Chapter V
Databases/Data Sources

Introduction

To a large extent, chemistry is still an empirical science, building its progress on an ever increasing flood of data and information. Thus, it was realized quite early on, that this flood can only be managed by storing it in electronic form. With more than 20 million compounds known, who could know them all? And this flood constantly increases, more than one million new compounds each year, more than 500 000 publications per annum that deal in one way or other with chemical information.

Chemistry was therefore one of the first scientific disciplines to start building databases to store its treasure trove of information. And, today, a effective overview of chemical information can be gained only by accessing databases.

In Chapter V, Section 1 Gary Wiggins introduces us to the variety of databases available in chemistry. Bibliographic databases are described in Chapter V, Section 2 by Andreas Barth. Because chemical structures are the language of chemistry, databases of chemical structures as presented by Gregory Paris in Chapter V, Section 3 play a central role. Greg also presents an interesting overview of retrieval strategies used to gain access to the information required, a topic that is further discussed in Chapter VI.

The most comprehensive chemical information system is that built by the Chemical Abstracts Service, outlined with its various components in Chapter V, Section 4 by Bill Fisanick and co-authors. The largest information system on organic compounds, which concentrates primarily on providing factual data is the Beilstein Database, described in Chapter V, Section 5 by Sandy Lawson.

The various databases available in inorganic chemistry are discussed in Chapter V, Section 6 by Jürgen Vogt and Axel Schunk.

Molecules are three-dimensional objects and, therefore, a knowledge of their three-dimensional structure is essential for many applications in chemistry. All small organic and organometallic molecules whose 3D structures have been determined experimentally are stored in the Cambridge Structural Database described in Chapter V, Section 7 by Frank Allen. The 3D structure of macromolecules such as proteins and nucleic acids are stored in the PDB database which is discussed, with a variety of applications, in Chapter X, Section 4.10.

An enormous amount of information on chemical reactions has been stored in reaction databases – the largest of which contain several million individual reactions. These are discussed in Chapter V, Section 8 by Engelbert Zass.

Spectra play a central role in structure elucidation. Despite this the number of spectra stored in spectral databases is still very small compared with the number acquired each day in analytical laboratories across the world. Such databases nevertheless include highly valuable information, as is shown in Chapter V, Section 9 by Antony Davies and Reinhard Neudert.

The conservation of the environment has become of increasing concern. Information on the properties and fate of chemicals introduced into the environment is therefore indispensable. Rapid access to this information can be provided by databases, as is illustrated in Chapter V, Section 10 by Kristina Vogt.

To file a patent it is essential to know whether a certain compound has already been made or claimed to have been made in another patent. For this task, patent databases have become indispensable tools; they are presented in Chapter V, Section 11 by Jürgen Vogt.

With the contribution in Chapter V, Section 12 on databases in biochemistry and molecular biology Alexander von Homeyer and Martin Reitz touch a topic that is also very important in bioinformatics. This contribution shows there is no clear-cut separation between chemoinformatics and bioinformatics – rather, there is a smooth transition between these two fields and each can benefit from methods and insights gained in the other.

The contribution ''Chemistry in the Internet'' in Chapter V, Section 13 by Alexei Tarkhov is a special case. The internet presents such a wide range of opportunities in chemistry – finding and sharing information, building databases, processing data, performing computations, etc. – that it is difficult to put this contribution into the rigid scheme chosen for this book. Because the sharing of information is probably the most important aspect of the internet, we have put this contribution in the chapter on databases and data sources.

The contribution on laboratory information management systems (LIMS) in Chapter V, Section 14 by Markus Hemmer emphasizes the importance of data quality and integrity, and how the management of different sources of data must be integrated into the daily workflow in industry and research institutes.

Unfortunately, other types of database could not be mentioned here, either because of lack of space or because we could not find an author. We particularly regret that databases of thermochemical or thermodynamic information are not mentioned. Databases on chemical engineering information or on chemical safety information are also missing. It is our hope that the interested reader will, nevertheless, find access to such sources of information if he/she needs it. Contributions can be found in the *Encyclopedia of Computational Chemistry* – by T. Droege on ''Databases on Chemical Engineering Information'' and by R. Sass and U. Westhaus on ''Databases on Chemical Safety Information''.

Nothing is said about databases containing results from quantum chemical calculations. This is not for lack of an appropriate author but because of the lack of such databases. Enormous amounts of computer – and human – resources are spent performing quantum mechanical calculations but the results from such calculations are not stored in a database with retrieval functionality. If this were done such results of quantum mechanical calculations could be disseminated to a wide

audience, and if such results could be retrieved according to different criteria (e.g. ''give me all structures containing an OH group'') one could perform systematic analysis (e.g. ''what is the range of polarity in the O–H bond of all those structures'').

It seems that the fields of theoretical chemistry and chemoinformatics are still worlds apart!

Gary Wiggins is Head of the Chemistry Library at Indiana University, Director of the Programs in Bioinformatics and Chemical Informatics in IU's School of Informatics, and Coordinator of the IU Bloomington Science Libraries.

1
Overview of Databases/Data Sources

Gary D. Wiggins

1.1
Introduction

That we are in the midst of a revolution in scientific publishing has been stated so often in recent years that it has become a cliché. Yet it is true that the influx of electronic journals and reference tools in certain areas of chemistry has completely obviated the need to physically visit a library to consult the literature. In chemistry there is a growing inter-connectedness of laboratory activities with computational chemistry, molecular modeling, chemical analysis, database searching, and information retrieval. The increasingly interdisciplinary nature of science in general has led to virtual collaborative efforts in which scientists take full advantage of the internet and new information technologies. These have enhanced and transformed the way research and scholarship is now conducted, allowing scientists to work in collaboratories or knowledge networks that draw on new approaches to investigating and analyzing concepts and phenomena.

Published research findings about modern chemistry date from the 18th century and have increased tremendously in volume since World War II. In 1946 the Chemical Abstracts Service (CAS) included in its flagship publication, *Chemical Abstracts* (*CA*), references to 39 578 documents; in 2001, the number was 755 915 [1]. Despite the enormous size and complexity of the chemical literature, there have been extraordinary contributions to its organization and accessibility in the last half of the twentieth century. In some cases there are also quite good and reliable free resources on the internet that chemists can use. However, the accepted opinion in chemistry is that the really valuable, comprehensive databases are products of the commercial sector.

1.2 Commercial Database Vendors and Databases

Control of chemical knowledge became much more feasible in the last third of the past century when there emerged commercial database vendors (information providers) who leased the databases that were beginning to appear in the commercial sector. Thus was created a partnership between the bibliographic abstracting and indexing (A&I) services and the vendors of online services that persists to this day. The computer search techniques developed by early vendors such as Lockheed Information Systems (later Dialog) and System Development Corporation (later Orbit) led to the rapid adoption of online searching in the 1970s.

1.2.1 Common Features of Vendor Systems

It is remarkable that the basic features of the database vendor systems have in some respects remained relatively static over the last several decades. All of them started with simple command-driven systems, with an increasingly larger range of command choices to build on the basic functions of choosing a database, entering a search, and displaying the results. The use of inverted dictionary files to enhance retrieval is common on all of the systems. By using a ''neighbor'' or ''expand'' command, one can choose the relevant options for a search strategy.

1.2.1.1 Front-End Search Software

Although command-language searching is often preferred by professional searchers, all of the major vendors have developed front-end search systems that make it easier for the novice or occasional searcher to use their systems. These products may be web-based or depend on client/server software. Examples include Questel·Orbit's QWeb and Imagination, MDL's CrossFire Commander, and from STN International, STN Express with *Discover!*, STN on the Web, and STN Easy, a web interface designed for the novice. There is considerable opposition in the library community to publisher-based searching and publisher-based table-of-contents delivery, but end-users do not seem to be as troubled by these options. Users continue to press for search systems that are easy to use, seeking consistent system interfaces and search capabilities across publishers. However, the simplistic search tools found on the web today are often incapable of multiple set creation, lack standard search features (truncation, proximity, etc.), and do not permit the precision of detailed field searching.

1.2.1.2 Database Search Costs

Typically the price the vendor charges the searcher for access to databases takes into account the time spent on the host computer, the royalty fees that must be paid to the database producer, and telecommunications charges. Additional charges are imposed on some systems for the number of search terms entered in the strategy or for special commands. Fees usually range from a relatively small cost for a minimal display of a record (perhaps only the abstract numbers) to a

quite expensive cost for the display of the full record. Many vendors also now offer subscription-based access. In many instances, the end users of these systems will be unaware of their costs, which are covered by their institution.

1.2.1.3 Data Analysis Tools

Increasingly, vendors are packaging their search options with tools for analyzing, visualizing, or interpreting the results. STN's ''analyze'' and ''tabulate'' commands are such options. In addition to an analyze option with SciFinder itself, the product can be linked to SpotFire Decisionsite for lead discovery.

1.2.2 STN International and CAS Databases

STN International, a cooperative service of CAS, FIZ Karlsruhe, and the Japan Science and Technology Corporation, now has over 200 databases, covering topics such as chemistry, engineering, life sciences, biotechnology, regulatory compliance, patents, and business [2]. Among the databases are some that allow searches for biosequence information, such as DGENE (Derwent Geneseq), GENBANK, and the CAS Registry File. STN International also offers CAS files that take full advantage of the many specialized chemical indexing techniques developed by CAS. For example, the CAS Registry Number, the unique number assigned to compounds and mixtures indexed by CAS, may be used to search over 60 different databases on the STN system. An example of this simple numeric search key is the CAS Registry Number for Isatin: 91-56-5. Other search systems also make use of CAS Registry Numbers.

CAS databases on STN include the abstract data from *Chemical Abstracts* in the CAplus and other databases, in addition to full structure searching capabilities in the CAS Registry File. Included in the Chemical Abstracts database are references to journals, patents, preprints, technical reports, dissertations, books, and conference proceedings in all fields of chemistry. Chemistry has often been called the central science because it overlaps with so many other scientific disciplines. Of course, many of the documents in the CA database are also of interest to scientists in medicine, geology, physics, etc.

Toward the end of 2002, the CA bibliographic files contained abstracts for over 20 million documents from 1907 onwards. CAS is now adding indexing from the printed CA volumes to the pre-1967 literature online, starting with the period 1962–66. Up to now, the earlier CA records were available, but without the thorough indexing of subjects and substances found in the printed CA volumes. The Registry File has more than 43 million substance records, over half of which are for biomolecular sequences (21.3 million nucleic acids and 1.5 million proteins in September 2002). The addition of pre-1967 indexing is expected to add many substances that were not in the Registry File before the start of that project. For over 800 000 compounds, experimental physical properties for boiling point, melting point, density, optical rotatory power, and refractive index have been added to the Registry File entries. In addition CAS has added millions of calculated properties using CAS substance connection tables and software developed by ACD.

Another CAS database, CASREACT, covers chemical reactions from 1907 onward. References from 1985 are supplied by CAS, but the file also includes earlier records supplied by InfoChem and the French organization INPI (Institut National de la Propriete Industrielle). CHEMCATS combines the catalogs of over 600 chemical suppliers, while CHEMLIST covers regulated chemicals from national inventories and national and international regulatory authorities. CAS also produces the database MARPAT, containing the Markush structures that lead to prophetic and generic structures in patents. Markush structures are imprecisely defined chemical structures found in the patent literature. This allows the inventor to claim patent protection for a large number of related compounds. Approximately 500 000 searchable Markush structures from 1988 onward are included.

Portions of the CA database are available from other vendors. On the Dialog system, records from 1967 can be searched in the CA Search files. In addition CAS Registry Numbers, nomenclature, and molecular formulas can be searched in the Chemsearch and Chemname files. Likewise on the Questel·Orbit system, the CAS database also dates from 1967. A CD-ROM version from 1977 to the present can be ordered directly from Chemical Abstracts Service.

1.2.3 SciFinder

With such a large and complex database, CAS saw the need to develop a software tool that would allow chemists who are infrequent or novice searchers of their databases to use the information sources effectively. The solution was SciFinder, an easy-to-use tool that now has an academic counterpart, SciFinder Scholar (Figure 1-1) [3].

SciFinder provides access to the databases produced by the world's most important abstracting and indexing service for chemistry and related disciplines – Chemical Abstracts Service (CAS) [4]. The ability to search for information on a chemical compound by drawing its 2D structure is one of the key benefits of the SciFinder interface to the CAS databases. Many electronic journal articles can also be accessed directly via the web from the references retrieved with SciFinder, provided the user is a subscriber to the journals. SciFinder is integrated with STN's ChemPort Connection, which in turn links to participating publishers' web sites for the full-text electronic journals. Thus, a search can lead from a reference in the CA database to the text of the original article without ever having to visit a library. Such links to electronic journals are becoming more common as abstracting and indexing databases and journal vendors (journal aggregators) forge new partnerships with primary journal publishers.

1.2.4 Other Vendors and Databases

Each vendor tends to specialize in certain areas. For example, Questel·Orbit has long been known for its chemical patents coverage, although STN in recent years has made significant strides in that arena. A special offering on the Questel·Orbit

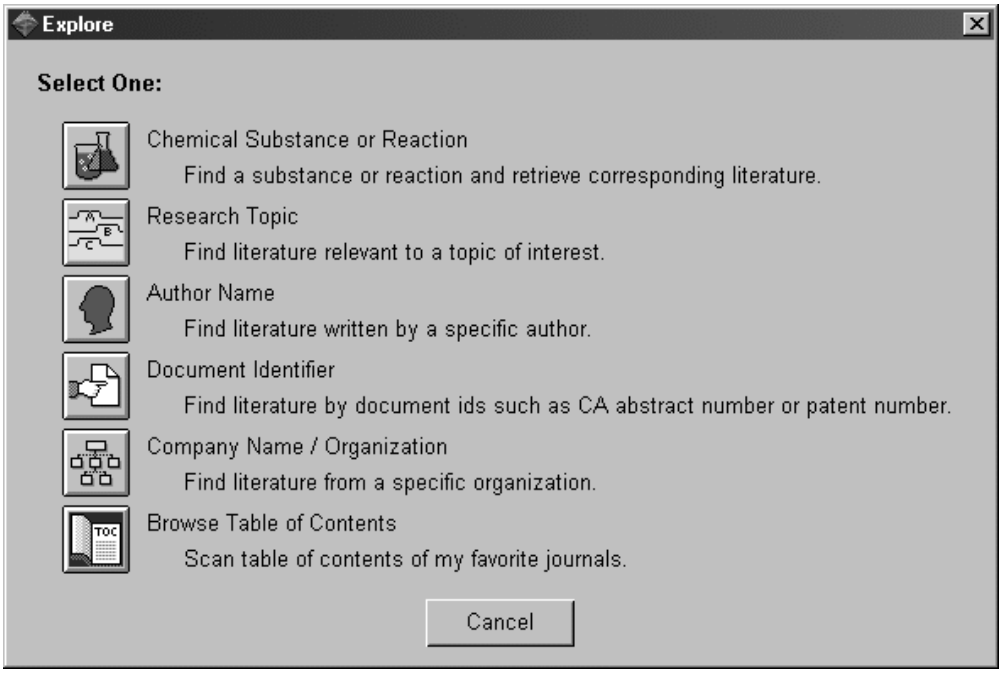

Fig. 1-1 Main search screen of SciFinder Scholar (reproduced by permission of Chemical Abstracts Service, a division of the American Chemical Society)

system is their Merged Markush Service that utilizes the structure search engine Markush DARC to search all chemical patents from 1987 and pharmaceutical patents from 1985. Markush DARC searches the Markush formulas databases MPHARM and IPAT (pharmaceutical and patent databases from INPI) and Derwent Information's comprehensive patent databases, WPIM and WPAT.

1.2.4.1 **Major Chemical Database Vendors**

Many chemical databases are found on search systems provided by other vendors. Other than STN and Questel·Orbit, those search systems most important for chemistry are:

- CSA (formerly Cambridge Scientific Abstracts) (*http://www.csa.com/*)
- Chemical Information System (*http://www.nisc.com/cis/*)
- Dialog (*http://www.dialog.com/*)
- knovel (*http://www.knovel.com/*)
- Ovid Technologies (*http://www.ovid.com/*)
- Questel·Orbit (*http://www.questel.orbit.com/index.htm*)
- SciFinder and SciFinder Scholar (*http://www.cas.org/*)
- STN International (*http://info.cas.org/stn.html*)

- Technical Database Services (*http://www.tds-tds.com/*)
- US National Library of Medicine (*http://www.nlm.nih.gov/*).

CSA has specialized environmental databases. The environment is also the focus of the Chemical Information System, the successor to the original US NIH/EPA CIS. Typical topics are site assessment, hazardous materials, material safety data sheets (MSDSs), chemical and physical properties, biodegradation and bioremediation, toxicology and carcinogenicity, regulations, pharmaceuticals, and spectroscopy. The system allows both structure and nomenclature searching. Dialog has traditionally been strong in news and business databases and patent searching. knovel's strength is in numeric database searches of hundreds of handbooks and other reference works. Ovid and NLM are best known for medical databases. TDS specializes in chemical engineering data. Included in their offerings is the American Institute of Chemical Engineers' DIPPR Pure Component Data Compilation, with 29 fixed-value properties and 13 temperature-dependent properties for about 1600 industrial chemicals. Taken as a whole, one can find on these vendors' systems files for subjects ranging from competitive intelligence to chemical reaction searching.

1.2.4.2 **Hybrid Publishers/Vendors**

A more recent development is the establishment of vendor-like systems that concentrate on the products of a particular publisher or database producer, while offering perhaps limited access to the offerings of other companies. An example is Elsevier's ScienceDirect, a system for searching the many electronic journals published by that company. Established in 1999, ScienceDirect also has a web search engine, Scirus, to expand the search to the internet [5]. (STN's eScience has a similar function and works with STN Easy, SciFinder, SciFinder Scholar, and CA on CD.)

The American Chemical Society now has all issues of its prestigious primary journals on the web. When the journals section of the ACS Publications Division web site is accessed, all issues of every ACS journal can be searched by authors' names or keywords [6]. Also searchable through this web site is DGRWeb, the Directory of Graduate Research in North American academic chemistry and related departments, and LabGuide Online, a buyer's guide for scientific laboratories.

Wiley is another company that is creating a search system to encompass all of its scientific and technical products, both journals and reference materials [7]. In the latter category are the venerable Kirk-Othmer Encyclopedia of Chemical Technology, Ullmann's Encyclopedia of Industrial Chemistry, the Encyclopedia of Reagents for Organic Synthesis (e-EROS), the Encyclopedia of Computational Chemistry, the Encyclopedia of Catalysis, and other standard reference works. Wiley InterScience's OnlineBooks collection includes a chemistry library among its offerings, as well as links to numerous electronic journals in chemistry.

Whereas the encyclopedias available from Wiley must be searched individually, a different concept underlies MDL and InfoChem's Integrated Major Reference

Works. There are links between MDL's synthetic methodology databases and reference works produced by other companies. Among these are Thieme's Science of Synthesis (successor to Houben–Weyl), Springer's Comprehensive Asymmetric Synthesis and their Glycoscience, Elsevier Science's Comprehensive Organic Functional Group Transformations, and Wiley's Encyclopedia of Reagents for Organic Synthesis, as well as links to primary journal literature. Reactions are indexed with InfoChem GmBH's Reaction Classification Code [8]. Results can be viewed in three groups – broad, medium, or narrow – on the basis of on the degree of specificity around the reacting center.

Another Elsevier company, Elsevier Engineering Information, has developed Ei ChemVillage to bring together products from Elsevier and other companies [9]. Included are Beilstein Abstracts and the Chimica abstracting service that covers over 600 of the top chemistry and chemical engineering journals. Furthermore CRC Press's CHEMnetBASE, with access to several important handbooks commonly found in chemistry libraries, is part of ChemVillage, as is the Chemical Business NewsBase.

ISI, the Institute for Scientific Information, has long produced an A&I database of interest to synthetic organic chemists, Index Chemicus. ISI has now moved more toward the role of a traditional vendor by creating the Web of Knowledge [10]. Because its parent company, Thomson, owns Derwent, a major patent A&I service, ISI can now draw on Derwent databases to enhance the Web of Knowledge offerings. Likewise, reaction searches can be done on the ISI Chemistry Server.

1.2.4.3 **Beilstein and Gmelin**

It should be noted that the coverage of most abstracting and indexing databases is limited to the last three decades or so (Chemical Abstracts being an exception). The CrossFire system, released in the mid-1990s by Beilstein, changed the perspective of many chemists: it made them rediscover chemistry before 1967 [11, 12]. Beilstein covers the literature of organic chemistry from its beginnings in the 18th century. With the CrossFire implementation, access is also provided to Gmelin, a database of inorganic and organometallic compounds that also extends to the 18th century. Users can now simultaneously search for facts and perform structure searches or reaction searches in both databases, covering over 8 million organic and approximately 1.4 million inorganic and organometallic compounds.

The original printed works that underlie the Beilstein and Gmelin databases bear the German word ''Handbuch'' in the title. A major use of the databases is, in fact, to discover if a property has been measured for a given substance since hundreds of such physical and chemical properties are indexed in the works. Thus, the files serve much the same function as one-volume handbooks, but on a far grander scale.

1.2.4.4 **knovel Databases**

Another source of numeric data is the knovel Engineering and Scientific Databases [13]. Very strong in engineering and polymer sources, knovel also has some standard works such as Perry's Chemical Engineers' Handbook, Lange's Handbook

of Chemistry, and the Royal Society of Chemistry's Dictionary of Substances and their Effects. knovel is especially good for searching by physical properties, enabling searches by a specific value or a range of values for a given property.

1.2.4.5 Handbooks, Encyclopedias, Physical Property Data Compilations

There are CD-ROM versions of various handbooks and larger compilations of data and facts, including the CRC Handbook of Chemistry and Physics and the Merck Index. Encyclopedias, such as the Kirk-Othmer Encyclopedia of Chemical Technology, originally released on CD-ROM, are now easily searched on the internet.

Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology is one of the largest collections of data in existence [14]. Containing data from the 19th century to the present, the online web file covers the material found in the hundreds of printed volumes that have appeared in the Basic and New Series. Planned is a pay-per-view option for accessing individual articles in the work.

Spectral databases are increasing in number [15a–d]. Aldrich offers both an FTNMR and an FTIR Library on CD-ROM. The IR collection has 11 000 compounds, and the NMR set contains 11 800 compounds in the standard library. Bio-Rad Laboratories offers NMR data on over 350 000 compounds in their HaveItAll NMR collection. Since 1996 Bio-Rad has no longer produced printed compilations of spectra.

1.2.4.6 Cambridge Structural Database

The Cambridge Structural Database (CSD) is the world's largest searchable database of experimentally determined crystal structures [16]. Produced by the Cambridge Crystallographic Data Centre, the CSD contains crystal structure information for over 250 000 organic and organometallic compounds analyzed using X-ray or neutron-diffraction techniques [17]. With the ConQuest interface to the CSD, there is now a very easy method of searching and displaying results on a PC.

1.2.5 Electronic Journals

By the end of 2002, it was rare to find a major primary journal in the chemistry, physics, biomedical and other bioscience areas not available electronically. Often even the very first issues ever published are among the web issues. The American Physical Society has digitized *Physical Reviews* to 1893, and the American Chemical Society's flagship *Journal of the American Chemical Society* is now totally online to 1879. All other ACS journals are also available back to volume one, issue one. Commercial publishers were quick to follow the lead of the professional societies. Elsevier plans to put backfiles of all of its journals online, and the chemistry journals are already done. Already links between references in online abstracting and indexing services and the original journal articles (and vice versa) are commonplace, and developing standards in the area of electronic journal linking provide avenues to an individual library's holdings in both print and electronic formats.

Utilizing metadata, the user may search for elements in an article that are not easily found today, such as the actual technique used in an experiment.

Publishers are now forging virtual journals that bring together articles on a particularly hot topic, e.g. the Virtual Journal of Nanoscale Science and Technology. Another example is the Royal Society of Chemistry's Chemical Biology Gateway (*http://www.rsc.org/chembiol*), with its virtual journal that provides free access to all relevant literature on this topic published in RSC journals for one month after publication. Much of the offerings on the so-called virtual communities (BioMedNet, ChemWeb, etc.) are in the form of electronic journal articles, sometimes enhanced with special features, such as Elsevier Science's Dymond (Dynamic Metadata ON Demand). Dymond links structures in articles in Tetrahedron or Tetrahedron Letters to the corresponding entries in Beilstein CrossFire. In addition, it can find similar compounds in the Combined Chemical Dictionary or locate NMR spectra or physical properties in ACD/Labs databases.

There is even a Chemistry Preprint Server on ChemWeb [18]. However, editorial stances vary by discipline or publisher on whether this is considered prior publication and therefore would result in a rejection of a manuscript. The Royal Society of Chemistry, the American Astronomical Society, and Elsevier Science allow it, whereas the American Chemical Society does not. A grassroots movement to force journal publishers to allow free web access to older issues of their journals after a period of time has had some impact, and the debate in this area will undoubtedly continue for some time to come.

The electronic journal is here to stay. The move toward the electronic version as the copy of record is inexorable. Some libraries already have such confidence in the future stability of the electronic journal that they have canceled the print versions of the journals and put their paper backfiles into storage, relying entirely on the web archive.

1.2.6 Free Internet Sources

There are still many examples of good web sites where reliable chemical information can be found at no cost [19]. The TORVS Research Group site features demo sites and prototypes for chemical information systems that link to databases and computational services [20]. Included is the NCI Database Browser Mark II to search the 250 000 compounds in the National Cancer Institute database. Among the offerings of the NLM is ChemIDplus, a chemical registry database with over 140 000 compounds of biomedical and regulatory interest [21]. Wiley's authoritative Organic Syntheses compilation is a free source covering over eight decades of synthetic knowledge [22].

A number of other reliable chemical databases can be searched free on the internet, but that situation may be changing. CambridgeSoft's ChemFinder searches hundreds of internet sites via a chemical name, CAS Registry Number, molecular formula, or molecular weight [23]. With their ChemDraw plug-in software, structure searching is also possible on ChemFinder. Now a commercial version of ChemFinder (ChemINDEX) is offered as an alternative to the free service.

The NIST Data Gateway provides easy access to scientific and technical data produced by the US National Institute for Standards and Technology, including free web databases [24]. Among the free databases is the NIST Chemistry WebBook which contains thermochemical data, reaction thermochemistry data, mass spectra, UV–Visible spectra, electronic and vibrational spectra, and constants of diatomic molecules. The ACS's Chemcyclopedia (available via *http://chemistry.org*) leads to commercial sources of chemicals by CAS Registry Number, chemical name or supplier name. Chemicals are also divided into categories such as oleochemicals, surfactants, or industrial and specialty gases.

A good source for patents on the internet is esp@cenet, the European Patent Office file with pdf images of patents, including US patents [25]. DEPATISnet is especially good for German patents [26]. Journal tables of contents are typically found on publishers' web sites, often including free access to the abstracts.

There are many other sources of chemical information on the internet. To discover them, use one of the comprehensive guides to chemistry resources on the net such as Chemdex, Links for Chemists, or ChemIndustry.com [27]. Mark Winter's Chemdex has over 7000 links in its database, whereas the University of Liverpool's Links for Chemists claims over 8000 sources. ChemIndustry.com's database has links for over 42 000 chemical industry entities [28–30].

1.2.7 The Future

It might not be long before computer databases and the internet eliminate most needs to consult a traditional chemistry library. The sources already available provide enough useful information that many chemists are turning to them first. The internet has fostered a new communication process that simply did not exist a decade or so ago. If a quick web search does not provide the answer, then news groups and listserves frequently give almost instantaneous answers to questions that would have taken days or weeks of research in the past. Nevertheless, there will still be times when a printed resource in a chemistry library is the only recourse. At such times, it is wise to consult a guide to the printed literature [31].

Even more exciting developments are appearing on the horizon, as the marriage of databases with molecular modeling and visualization techniques becomes more widely applied in internet chemical information sources. That trend is bound to accelerate as scientists and librarians come to realize the value added to the electronic resources as compared with their printed counterparts.

References

1 CAS Statistical Summary. *http://www.cas.org/EO/casstats.pdf*

2 STN: Your Connection to Science and Technology. *http://www.cas.org/stn.html*

3 K. Schwall, K. Zielenbach, *Chemical Innovation* **2000**, *30*(*10*), 45–50.

4 D. D. Ridley, *Information Retrieval: SciFinder and SciFinder Scholar*, Wiley, Chichester, England; New York, NY, **2002**.

5 Elsevier ScienceDirect and Scirus. *http://www.sciencedirect.com and http://www.scirus.com*

6 ACS Publications. *http://pubs.acs.org*

7 Wiley Interscience. *http://www.wileyinterscience.com/*

8 The InfoChem Reaction Classification Program V.2.5. *http://www.infochem.de/prod6unten.htm*

9 Ei ChemVillage. *http://www.chemvillage.org*

10 ISI Web of Knowledge. *http://www.isiknowledge.com*

11 MDL Products Overview: CrossFire. *http://www.beilstein.com/beilst_2.shtml*

12 P. Meehan, H. Schofield, *Online Information Review* **2001**, *25*(*4*), 241–249.

13 knovel Engineering and Scientific Databases. *http://www.knovel.com*

14 Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology. *http://www.springer.de/phys/laboe/*

15 (a) W. A. Warr, "Spectral databases: a review." *Analytical Proceedings* **1993**, *30*(*4*), 200–201; (b) W. A. Warr, *Anal. Chem.* **1993**, (*65923*), A1045–A1050; (c) W. A. Warr, *Anal. Chem.* **1993**, (*65923*), A1087–A1095; (d) R. Neudert, *Spectroscopic Databases in Encyclopedia of Computational Chemistry*, P. von R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, P. R. Schreiner (Eds), Wiley, Chichester, **1998**, *4*, 2632–2646.

16 F. H. Allen, *Acta Crystallogr., Section B: Structural Science* **2002**, *B58*(*3, No. 1*), 407–422.

17 Cambridge Crystallographic Data Centre. *http://www.ccdc.cam.ac.uk/*

18 J. R. Weeks, J. Kuras, W. G. Town, B. A. Vickery, *J. Chem. Inf. Comput. Sci.* **2002**, *42*(*3*), 765–766.

19 K. Voigt, G. Welzl, *Online Information Review* **2002**, *26*(*3*), 172–192.

20 TORVS Research Team. *http://www2.chemie.uni-erlangen.de/services/*

21 NLM's ChemIDplus. *http://chem.sis.nlm.nih.gov/chemidplus/*

22 Organic Syntheses. *http://www.orgsyn.org/*

23 ChemFinder.Com Database and Internet Searching. *http://chemfinder.cambridgesoft.com/*

24 The NIST Data Gateway. *http://srdata.nist.gov/gateway/*

25 esp@cenet. *http://www.european-patent-office.org/espacenet/info/index.htm*

26 DEPATISnet. *http://www.depatisnet.de*

27 C. Glander-Höbel, *Online Information Review* **2002**, *26*(*3*), 146–163.

28 Chemdex. *http://www.chemdex.org/*

29 Links for Chemists. *http://www.liv.ac.uk/Chemistry/Links/links.html*

30 ChemIndustry.com. *http://www.chemindustry.com/*

31 R. E. Maizell, *How to Find Chemical Information: A Guide for Practising Chemists, Educators, and Students*, 3rd edn, John Wiley and Sons, New York, **1998**.

32 A. Barth, "*Online Databases in Chemistry.*" in *Encyclopedia of Computational Chemistry*, P. von R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, P. R. Schreiner (Eds.), Wiley, Chichester, **1998**, *3*, 1968–1980.