Chemical Structure Association Trust

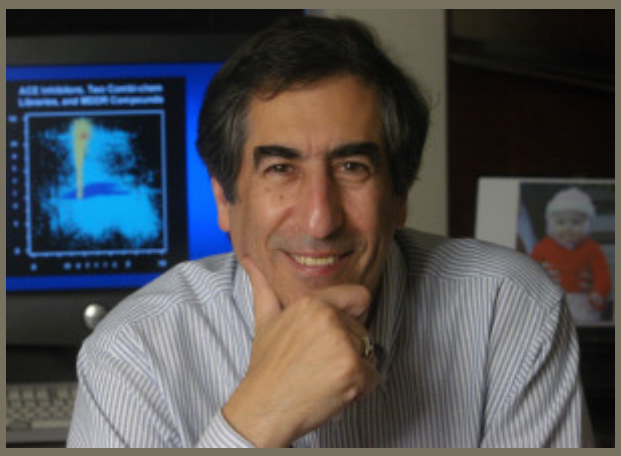
NEWSLETTER

Winter 2006 Issue 14

Our new Chairman, Andreas Barth, presided over his first CSA Trust AGM at the SCI, Berkeley Square, London, on 27th November. The AGM is always a time of change, and this year we said farewell to two of our longest standing Trustees: Mike Lynch and Peter Johnson. Thanks very much to them, and to the other Trustees whose term of office has come to an end, for their contributions to the work of the Trust. New Trustees to replace them will be appointed in due course.

Trustees enjoying the traditional English fare at Bumbles restaurant after the AGM

2007 Herman Skolnik Award Winner



Robert S. Pearlman, of the University of Texas at Austin, will receive the 2007 Herman Skolnik Award. This award is given by the Chemical Information Division of the American Chemical Society, in recognition of outstanding contributions to and achievements in the theory and practice of chemical information science. The award ceremony will be held at the American Chemical Society National Meeting at Boston in September 2007.

Dr Pearlman is best known for developing CONCORD, a tool for conversion of 2D connection tables into 3D structures. He received the Society for Biomolecular Sciences Accomplishment Award 2006 for his work on the DiverseSolutions package, which introduced novel concepts and descriptors related to clustering and diversity in chemistry-space. His BCUT-descriptors are widely used to help focus lead discovery efforts and guide the growth of corporate screening collections.

His more recent work includes innovative software for addressing both tautomerism and stereoisomerism in the contexts of cheminformatics and computer-assisted molecular discovery.

2008 Herman Skolnik Award – Call for Nominations

The ACS Division of Chemical Information established this Award to recognise outstanding contributions to and achievements in the theory and practice of chemical information science. The Award is named in honour of the first recipient, Herman Skolnik.

By this Award, the Division of Chemical Information is committed to encouraging the continuing preparation, dissemination and advancement of chemical information science and related disciplines through individual and team efforts. Examples of such advancement include, but are not limited to, the following:

- Design of new and unique computerised information systems
- Preparation and dissemination of chemical information
- Editorial innovations
- Design of new indexing, classification and notation systems
- Chemical nomenclature
- Structure-activity relationships
- Numerical data correlation and evaluation
- Advancement of knowledge in the field.

The Award consists of a $3000 honorarium and a plaque. The recipient is expected to give an address at the time of the Award presentation. In recent years, the Award Symposium has been organised by the recipient.

Nominations for the Award should describe the nominee's contributions to the field of chemical information and should include supportive materials such as a biographical sketch and a list of publications and presentations. Three seconding letters are also required. Nominations and supporting material should be sent by email to me (ggrethe@comcast.net). Paper submissions are no longer acceptable. The deadline for nominations is June 1, 2007.

Guenter Grethe, CINF Awards Chair

People and Places

Peter Gregory, ex Director of Publishing at RSC, has re-joined Wiley-VCH, where he is now editorial director and a member of Wiley's Global STM Board.

Ian Russell will be the next CEO of the Association of Learned and Professional Society Publishers (ALPSP). He took up office in October and is spending three months working in parallel with Sally Morris, the current CEO, who retires after more than eight years with ALPSP.

Eugene Garfield is the recipient of the Online Information Lifetime Achievement Award for 2006, in recognition of more than 50 years in the information industry. Dr Garfield is founder and chairman emeritus of the Institute for Scientific Information (ISI), now Thomson Scientific. Dr Garfield began publishing the Science Citation Index in 1964. The SCI became an important tool for navigating the scientific literature, and is now part of Web of Science. At 81, Dr Garfield continues to be active in scientific communication and information science.

Aureus Pharma has appointed Jason Theodosiou as CEO. He will replace André Michel, founder of Aureus who has been appointed Chairman of the Supervisory Board.

Hansruedi Kottmann joined Vertical*i as CEO in March of 2006. He has previously worked for Accelrys, Symyx, Base4, and MDL.

Eric Jamois, formerly of Accelrys, is now Vice President, Business Development, at Strand Life Sciences.

2nd German Conference on Cheminformatics, Goslar

This conference was also the 20th in the CIC Workshop series started by Johnny Gasteiger, and it was dedicated to him, since he plans to retire in April 2007. The series changed its focus in 2005. In particular, the language of the conference was changed to English and a determined effort has been made to attract an international audience, and a higher proportion of attendees from industry, while not lowering the high standard of research presentation that the audience has come to expect.

The signs are optimistic. In 2006, out of 150 attendees, about 50% were from industry; about 20% of all delegates were not based in Germany. There were 40 posters, 23 talks, two pre-conference tutorials and a small exhibition. The meeting proper opened with presentations of the FIZ CHEMIE Berlin awards (by CSA Trustee René Deplanque) to the two students whose dissertations had achieved most acclaim.

Keynote papers opening the three technical sessions were given by Gisbert Schneider ('Tinkering with Molecules'), Henry Rzepa ('Mechanism for Semantic Intertwingling in Molecular Sciences') and Mathias Rarey ('About Screening and Searching: Novel Computational Approaches to Structure-based Design'). There were two evening lectures, one by Johnny Gasteiger himself on software development in chemistry (covering also some of the history of the 20 CIC workshops) and one by myself: a tribute to Johnny as a pioneer in cheminformatics. This informal photograph of us was taken by Hilary Warr after that final lecture. The slides used by speakers will be mounted on the conference website (http://www.gdch.de/strukturen/fg/cic/tagungen/2gcc__e.htm); a great many 'official' photographs are already online.

Dr. Wendy A. Warr
http://www.warr.com

18th Annual ICIC meeting, Nîmes

During this year's 'Nîmes' meeting, most attendees appeared to feel that, beyond the magnificent food and wine, it was a good year for the talks. In *Information World Review*, Wendy Warr commented that 'It was a pleasure to see more chemical structures on the menu this year.'

The first morning was devoted to the future: what are the trends; how technology impacts our business; how information and communication are becoming more intricate. The meeting started with a duet, Pierre Buffet (Questel-Orbit) trying to understand why our information world is dominated by production and not distribution, and Harry Collier (Infonortics) surfing on the new Web 2.0 wave predicting free lunches and how the concepts of Web 2.0 will change our all professional lives.

Then came interesting highlights on new tools: Karen Blackman (RBA) spoke on what professionals can expect from RSS, Blogs and Wikis, Jorge Manrique (Prous Science) explained the impressive plans of Prous Science using Internet-2 capabilities, while Humberto Montenegro of Henkel gave us a tour in a new world for most of us: the world of industrial designs. Gary Horrocks (King's College, London) gave a well-crafted comparative evaluation of Scopus, Web of Science and Google Scholar. Two further overviews attracted enthusiastic comments: Valentina Eigner-Pitto (InfoChem) discussing a new approach to retrosynthesis, and Martin Brändle (ETH, Zürich) overviewing and evaluating chemicals catalogues databases.

Speaker Bob Stembridge of (Thomson Scientific) and Paul Peters (CAS).

Space limits discussion of the ten papers on the second day, but perhaps we can pick out Hansruedi Kottmann (Vertical*i) on Alliance Management, Diederick Braam-van-Vloten (ZyLAB) on searching large email collections, and Charles Huot (TEMIS) on recognition of chemical entities in scientific literature (entity extraction).

The last half day is often a difficult one. This year, starting with an update on the International Patent Classification, could have been even worse, but Bob Stembridge (Thomson Scientific) succeeded in making this topic interesting, highlighting the benefits of worldwide standards in patent description, but also pointing out the limits of any classification system compared to other indexing mechanisms, especially in terms of costs generated in classification systems by the need to re-index constantly as technology evolves.

In the first of two talks on machine translation, Enrique Filloy-Garcia explained how MT has become a reality at the EPO. Then Pierre Bernassau (Systran) gave an overview of its success and of what can or cannot be expected from MT, describing various levels of use, e.g. instant translation of web pages or pre-translation of technical documentation in a specific world. Daniel Keesman (GeneData) was excellent on semantic data integration; 'A breath of fresh air' declared Wendy Warr, confessing to advance trepidation regarding this complex subject area. All presentations can be found on the ICIC area of the Infonortics website (www.infonortics.com).

For a little variety, the meeting moves to Spain next year, to Sitges (Barcelona). Dates are 21–24 October 2007 and the venue is the luxury Hotel Melia in Sitges.

Pierre Buffet (Questel-Orbit) and Harry Collier (Infonortics)

Drug Discovery Technology & Development

The Drug Discovery Technology & Development (DDT) conference and exhibition returned as usual to Boston in August, but was held in the third location in as many years, and this may have contributed to the lower than usual attendance. The conference comprised six dedicated (read 'overlapping') conferences on *Targeting Disease and Evaluating Disease-related Targets*; *Lead Discovery and Optimisation*; *Discovery to Development; Biomarkers*; *R&D Strategies and Business Alliances*; and *The Interface between Drug Discovery and Informatics*. There were also workshops and tutorials, poster sessions and four keynotes from marquee speakers: Peter Corr (Pfizer), Susan Hockfield (MIT), Andrew von Eschenbach (FDA) and Steven Paul (Eli Lilly).

The more scientifically-oriented conference tracks were the best attended, and the hot topics this year included biomarkers, high content screening, RNAi, pharmacogenomics, drug safety assessment and translational medicine. But as your reporter's company had paid good money to be a sponsor of the Informatics conference, he felt duty-bound to hang out there, even though the excitement was clearly elsewhere.

The Informatics conference had sessions on *Workflow and Data Standards*; *Data Integration, Knowledge Management and the Semantic Web*; *Data Visualisation*; *Predictive Informatics, Modeling and Systems Biology*; *Structural Information*; and a catch all *Technologies* session. Each session had two or three papers, and then a panel discussion, and the quality of these varied somewhat, depending on the size and alertness of the audience, and how much time was left for discussion.

The conference started with Michael Haerter and Roger Brunne of Bayer Healthcare describing their PIX Pharmacophore Identification decision support system. This provides integrated access to data, workflow support, and project and portfolio management tools. It is deployed to 800 scientists and met the majority of expectations for things like meaningful decisions sooner, increased efficiency in lead identification, and reduced attrition.

Another interesting deployed system is in operation at Array Biopharma. Daniel Weaver stressed the need to focus not just on the scientific data, but also to include business data (costs, resources) and soft data (thoughts, opinions, rationale). The Array system is based on a semantically rich ontology, and uses a number of commercial tools from CambridgeSoft, Teranode and GD Viz: the latter had an interesting way of allowing data to 'self-present' itself on forms with its underlying properties and meaning preserved. The Semantic Web panel discussion was taken up with overlong presentations from the panelists, so there was little time for actual discussion.

The afternoon session started with Jeffrey McDowell of Abbott Laboratories talking about the use of Web Services as a low maintenance way to serve up functionality to scientists, and he described simple applications to send a search to PubChem, and an Excel-based Lead Selector application.

Your reporter took part in a panel discussion on the benefits (or not) of integrated information systems. Fellow panellists were Otto Ritter (AstraZeneca), Rainer Fuchs (Biogen IDEC) and John Weinstein (NCI). Otto described efforts at AZ to quantify the benefits of projects through a combination of time/cost/risk/return. Rainer was more sceptical and stressed the need for human intervention to make any sense of 'integrated' data. He also posited that while chemists may benefit from integration, it does not seem to be helping biologists. John characterised many scientists as still in the 'hunter/gatherer' phase when it comes to information retrieval, and felt the need to switch the focus from data access to problem solving.

The second day of the informatics conference started with William Egan from Novartis describing predictive toxicology tools, which are deployed to chemists to help them assess which compounds to kill and which to advance. John Weinstein of NCI coined (and trademarked) the term Integromic to describe the suite of tools for drug and biomarker discovery at NCI.

The panel discussion on systems biology technology options was wide ranging and illustrated that this discipline – if one can call it that – is still evolving, with different camps and theories, and tension between the 'look at the big picture' people and those who want to focus on specific targets in detail. Eric Martin of Chiron contributed the second neologism of the day, describing his work on kinases as 'Chemo-Kinametrics'.

The conference finished with an open-ended question and answer session, and Next Steps panel discussion, and the three panellists, Otto Ritter (AstraZeneca), Joseph Cerro (Bayer Pharmaceuticals) and Seth Pinsky (Abbott Laboratories) did well to keep the attention of the rapidly dwindling audience. Some of the trends they saw emerging were the increasing need to measure the business value of assets ('How much is an ELN worth?'); to manage identity (Life Science IDs); to add semantics and meaning to data; to learn from other businesses and academia; and to observe what scientists actually do, not what they say they do.

DDT continues to be a well respected fixture in the science/informatics conference landscape, but somehow this year felt flatter. It will be interesting to see whether the change in ownership of the meeting organisers (IBC is now owned by Informa) will make a difference in the level of attendance or the quality of the conference.

Phil McHale
Elsevier MDL

12th Society for Biomolecular Sciences Annual Conference and Exhibition

The Society for Biomolecular Sciences (SBS) was until recently the Society for Biomolecular Screening, and this name change reflected a desire to broaden its outlook from just screening to the science around screening. This was very much in evidence at the 12th Annual SBS Conference held in Seattle in September: in addition to plenary sessions, the program included poster sessions, short courses, workshops and tutorials on a wide range of topics around screening with titles like Chemical Genomics for Dummies and HTS 101. Themes addressed in the conference included structure-based and rational drug design, biomarkers, ADME/Tox in early drug discovery, systems biology and high-content assays.

One particular session, on 'Knowledge Management: Extracting Value from Large Data Sets', was jointly chaired by your intrepid reporter and Leo Bleicher (SciTegic) and may be of interest to Newsletter readers.

The session introduction suggested that advances in high throughput screening, automation, and combinatorial and parallel chemistry have led to sample and results data sets of ever increasing size; and with cell-based and other assay types, the complexity of the data in the result sets is also growing. Faced with larger and more complex data sets, what are scientists and project teams to do? How can they extract meaningful and actionable information that will enable better decisions on structural modifications, secondary screening and hit list triaging? This session heard from five speakers from academia and industry who described a variety of approaches to these challenges and the results obtained.

The plenary lecture was given by Debra Toburen of Elsevier MDL (speaking for Trevor Heritage), and described the high level informatics challenges associated with lead identification and lead optimisation, and some of the technologies and applications that are available to help solve these. In addition to data input, analysis and visualisation tools used with internally generated data, it is also important to examine and exploit the information in the scientific literature, and the speaker illustrated this with a brief case study on a lead optimisation candidate, for which there was a wealth of published information which could be used to make better informed decisions on next steps.

Joergen Nielsen of F. Hoffmann-La Roche described methods for systematically processing HT dose response screening data and making cross-assay comparisons. The system uses Genedata and provides similarity searching for bioactivity profiles, clustering of compounds, and a 'hit profiler', an annotated spreadsheet for analysing and presenting results. Some 240K HT results across 23 projects have been processed, and have resulted in EC50 values two to three months earlier. There is now complete selectivity for 1000s of compounds across several assays, and lists with 1000s of hits have been reduces to 10s to 100s, with desired profiles.

Magnus Kjellberg of AstraZeneca is an 'HTS Informatician' and his group has created a modular screening data management system, using in-house and third party components building on industry standard technologies. This is used for data entry QC, hit selection, and data mining. Spotfire is used for visualisation and analysis, and Pipeline Pilot to bring in structure-based fingerprints and molecular properties. The system has resulted in improved data quality, time efficiency and user satisfaction, and is cost efficient to maintain.

Nicholas Tsinoremas of the Scripps Research Institute runs a large and multicomponent screening centre which can generate 1.5M data points per day. The core of the informatics system which manages the structures and assay data is a combination of MDL ChemBio, Plate Manager and Assay Explorer, and these are complemented by tools to automate procedures and for analysis, visualisation and reporting.

John Davis of Novartis emphasised the importance of bringing computational chemistry knowledge to bear on primary screening results at an early stage. Using these techniques a set of 1M compounds which generates 100–30,000 hits can be reduced to 2000 compounds worthy of validation. Undesirable compounds are modelled based on medicinal chemists' knowledge, and Bayesian models are used to recover false negatives. 3D similarity searching is used to scaffold hop to new chemotypes. The part of the talk that aroused most audience interest was a hit list triaging tool built using Pipeline Pilot which captures the decisions used to 'rescue' and 'evict' compounds, and allows annotations of the decisions with reasons.

In an interesting organisational ploy, the conference organisers kept the two main keynote addresses to the last morning, and a large crowd was treated to two excellent talks: Chris Lipinski spoke on 'Drug-like or Tool-like Compounds: Why does it matter in HTS?' and Leroy Hood of the Institute for Systems Biology talked about 'A Systems Approach to Disease' and painted an inspiring picture of a future with predictive, preventive and personalised medicine.

The SBS conference certainly has evolved, and now offers an excellent cross-discipline view of the tools, techniques and science being used to 'advance the science of drug discovery'. The next annual meeting will be in Montreal in April 2007.

Phil McHale
Elsevier MDL

CSA Trust Jacques-Émile Dubois Grant 2007 Award winner

Congratulations go to Rajarshi Guha of Indiana University, who has been awarded the CSA Trust Grant for 2007, in memory of Professor Dubois. His research is the use of chemical information to enhance chemical research and his specific area is in the development of QSAR models with interpretable descriptors. Rajarshi intends to use his award to attend the Gordon Research Conference on Computer Aided Design in August 2007.

Quantum Pharmacology – 30 Years on

One of a regular calendar of meetings run by the Molecular Graphics and Modelling Society (MGMS), this conference was a Festschrift to honour the imminent retirement of Professor Graham Richards, past Chair of Chemistry at Oxford, and an extremely influential pioneer in the field of computational drug design. Graham also founded Oxford Molecular. He was elected as Chairman of Chemistry in Oxford in 1997 and was instrumental in the building of the new chemistry department there, which was opened in 2004. Graham is the author of over 330 papers and 15 books. He is the recipient of a number of awards including the Mullard Award in 1998, the Italgas Prize in 2001 and the 2004 ACS Award for Computers in Chemical and Pharmaceutical Research. In 2001 he was awarded the CBE for his services to chemistry.

The conference, held at St Catherine's College, Oxford UK in September 2006, was attended by about 120 participants, and comprised a diverse and very 'cutting edge' list of topics. The speakers were from the research groups of the large pharmaceutical companies, and the best academic groups across the world. The topics presented, and those followed up during the networking out of sessions, identified the key problems in rational design of bioactive agents and materials, and the latest research aimed at overcoming these problems.

The main themes in the talks were:

- Docking is very difficult and unreliable in many cases
- QSAR can often produce a better explanatory and predictive model more easily than docking
- Coarse graining of biological systems (describing them at a lower level of detail) can often produce more effective models (better predictive performance and less effort) than more detailed methods because they ignore much noise or confounding detail that obscures the model
- Some of the simpler methods that involve coarse-grained descriptions of molecules are surprisingly effective at virtual screening, similarity searching, database searching and building 3D QSAR models
- Most problems in drug design require multi-objective optimisation methods (e.g. Pareto methods).

Robin Ganellin (UCL, inventor of first H2 antagonist, Tagamet, the first $1Bn drug) spoke on the evolution of QSAR. He spoke on the activity of chlorhexidine-like molecules (currently used as antiseptics) against potassium channels.

Kate Burt (Pfizer) discussed new computational methods for antiviral (HCV and HIV) discovery. The key message here was integration of data, and making it available to chemists along with structural biology and modelling results. Advances in web technology may make it reasonably easy to implement such a system within the Division. She also mentioned that the ROCS method of similarity searching from OpenEye seems to be quite effective in finding molecules in databases that have similar biological activity to a lead query.

Chris Baker (Oxford and Cambridge) looked at ways of producing multipole representations of electrostatic fields around molecules using carefully placed partial charges. This suggested a method for generating useful QSAR descriptors involving aromatic interaction based on these moments.

Sir Tom Blundell (Cambridge and Astex) examined developing drug using high throughput structural biology. This seems to be the most reliable method of generating highly active, small drug candidates (especially given the unreliably of docking methods).

David Clary (Oxford) spoke on rewriting the Schrödinger equation in a form that made it equivalent to the diffusion equation. This allows it to be solved by a random walk algorithm. He could then use it to apply quantum mechanics to very large molecules like proteins and materials clusters.

Bill Jorgenson (Yale) is very well known for developing force fields for modelling proteins and their interactions with ligands. He spoke on the Free Energy Perturbation method for generating more accurate models on such interactions, and for rapid chemical library design and lead optimisation.

Garrett Morris (Scripps) compared the performance of the most commonly used docking programs (none were particularly accurate). He developed a new method called a Lamarck genetic algorithm that seems to produce a better outcome than the current methods.

Garland Marshall (Washington University St Louis and founder of Tripos Associates) presented studies of the rhodopsin molecule, the prototype structure for G protein-coupled receptors (GPCRs) that are currently the most successful pharmaceutical drugs and targets. He developed small molecule mimics of peptides relevant to binding to rhodopsin, via an analogous method that we used for the betabiotics project.

Frank Blaney (GSK) is an expert on modelling GPCRs. He posed the question as to whether crystal structures of proteins are better than models. In spite of most people expecting the answer to be positive, he showed number of cases where the x-ray structures were wrong or misleading as they showed ligands bound in non-productive poses, or poses that were not consistent with the known SAR of drugs binding to the targets, or with site directed mutagenesis experiments. In some cases, model structures were more useful for design than the x-ray structures.

Nicolas Foloppe (Vernalis) spoke on ligands for the Chk1 protein that is a possible cancer target. He quoted useful references on lead-like compounds.[1]

Richard Catlow (UCL, pioneer on inorganic modelling and zeolite design) spoke on novel computational methods based on genetic algorithms and molecular topology, to design zeolites, micro and mesoporous materials, and catalysts. These methods allowed materials to have many desired properties designed in. His GULP program, which used genetic methods, is available for materials design. He also studied mechanisms of self-assembly of these materials.

Richard Lavery (IBPC Paris) discussed a very effective coarse-grained representation of proteins based on an elastic network. Such networks have the interesting property of identifying the most important residues in a protein (ligand binding sites, hinge regions) by means of forces.

David Chandler (Berkeley) used a coarse-grained representation of proteins and water to develop fast and effective methods of studying the self-assembly of viral capsids, microsomes, and so on. His models contained few parameters but could reproduce the complex viral capsid structures very well. This has implications for other themes and streams within the Division involved with self-assembly of surfaces and nanodevices, as these methods could be adapted to model and design novel systems of this type.

Adrian Elcock (University of Iowa) described simulations of protein clusters and the cytoplasmic regions of cells at full atomic detail. His methods would be very applicable to dynamical simulation of large cellular, materials or biomaterials systems.

Chandra Verma (A-Star) spoke on the Biopolis developments in Singapore that has major input and infrastructure from several large pharma companies. They have curated a large stem cell database that is potentially useful for our stem cell modelling project.

The abstracts may be accessed at http://www.chem.ox.ac.uk/mgms/abstracts.html.

David Winkler

[1] Foloppe *et al*., *Biorg. Med. Chem*., 2006, 14, 1792; 4792–4802

Cyberinfrastructure in Chemistry, Information and Education: New Emerging Technologies

This symposium was organised by the CIC–CINF Collaborative Working Group, and held at the 232nd ACS National Meeting in San Francisco in September 2006. Information scientists from Germany, the UK and the US discussed the impact of cyberinfrastructure on the chemical information community.

Evan Bolton (NIH) offered an excellent overview of PubChem. He discussed the role of this much talked-about online public information resource, the present status and the various links to important biological data. In 'Finding the right stuff', Debra Danville from AstraZeneca addressed the challenges facing researchers using text mining to find information in chemical and biological documents. René Deplanque discussed the efforts in Europe to create cyber networks to facilitate collaboration of research centres across Europe. He reviewed a national access system that deals with research information on ultra-fast laser optics. Scientists from 18 research institutions in 10 European countries combined the tools of their infrastructures to build a very large distributed superstructure.

Technical and social aspects of collaboratively developed information systems were discussed by Christoph Steinbeck from Cologne University. He talked about emerging technologies, such as Wikipedia, that allow communities to develop open information systems and how these developments can be adapted to the specific requirements of chemical information systems. Semantic web applications were the topic of presentations from three UK scientists. Omer Casher from GlaxoSmithKline talked about SemanticEye. Developed with Henry Rzepa from Imperial College, this application is based on the digital music metaphor and is intended to rationalise and enhance chemical electronic publishing. Jeremy Frey (University of Southampton) talked about CombeChem, an application providing e-science semantic support for the chemical life cycle, i.e. from inception in the laboratory to dissemination of data. In the absence of high-quality metadata, human intervention is still required, but it is hoped that the semantic approach will alleviate this by treating data and ideas by parallel but linked methods as in the concept of Publication@ Source. Peter Murray-Rust deplored the slow adoption of the cyber-revolution by the chemical community. Generating and disseminating datuments, documents semantically enriched through markup languages, should speed up adoption by both researchers and publishing houses. He mentioned early adopter projects in creating semantic chemistry, including the 'journal-eating robot'.

In the afternoon, speakers turned the attention to education. Jost Bohlen from FIZ CHEMIE Berlin described CHEMGAROO, an umbrella brand name for educational offers from FIZ. These include ChemgaPedia, an interactive encyclopedia for education in chemistry and related sciences. A large number of 3D animations, video and audio files enhance the value of the content of the individual, fully linked chapters. Barlow Culp (Mellon Library of Chemistry at Purdue) dealt with the need for a digital depository of chemical instructional material. He combined the discussion about the requirements with an outline on the progress of the pilot project being undertaken by the CIC–CINF Collaborative Working Group. The project aims to create a standardised, useful and adaptable digital depository of instructional material.

Irina Sens and Grace Baysinger from the German National Library of Science and Technology in Hannover and Stanford University Libraries, respectively, examined the impact of cyberinfrastructure on large research libraries. Irina discussed two portals, GetInfo and Vascoda. The first offers an integrated information infrastructure that is readily available, and the second is an interdisciplinary portal for scientific and scholarly information representing a strategic alliance between virtual libraries, information networks and the electronic journal library. Grace discussed the tremendous changes taking place in libraries worldwide and how technologies will shape research libraries of the future. Access to information is being revolutionised through mass digitisation and computing; the resulting changes are already seen in user behaviour and the roles and services offered by libraries.

The symposium was very well attended and the attendees' high interest in the presented topics was evident in the lively discussions after nearly every paper. Most of the presentations can be viewed on the CINF homepage at http://www.acscinf.org/.

Guenter Grethe
Co-chair, CIC–CINF CWG

An Introduction to Chemical Information

This one-day training course, held in November 2006 at Burlington House, London, was organised jointly by the CSA Trust and RSC-CIG. It continued the successful meetings held in previous years, with a series of presentations providing an impartial and critical overview of chemical information products, given by specialists from academic information departments, vendors and industry.

What makes chemical information different

David Walsh from Pfizer highlighted the characteristics of chemical information which distinguish it from many other disciplines – its timelessness, the sheer volume of information and, in particular, the need to record and process chemical structures, which exist in three dimensions. He illustrated these aspects by briefly reviewing the history of chemical information and the staggering statistics on the numbers of papers and compounds being produced. He outlined the various ways of recording chemical structures, including molecular formulae, notations, systematic nomenclature, trade names and connection tables, and touched on the complexities introduced by, *inter alia*, salts, mixtures, polymers, biological molecules, stereochemistry, isomerism and Markush structures.

Searching for chemicals on the internet

Teresa Loughbrough from Unilever covered searching on the internet. Teresa stressed that the internet was not always the best solution, though becoming more and more accepted. It is good for getting a feel for a new area or searching for very specific information. It is always worth trying different search engines, as they can give different results, and searching iteratively, i.e. using the results of initial searches to refine search methodology. There are useful resources on the web for evaluating websites.

Chemical Structure Drawing Packages

Don Parkin of the Chemical Database Service at Daresbury gave an overview of structure drawing packages. Each package of chemical information processing software tends to include its own drawing package. They all provide essentially the same functionality, such as drawing of bonds and atoms, selection from a set of templates, stereochemical indicators such as wedges, and the ability to fuse rings. However, each one has its own way of doing things and it can be tricky remembering how to perform a function in each package. It is particularly important to ensure the cursor is clicked in the correct position as even a small error can ruin a structure. Structures input can be subject to 3-D visualisation using packages such as Rasmol.

Overview of structure databases on STN

Diane Smith (STN) took the opportunity to introduce STN, the Science and Technology Network, which currently has 210 databases in the area of science and technology. STN is the major chemistry resource for information professionals, covering both patent and non-patent sources, and many of the databases are structure searchable. In addition, there are a number of databases that search for chemical reactions. Slides showed the running of searches by each search system (STN Express, STN on the Web and STN Easy) and the post-processing of results from the searches was described with an overview of the advantages of each. A list of the dates of future courses is available on http://www.rsc.org/Publishing/CurrentAwareness/STN/public.asp.

How to search for chemical structures – substructure searching

Barry Dunne from CAS gave an overview of chemical structure searching in the STN databases, and reminded us that, although each chemical has a unique structure, that structure can also be a member of a class of compounds, or an enantiomer or isotope, it can be part of a bigger molecule, or a salt; by structure searching, all occurrences of the structure can be found. The REGISTRY database, produced by CAS, is the world's largest collection of chemical substance information containing structures of organic compounds, inorganic compounds, sequences, polymers, metals and alloys; a number of STN databases can be structure searched. Most chemical structures can be constructed by drawing chains, rings or ring systems, then connecting drawn fragments and specifying the atoms and bond types. The package allows atoms and bonds to be easily defined and altered if appropriate. Barry went on to explain the uploading and searching of structures.

The RSC Virtual Library

Nazma Masud, a chemical information specialist based at the RSC Library and Information Centre (LIC) at Burlington House in London, gave an overview of the RSC's Virtual library. The library contains nearly 500 full text e-books as well as four bibliographic/abstract databases. For RSC members, it is free remote access to chemical information. Nazma demonstrated the access to the resource together with the searching and productivity tools from Knovel. An overview of the RSC library services was also given. This abstract does not give justice to Nazma's presentation nor to the services provided to the RSC members. The author recommends that readers should access and browse the Virtual Library themselves.

Introduction to reaction searching

Jeanette Eldridge (AstraZeneca) gave an introduction to Reaction concepts and the common terminology, explaining why the information is needed. Although some reactions could be searched textually ('Beckman rearrangement', Grignard), this is not always a reliable search method and further refinement may be needed such as the solvent or temperature appropriate to the compounds in question. There are many databases with reaction data, apart from those available from STN which may or may not need passwords and may or may not be free! Some of these were reviewed and methodology of collating search terms described.

We were pleased to welcome new faces and new speakers to this joint meeting of the Chemical Information Group of the RSC and with the CSA Trust. The sessions were all very different and well received and the organisers would like to publicly thank everyone for coming.

The CIG is merging with the Chemistry Computer Applications Subject Group in the New Year (Chemical Information and Computer Applications Group – CICAG). Slides covering these brief session overviews can be found at http://www.rsc.org/Membership/Networking/InterestGroups/ChemicalInformation/meetings.asp .

Doug Veal

SPONSOR'S SPOTLIGHT

Elsevier MDL

For over 25 years, Elsevier MDL has been providing the informatics infrastructure, workflow applications, reference databases and tools for decision support that scientists need for efficient R&D. MDL solutions — for collecting, cataloguing, analysing and sharing data, managing experimental workflows and supporting decisions — help scientists bring new compounds quickly and cost-effectively to market.

With the recent consolidation of the Elsevier MDL UK User Group with the European Users' Group, the UK User Group has donated over £4000 to the CSA Trust to support the Trust's work in promoting the use, development and training of students and others in chemical informatics.

Data on demand. Power for decisions
The MDL Isentris informatics system integrates information and applications into scientific workflows to improve R&D productivity. Organisations can use Isentris to give scientists access to internal and external data sources, integrate with specialised tools and develop and extend software applications, creating research environments that support workflows and empower decision makers.

Data access and analysis
The MDL Isentris client enables scientists to browse, collate and share molecule and reaction information in an efficient, time-saving manner that is integrated with their workflows. A unique reaction planner enables the researcher to create retro-synthetic schemes and quickly evaluate the utility of given reactions in the synthesis of new target molecules. The DiscoveryGate content platform is key to integrating MDL, Elsevier, third-party and customer proprietary content in the context of researchers' workflows. With a single query, DiscoveryGate provides access to 22 databases covering synthesis, bioactivity, physical properties, pharmacology, metabolism, toxicology and chemical sourcing. For researchers engaged in drug safety, the PharmaPendium database combines the best sources of preclinical animal and clinical human data, including FDA approval packages on approved drugs, enabling researchers to search and sort these data in a workflow-intelligent manner.

Synthesis planning and experiment management
The DiscoveryGate platform also gives chemists a comprehensive picture of the available reaction information needed to solve synthetic problems or plan suitable synthetic routes. The online platform provides access to numerous synthetic methodology databases and authoritative reference works on synthetic methodology. MDL Notebook is an electronic lab notebook that lets chemists plan experiments, capture data and share research results in a consistent, legible fashion utilising MDL's premier chemical representation and reaction searching capabilities.

Procurement and logistics
By optimising chemical inventory management, research organisations can drive down operating costs, reduce cycle times, satisfy regulatory requirements and increase candidates in the pipeline. MDL Logistics is a complete solution for managing the reagent procurement and inventory process. It incorporates in-house inventory data with the MDL Available Chemicals Directory (the world's largest electronic collection of chemical supplier catalogues) and the MDL OHS Pure Substances Database (an electronic encyclopedia of product safety information and SDS sheets).

Bioassay planning and management
MDL solutions improve biological research by providing simplified access to chemical, biological, pharmacological and ADME/Tox information and powerful applications for managing experiments and analysing data. The MDL Assay Explorer biological data management system enables biologists to capture, analyse, store and integrate biological and chemical data, along with experimental results and methodologies, in a central location. Likewise, MDL Plate Manager is a central repository for plate and sample information that integrates with MDL data management and chemical registration tools.

Application development
Efficient application delivery supports ever-changing scientific processes. With Isentris controls, developers can customise workflows and quickly build applications that have a consistent look-and-feel, are easy to learn and include powerful functionality such as list logic, history trees and user-generated forms.

MDL Isentris addresses the integration challenges of modern scientific research by bringing together workflow processes, internal/external data and applications from different vendors to help scientists collaborate, work smarter and make informed, timely decisions. MDL Isentris 2.0 will be released in December 2006. For more information, the MDL website is at http://www.mdl.com.

EVENTS 2006–2007

2007

January		
22–24	International Conference on Chemoinformatics, National Chemical Laboratory, Pune, India Postponed from August 2006	http://moltable.ncl.res.in/icci/index.html E-mail: icci2006@ncl.res.in
29–31	15th BOBCATSSS Symposium, Prague: Marketing of Information Services	http://www.bobcatsss.org/html/index-htm.htm
February		
26–27	Elsevier MDL 2007 European Users' Group Meeting, Hilton Strasbourg, Strasbourg, Alsace, France	http://www.mdl.com/company/events/user_conferences/2007/eugm07/index.jsp
27–28	Smart Lab Exchange 2007 Millennium Mayfair, London, UK	http://www.iqpc.co.uk/cgi-bin/templates/genevent.html?topic=233&event=11639
March		
4–8	International Patent Information Conference & Exposition, Hilton Sorrento Palace, Sorrento, Italy	http://www.IPI-ConfEx.com E-mail: searches@tprinternational.com
12–15	Drug Discovery Technology Europe 2007 Novotel London West, London, UK	http://www.drugdisc-europe.com/ E-mail: simon.lau@informa.com
25–29	233rd American Chemical Society Meeting and Exposition, Chicago, USA. Includes CINF–CSA Trust symposium, 'Advanced mining and use of life science information'	http://www.acs.org/meetings/ E-mail: natlmtgs@acs.org
April		
24–25	Discovery Knowledge & Informatics 2007, The Grand Amsterdam Sofitel Demeure Hotels, Amsterdam, The Netherlands	http://www.iqpc.co.uk/cgi-bin/templates/genevent.html?topic=237&event=12039&
June		
18–20	CSA Trust/Molecular Graphics and Modelling Society 4th Joint Sheffield Conference on Chemoinformatics University of Sheffield, UK (see page 12)	http://cisrg.shef.ac.uk/shef2007/ E-mail: cheminf2007@sheffield.ac.uk
July		
8–11	AIMECS07 Congress, Istanbul, Turkey	http://www.aimecs07.org
August		
5–11	41st IUPAC World Chemistry Congress, Turin, Italy	http://www.iupac2007.org/
19–23	234th American Chemical Society Meeting and Exposition, Boston, USA	http://www.acs.org/meetings/ E-mail: natlmtgs@acs.org
23–25	12ACC, Chemistry for Development, Environment and Sustainability in Asia, Kuala Lumpur, Malaysia	http://www.ikm.org.my/12acc.htm
October		
21–24	ICIC International Conference for Science & Business Information, Sitges, Barcelona, Spain	http://www.infonortics.com/chemical/index.html E-mail: contact@infonortics.com

Product News

Elsevier MDL introduces NEMA

Elsevier MDL has developed a new algorithm for the determination of stereochemical configurations. The new method, NEMA (Newly Enhanced Morgan Algorithm), offers a major improvement in stereochemistry perception, by extending the recognition to atropisomers, for example, biaryls and allenes. NEMA also improves on existing capabilities for handling tetrahedral stereogenic centres and provides more reliable recognition of stereoisomers. The new method can be used to calculate equivalence classes for a given molecular structure, to perceive tetrahedral and geometric stereogenic centres, to verify non-tetrahedral stereogenic centres and to identify meso stereoisomers correctly, supporting both 2D and 3D stereochemistry perception.

The NEMA algorithm is used in the MDL Direct 6.0 data cartridge for chemical structure and reaction searching. The NEMA method will be the standard approach to stereochemical recognition used by all MDL products, including the soon-to-be-released MDL Cheshire 4.0. It is fully backwards compatible with existing representations and retains all search and retrieval characteristics of current systems.

MDL Direct 6.0

Elsevier MDL has released the MDL Direct 6.0 chemistry data cartridge, enabling researchers to register, search and retrieve structures and reactions stored in large databases entirely in the Oracle relational environment. The MDL Direct data cartridge is the database component of the MDL Isentris 2.0 system. This latest release also extends support to the Linux platform, thereby expanding the availability of MDL Direct data cartridge technology to a growing community of scientific researchers.

MDL Direct technology incorporates Elsevier MDL's premier chemistry searching capabilities including Rgroup, Sgroup, flexmatch, 2-D and 3-D exact and molecule/reaction substructure searching; structure similarity, super-similarity and sub-similarity searching; reaction similarity searching; comprehensive registration; and searching of tetrahedral stereoisomers and non-tetrahedral stereoisomers. New, improved indexing provides an impressive improvement in performance for scientists searching large reaction databases.

MDL Direct combines full reaction transformation searching, unsurpassed chemistry sophistication and proven performance on databases with over five million reactions and at least 20 million structures. As the database tier of the MDL Isentris system, MDL Direct accelerates structure and reaction searching for the Isentris system and for related applications, including MDL Registration (for building compound registries) and MDL Notebook (for managing laboratory workflows and data entry in an e-R&D environment).

To ensure a smooth transition to the new technology, straightforward conversion tools will simplify upgrading to MDL Direct 6.0. Further information can be fond on the website at http://www.mdl.com.

Advanced Protein Modeling

Advanced Protein Modeling is an innovative new discovery software product, which combines the best methods for homologue identification, sequence alignment and fast interactive comparative protein structure modelling. It is now available as an optional component of Tripos' SYBYL 7.3 product line, and is based on the FUGUE and ORCHESTRAR technologies, developed by Professor Sir Tom Blundell and his research team at the University of Cambridge, UK.

Comparative protein modelling predicts a protein's unique 3-D structure, representing the shape into which the protein naturally folds, and also indicates the protein's biological function. These predicted 3-D structures are essential for structure-based drug design when experimental data are unavailable.

Those wanting more information about Advanced Protein Modeling should visit http://www.Tripos.com/APMod.

Chemistry Central

Chemistry Central is a new service, publishing peer-reviewed open access research in chemistry from BioMed Central, the leading biomedical open access publisher. The Chemistry Central website currently features chemistry-related articles published in BioMed Central journals and independent journals utilising BioMed Central's open access publishing services.

Chemistry Central has just launched Chemistry Central Journal. This broad ranging open access chemistry journal is now accepting submissions, and the Editors invite authors to submit manuscripts using the online submission system.

Chemistry Central is planning to launch further chemistry-specific journals in the near future. If readers are interested in starting a new open access journal in chemistry please contact http://www.chemistrycentral.com/

SciFinder 2007

The new version of SciFinder from CAS offers users a range of new features. It is now possible to combine a current answer set with a saved answer set of the same type: reference, substance, or reaction. Users can move catalogue information for commercially available chemicals into spreadsheets for further manipulation, sorting, and filtering. Users will be able to capture a chemical structure from a substance answer display, then use it to search by structure in the CAS Registry database. Printing structures in thumbnail display format allows multiple chemical structures to be viewed in a grid arrangement for a comparative view. In journal literature searches, SciFinder users have new options to view full journal titles and use the full or abbreviated titles in the bibliographic software packages Reference Manager, EndNote and ProCite.
The website can be found at http://www.cas.org/SCIFINDER/scicover2.html.

CSA Trust & the Molecular Graphics and Modelling Society

4th Joint Sheffield Conference on Chemoinformatics

The Octagon Centre, University of Sheffield, UK,
18th to 20th June 2007

The conference will be held at the Octagon Conference Centre, with ensuite accommodation near the venue. The conference dinner will be held at Chatsworth House, Derbyshire (home of the Duke and Duchess of Devonshire) and includes a tour of the House.

Call for Papers

Papers are welcomed in all aspects of chemoinformatics. Possible topics include (but are not limited to):

- *High-Throughput Screening*: assay quality control; design of screening collections; systems based design
- *Virtual Screening*: docking and pharmacophore analysis, similarity and clustering methods; machine learning
- *Computational Methods for Lead Identification and Optimisation*: modelling and structure-activity methods; structure-based design; ADMET prediction
- *New Algorithms and Technologies*: data mining; searching methods; distributed processing; data handling and visualisation;
- *Case Histories*, incorporating practical experience of any of the above.

The programme will include about two dozen oral presentations, with extensive opportunities for poster presentations. Authors wishing to submit a paper should send a title and abstract (minimum 500 words) by 31 January 2007 to cheminf2007@sheffield.ac.uk, stating whether they wish to be considered for an oral or for a poster presentation. Successful submissions will receive notification of acceptance by 28 February 2007. The Organising Committee will seek to achieve a balance between the various areas of the subject and between new methodologies and successful applications of existing techniques.

Further details can be found at **http://cisrg.shef.ac.uk/shef2007/**

Thanks to our Corporate Sponsors

AstraZeneca
Cambridge Crystallographic Data Centre
Chemical Computing Group
ChemIndustry.com, Inc.
Digital Chemistry
FIZ-Chemie
GlaxoSmithKline
Hampden Data Services
InfoChem
Kilmorie Clarke Ltd
Elsevier-MDL
Pfizer
Questel.Orbit
Rusch Consulting Group
Sheffield Chemoinformatics Research Group
Thomson Scientific
Wendy Warr & Associates

Contact us

Chair of Trustees
Dr Andreas Barth
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen
Germany
Tel: +49-7247-808-450
Fax: +49-7247-808-133
E-mail: andreas.barth@FIZ-Karlsruhe.de

Secretary
David J Walsh
Pfizer Global Research and Development
Sandwich
Kent CT13 9NJ
Tel: +44-(0)1304-645619
E-mail: david.j.walsh@pfizer.com

Donations Secretary and Website Administrator
Dr John Holliday
Department of Information Studies
University of Sheffield
Sheffield, S10 2TN, UK
Tel: +44-(0)114-2222685
E-mail: j.d.holliday@sheffield.ac.uk

US Donations
Dr Peter Rusch
162 Holland Court
Mountain View
CA 94040, US
Tel: +1-(650)-961-8120
E-mail: pfrusch@aol.com

Treasurer
Dr Geoff Downs
30 Kiveton Lane
Todwick
Sheffield, S26 1HL, UK
Tel: +44-(0)1909-515444
E-mail: geoff.downs@digitalchemistry.co.uk

Chair – Public Relations
David Wild
Tel: +1-(812)-856-1848
E-mail: djwild@indiana.edu

Chair – Grants
Bonnie Lawlor
276 Upper Gulph Road
Radnor
PA 19087, US
E-mail: blawlor@nfais.org

Chair – Meetings and Training
Vacant

Chair – Fund Raising
Vacant

Published by the CSA Trust
Views expressed in the Newsletter are those of individual authors and not of the Editor

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

NEXT COPY DATE
31 March 2007