about RSS? Well, think of a published chemistry journal. It is really just a very high class Blog (which is enhanced with peer review, etc). As a reader, you can perhaps visit five e-journals a day and check to see what is new, but often, you will only do two things:

a) look at the graphical abstract, or table of contents
b) open an Acrobat file to see if you are really interested in it.

The latter has the undesirable consequence that after several years of doing this, you will have perhaps 1000 unmanageable Acrobat files on your hard disk (yes, I know you can organise them using EndNote, but in truth, it's a pretty primitive system).

It is far better to create an RSS entry for each new journal article, and have the user subscribe to the journal "channel". So here I issue my first rallying call:

Can I urge all chemistry journals to offer an RSS feed or channel for their readers?

If you have managed to read this far, you are probably wishing for some examples. Click on: http://purl.org/net/syndication/subscribe/?rss=http://www.ch.ic.ac.uk/rzepa/index.rss

This takes you to a syndicator, and it will check (actually XML validate) http://www.ch.ic.ac.uk/rzepa/index.rss, which is the RSS feed for my Blog (well, previously known as Web page). It will offer you information about 9 RSS clients or aggregators, and if you have them installed, it will insert this feed into your local copy. You do this to all RSS feeds you wish to have (say about 100?), including the journals I encouraged above! You leave this program running, and it will periodically check to see what is new and alert you.

Some RSS clients will do more – much more is promised! NewsMonster will index and archive the RSS feeds, so that you will be able to search the entire aggregate, and also archive so that entries that scroll off the bottom remain on your hard disk. Arguably, you need no longer download those Acrobat files, but merely ensure you have a pointer to them so that you can access them whenever you need.

This is not the place to go into what else can be achieved with RSS (being XML, it can be manipulated in astounding ways), how it is created, how it can be extended etc. For further information, see the full article, which has been submitted for publication to the *Internet Journal of Chemistry*: "Towards the Chemical Semantic Web. An introduction to RSS". Peter Murray-Rust and Henry S. Rzepa. It can be accessed as a preprint at http://www.ch.ic.ac.uk/rzepa/rss/

So, to find out how/whether RSS is useful to chemists, I would like to issue my second rallying call:

Will all chemistry content providers who have new and interesting announcements to make on websites consider putting up an RSS feed?

For example, the following will announce new entries for our local Molecule of the Month pages: http://xml.mfd-consult.dk/syn-sub/?rss=http://www.ch.ic.ac.uk/motm/index.rss.

Henry Rzepa

Department of Chemistry
Imperial College
London
SW7 2AZ, UK

If you would like to comment on this article, please email Henry on h.rzepa@imperial.ac.uk

The Future of Chemical Information in the UK

The Royal Society of Chemistry Chemical Information Group held their Spring Meeting and AGM at Burlington House, London, on 27th May 2003.

With an emphasis on the situation in academia, the four speakers in the morning gave interesting presentations on the changing nature of chemical information provision in the UK.

The speakers were Nigel Lees, Manager of Library and Archival Services at the RSC; Julia Stocken, Head of Science, Technology and Innovation at the British Library; Joanne Yeomans, Faculty Team Librarian at the University of Leeds, currently on secondment to CERN; and Gillian Sinclair, the JISC Resource Guide Adviser for the Physical Sciences, based at the University of Manchester.

There is a substantial decline in the number of students studying chemistry in the UK, and a consequent reduction in information scientists who have a background in the chemical sciences. An open forum in the afternoon looked at the problem of Training and Education of chemists/chemical information scientists, and considered ways in which the profile of chemistry could be raised.

Preliminary announcement

One-Day Training Course on Chemical Information

at the Royal Society of Chemistry, Burlington House, Piccadilly, London

Thursday 30th October 2003

Organised jointly by CSA Trust and RSC-CIG

This one-day training course will contain a series of presentations, which will provide an impartial and critical overview of Chemical Information Products. It will be given by specialists from academic information departments, vendor representatives and the pharmaceutical industry.

The theme of the presentations will cover structure searching in relation to chemical reactions, chemical properties and chemicals reported in patents.

The cost of this training course is likely to be £25. Further details will be available from:

David Walsh (Pfizer Ltd, david_j_walsh@sandwich.pfizer.com)

or **Chris Evans** (Chirex Ltd, chris.evans@rhodiachirex.com)

Bioactive Discovery in the New Millennium

February 5–9 2003
Lorne, Victoria, Australia

The Royal Australian Chemical Institute's Division of Biomolecular Chemistry held a very successful international conference in Lorne, a beach resort on the beautiful, world recognised 'Great Ocean Road'. The meeting focused on cutting edge methods of discovering and developing new bioactive agents. It was a Gordon conference-style residential meeting to maximise the interactions between the delegates. This proved very successful, especially for the student participants.

The conference was a joint meeting of:

- the Biomolecular Chemistry Division of the RACI
- the European Chapter of the Molecular Graphics and Modelling Society (MGMS)
- the Australian Molecular Modelling Workshop (MM2003)
- the Medicinal Chemistry project of the Federation of Asian Chemical Societies

It aimed to facilitate dialogue on multidisciplinary approaches to the design and development, not only of pharmaceuticals but also veterinary drugs, agrochemicals and other bioactive agents. Given the involvement of several specialist modelling societies, and the overlap with the structural biology conference, there was a focus on the use of computational methods and molecular modelling to design new bioactive agents using protein structure information. The conference attracted 175 participants from 19 countries, and 11 trade exhibitors. Approximately 100 papers were presented, roughly half oral and half posters. The conference abstract book and program can be downloaded from http://www.chem.csiro.au/raci/biomolecular.

Conference topics included: natural bioactive compounds; *de novo* design; molecular docking methods; novel organic and biomimetic synthesis, chemistry-driven lead generation and optimisation; quantitative structure-activity relationships (QSAR); soft computing methods and complex systems; measurement and modelling of ADME properties; informatics; quantum chemical methods and applications; high-through put structural biology, NMR studies of ligand-receptor interactions, molecular mechanics and dynamics; pharmacology; toxicology.

The conference had a very strong line up of 20 plenary speakers including Nobel Laureate Professor Peter Doherty. The meeting opened on Wednesday evening with a cocktail party in the beautiful gardens of the Cumberland, followed by the first two scientific sessions. Dr Frank Blaney (Glaxo Smithkline) gave a stimulating opening plenary lecture on the state of the art in design of drugs targeting the very important G-protein coupled receptors. Dr Regine Bohacek (Boston Denovo) then kindly swapped time slots with Professor Mark von Itzstein, who had to delay his arrival due to illness. She spoke on the latest molecular docking and scoring algorithms for *de novo* design and virtual screening. Dr John Tallarico (Harvard Medical School) then gave a fascinating talk on the new field of chemical genomics, which blends medicinal chemistry with informatics to find new drug leads.

Dr Dick Cramer (Tripos Associates) started the Thursday morning session with a talk about a new, very computationally efficient method of carrying out virtual screening on extremely large databases. After coffee, Professor Paul Alewood (IMB) was presented with the 2002 Adrien Albert award, and gave an excellent lecture on the use of toxins as drug leads. Professor Martyn Ford (University of Portsmouth) presented a plenary paper on the use of neural networks for high-throughput screening after an excellent lunch, and Dr Linda Brinen (Stanford Linear Accelerator) gave an inspiring multimedia talk on high-throughput structural biology.

Professor Graham Richards (Oxford University) presented his recent work on setting up a world-wide grid computing network using idle time on some 2 million PCs to find new drug leads against cancer and bioterrorism agents such as smallpox. During the conference, there was a public launch of this initiative in the US which was widely reported by the world's media sources. Many press reports (e.g. Forbes, NY Times Evening Standard, Straits Times) can be viewed at the following website, http://www.chem.ox.ac.uk/smallpox/news.html. Professor Glen Kellogg of Virginia Commonwealth University (home to the latest Nobel Prize-winner in Chemistry) finished off the day sessions with a thought-provoking talk on the role of lipophilicity in molecular design.

Friday morning opened with a plenary lecture by Dr Jill Gready (ANU) on the difficulties of carrying out hybrid quantum mechanics-molecular mechanics simulations of enzymatic reactions. Professor Paul Wender followed Jill with one of the stellar talks for which he is renown. He showed how synthetic medicinal chemistry and molecular modelling can be married to discover some exceedingly potent anticancer drugs.

The conference held a 'Startups and Spinoffs' workshop exploring the path between scientific discovery,

invention and ultimate spinoff company. The aim was to allow researchers who had made the jump from pure research to a startup or spinoff company to explain how this was achieved, and to discuss the advantages and pitfalls of commercialising research outputs in various ways. The panellists included: Professor Paul Wender, (CellGate, Libraria); Professor Graham Richards, (Oxford Molecular); Professor Glen Kellogg, (EduSoft); Professor Paul Alewood, (Auspep, Xenome, Australian startup companies); and Dr Seb Marcuccio, (Boron Molecular, a recent Victorian startup company). The speaker presentations were followed by a panel and floor discussion and questions.

After dinner, Dr Bob Clark (Tripos Associates) opened the session on ADMET prediction and drug discovery with an excellent talk on the pitfalls of ADME modelling. Dr David Manallack (Denovo Pharmaceuticals) opened the final session of the evening with a talk on his new pharmacophore generation methodology, Quasi2.

After a well-earned sleep, the delegates reconvened for the Saturday morning session on ligand-based design with a plenary paper by Dr C. M Venkatachalam (Accelrys). After coffee, Professor Doherty presented his stimulating special conference lecture on virus-specific T-cell response.

At Saturday night's gala dinner, the audience were extremely well entertained by a light after-dinner speech by Professor Doherty. His speech was followed by the presentation of the student prizes kindly donated by CSIRO Publishing/Australian Journal of Chemistry. The prize for best oral presentation ($200 plus an journal subscription) was awarded to Ben Tehan from Monash University. Two poster prizes ($150 plus subscription) went to Catherine Hemley (University of Woollongong) and Jerome Wielens (Monash University). The conference dinner was followed by a party jointly sponsored by Tripos Associates and Accelrys which lasted into the small hours.

Professor Mark von Itzstein (Griffith University) opened the Sunday session with an excellent paper on the role of sialic acid recognising proteins in drug discovery, and Dr Ajay Royyaru (IBM Life Sciences/ Blue Gene Project) gave a very interesting insight into the computational requirements for Blue Gene. This segued well into an invited paper by Dr Dan Stevens (SGI) on the future of high performance computing in life sciences, and the conference ended with a very interesting paper by Dr Tim Clark (University of Erlängen) on the role of quantum mechanical calculations in QSAR.

The organisers made every effort to encourage student participation. The conference budget provided $500 bursaries to students who presented talks or posters at the meeting and were RACI members. Additional bursaries were provided through the kind sponsorship of Biota Ltd. Due to the conference being designed to maximise interactions between delegates, students were able to talk to and mix with some of the 'legends' of their field over meals or at social functions. This provided a very valuable inspirational component for the students.

A major factor in the success of the conference was the excellent support from sponsors. They are always the difference between a great conference and an ordinary one. The conference committee would like to thank them for their great support:

Principal sponsors

- Tripos Associates
- Accelrys
- Victorian Department of Innovation, Industry and Regional Development;

Major sponsors

- APAC
- Davies Collison Cave
- CSIRO
- Elsevier Science;

Sponsors

- Glaxo Smithkline
- Pfizer
- Boron Molecular
- VPAC
- FB Rice and Associates
- IBM
- Denovo Pharmaceuticals
- Advent One
- CSIRO Publishing
- Biota
- Silicon Graphics
- SPECS
- Medos
- DA Information Services.

David Winkler
Chair Organising Committee
Past Chair, Division of
Biomolecular Chemistry

3rd Joint Sheffield Conference on Chemoinformatics

21–23 April 2004

The Chemical Structure Association and the Molecular Graphics and Modelling Society announce their Third Joint Sheffield Conference on Chemoinformatics. The conference will be held at The Octagon Centre and Tapton Hall, University of Sheffield, UK from 21st to 23rd April 2004.
Further details will be available on the conference website at http://cisrg.shef.ac.uk/shef2004 in due course. A Call for Papers, including a list of themes, will follow shortly.

International Chemical Information Conference
Nîmes, France
19–22 OCTOBER 2003
http//:www.infonortics.com/chemical/index.html

PROGRAMME

SUNDAY 19 OCTOBER

Welcome cocktail and buffet dinner

MONDAY 20 OCTOBER

Opening Keynote
Speaker to be announced

Session 1: Changing Technology in Information Services

Thomas Lorenz, BASF, Germany
Integration of corporate information systems: the impact of changing technology (provisional title)

Robert Swann, CAS, Ohio, USA
The impact of changing technology on delivering products and services in chemical information

Karl Harrison, Central Chemistry, University of Oxford, UK
Chemistry computing on a massive scale

New Product Information Session: CAS / Springer-Verlag / FIZ Karlsruhe/ Search Technology

Session 2: Informatics and Analytics

Manuel Peitsch, Novartis Pharma
Informatics and knowledge management in pharma research

Holger Ernst, Wissenschaftliche Hochschule für Unternehmensführung (WHU), Germany
Using patent data for competitor intelligence

New Product Information Session: John Wiley / STN / Questel-Orbit / Recommind

Jacques Smeets, Xerox Global Services, The Netherlands
Trends in intellectual asset management

Shirley Bailey-Wood and Bob Stembridge, Thomson Derwent, UK
Patent analytics: tips, traps and techniques

Conference Cocktail

TUESDAY 21 OCTOBER

Session 3: The New Information Metrics. Why Pay?

Liz Blankson-Hemans, Quantum Dialog, UK
Defining the value of information: beyond ROI

Christopher Forbes, Knovel Corporation, New York, USA
Decreasing time to market through linking reference information and analysis tools to workflow

Minoo Philipp, Henkel, Germany
Why pay for value added information?

Jason Ward, Chemical News & Intelligence, UK
Why pay for information? A view from the value-added information provider

New Product Information Session: Micropatent / BizInt Solutions / Derwent Information / Knovel

Session 4: Search and Retrieval

Chahab Nastar, LTU Technologies, France
Content-based image retrieval: a state of the art

Stephen E Arnold, AIT, Kentucky, USA
Building search systems for sci-tech retrieval: it is not always as easy as it appears

Gérard Giroud and Barrou Diallo, EPO, The Netherlands
Reference linking: experiences from the European Patent Office

Session 5: Information and Publishing Industries

New Product Information Session: Accelrys / Elsevier / Temis

Steven Bachrach, Trinity University, Texas, USA
Publishing at a cross-roads: who is going to be doing what?

Michel Vajou, Bureau van Dijk Information, France
Concentration and integration in the information industry

Peter T. Shepherd, COUNTER, UK
Standards for the recording and exchange of online usage data: towards fairness for publishers and customers

Conference Dinner and Entertainment (Le Mas de Peint)

WEDNESDAY 22 OCTOBER

Session 6: The New Patent World

Ebe Campi, EPO, The Netherlands
A chemical structures database at marginal cost: a dream that could become reality

David Dickens, Pierre Buffet and Yuji Takashima, Questel-Orbit, France and Patolis, Japan
A guide to the Japanese patent information galaxy: who are the stars?

Session 7: Classification and Categorisation

Caspar Fall and Patrick Fievet, ELCA Informatique, and WIPO, Switzerland
Computer-assisted categorisation of patent documents in the International Patent Classification

Anthony Rowe, Wellcome Trust Biological Atlas of Insulin Resistance Project, UK
Towards domain-mapping systems to integrate biological and chemical analysis

Kirill Degtyarenko, European Bioinformatics Institute, UK
Vocabularies and ontologies for bioinformatics

Stephen Stein, National Institute of Standards and Technology, Maryland, USA
An open standard for chemical structure representation: the IUPAC Chemical Identifier

End of programme at about 13.00

CSA Trust Grants for 2004

Purpose of the Grants:
The Grant Program aims 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 on chemical structures, reactions and compounds. Grants are awarded up to a maximum of $1,000 each. Grants are awarded for specific purposes, and within a year each grantee is required to submit a brief written report detailing how the grant funds were allocated.

Who is Eligible?
Applicant(s), age 35 or younger, who have demonstrated excellence in their chemical information related research and who are developing careers with the potential to have a positive impact on the utility of chemical information relevant to chemical structures, reactions and compounds, are invited to submit applications. While the primary focus of the Grant Program is the career development of young researchers, additional bursaries may be made available at the discretion of the Trust. All requests must follow the application procedures noted below and will be weighed against the same criteria.

What Activities are Eligible?
Grants may be awarded to acquire the tools necessary to support research activities, or for travel to collaborate with research groups, to attend a relevant conference, to gain access to special computational facilities, or to acquire unique research techniques in support of one's research.

Applications must include the following documentation:

1. A letter that details the work upon which the Grant application is to be evaluated as well as details on research recently completed by the applicant;
2. The amount of Grant funds being requested and the details regarding the purpose for which the Grant will be used (e.g. cost of equipment, travel expenses if the request is for financial support of meeting attendance, etc.). The relevance of the above-stated purpose to the Trust's objectives and the clarity of this statement are essential in the evaluation of the application);
3. A brief biographical sketch, including a statement of academic qualifications;
4. Two reference letters in support of the application. Additional materials may be supplied at the discretion of the applicant only if relevant to the application and if such materials provide information not already included in items 1-4. Three copies of the complete application document must be supplied for distribution to the Grants Committee.

Deadline for Applications:
Applications must be received no later than October 17, 2003. Successful applicants will be notified by December 19, 2003.

Three copies of the documentation should be forwarded to:

Bonnie Lawlor, CSA Trust Grant Committee Chair, 276 Upper Gulph Road, Radnor, PA 19087, USA.

E-mail submissions, if complete, may be forwarded to the Grant Committee at chescot@aol.com.

Sheffield Short Course: A Practical Introduction to Chemoinformatics

June 17-20 2003

This year's course was the third to be run in the Department of Information Studies, Sheffield University and attracted 16 delegates from the UK, the US, Holland, Spain and Poland. The topics covered were 2D databases and substructure searching, similarity searching, diversity and compound selection, 3D conformational analysis, 3D molecular interactions, pharmacophore generation, 3D database searching, structure based design and docking, combinatorial libraries, and high-throughput screening.

Each of these areas was covered by a lecture, followed by a practical session using industry-standard software. There were also discussion group sessions with each group summarising their findings in a presentation.

On the social side, the first evening began with dinner in Halifax Hall, followed by an introduction to the course and the tutors, two of whom were our usual industrial experts, Robin Taylor from CCDC and Andrew Leach from GSK. Once again, we thank them for their invaluable contributions. The second evening started with a two mile walk in the Peak District, followed by dinner in the picturesque setting of Monsal Head. A fine meal at the Botanical Gardens rounded off the social side of the course on the final evening.

The course was a great success (despite a few technical gremlins with the university network), and all of the delegates seemed to have learnt a great many new techniques as a result of their short stay in Sheffield. There was a good mix of people from industry and academia both amongst the delegates and the tutors, which meant there was never a dull moment, and the food was, as ever, of the highest standard.

Two student bursaries were awarded, one by the CSA Trust and one by the Molecular Graphics and Modeling Society. The CSA Trust's bursary was awarded to Phil Evans, a PhD student in Martyn Ford's group at Portsmouth University. The next newsletter will include a report on the course from an attendant's point of view, courtesy of Phil.

John Holliday
Research Manager
Dept. of Information Studies
University of Sheffield

EVENTS 2003–2004

August

10–15 August: Drug Discovery Technology 2003, The Hynes Convention Center, Boston, MA, USA. Website: http://www.drugdisc.com/section.asp

10–15 August: 39th IUPAC Congress and 86th Conference of The Canadian Society for Chemistry, Ottawa, Canada. Contact: 39e Congrès de l'UICPA et 86e conférence de la Société canadienne de chimie, Conseil national de recherches Canada, Bureau des services de conference, Édifice M-19, Chemin Montréal, Ottawa K1A 0R6, Canada. Tel. 613-993-0414; fax: 613-993-7250; e-mail: iupac2003@nrc.ca; website: http://www.nrc.ca/confserv/iupac2003

September

2–6 September: Proteomic, Combinatorial & Other Strategies for Drug Discovery, Development & Delivery, Imperial College, London, UK. Contact: Professor Roger Epton, Symposium Organiser, Mayflower Worldwide Ltd, P.O. Box 13, Kingswinford, West Midlands, DY6 0HR, UK. Tel: +44-(0)1384-279324; fax: +44-(0)1384-294463; e-mail: r.epton@mayflower.demon.co.uk

7–11 September: 226th ACS National Meeting New York, with CINF session entitled "Advances in Reaction Searching". This session is co-sponsored with the ORGN and MEDI divisions. Contact: Office of National Meetings, Meetings and Expositions Department, American Chemical Society, 1155 Sixteenth Street NW, Washington, DC 20036, USA. Tel: (202)-872-4374; fax: (202)-872-6128; e-mail: natlmtgs@acs.org

October

14–17 October: Fifth AFMC International Medicinal Chemistry Symposium, Kyoto Park Hotel, Kyoto, Japan; http://www.acplan.jp/aimecs03/

19–22 October: International Chemical Information Conference, Nîmes, France. Programme on page 8. Website: http://www.infornortics.com/chemical/index.html

19–22 October: EUSIDIC Annual Conference. Branding–Establishing Online Visibility and Marketing Web-based Services, Prague, Czech Republic. Contact: EUSIDIC Secretariat, c/o FIZ Karlsruhe, 76344 Eggenstein-Leopoldshafen, Germany. Tel. +49-(0)7247-808-403; fax: +49-(0)7247-808-114; e-mail: eusidic@fiz-karlsruhe.de; website: http://www.eusidic.org

21–23 October: 10th Asian Chemical Congress, Hanoi, Vietnam. Contact: Dr. David A. Winkler, CSIRO Molecular Science, Private Bag 10, Clayton South MDC 3169, Australia. Tel. +61-3-9545-2477; fax: +61-3-9545-2446; e-mail: dave.winkler@csiro.au; website: http://www.facs-as.org/meetings/10acc-index.htm

December

2–4 December: Online Information 2003, Olympia, London, UK. Contact Fiona Ashton, Online Information, c/o Circulation Data Services (CDS), PO Box 6009, Thatcham, Berkshire, RG19 4TT, UK. Tel: +44-(0)1635-588863; fax: +44-(0)1635-868594; website: http://www.online-information.co.uk

2004

February

11–13 February: ADMET 1, Town & Country Hotel, Hotel Circle, San Diego, CA 92108, USA. Contact Darrin Scherago, Scherago International, 11 Penn Plaza, Suite 1003, New York 10001, USA. Tel: +1-(212)-643-1750, ext. 20; fax: +1-(212)-653-1758; e-mail: darrins@scherago.com; website: http://www.scherago.com/admet/

March

7–10 March: International Patent Information Conference and Exposition, IPI-ConfEx, Lisbon, Portugal. Website: http://www.ipi-confex.com/

March 28–April 1: 227th ACS National Meeting, Anaheim, California. ACS Meetings, 1155 16th St., NW, Washington, DC 20036-4899, USA. Tel: 1-800-227-5558; fax: +1-(202)-872-6128; e-mail: natlmtgs@acs.org

April

18–21 April: Drug Discovery Frontiers in Cancer and Neuroscience. The Application of Chemistry to Diseases, Leavey Conference Center, Georgetown University Medical Center, Washington, DC, USA. Contact: Jessica Bauman, Research Assistant, Drug Discovery Program, Research Building, Rm. EP09, 3970 Reservoir Rd, NW, Washington, DC 20057. Tel: +1-(202)-687-0151; fax: +1-(202)-687-0738; e-mail: drugdiscovery@georgetown.edu; website: http://drugdesign.georgetown.edu/ddc2004/index.htm

21–23 April: Third Joint Sheffield Conference on Chemoinformatics. CSA Trust/MGMS, University of Sheffield, Sheffield, UK. Website will be: http://cisrg.shef.ac.uk/shef2004

PRODUCT NEWS

Patent ProfilesNet

Patent ProfilesNet is the new Web-based patent alerting service from Thomson Derwent which provides essential alerts based on original patent documents from major patent offices worldwide. Presented in HTML format and delivered within days of publication, records consist of the patent's bibliographic data, main claim, abstract, and most important drawing. Patent ProfilesNet also links to the full patent documents in PDF format. Users have the option of receiving alerts weekly, bi-weekly, or monthly. Up to six months of alerts can be archived on the Patent ProfilesNet server, allowing users to track and report on usage through the statistics screen. Additional Patent ProfilesNet features include:

- Screen optimized layout
- Zoomable drawing
- User-friendly navigation
- Printing and forwarding capability
- Online help system

For a free trial of Patent ProfilesNet, visit http://www.patentprofilesnet.com/demo.

Strategic Drugs database

According to industry reports, $73 billion in major drug patents will expire between 2002 and 2007. There is intense pressure and competition in research and development and the introduction of new drugs. Strategic Drugs database (SDdb) is a business analysis and planning solution from Current Drugs for strategists in the international bio-pharmaceutical industries. It provides comprehensive and strategic information available on drugs in development and on the market. SDdb allows business and marketing executives to analyse current and predicted sales of the leading prescription drugs using graphical analysis tools. In addition to these financial analyses, one can examine the underlying fundamental scientific data that defines the sales prospects of drugs on the market today and in the R&D pipeline.

A premium-service version of the SDdb is also available to customers, combining all of the above with a daily monitoring service, covering all major scientific journals, news and business sources. Information captured by SDdb editors is then assessed against a SWOT (Strengths, Opportunities, Weaknesses, Threats) analysis of the drug and its nearest competitors. Customers are then alerted to any new information that is likely to affect this analysis. The website is at http://thomson currentdrugs.com/products/sddb/.

Patents from CAS

Chemical Abstracts patent coverage has been expanded by the addition of patents from six more countries. CAS has added the following national patent authorities to its regular patent coverage: New Zealand, Estonia, Monaco, Bulgaria, Slovenia and Hong Kong, in addition to the African Regional Industrial Property Organization (ARIPO). In total, CAS now offers patent information for 44 active patent issuing authorities.

CAS now offers ultra fast delivery of information from the UK and French patent offices, including them in a group of 7 core patent issuing authorities, which also include the US Patent & Trademark Office (USPTO), German Patent and Trademark Office (GPTO), Japanese Patent Office (JPO), European Patent Office (EPO) and the World Intellectual Property Organization (WIPO). For these offices, patent bibliographic information and abstracts are available in CAS electronic services within 2 days of the patents' issuance. Fully indexed records for these patents appear within 27 days. Additionally, CAS is extending its patent selection criteria by including more classification codes: the US NPC, which began in January 2003, and the ECLA codes and Japan F-Terms to be added to patents in the CA/CAplus files in the near future.

OmniViz v3.6

OmniViz, Inc. has released OmniViz v3.6, which offers improved text analysis, chemical fingerprint generation and clustering from Barnard Chemical Information (BCI), and Statistical Analysis of Microarrays (SAM) from Stanford. The integration of BCI's fingerprint generation and Ward's hierarchical clustering components into OmniViz enables the researcher to find positive leads and eliminate unproductive ones faster. OmniViz 3.6 also introduces the power of the Significance Analysis of Microarrays tool, based on Stanford University's SAM algorithm. SAM's approach enables the researcher to eliminate false positives quickly. In particular, SAM empowers the researcher to identify genes that are differentially expressed under various experimental conditions while controlling the false discovery rate. OmniViz's implementation allows the analyst to apply the SAM algorithm to any kind of analysis that involves the comparison of numeric measurements, and is not limited to microarrays. The BCI chemical informatics capabilities and the SAM algorithm are available as part of the new OmniViz Titanium Package. The website is at http:// www.omniviz.com/about/ pr_2003_ 04_25.htm.

Contact Details

Chair of Trustees

Guenter Grethe
Tel: +1-(510)-865-5152
e-mail: ggrethe @comcast.net

Treasurer

Geoff Downs
Tel: +44-(0)1909-515444
e-mail: geoff@ bci.gb.com

Secretary

Clive Weeks
Tel: +44-(0)1252-674890
e-mail: Drcliveweeks@ aol.com

Sponsorship Secretary

John Holliday
Tel: +44-(0)114-222-2685
e-mail: J.D.Holliday@ sheffield.ac.uk

Newsletter Editor

Frances Daniel
Tel: +44-(0)1949-844837
e-mail: francesdaniel@ bottesford65.fsnet.co.uk

Next Copy Date

September 30 2003

Thank You to Our Donors!

We would like to thank all of you who have generously donated to the Trust this year, and particularly wish to thank our Corporate Donors:

AstraZeneca

Barnard Chemical Information

Cambridge Crystallographic Data Centre

Kilmorie Consulting

MDL Information Systems (San Leandro, US)

Sheffield Chemoinformatics Research Group

Syngenta, Jealott's Hill

Wendy Warr & Associates

Your donations enable the Trust, which is an internationally recognised charity, to continue to promote education, research and development in the field of processing and retrieval of information about chemical structures, reactions and compounds.

Many students and researchers worldwide, who work in this area, have benefited from a Trust Grant to further their research work. Grants of up to $1000 are awarded annually to young research workers and may be used to assist with attending a relevant conference, or to collaborate with other research groups, or to acquire tools necessary to support the research.

Additional bursaries are also made available from time to time, at the discretion of the Trust. If you know of anyone who would benefit from a grant or bursary, please pass on the details given on page 9 of this Newsletter. The deadline for application for a grant in 2004 is October 17th 2003.

If you are reading this and have not contributed to the Trust this year, we would like to encourage you, or your organisation, to make a donation. As a donor, you will continue to receive the lively CSA Trust Newsletter, keeping you up-to-date with conferences, meetings and training related to chemical information activities, as well as interesting reports and articles.

Whilst we suggest a minimum donation of £25 ($35 or 42 Euros) for individuals and £100 ($150 or 165 Euros) for corporate donations, we would be very grateful for any amount. A donation form is available on the web at http://www.csa-trust.org/donors.htm. Not only will you help to further the work of the Trust, but we hope you will also see the benefit of making a contribution.