



The Royal Society of Chemistry

**Chemical
Information
Group**

NEWSLETTER

January 2002

Wednesday 14 April 2002 Burlington House, London

The Changing Face of Information Provision in an Ever Changing Chemical Industry

There has been much rationalization throughout the chemical industry over the past year or two. Some of this has resulted in the closure of information departments and the loss of centralized facilities such as libraries. In other organizations there has been a merger of information departments. This meeting will look at these changes from various perspectives.

This meeting was originally scheduled for October 2001. Full details will be mailed at a later date together with details of the Annual General Meeting which will be held

6th International Conference on Chemical Structures

Sunday 2 June to Thursday 6 June, 2002

CIG is again a co-sponsor of the latest meeting in this important series which will once again be held at Noordwijkerhout in The Netherlands. Further details are given elsewhere in this Newsletter.;

CIG Web Site

CIG has a web site which gives up-to-date information on CIG, including forthcoming meetings. The web site, which is organised by Don Parkin, can be accessed on:

<http://cds.dl.ac.uk/cds/CIG/cig.html>

This edition of the CIG Newsletter has been put together by Peter Rhodes.

Contributions for the next issue should be sent to: Peter Rhodes, 122 Long Lane, Attenborough, Nottingham, NG9 6BW or by electronic mail to pete.rhodes@ntlworld.com

6th International Conference on Chemical Structures

The 6th International Conference on Chemical Structures brings together an international group interested in handling chemical structures and related topics. Participants discuss research and development in the processing, storage, retrieval and use of chemical structures. The conference fosters cooperation among organizations and researchers involved in the merging fields of cheminformatics and bioinformatics. The conference opens with a keynote address on Sunday afternoon, June 2, and continues until lunchtime on Thursday, June 6. The main technical program is divided into separate plenary sessions, taking place

each morning. A new-product review session for commercial presentations and an extended poster session are part of the scientific program. An exhibition featuring both commercially available software and also software from research projects and Special Interest Group (SIG) sessions forms an integral part of the conference. The official conference language is English. The conference takes place at the Golden Tulip Conference Hotel, Leeuwenhorst, Noordwijkerhout, The Netherlands. This is a modern, comprehensive centre, which is easily reached from Schiphol airport. The Centre is in a quiet rural setting only 2 km from the dunes and 4 km from the beach. There are a number of recreational facilities available, including tennis courts, an indoor swimming pool, and bicycle rentals. Accommodations at the Center include single and twin-bedded rooms. All rooms have telephones.

Registration for the entire conference, including full board and four nights lodging, excursion and conference dinner, and conference proceedings is Euro 975 (currently about £600) for single room accommodations. Student and early registration discounts are available. Full details, including a registration form, can be found at <http://www.chemweb.com/docs/6iccs/6iccs.html>

Tony Kent Strix Award

The Strix Award is in memory of Dr Tony Kent who died in 1997. Tony made a major contribution to the development of information science and information services in the UK and internationally, particularly in the field of chemistry. The Award is given in recognition of an outstanding practical innovation or achievement in the field of information retrieval. This could take the form of an application or service, or an overall appreciation of past achievements from which significant advances have emanated. The Award is open to individuals or groups from anywhere in the world.

Nominations for the 2002 Award are now invited, and should be sent, with full supporting documentation to: Alan Gilchrist, Editor, Journal of Information Science, 32 Friar Road, Brighton BN1 6NH. The closing date for nominations is Friday 16th June.

Nominations will be judged by a panel of experts, and the statuette of an owl will be presented to the winner in September. The name Strix was chosen both to reflect Tony's interest in ornithology, and the name of one of the last and most successful information retrieval packages which he created.

The Award for 2001 was awarded to Professor Peter Willett of the Department of Information Studies, Sheffield University, who is one of the best known people working in information retrieval today. He is the author or editor of 14 books in the field, as well as more than 300 other publications. He is equally active in teaching and administration. His interests within Information Retrieval are extremely wide-ranging, so that there is almost no area of research which he has not at some time investigated.

Particularly noteworthy has been his work on document clustering, linguistic analysis, term extraction and indexing, and basic algorithm design for retrieval processes.

His outstanding contribution however has been in chemoinformatics, in which subject he stands, perhaps, as the world leader. His PhD work in the mid-seventies was in this field, and he has devoted the past quarter century to its study. His work combines theoretical achievement with practical utility. The chemical and pharmaceutical industries have benefited through his collaboration with them, and these are industries which bring benefits to us all. His teaching in Chemoinformatics has been noteworthy. Over 40 research students have gained higher degrees under his supervision, and he has recently established at Sheffield an MSc in Chemoinformatics, the first such course of its kind.

Peter's name is indissolubly connected with Sheffield University, where he has been since 1976. The rise of its Department of Information Studies to become the leading centre for IR studies in England is in great measure due to his continued presence there.

New from Chemical Abstracts Service (CAS)

CAS COMPLETES "SCIENTIFIC CENTURY" PROJECT TO PROVIDE EASY ACCESS TO NEARLY 100 YEARS OF RESEARCH REPRESENTED IN CHEMICAL ABSTRACTS

CAS has significantly expanded the digital research environment for scientists by making the bibliographic and abstract information from the entire Chemical Abstracts (CA) collection back to 1907 available for searching through STN, SciFinder and SciFinder Scholar research tools. CAS has made the historic literature available in the same online databases that cover the latest research. Now available in the CA and CAPlus files are 3.8 million records from CA issues prior to 1967. This material brings the total number of online records to 20.5 million spanning 1907 to the present. About 789,000 of the earlier records are for patents while 2,855,000 are journal material. The balance are for books, technical reports, conference proceedings, and dissertations.

CAS began in 1999 adding material from pre-1967 CA issues to CAS' online databases. The massive undertaking required converting to electronic form a huge volume of information that was previously available only in print. With the completion of the project, researchers can use STN or SciFinder services to search this earlier literature via words in the abstract text and title, the publication title, author names, publication year and more.

CAS LINKS WEB JOURNAL CITATIONS TO RELATED CAS DATABASE INFORMATION

Finding chemical substances and citations related to scientific journal articles on the Web has become much easier with the introduction of ChemPort's new "Enhanced Reference Linking" service. With the latest ChemPort enhancement, researchers can derive much more value from the articles cited in journal articles on the Web sites of participating publishers. In addition to viewing a CAS record for the article or its full text, users of the new CAS service can choose to see the chemical substances discussed in the cited article and indexed by CAS scientists or a list of the "citing" documents that refer to the article.

Publishers for which Enhanced Reference Linking is now available are ACS Publications, Academic Press, Blackwell Publishers, the American Institute of Physics (AIP), Institute of Physics (IOP), International Union of Crystallography (IUCR), the Royal Society of Chemistry (RSC), Springer-Verlag, and Catchword. In total, 135 publishers and more than 2700 journals participate in ChemPort. The ChemPort Connection is incorporated into all CAS electronic search services, including STN Express, STN Easy, STN on the Web, SciFinder, SciFinder Scholar and CA on CD. Researchers identifying relevant articles or patents by searching CAS and other databases on the STN International network can immediately link from the database record to the associated full-text document on the Web sites of participating journal publishers or patent offices for no additional charge. A list of those participating in ChemPort is available at <http://www.chemport.org>.

The Chemical Information Sources Discussion List, CHMINF-L

CIG continues to be a sponsor of The Chemical Information Sources Discussion List, CHMINF-L, a discussion list for all those involved in any way with chemical information. The list is organised by Gary Wiggins, Head of the Indiana University Chemistry Library. The list serves as a forum for discussion of, and an information source for, chemistry reference questions and the sources used to find information needed by chemists. Members of the CIG are entitled to use it. You need to register, but it is free.

To subscribe, send mail to LISTSERV@LISTSERV.INDIANA.EDU with the command: SUBSCRIBE CHMINF-L.

The Cambridge Structural Database:

A quarter of a million crystal structures and rising

On October 5th 2001, the 250,000th crystal structure was archived to the Cambridge Structural Database (CSD), marking another milestone in the ever-accelerating growth of the world's repository of small molecule organic and metal-organic crystal structures. The CSD occupies a central place in the "spectrum" of structural databases. Complementary to the CSD are the ICSD, covering inorganic and mineral structures, the MDF, containing metal and alloy structures, and of course the PDB, in which protein and protein-ligand complex structures are deposited.

The Cambridge Structural Database began to be compiled in 1965, by the research group of Dr Olga Kennard in the Chemistry Department, University of Cambridge. In those days, there were perhaps a couple of hundred relevant structures reported in the literature each year, and the Database was collected together relatively easily and made available to academics worldwide. However, the number of structures entering the Database each year grew quickly, as did the interest in it from both academia and industry, and as did the effort required to keep it current, accessible and fully searchable. Ultimately, in 1989, a new organisation, the Cambridge Crystallographic Data Centre (CCDC), was spun out of the University as an independent not-for-profit institution, to continue collating and distributing the CSD. Funding for the CCDC comes not from government grants and subsidies, but very largely from the industrial subscriptions to the CSD – companies pay an annual fee to obtain local copies of the CSD System on each of their research sites. This enables academics to obtain copies for a much reduced fee.

Now, nearly 800 published papers describe uses of the CSD to study chemical bonding, molecular conformations, hydrogen bonding and other intermolecular interactions, with particular emphasis on molecular and drug design. The CSD has subscribers at almost 120 industrial sites and nearly 1000 academic institutions, covering 56 countries worldwide. CCDC also, alone or in collaborations, now produces and markets some applications programs, which make use of crystallographic information, either directly or culled from the CSD, to approach real research problems of academic and industrial relevance.

The 250,000th CSD entry describes the structure of an organic compound, the product of a photocyclization reaction. It has been assigned the CSD reference code IBEZQA and was one of two structures published in *Angewandte Chemie (International Edition)*, 2001, 40, 577-579. The paper's authors are A.G. Griesbeck, W. Kramer and J. Lex, of the University of Köln, Germany.

Growth of the CSD has remained remarkably constant at about 10% per annum over recent years.

Assuming that this level of growth is maintained, the CSD will reach half a million entries in 2010. It is not unreasonable to expect this rate of growth to accelerate, however, with the advent of faster, electronic publishing mechanisms for crystal structures such as that employed for *Acta Crystallographica E*, and ever-improving technology in terms of detector hardware and computing, vastly increasing the productivity of diffractometers. It is an ongoing challenge for the CCDC to keep pace with these developments, but one which it has thus far proved equal to. So it is by no means impossible that the CSD will double in size once again in rather less than the 9 years mentioned above!

<p>We regret to announce the death of Bob Wood who served for many years on the CIG Committee including a spell as treasurer. Bob lost his fight against prostate cancer on January 6th, 2002.</p>
--