Kirk-Othmer Encyclopedia of Chemical Technology.

INFORMATION RETRIEVAL

The literature of chemistry and associated fields has increased enormously since 1980. Establishment of subspecialties and newly defined disciplines as well as increased research output have led to an explosion of journals, books, and on-line databases, all of which attempt to capture, record, and disseminate this plethora of knowledge (1). Tertiary reference tools in chemistry and technology (eg, *Kirk-Othmer*, 4th ed.) help track the primary literature. Excellent references that discuss basic chemical information tools are *The Literature Matrix of Chemistry* (1), *Chemical Information Sources* (2), and *How to Find Chemical Information* (3).

Retrieval of chemical information will continue to be an issue of accessing what is available in the fastest, most cost-effective manner. Changes in the ways information is located and retrieved will be driven by technological advances in computer hardware, development of software, and progress in telecommunications. The resources available through Internet are increasing daily. Content of electronic databases has remained basically the same; what has changed are the tools to access the information. Publishers continue to explore the possibilities of electronic media. Ease of information access, for example, through the use of a natural language interface, or the ability to query multilingual databases using a single language, is another emerging issue. The Special Interest Group on Information Retrieval of the Association of Computing Machinery (SIGIR ACM) meets annually to address these and other information issues. Proceedings of their meetings provide an overview of advances in information technology and access. As technology becomes more complex, issues of ownership, copyright protection, reuse of retrieved information, and access costs will need to be examined and resolved (see Copyrights and trademarks, copyrights).

Libraries and information centers are rapidly moving from purchase and ownership of print and on-line resources to rapid access to these resources; a change in philosophy from "just in case" to "just in time." Libraries are making hard decisions about what to purchase and what to exclude because of the magnitude and cost of a complete collection of available information. These factors have forced severe cutbacks in what is actually bought as well as increases in cooperative purchasing and loan agreements between libraries on the local, regional, national, and even international levels. Increased computer power and technology have made geographic boundaries and limitations obsolete (4). Documents, such as journal articles, can be obtained quickly from other libraries and commercial document delivery vendors (5). Paper copy remains the preferred format for delivery of such documents. Documents can be delivered by regularly scheduled mail, by overnight delivery, or by facsimile transmission. Fax is the method of choice if extremely short delivery time is required; it is relatively inexpensive, ie, generally the cost of a phone call, paper, and supplies, when compared to courier and overnight delivery options. Documents created in an electronic format not only can be easily edited or updated but also can be transmitted rapidly via electronic mail to multiple users. Electronically created files can be loaded onto diskettes for delivery of information if an electronic mail system is not available. Another method of transfer of documents is via modem, which eliminates the physical transfer of diskettes, allows transmission of word-processed, non-ASCII (American Standard Code for Information Interchange) documents with no loss of format, allows transmission of non-English language documents with non-Roman alphabets, and permits rapid, unattended transmission. Speeds in excess of 14,400 bps are currently available.

Although electronic publishing of journals is in its infancy, more and more full text journals are becoming available on-line, allowing printing of a document from the user's computer. The limitation of ASCII format (text only, no graphics) is being addressed by vendors. The American Association for the Advancement of Science's publication *Online Journal of Current Clinical Trials*, which debuted in July 1992, supports text and nontext and publishes a paper within days of acceptance (6). The Research Libraries Group has produced ARIEL, a software package that allows image scanning of a document transmitted through Internet. ARIEL provides images and text of greater resolution than fax and uses standard personal computer (PC) hardware (7).

Location of and access to chemical and technical information other than journal articles is available through computerized information networks. Electronic bulletin board systems (BBS) provide a telecommunications tool to anyone who has a computer and a modem. Questions can be posted and read by thousands of bulletin board users worldwide, and files and software are easily transferred from virtually anywhere to one's computer.

1. Networks

The rise in popularity and use of Internet has dramatically changed the way information is disseminated. Internet is a worldwide link of thousands of separately administered computer networks of many sizes and types. Each of these networks is connected to as many as tens of thousands of computers; the total number of individual Internet users is in the millions. This high level of conductivity fosters an unparalleled degree of communication, collaboration, resource sharing, and information access. Electronic mail, or e-mail, is a fast, easy, and inexpensive way Internet users around the world can communicate with each other and with users of other independent networks such as CompuServe, Applelink, and the WELL. In the United States, the National Science Foundation Network (NSFNet), a very high speed network that connects key regions across the country, comprises the Internet backbone. The NSFNet will likely evolve into the National Research and Education Network (NREN) as defined in the High Performance Computing Act of 1991 (P. L. 102–194) (8).

Remote login is the ability of a computer user in one location to establish an on-line connection with another computer elsewhere. Once the connection is established, the remote computer is used as if it were a hard-wired terminal of that system. Within the Transmission Control Protocol/Internet Protocol (TCP/IP) suite, this facility is called Telnet. Using Telnet, an Internet user can establish connections with a multitude of library catalogues, other bibliographic databases, university information systems, full text databases, data files (eg, statistics, oceanographic data, meteorological data, and geographic data), and other on-line services. Many of these connections are available to any Internet user and can be accessed without an account.

Internet is remarkable in that ease and speed of access are not dependent on proximity. Users can connect to a network on the other side of the globe as easily as, and almost as quickly as, they can connect to a system in the next building. In addition, because many Internet users are not currently being charged according to their level of use, cost seldom inhibits usage. Therefore the barriers of distance, time, and cost, which are often significant when using other forms of electronic communication, are reduced in the Internet environment. Disadvantages include high initial costs for Internet connection and access that requires a computer and telecommunications.

Internet can also be used to transfer files from one computer to another. This function is provided by the File Transfer Protocol (FTP) of the TCP/IP suite. In a method similar to that used with Telnet, an on-line connection is initiated with another Internet computer via FTP. Unlike Telnet, this on-line connection can perform only functions related to locating and transferring files, such as changing directories, listing files, and retrieving files. Every kind of file that can be stored on a computer can be transferred using FTP: text files, software programs, graphic images, sounds, and files formatted for particular software programs, eg, files with word processing formatting instructions. Many computer administrators have set aside archives of files on

their machines that anyone on Internet can retrieve. These archive sites support anonymous logins, called anonymous FTP sites, which do not require an individual account to access. To locate files, Internet users can use the Archie service which indexes files from over 900 separate anonymous FTP sites (9).

The three basic Internet applications of remote login, electronic mail, and file transfer are also building blocks of more sophisticated applications that offer increased functionality and ease of network use. Tools such as Gopher, Wide Area Information Servers (WAIS), and World Wide Web (WWW) go beyond the three basic Internet functions to make information on the network easier to locate and use. Detailed descriptions of these tools are available (10). This trend toward more powerful, user-friendly networked information resource access systems should continue as Internet grows and matures.

News groups are a feature of Internet that allow for rapid, worldwide exchange of ideas with others interested in the same field. Most new groups are part of Usenet, the global news service whose topics encompass many areas of science, recreation, and social issues. Users subscribe to new groups of interest, then read and post messages to those groups. The news group "Sci" posts discussions of research marked by special and practical knowledge relating to established scientific disciplines.

1.1. Copyright

Any discussion of emerging technologies in storage and retrieval of information leads to a question of copyright compliance. Copyright in the electronic and computer age, using the internal and external networks and document delivery mechanisms outlined above, is a complex and unresolved issue. One of the primary reasons for using document delivery services, other than for obtaining information that is not locally available, is copyright compliance. Document suppliers generally handle payment of any required copyright fees. In the electronic and computer age, establishing who owns a particular piece of information can be difficult, as can determining what can legally be done with information once it is obtained.

Information stored electronically can be both textual and nontextual. It can be linked with other stored files, downloaded from a commercial database, and edited to create entirely new documents, which can be forwarded to others by use of electronic mail. Clearly technology is moving ahead of the law. Traditionally, copyright protection is extended to the original creative expression of an idea fixed in tangible media (11). Because of expanding multimedia use, it is becoming harder to define and apply existing copyright law to protect the original idea.

AT&T Bell Laboratories has built RightPages, a prototype electronic library, as a current awareness alerting tool. This pilot project illustrates many of the copyright problems encountered in the use of advanced technologies. Identified issues include time and cost involved in securing permissions from individual publishers, pricing issues, complexity of licensing agreements and the restrictions they impose, and administrative costs incurred in obtaining and managing all such licensing agreements (12).

The Copyright Clearance Center, which handles licenses for photocopying journals in paper format, has initiated a small electronic copyright pilot project in an attempt to address problems encountered in complying with the law in the computer age (13). Expansion of this initial project is expected. Changes to the traditional copyright law will be market-driven. Publishers and vendors of information who market products in nonpaper formats will need copyright protection while providing an affordable product to make information widely available (14).

1.2. Budgeting

These changes in the storage and retrieval of chemical information require that libraries and information centers now consider not only what should be purchased but also what monies should be allocated for the purchase of information in nonprint formats such as CD-ROMs (compact disk read-only memory) and on-line databases. Coupled with this is budgeting for the cost of hardware and software to enable the rapid and

cost-effective delivery of needed information (15). The geometric increase in sources, both printed and on-line, has increased the role of information specialist as an expert in the delivery of chemical information. Retrieval from increasingly diverse and complex sources becomes the paramount issue for searchers of chemical literature in the 1990s.

2. On-Line Database Resources

The on-line information industry has grown dramatically since 1972 when Dialog Information Services, Inc. (Dialog) offered the first publicly available commercial databases (qv). This service, comprising two databases, had fewer than 20 subscribers (16). The 1980s saw incredible growth in the number and range of on-line databases. Databases covering virtually all important subject areas were developed, and significant publications became available via thousands of bibliographic, abstract, textual, directory, and numeric databases. Although databases have existed in the form of bibliographic and legal files since the 1600s, the term database was not coined until the 1950s (17). For electronic databases to be made available publicly, development was required in three primary technologies: computers, communications, and databases themselves (18).

Though unstable, the computer industry grew rapidly during the 1960s, and the final piece of the computer development puzzle, time-sharing, came about late in the decade. Efforts to develop and commercialize time-shared computers were led by General Electric's computer department, which was quickly overtaken by IBM, UNIVAC, and Digital Equipment Corp.

The U.S. government, a primary sponsor of scientific and technological developments that fostered the computer and communications technologies needed by the on-line database industry, also sponsored database development projects, information usage studies, and combined computer database development–usage projects. The successors of some of these projects continue to be prominent and include DIALOG, MEDLINE, BRS, LEXIS, and the Chemical Abstracts Registry System.

During the early 1970s, the necessary telecommunications technology became available with packet switching. ARPANet, the first operational packet-switched digital communications network, was implemented by the U.S. Department of Defense. Commercial systems (eg, Telenet, TYMNET, and GENet) became available shortly thereafter.

In 1981, IBM introduced a low cost PC, which provided avenues for access to on-line databases by end users. In 1986 the president of Dialog noted that, although 85% of DIALOG's customers were information specialists or librarians, 80% of new DIALOG accounts were established for end users (18, 19). Users wanted the on-line industry to accommodate their needs and expectations, but the on-line industry did not recognize that the availability of large amounts of on-line information would not, of itself, induce people to use the information.

2.1. Database Producers

Producers of databases, also known as database publishers or information providers, determine the content of the databases, produce them, and typically lease or license them to private organizations or database vendors. Database producers may be categorized as government, not-for-profit, commercial/industrial, and mixed.

Government examples include the National Library of Medicine (NLM), and National Agricultural Library (NAL). Governments have long been sponsors of scientific and technological developments needed for nurturing the database industry (18). During the 1960s and 1970s, the majority of database producers were government organizations (20). Not-for-profit examples are Chemical Abstracts Service (CAS) and Biosis. Databases produced by academia and professional society-based not-for-profit (NFP) organizations became widely known during the late 1960s and 1970s. Together with government databases, they continue to maintain their value as resources, especially in the sciences. Additionally, some NFP organizations are chartered to

disseminate information at little or no cost (21). Some of the commercial databases are founded on government data, often collected by a government agency at substantial cost. They began collecting data to fill the needs of various industries. The United Nations and the European Economic Community offer mixed examples. Many government and nongovernment information sources are coordinating efforts to disseminate information (22).

The percentage of both government not-for-profit databases decreased between 1977 and 1992, whereas the percentage of commercial databases increased. In 1977, 56% of all databases were produced by government agencies, 22% by not-for-profit groups, and 22% by commerce/industry. In 1992, 75% of all databases were produced by commerce/industry, 15% by government groups, and 9% by not-for-profits. Databases produced by the mixed sector were 11% of the total in 1985 but only 1% in 1992 (see Databases).

2.2. Database Vendors

Database vendors, sometimes referred to as on-line service providers, often lease or license databases from producers and then add value by putting the information or data into a retrievable form. This is done by processing and preparing the databases for eventual loading onto their computers or time-sharing systems, providing the unique capabilities of the vendor's search software, and making available an audience of potential searchers. Thus database vendors make available to database producers a medium that facilitates speed, searchability, organization, ease of use, full-text searching, use of illustrations and commentary, and links to other pertinent information (20). Additionally, database vendors often provide related services, such as on-line document ordering, selective dissemination of information (SDI), current awareness, and search services, as well as distribution of CD-ROM products to the database users. Some database producers, eg, CAS and NTIS, also act as vendors by creating and operating their own time-sharing systems in order to deliver information directly to customers, thus retaining control over pricing and delivery of their information (20).

Most commercial on-line services did not originally market to end users; their databases were designed from a system's point of view rather than a user's (23). Eventually, the on-line pioneers added menus to facilitate information retrieval. In 1981, Dialog and Bibliographic Retrieval Service (BRS) offered Knowledge Index and BRS/After Dark, respectively, but it was not until the mid-1980s that on-line service providers recognized that the use of software to manage information was critical if the industry was to attract customers.

In 1982, the European Space Agency's Information Retrieval Service (ESA/IRS) introduced the ZOOM command, providing users with a mechanism to analyze retrieved sets. In 1984, service at a baud rate of 2400 was made available by Tymnet and Telenet for public access to on-line databases. In 1985, the first commercial CD-ROM drives for personal computers became available, along with the first commercial CD-ROM databases. In 1986, Grateful Med software was designed for the National Library of Medicine (NLM). In 1987, Tymnet made service at 9600 baud rate available for accessing public on-line databases, and Dialog introduced OneSearch, a multiple database searching capability (24). Several other systems had multiple-file searching capabilities in place at this time, but Dialog's implementation made a greater impact on the searching community (25). In 1988, Dialog offered image searching and retrieval from the TrademarkScan database, and in 1991 vendors added a host of sources and databases with international coverage in anticipation of the inception of the European Economic Community in 1992 (26). Despite all these enhancements, since the emergence of on-line databases, difficult command languages and the lack of common commands among the various vendors' systems continue to inhibit growth of end user searching.

Existence of the end user was recognized in the 1980s; it was not until the 1990s that vendors made concerted efforts to accommodate end user needs by providing faster, easier, and more powerful ways to retrieve relevant information. In electronic information retrieval, the term accessibility implies that data exist in electronic form, that data retrieval is cost effective, intuitive, and easy, and that the electronic medium contributes to the quality and usability of the information (27).

Menu interfaces continued to be developed, facilitating access to on-line databases. In late 1989, Dialog introduced HOMEBASE, a menu-driven service designed to guide the user to information about DIALOG services

(28). In 1991, the Materials Property Data Network on STN provided menu-driven access; DIMDI (Deutsches Institut für Medizinische Dokumentation und Information) introduced GRIPS-Chem, which accounted for Registry Numbers, synonyms, and other chemical nomenclature when retrieving chemical information from several medical and life science files that did not employ the structure terminology of the predominantly chemical databases; Biosis introduced its Life Science Network whose menu-driven search engine did not require users to either specify which databases were to be searched or to enter search commands; Dialog extended its Corporate Connection menu interface to enable all users to search a large subset of Dialog's databases; Data-Star's Business Focus began providing menu access to several business and marketing files; and Thomson Financial network introduced CORIS, a menu-driven service, to company and industry information sources (27).

Other significant enhancements include the 1991 introductions of DIALOG JOURNAL NAME FINDER, which helped users locate the databases that indexed a particular journal title and the number of records for that title in each file (25), plus two companion files for locating companies and products in DIALOG databases: DIALOG COMPANY NAME FINDER and DIALOG PRODUCT NAME FINDER (29). Orbit's GET command, originally developed by Pergamon InfoLine to analyze patent data, was introduced in 1988; it permitted quick and simple on-line statistical analysis of any printable field in any file within a previously created search set of records and generated output ranked by criteria specified by the searcher (30).

The rapidly changing environment of the on-line industry is reflected in its business transactions. In 1980 BRS was acquired from its founders by a Dutch company, the Thyssen-Bornemisza Group (TBG) (31). In 1983, STN International became available as a commercial on-line database service, created by Chemical Abstracts Service (CAS) to prevent a monopoly by DIALOG in distributing its Chemical Abstracts database (31). In 1986, entrepreneur Robert Maxwell bought Orbit from System Development Corp. In 1988, Dialog was sold by Lockheed Corp. to Knight-Ridder, Inc. In 1989 BRS was sold by TBG to Robert Maxwell (29). In 1991, Ziff Communications acquired Predicasts, Inc. from Information Handling Services Group. In 1993, Data-Star was acquired by Knight-Ridder, Inc., the parent company of Dialog, from Motor-Columbus, and MDL Information Systems Inc., formerly Molecular Design Limited, became a publicly owned company. In 1994 BRS Online Products was acquired by CD Plus Technologies from InfoPro Technologies, and Orbit Online Products was acquired by Questel, also from InfoPro Technologies. In June 1990 Dialog filed a lawsuit against the American Chemical Society (ACS), charging the ACS with trying to monopolize the chemical information business (16). In August 1990 the ACS filed a countersuit, charging Dialog with fraudulent and deceptive accounting procedures, which led to underpayment of royalties (16). In October 1993 ACS and Dialog settled the suit and countersuit, releasing all claims against each other (32).

Advances in technology have led to significant performance improvements in the on-line industry since its birth in the early 1970s. With the lower costs of PCs and modems and with higher modem speeds, ie, 9600 and now 14,400 baud, becoming more common, access to information continues to become more convenient and economical to users. Wireless telecommunications technology facilitates access to electronic information by travelers using portable PCs. Because users are becoming dependent on electronic information, they want to influence database vendors to make interfaces with their systems faster, easier, and more powerful than existing interfaces (33).

As the end user population continues to expand, the issue of cost of on-line access is moving to the forefront. The end user market has been nurtured by on-line instructional programs provided by vendors at reduced prices for educational institutions. The skilled users from these programs expect database vendors and producers to review their pricing algorithms. Alternative pricing strategies based primarily on connect time have been implemented by most vendors and fixed price subscription services have become readily available to end users. Primary database vendors that offer or are developing fixed price options include Mead Data Central, Dow Jones News/Retrieval and DataTimes, NewsNet, Dialog, and OCLC (34). The environment of the 1990s is one in which database vendors and producers are competing for survival, and end users are leading the way to what could represent significant growth in the on-line database industry.

2.2.1. BRS Online Products

This vendor comprises BRS Online Service, BRS/Colleague, BRS/After Dark, and BRS/Morning Search (35). The strength of BRS is in medical, physical, and social sciences as well as business and news databases of value to the health care and pharmaceutical industries (36). BRS Online Service contains over 150 bibliographic and full-text databases in the areas of biomedicine, science, technology, business, economics, humanities, and social sciences (37). BRS/After Dark is an after-hours PC oriented version of BRS Online Service, offered at reduced rates (37). BRS/Morning Search is available only in Europe and retrieves information from the BRS Online Service databank. BRS/Colleague provides access to the BRS Online Service databank, but it is a menu-driven on-line service designed for use by health professionals with or without on-line search experience (37).

2.2.2. Cambridge Crystallographic Data Centre

CCDC focuses on collecting, evaluating, and disseminating crystal and molecular structure data obtained by diffraction methods. CCDC maintains the Cambridge Structural Databases, which contain about 110,000 bibliographic, chemical connectivity, and numeric data entries (37) for crystal structures of organic and organometallic compounds analyzed by x-ray or neutron diffraction methods and reported since 1935.

2.2.3. Chemical Information Systems

CIS is a collection of approximately 30 publicly accessible on-line databases of numeric data, bibliographic references, and some full text. The databases contain information on specific chemical substances, including toxicological and carcinogenic research data, hazardous materials handling, chemical and physical properties, safety and health effects, and spectroscopic and pharmaceutical data. Also accessible on Chemical Information Systems (CIS) are databases that provide regulatory information on events or actions at specific sites. The system was originally developed during the 1970s and 1980s under contract to the U.S. Environmental Protection Agency (EPA) and National Institutes of Health (NIH) and has been made available since 1984 to the public from CIS, now a division of PSI International, Inc. (38).

2.2.4. Data-Star

This is Europe's leading on-line database service (39) and covers worldwide business news, financial information, market research, trade statistics, business analysis, healthcare/pharmaceuticals, chemicals/petrochemicals, chemical industry, biomedicine/life science, biotechnology, and technology, with an emphasis on Europe. It was originally formed as a joint venture among BRS, Predicasts, and Radio Suisse (the Swiss telecommunications company) (37). Data-Star offers access to about 300 bibliographic, abstract, directory, and full-text on-line databases, of which approximately 150 are also available on Dialog (40).

2.2.5. DIALOG Information Retrieval Service

DIALOG focuses on business, scientific, technical, and professional information including chemistry, current events, economics, engineering law, medicine, and social sciences, and provides access to over 400 bibliographic, numeric, full-text, and directory databases. All DIALOG databases are searchable by commands, and most have menus (41). DIALOG also includes most of the full-text newspapers formerly offered by VU/TEXT Information Services, Inc., a division of Dialog that closed in December 1992. DIALOG is the oldest scientific/technical on-line service and the world's largest on-line information retrieval service.

2.2.6. ESA-IRS

The European Space Agency's Information Retrieval Service covers chemistry, aerospace/astrophysics, agriculture/food science, biomedicine, physics, health and safety, data processing, earth and environmental sciences, education research, electronics, energy, management science, metallurgy, remote sensing, finance, and news. ESA-IRS offers access to over 200 bibliographic and factual scientific and technical databases. This European

on-line database vendor was established in 1966 to meet the needs of the ESA, which promotes cooperation among European states in space research and technology (27).

2.2.7. MDL Information Systems, Inc

MDL provides modular software systems for managing chemical information, as well as related molecular and reaction databases for use with the software. MDL's database management programs, MACCS-II and REACCS, provide access to compound and reaction databases and also have the capability to manage user-created databases (37). Although MDL is not considered to be an on-line database vendor, it is mentioned here because of the value of its information products and services to the chemical industry.

2.2.8. Mead Data Central

MDC provides access to electronic legal, news, and medical information via its LEXIS, NEXIS, and MEDIS services, respectively. LEXIS, available since 1973, contains archives of federal and state case law, codes, and regulations, as well as specialized libraries covering various fields of legal practice, such as tax, securities, banking, environmental, and insurance. NEXIS, available since 1979, is a news and business information service providing on-line access to over 750 full-text and 2000 abstract sources. LEXPAT, accessible via NEXIS, contains the complete text of patent and trademark information for U.S. patents issued since 1975. MEDIS contains bibliographic and full-text entries of over 40 current medical journals and textbooks.

2.2.9. Medlars

The U.S. National Library of Medicine's (NLM) Medical Literature Analysis and Retrieval System contains over 30 bibliographic databases covering medicine, dentistry, nursing, pharmacology, toxicology, cancer, veterinary medicine, and allied health professions (37). The MEDLARS electronic storage and retrieval system was established at NLM to provide bibliographic access to NLM's biomedical literature collection.

2.2.10. ORBIT Online Service

Orbit is known for its coverage of patents, chemistry, engineering, and occupational health and safety (36). It provides access to about 100 bibliographic, numeric, full-text, and directory databases (42). ORBIT became available in 1972 as the second commercial provider of on-line information, after Dialog. In 1992, Maxwell Online, the parent company of ORBIT Search Service and BRS Information Technologies, changed its name to InfoPro Technologies. It also refocused its two on-line divisions: ORBIT Online Products, featuring ORBIT Online Service, is concentrating on providing patent and patent-related information; BRS Online Products is concentrating on providing comprehensive medical information. In 1994, ORBIT Online Products was acquired by, and became a division of, Questel (a subsidiary of the France Telecom Group). Questel has since changed its name to Orbit-Questel, Inc.

2.2.11. Questel

This is a division of Orbit-Questel, Inc. and a subsidiary of France Telecom (37) and provides on-line access to over 70 bibliographic, abstract, full-text, and structure databases focusing on science/technology, chemistry, European business and news, patents, and trademarks (43). Questel provides two on-line search systems for retrieving chemical structures in databases: Generic DARC, the integrated chemical structure search system, which operates on individual specific compound databases, and Markush DARC, which is designed for inputting, storing, and retrieving compounds included in the definition of generic structure representations commonly used in patents (37) (see Patents and trade secrets).

2.2.12. STN International

The Scientific and Technical Information Network provides access to approximately 160 bibliographic, chemical structure, numeric, reaction, full-text, and directory databases. Topics addressed include chemistry, engineering, health and safety, math, physics, geology, biotechnology, medicine, energy, materials science, pharmacology, and government regulations (44). STN International is operated worldwide by three nonprofit organizations: Chemical Abstracts Service in North America, a division of the American Chemical Society (ACS), the Japan Information Center of Science and Technology in Tokyo, and FIZ Karlsruhe in Germany.

3. Chemical Information and Search Methods and Services

Chemical information is reported and recorded in many forms, and a wide variety of databases have evolved to collect the various types of information. The following tables outline the bibliographic, business, structure, numeric, spectra, and reaction databases currently available; their producers and vendors; and the subject matter they cover.

3.1. Bibliographic/Technical

The principal databases in which bibliographic chemical information is stored are listed in Table 1. Examples of the use of these databases include searching the CA file to find the published work of a certain author, or the World Textiles file to determine the extent of weft knitting machinery use in Europe.

Table 1. Bibliographic/Technical Databases

Database	Producer	Vendor	Subject coverage
Agrochemical Handbook	Royal Society of Chemistry (RSC)	Dialog	information related to ingredients used for pest control
Analytical Abstracts (AA)	RSC	Dialog, Orbit, STN	literature on analytical chemistry
APILIT	American Petroleum Institute	Dialog, Orbit, STN	petroleum/energy industry
APIPAT	American Petroleum Institute	Dialog, Orbit, STN	patents of interest to the petrochemical industry
BIOSIS Previews	Biosis	BRS, Dialog, ESA-IRS, Data-Star, Orbit, STN	life sciences
CAFile	Chemical Abstracts Service (CAS)	STN	chemistry and chemical engineering abstracts
CA Registry File	CAS	Orbit (dictionary), Dialog (dictionary), Questel (dictionary and structure), STN dictionary and structure)	chemical substance information and identification
CA Search	CAS	BRS, Data-Star, Dialog, ESA-IRS, Orbit, Questel, STN	chemistry and chemical engineering citations
CAB ABSTRACTS	CAB International	BRS, Data-Star, Dialog, ESA-IRS, STN	agricultural science and related areas of biology
Ceramic Abstracts	American Ceramic Society	Orbit, Dialog, STN	ceramics
Chemical Engineering and Biotechnology Abstract (CEBA)	RSC	Dialog, Orbit, STN	plant and process chemical engineering
Chemical Journals of the American Chemical Society (CJACS)	American Chemical Society (ACS)	STN	full-text journals published by ACS
Chemical Journals of the Royal Society of Chemistry	RSC	STN	full-text journals published by RSC

Table 1. ***Continued***

Database	Producer	Vendor	Subject coverage
Chemical Safety NewsBase (CSNB)	RSC	Dialog, ESA-IRS, Orbit, STN	health and safety in chemical industry
Chinese Patent Abstracts in English Database	European Patent Office (EPO)	Dialog, Orbit	Chinese patent (English abstracts)
CLAIMS	IFI Plenum Data Corp.	Dialog, Orbit, Questel, STN	U.S. patents
COMPENDEX PLUS	Engineering Information, Inc.	Data-Star, Dialog, Orbit, STN	engineering
CORROSION	InfoPro Technologies	Orbit	corrosion
Current Biotechnology Abstracts (CBA)	RSC	Data-Star, Dialog, ESA-IRS	biotechnology, including legal and safety issues
Current Patents Evaluation/Fast-Alert	Current Patents Ltd.	Data-Star	pharmaceutical patents (U.S., British, and European)
Dissertation Abstracts Online	University Microfilms International (UMI)	BRS, OCLC, Dialog, Data-Star, STN	doctoral dissertation (accredited North American universities)
EMBASE	Elsevier Science Publishers	BRS, Data-Star, Dialog, STN	biomedicine, human medicine
Energy Science & Technology	U.S. Department of Energy	Dialog, STN	energy
EPAT	France Institut National de la Propriete Industrielle (INPI)	Questel	patents applied for and published in the European Patent Office
European Directory of Agrochemical Products	RSC	Dialog	European agrochemical products
GenBank	National Institute of Health	STN	bio-sequences, DNA/RNA sequences
INPADOC	European Patent Office	Dialog, Orbit, STN	patent family and legal status
INSPEC	The Institution of Electrical Engineers (IEE)	Dialog, STN	literature on physics, electrical engineering and electronics, control theory and technology, computers and computing
JAPIO	Japan Patent Information Organization	Dialog, Orbit, Questel	patents abstracts of Japan
LEXPAT	Mead Data Central, Inc.	MDC	full-text patents
MEDLINE	U.S. National Library of Medicine	BRS, Data-Star, Dialog, Medlars, STN, Questel	medicine, life sciences
NEWCRYST	FIZ Karlsruhe	STN	inorganic and organic crystal structures
NTIS Bibliographic Database	U.S. National Technical Information Services	BRS, Data-Star, Dialog, ESA-IRS, Orbit, STN	government research
Paperchem	Institute of Paper Science and Technology	Dialog	pulp, paper, board
PATDD	Deutsches Patentamt	STN	citations and abstracts from former German Democratic Republic
PATDPA	Deutsches Patentamt	STN	reference, abstract, and illustration patents published by the Deutsches Patentamt
PATOSDE	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents from the German Patent Office (Deutsches Patentamt)
PATOSEP	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents granted by the EPO
PATOSWO	Wila Verlag Wilhlm Lampl GmH	STN	citations to patents published by the World Intellectual Property Organization (WPIO)
PHARMSEARCH	INPI	Questel	patents' citation to French, European, and U.S. pharmaceutical patents
Pira Abstracts	Pira International	Data-Star, Dialog, Orbit, STN	paper, pulps

Table 1. *Continued*

Database	Producer	Vendor	Subject coverage
Rapra Abstracts	Rapra Technology Ltd.	Data-Star, Dialog, Orbit, STN	rubber and plastics industries
SciSearch	ISI	Data-Star, Dialog	science, biobusiness technology
Thomas Register Online	Thomas Publishing Online	Dialog	North American companies and their products
U.S. Patents Fulltext	U.S. Patent and Trademark Office (USPTO)	Dialog	full-text U.S. patents
World Patent Index (WPI)	Derwent Publications, Ltd.	Dialog, Orbit, Questel, STN	chemical, electrical, mechanical patents
World Surface Coating	Paint Research Associate	Orbit	coating, paints
World Textiles	Elsevier Science Publishers	Dialog	fibers, fabrics

3.2. Business/Industrial

The principal databases in which business and industrial information is stored are listed in Table 2. Examples include finding a phone number for a company in Wyoming involved in health care or determining the potential market for a new herbicide in the Far East.

Table 2. Business/Industry Databases

Database	Producer	Vendor	Subject coverage
ABI/INFORM	UMI/Data Courier Management	BRS, Data-Star, Dialog, ESA-IRS, MDC, Orbit, STN	new product, business management, electronic data processing
BIOBUSINESS	Biosis	BRS, Data-Star, Dialog, STN	biomedical research
BUSINESSWIRE	Business Wire	Dialog, Dowjones, MDC, Newsnet	industrial news, press releases
CENDATA	U.S. Census Bureau	Compu-Serve, Dialog	census data
Chemical Industry Notes (CIN)	CAS	Dialog, Orbit, STN	chemical business news
Commerce Business Daily	U.S. Department of Commerce	Dialog, Newsnet	government services, defense contracts
Conference Papers Index	Cambridge Scientific Abstracts	Dialog, STN	technical papers, conference papers
Dialog Journal Name Finder	Dialog Information Services, Inc.	Dialog	Dialog search aid
Dialog Product Name Finder	Dialog Information Services, Inc.	Dialog	Dialog search aid
DISCLOSURE DATABASE	Disclosure Inc.	BRS, Data-Star, Dialog, Dowjones, Lexis	financial statement
Dun's Market Identifiers	Dun & Bradstreet Information Services	Data-Star, Dialog, DowJones	corporate data
ERIC	U.S. Department of Education	BRS, Data-Star, Dialog	education programs and projects
Federal Register	U.S. Printing Office	Dialog, MDC, Lexis	U.S. government rules and regula-tions, premanufac-turing notices
Food Science and Technology Abstracts	International Food Information Services	Data-Star, Dialog, Orbit, STN	food science
Harvard Business Review	John Wiley & Sons, Inc.	BRS, Data-Star, Dialog, MDC	general management, business reviews
Health Periodicals Database	Information Access Company	BRS, Data-Star, Dialog, Compuserve	health, nutrition, and fitness
Health Planning and Administration	U.S. National Library of Medicine	BRS, Data-Star, Dialog, Medlars	health care delivery

Table 2. *Continued*

Database	Producer	Vendor	Subject coverage
ICC International Business Research	ICC Stockbroker Research Ltd.	Dialog	industry and stock reports
International Pharmaceutical Abstracts (IPA)	American Society of Hospital Pharmacists	BRS, Data-Star, Dialog	pharmacy literature drugs
INVESTEXT	Thomson Financial Networks	Compuserve, Data-Star, Dialog, DowJones, MDC, MDL, Newsnet	investment analyst report
MANAGEMENT CONTENTS	Information Access Company	BRS, Data-Star, Dialog	business and management topics
Marquis Who's Who (MWW)	Marquis Who's Who, Inc.	Dialog	notable Americans
MATERIALS BUSINESS FILE	Materials Information	Data-Star, Dialog, Orbit, STN	ceramics, iron, composites, and steels
NTIS Bibliographic Database	U.S. National Technical Information Services	BRS, Data-Star, Dialog, ESA-IRS, Orbit, STN	government research, reports, and projects
Nursing and Allied Health	CINAHL Information Systems	BRS, Data-Star, Dialog	nursing, health professions
PHARMACEUTICAL NEWS INDEX (PNI)	UMI/Data Courier, Inc.	BRS, Dialog, Orbit, STN	pharmaceuticals, health
Pharmaceutical Business News	FT Business Enterprises Ltd.	MDC	new pharmaceuticals
Pharmaprojects	PJB Publications Ltd.	BRS, Dialog, STN	drug development and licensing
Pollution Abstracts	Cambridge Scientific Abstracts	Dialog	pollution
PR Newswire	PR Newswire Association, Inc.	Dialog, MDC	business and financial news
PTS News	Predicasts	Dialog	new products
PTS PROMPT	Predicast	Data-Star, Dialog	marketing, new technology
SEC Online	SEC Online, Inc.	Dialog, MDC	companies on U.S. Securities and Exchange Commission
Textile Technology Digest	Institute of Textile Technology	Dialog	textiles
Textline	Reuters Limited	Dialog	European news
THOMAS REGISTER ONLINE	Thomas Publishing Co.	Dialog	information manufacturers
TOXLINE	U.S. National Library of Medicine	BRS, Dialog, MDC	drugs, toxicology
Trade and Industry	Information Access Company	Dialog	trade news
World Textiles	Elsevier Science Publishers	Dialog	textiles

3.3. Structure

Structure searching involves matching a query compound against a machine-readable file of chemical structures. Structure searching determines if a compound is present in a file and retrieves it along with any associated information (45, 46). Chemical structure files are compiled as novel chemicals, and compounds are registered and given unique identifiers, eg, the CAS Registry Number, which is assigned sequentially to each new structure entering the system. Such identifiers link the structure with all related information in a system, such as chemical names and bibliographic data.

Substructure searching involves retrieval of all the compounds in a file containing some specified portion of a chemical structure, irrespective of the rest of the molecule in which the query substructure occurs (47). Substructure searching is much more complex than structure searching because of the incomplete specification of attachments to or within the substructure (48, 49). A two-stage process is normally used for substructure searching. First a file is rapidly scanned for characteristics of the query structure; this eliminates all structures that cannot match the query and produces a subset of query-matching structures. Second, the query substructure is compared to the subset of the file that passed the initial screening to determine if the query substructure is present.

3.3.1. Beilstein File

Beilstein first went on-line on STN International in December 1988 with information on 350,000 compounds, and several months later it was also on Dialog. Beilstein Online now comprises data on five million compounds. The organic substance records contain the critically reviewed and evaluated documents from the *Beilstein Handbook of Organic Chemistry*, main volume and supplements 1–5, which cover the chemical literature from 1779 through 1979. These evaluated data are indicated as Handbook Data in the notes of literature references. The Beilstein File also contains organic substance records for unreviewed excerpts from the primary literature from 1980 to 1991.

The records in the Beilstein File contain structural data and corresponding structural images, numeric data for chemical and physical properties, Beilstein Registry Numbers, CAS Registry Numbers for most substances, and bibliographic data for references to the primary literature. All information is searchable with the exception of stop-words. The stop-words are defined as articles and prepositions (an, by, for, from, of, the, to, and, with) and other frequently occurring words which do not form useful indexing entries and are not directly searchable in an on-line system. The database is in English, but text descriptions of the following fields are in German: Biological Function (BF), Crystal Property Description (CPD), Ecological Data (ECOL), Isolation from Natural Product (INP), Purification (PUR), Toxicity (TOX), Use of Compound (USC), and most Notes (NTE). The Basic Index contains the Beilstein Registry Number, CAS Registry Number, molecular formula, and single words from selected fields.

Structure searching and display software are host-specific. The Softon Substructure Search System (S4) was developed by the Beilstein Institute and Softon of Graefelfing Germany (50). It is a full structure and substructure searching module. The S4 is used in-house by the Beilstein Institute and is operated by DIALOG. STN uses CAS ONLINE's messenger software for on-line structure searching of the Beilstein on-line database (51).

3.3.2. Gmelin File

This file became available on STN International in December 1991, and is comprised of information on 277,458 compounds, a number that is expected to increase to about 700,000 by 1997. The inorganic substance records contain the critically reviewed and evaluated documents from the *Gmelin Handbook of Inorganic and Organometallic Chemistry*, main volume and supplements, covering the chemical literature for the period 1817–1970. Also included are selected data from a pool of 112 journals of inorganic, physical, and organometallic chemistry plus other journals of physics from 1988 to the present. The records in the Gmelin file contain structural data and corresponding structural images, numeric data for chemical and physical properties, Gmelin Registry Numbers, CAS Registry Numbers for most substances, and bibliographic data with references to the primary literature. The database language is English. The Basic Index contains the Gmelin Registry Number, the CAS Registry Number, molecular formula, and single words from CT (control term) and CTM (control term to multicomponents system) (52).

3.3.3. Registry File

This CAS file contains more than 11.8 million chemical substance records. About 8,000–14,000 records are added each week as new substances are identified by the CAS Registry System. The substance records contain CAS Registry Numbers, chemical names, structures, molecular formulas, ring data biosequence information, and classes for polymers. All of this information may be displayed.

Substance information in the Registry file may be searched in a variety of ways, eg, structure information may be searched using structures built on-line with the Structure command or with codes (screen numbers) for predefined structural fragments or class identifiers. Another option is to upload structures drawn using STN Express or other structure-building software, eg, ChemDraw. Protein and nucleic acid sequences may be searched using codes for amino acids or nucleotides. Substance names may be searched using complete names

or name fragments. Complete molecular formulas, molecular formula fragments, and information derived from these formulas, including element counts, atom counts, formula weights, and element symbols, may be used to retrieve compounds from the file. Ring identifiers and ring analysis terms may be used to retrieve substances containing ring systems. Polymers may be retrieved using polymer class terms. Alloys may be retrieved using weight percentages and relative compositions.

Answers from all of these searches contain CAS Registry Numbers. Answer sets may be combined, using the Boolean operators AND, OR, or NOT, with other answer sets or with text terms, such as names or molecular formulas. Any answer set also may be used to define subsets of the file for subsequent structure searching. Answer sets of up to 10,000 Registry Numbers from any type of search in this file may also be used as search terms in other files, such as the CA or CAOLD files (53).

3.3.4. CAS/STN International

CAS/STN offers structure searchable files such as Registry, Beilstein, MARPAT, CASREACT, and Gmelin; a variety of learning files, eg, LRegistry, LBeilstein, LMARPAT, and LCASREACT; and software products such as STN Express for on-line structure and substructure searching. Chemical Abstracts Service, a division of the American Chemical Society, has published *Chemical Abstracts* since 1907 and jointly operates STN International with FIZ Karlsruhe and the Japan Information Center of Science and Technology.

A number of files under the generic title CAS ONLINE are available on-line on STN International. The system software, MESSENGER, includes chemical substructure, text, and numeric data searching facilities. Chemical structures and Registry Numbers are contained in the CA Registry file. The four ways to search the structures are EXA, FAM, SSS, and CSS.

EXA (exact) search retrieves the input structure and its stereoisomers, homopolymers, ions, radicals, and isotopically labeled compounds. FAM (family) search retrieves the same structures as EXA, plus multicomponent compounds, copolymers, addition compounds, mixtures, and salts. SSS (substructure) search uses a range of possible substituents and bonds in the input structure. CSS (closed substructure) search is a more restrictive substructure search with limitation on allowed substitution. Structures can be input textually, using alphanumeric characters, or graphically, using a graphics terminal or personal computer.

A number of graphics front-end packages can be used. STN Express, marketed by STN International, is a completely integrated software package that enables the user to perform on-line structure and substructure searches. This sophisticated software program allows the searcher to draw actual chemical structures to be used in a query on STN. Logon procedures as well as structure uploads to STN are automated. Structure drawing is done off-line, allowing the user to build the structure without incurring connect charges. A sample search of 5% of the file can be run at no cost, other than connect charges. This projects whether the search will succeed or whether it is too broad. Once a Registry file search is complete, it is possible to switch to the CA file for retrieval of the corresponding bibliographic information.

3.3.5. Description, Acquisition, Retrieval, and Correlation File

This is the only other public substructure search system, apart from CAS Online, that provides full access to the CAS Chemical Registry File. The DARC file, commercially available on-line from Telesystems-Questel, offered the first public on-line implementation of substructural searching of the CAS Chemical Registry System. The advantages and disadvantages of the CAS Online and DARC systems have been discussed (49).

3.3.6. Structure and Nomenclature Search System

This system links the collection of chemical databases found in the Chemical Information System (CIS), one of the first interactive systems for structure and substructure searching. References from the separate files can be retrieved by SANSS using CAS Registry Numbers, and the database of structures may be searched

for structures or substructures. An adaptation of the SANSS software for substructure searching has been incorporated in the Drug Information System of the National Cancer Institute for its own use (54).

3.4. Numeric

Researchers routinely use reported numeric measurements and data in their work. Handbooks have been the primary source for locating this type of information, but numeric databases are now increasing in availability. Advantages of searching numeric databases on-line include ease of use, direct access to desired data, and ability to manipulate the information in the answer set.

3.4.1. Beilstein Handbook of Organic Chemistry

This reference (55) is one of the most significant collections of data in organic chemistry. The physical and chemical properties of organic compounds are tabulated in more than 500 fields. Most of these fields are searchable, and a sample of the record for chlorobenzene [108-90-7] is shown in Table 3.

Table 3. Sample Fields for Chlorobenzene in Beilstein

Name	Code	Number of references	Name	Code	Number of references
adsorption	CTADSM	14	infrared spectrum	IRS	43
association	CTASSM	47	ionization potential	IP	28
autonom name	AUN	1	kinematic viscosity	KV	5
azetrope	AZE	47	Lawson number	LN	1
Beilstein citation	SO	7	linear expansion coefficient	LEC	6
Beilstein preferred RN	BPR	1	liquid–liquid systems	CTLLSM	29
boiling point	BP	103	liquid–solid systems	CTLSSM	14
bond moment	BM	2	liquid–vapor systems	CTLVSM	74
boundary surface phenomena	CTBSPM	19	liquid phase	CTLIQ	4
calorific data	CTCAL	4	liquid transition point	LTP	1
CAS Registry Number	RN	1	magnetic susceptibility	MSUS	12
chemical derivative	CDER	38	mass spectrum	CTMS	38
chemical name	CN	2	mechanical properties	CTMEC	15
chemical reaction	REA	1208	melting point	MP	24
circular dichroism	CDIC	1	molar polarization	MPOL	2
conformation	CTCFM	1	molar volume	MVOL	16
coupling phenomena	CTCPL	2	molecular energy	CTMEN	3
critical density	CRD	1	molecular formula	MF	1
critical pressure	CRP	4	molecular rotational constant	MRC	3
critical temperature	CRT	6	moment of inertia	MI	3
critical volume	CRV	3	nmr absorption	NMRA	69
cross-file reference	XREF	15	nmr data	CTNMR	6
crystal lattice parameter	CLP	4	nmr spectrum	NMRS	6
crystal phase	CTCRY	1	nuclear quadrupole coupling constant	NQC	4
crystal space group	CSG	4	nuclear quadrupole resonance	CTNQR	1
crystal system	CSYS	3	optical anisotropy	OA	8
crystal transition point	CTP	1	optical rotation dispersion	ORD	1
density (crystal)	DEN	1	optical rotatory power	ORP	1
density (liquid)	DEN	131	optics	CTOPT	15
dielectric constant	DIC	128	other source	OS	5

Table 3. *Continued*

Name	Code	Number of references	Name	Code	Number of references
dielectric static constant	DISC	17	other spectroscopic methods	CTOSM	5
dipole moment	DM	111	polarographic half-wave potential	PHWP	11
dynamic viscosity	DV	26	preparation	PRE	127
electrical data	CTELE	17	purification	PUR	1
electrical polarizability	CTELP	2	Raman maximum	RAM	5
electrochemical behavior	CTECB	4	Raman spectrum	RAS	19
electronic absorption maximum	EAM	41	redox potential	RDXP	1
electronic absorption spectrum	EAS	40	refractive index	RI	129
electronic spectrum	CTESP	1	rotational spectrum	CTROT	4
emission spectrum	CTEMS	11	self-diffusion	SDIF	10
energy barrier of conformation	EBC	1	skeletal characteristics	CTSKC	4
energy of dissociation	EDIS	5	solubility	SLB	23
energy of MCS	CTENEM	44	solution behavior	CTSOLM	71
enthalpy of combustion	HCOM	3	stereo family	SF	1
enthalpy of formation	HFOR	5	structural data	CTGEN	1
enthalpy of fusion	HFUS	4	surface tension	ST	45
enthalpy of vaporization	HVAP	8	synonym	SY	1
entropy	SREF	3	thermal conductivity	TCND	7
ESR data	CTESR	3	transport phenomena	CTTRAM	38
formula weight	FW	1	unchecked data	CTUNCH	33
gas-phase behavior	CTGASM	16	use of compound	USC	2
heat capacity, C_p	CP	8	vapor pressure	VP	18
heat capacity, C_v	CV	1	vibrational spectrum	CTVIB	10
infrared maximum	IRM	20			

3.4.2. Gmelin Handbook of Inorganic and Organometallic Chemistry

This provides data similar to Beilstein for both inorganic and organometallic chemicals (56). A sample of the record for sodium metabisulfite [7681-57-4] is summarized in Table 4.

3.4.3. Property Data Networks

These include the Materials Property Data Network, Inc. (MPD) (57) and Chemical Property Data Network (CPDN) and are available on STN. These networks provide menu access to numeric data on the performance of different materials and chemicals. Tables 5 and 6 summarize using the numeric files available on STN. NUMERIGUIDE is a data directory and property hierarchy support file produced by STN; it contains information on all properties available in the numeric files on STN.

3.4.4. TDS Numerica

This is another source for numeric databases (58). This company provides different on-line databases and software for chemistry, engineering, and environmental data. A summary of its databases is contained in Table 7.

Table 4. Sample Fields for Sodium Metabisulfite in Gmelin

Name	Code	Number of references
CAS Registry Number	RN	3
chemical name	CN	1
component molecular formula	CMF	1
conformation and bonding models/description of structure	CTCFM	1
crystal property description	CPD	2
enthalpy of formation	HFOR	1
formula weight	FW	1
general data	CTGEN	5
infrared spectrum	IRS	3
isotope search data	IFOR	1
linearized structure formula	LSF	1
molecular formula	MF	1
reaction	REA	48
solubility	SLB	58
spectroscopic information	CTSPE	4
uv and visible spectrum	UVS	2

Table 5. Materials Property Data Network

Database	Producer	Subject coverage
AAASD	The Aluminum Association, Inc.	mechanical and physical properties of commerical aluminum alloys
ALFRAC	U.S. Department of Commerce	testing procedures
ASMDATA	ASM International	specifications for composites, plastics, ferrous and nonferrous alloys, and metals
COPPERDATA	Copper Development Association, Inc.	numeric data for coppers and copper alloys
IPS	International Plastics Selector, D.A.T.A. Business Publishing	manufacturer-supplied data on commercial plastics
MARTUF	The National Materials Property Data Network, Inc.	toughness of steels
MDF	Materials Information ASM International	ferrous and nonferrous alloys
METALCREEP	The National Materials Property Data Network, Inc.	creep and rupture stress of aluminum alloys and steels
MH5	U.S. Department of Defense and Federal Aviation Administration	mechanical and physical properties for metallic aerospace materials
NISTCERAM	National Institute of Standards and Techology Gas Research Institute, Ceramics Division	mechanical, physical, electrical, thermal, corrosive, and oxidation properties for alumina nitride, beryllia, boron nitride, silicon carbide, silicon nitride, and zirconia
PDLCOM	William Andrew, Inc., Plastics Design Library	test data on the chemical compatibility and the environmental stress crack resistance of plastics
PLASPEC	D&S Data Resources	data on commercial plastics
STEELTUF	Electric Power Research Institute (EPRI) and PROD Materials Properties Council	toughness of more than 100 grades of steels used in the power industry
MPDSEARCH	The National Materials Property Data Network, Inc.	the *MPD Guide to Materials and Substances Data Sources* provides information about the materials property databases available on STN International
PLASNEWS	D & S Data Resources/Plaspec	prices, market statistics, critical plastics industry news

Table 6. Chemical Property Data Network

Database	Producer	Subject coverage
DIPPR	American Institute of Chemical Engineers and Design Institute for Physical Property Data	numeric physical property data for commercially important chemicals and substances
HODOC	CRC Press, Inc.	numeric file representing the nine-volume 2nd ed. of the CRC *Handbook of Data on Organic Compounds*
HSDB	National Library of Medicine's Toxicology Information Program	toxicology and the environmental effects of chemicals
JANAF	U.S. Department of Commerce National Institute of Standards and Technology	chemical thermodynamic properties of inorganic substances and of organic substances containing only one or two carbon atoms
NISTFLUIDS	U.S. Department of Commerce National Institute of Standards and Technology	critically evaluated thermophysical and transport porperties for 12 important industrial fluids
NISTTHERMO	U.S. Department of Commerce National Institute of Standards and Technology	evaluated chemical thermodynamic properties of inorganic and organic substances containing one or two carbon atoms
POLYMAT	DKI, Deutsches Kunstoffinstitut and Fachinformationszentrum Chemie GmbH	data on commercially available plastic materials
RTECS	National Institute for Occupational Safety and Health (NIOSH)	registry of toxic effects of chemical substances; contains toxicity data and references commercially important substances
SPECINFO	Chemical Concepts GmbH	spectral data for a representative section of organic chemistry
TRCTHERMO	Thermodynamic Research Center, Texas A&M University, College Station, Texas	thermodynamic data

Table 7. TDS Numeric Databases

Database	Producer	Subject coverage
CHEMIST	Arthur D. Little, Inc.	predicts physical and environmental properties
CHEMSAFE	Physikalisch-Technische Bundesanstalt (PTB) (also available on STN)	data for plant safety and accident prevention
DETHERM	Dechema eV (also available on STN)	thermophysical properties and phase equilibrium data
Log P Database	Medicinal Chemistry Project at Pomona College	partition coefficients
PPDS2	National Engineering Laboratory (NEL, U.K.)	thermodynamic and phase equilibrium data, provides a modeling system based on several groups of physical properties data (six files)
TRC Vapor Pressure	Thermodynamic Research Center (also available on STN)	vapor pressure and boiling points

3.4.5. Other Databases

Available from different vendors (Table 8). For example, the researcher can obtain physical properties by using the *Merck Index Online* or the Dictionary of Organic Compounds available by Chapman and Hall Chemical Database. In DIALOG, numeric databases are collected under the name of CHEMPROP.

3.5. Cambridge

The Cambridge Structural Database is an integrated system of programs for searching, retrieving, and analyzing data on more than 96,000 organic and organometallic structures, which were determined by x-ray and neutron diffraction (59). About 15,000 compounds a year are being added to the database. The development

Table 8. Other Numeric Databases

Database	Producer	Vendor	Subject coverage
Chapman and Hall Chemical Database	Chapman and Hall, Ltd.	Dialog	*Dictionary of Organic Compounds* (5th ed.), *Dictionary of Organometallic Compounds, Carbohydrates, Amino Acids, Peptides, Dictionary of Antibiotics and Related Compounds*, and *Dictionary of Organophosphorus Compounds*
Computerized High Temperature Materials Properties Data-base	Purdue University	Purdue University (CINDAS)	compiled from literature and evaluated reference data for high temperature materials
EGIN-PLAST	Guide de Choix Engin-Plast (France)	Mintel International Group	mechanical, thermal, electrical, physical, processing, and flammability properties and applications for more than 4500 commercial plastics
Encyclopedia of Polymer Science and Engineering Online	John Wiley & Sons, Inc.	Dialog	covers natural and synthetic polymeric materials
Material Properties Bibliographic Data	Purdue University	Purdue University (CINDAS)	thermophysical, mechanical, and electronic properties of materials; bibliographic references and author index
Material Properties Numerical Data System	Purdue University	Purdue University (CINDAS)	evaluated data compiled, correlated, analyzed, and synthesized to generate values for the thermophysical, mechanical, and electrical properties of materials
MATUS	Engineering Information Co. (U.K.)	IPS, STN	mechanical, electrical, thermal processing properties
NAPRALERT (Natural Products ALERT)	University of Illinois at Chicago	STN	information on the pharmacology, biological activity, taxonomic distribution, medicine and chemistry of plant, microbial, and animal (including marine) extracts
NIST Update	High Tech Publishing Co.	NewsNet	news and information on activities of the U.S. National Institute of Standards and Technology (NIST)
NISTFLUIDS	NIST	STN	programs for calculating thermophysical and transport properties of cryogenic fluids
Plaspec Material Selection Database	Data Resources, Inc.	Dialog, STN	detailed engineering and design data, chemical descriptions, and trade names for over 11,500 grades of plastics materials
SPAO	Laboratorire National d'Essais (France)	Teletel	mechanical, electrical, thermal processing properties, and chemical resistance (Europe)
Agrochemical Handbook	The Royal Society of Chemistry, Cambridge (U.K.)	Dialog, Data-Star, Knowledge Index	information on the active components found in agrochemical products used worldwide, chemical names, including synonyms and trade names, CAS Registry number, molecular formula, molecular weight, manufacturers' names chemical and physical properties, toxicity, mode of action, activity, health, and safety
The Merck Index Online	Merck & Co., Inc.	Dialog, CIS, BRS, Knowledge Index, Questel	the online counterpart to the printed 11th ed. of *The Merck Index*; records contain molecular formulas and weights, systematic chemical names, physical and toxicity data, therapeutic and commercial uses, and bibliographic citations
TSCA	U.S. Environmental Protection Agency	Dialog	information regarding chemical substances in commerce covered in the Toxic Substances Control Act

of highly sophisticated x-ray diffraction equipment has contributed to greater precision in defining crystal structures and to more published crystallographic work. This database was developed by the Cambridge Crystallographic Centre of Cambridge, U.K. CAMBRIDGE has three basic files: classified bibliographic information, retrospective to 1935; evaluated numeric data for structures, since 1959; and chemical connectivity records, retrospective to 1935.

Because CAMBRIDGE is a compilation of all published crystallographic data, it is ideal for structure building and analysis. Text searching can be done in the bibliographic file. Substructure searching can be done using manual or graphic input. Once a structure or structure fragment has been found, related molecular and statistical data, such as bond lengths, bond angles, and coordinates, can be captured. The geometrical calculations and statistical analysis (GSTAT) feature can be used to generate further structural data, such as geometry of coordination spheres around specified atoms, calculation of centroids, vectors, and planes, user definition of atomic radii, fragment geometry and systematic tabulations, output of histograms, and dimensional scattergrams. CAMBRIDGE can also use PLUTO to generate a three-dimensional graphic output.

Examples of the versatility of CAMBRIDGE include its use in drug design to search for specific pharmacopores and the three-dimensional arrangement of chemical groups essential for biological activity (60). It has also been applied in molecular modeling (61) and amino acid and peptide studies. Development work is taking place to link CAMBRIDGE to MACCS-II to enhance the usefulness of both databases.

3.6. Spectra

The ability to consult collections of standard spectra is crucial in the analysis of unknown compounds. A long history of data collection efforts has been aimed at these applications. Among the best known of the published handbooks are the *Sadtler Spectral Data Sheets*, which include ir, Raman, and nmr spectra. An extensive bibliography of older hard-copy ir spectra is given in *The Coblentz Society Desk Book of Infrared Spectra* (62). Since the mid-1980s, comprehensive databases have been available in computerized form where the spectra themselves, not merely the bibliographic references, are searchable and displayable. The search algorithms vary considerably among the available systems; no algorithm standard exists (ca 1994), but several are under development (63, 64). Expert systems, which assist in the automatic interpretation and identification of spectra, have existed for many years but are not commonly used (65). Computerized spectral databases are either local, PC-based, or public.

3.6.1. Local and PC-Based Databases

Local databases can be either free-standing or integrated with spectrometers. Many of the primary equipment manufacturers publish their own collections for use with their own instruments. Many free-standing local database systems are available for use on a PC or mainframe computer. Both types can have a database management system forbuilding a user's personal database. Unless high standards are maintained in the accuracy of interpretation or calibration, the quality of these privately built databases can suffer. Also, disk space can limit the number of spectra stored. These disadvantages can be overcome by using on-line or CD-ROM databases, with their nearly high capacities and high quality; but these databases cannot currently be used to store and search personal data.

3.6.1.1. Nuclear Magnetic Resonance Spectroscopy. Bruker's database, designed for use with its spectrophotometers, contains 20,000 ^{13}C-nmr and ^{1}H-nmr, as well as a combined nmr-ms database (66). Sadtler Laboratories markets a PC-based system that can search its collection of 30,000 ^{13}C-nmr spectra by substructure as well as by peak assignments and by full spectrum (64). Other databases include one by Varian and a CD-ROM system containing polymer spectra produced by Tsukuba University, Japan. CSEARCH, a system developed at the University of Vienna by Robien, searches a database of almost 16,000 ^{13}C-nmr. Molecular Design Limited (MDL) has adapted the Robien database to be searched in the MACCS and ISIS graphical display and search environment (63). Projects are under way to link the MDL system with the Sadtler library and its unique search capabilities.

A PC-based ^{1}H-nmr database, which includes full spectrum search capability, is being constructed by the Toyohashi University of Technology (67). SpecInfo, owned by Chemical Concepts, offers a 150,000 spectra library and database system for mainframe computers, which includes ^{1}H, ^{15}N, ^{19}F, ^{17}O, ^{31}P-nmr, and a large

collection of ^{13}C-nmr spectra compiled by Bremser at BASF (68, 69). It also offers ^{11}B-nmr spectra compiled by Nöth at the University of Munich.

The National Chemical Laboratory for Industry (NCLI), Japan, has developed an integrated Spectral Database System (SDBS) which is available to users in Japan. All spectra were determined at NCLI under controlled conditions and are available on a PC/CD-ROM or magnetic tape. The system has both ^{1}H-nmr (6000 compounds) and ^{13}C-nmr spectra (5700 compounds), along with searching software. NCLI has also developed an integrated ^{13}C–^{1}H-nmr system that can be used for two-dimensional data elucidation (70, 71).

The Novosibirsk Institute of Organic Chemistry has developed a method for computer-aided retrieval of structural information from ^{1}H-nmr using its database of 50,000 spectra (72). Fraser Williams Ltd. (Scientific Systems) has special software to search its ^{19}F-nmr database (73). Protein nmr data have been compiled into a relational database at the University of Wisconsin (74).

3.6.1.2. Infrared Spectroscopy. The Sadtler collection is the largest commercially available system with over 60,000 spectra, largely from prism and grating spectrometers. Fourier transform infrared (ftir) data are currently being added and will be searchable by substructure (63). Nicolet, in collaboration with Aldrich and Sigma, has developed separate databases, comprising over 10,000 ftir spectra each, of the compounds in the catalogues of the two companies. The EPA vapor-phase spectra collection is available through various ir instrument companies, and the SpecInfo system includes over 50,000 spectra from BASF and the Hummel ir Standards from Cologne University.

A computer file of about 19,000 peak wavenumbers and intensities, along with search software, is distributed by the Infrared Data Committee of Japan (IRDC). Donated spectra, which are evaluated by the Coblentz Society in collaboration with the Joint Committee on Atomic and Molecular Physical Data (JCAMP), are digitized and made available (64). Almost 25,000 ir spectra are available on the SDBS system developed by the NCLI as described. A project was initiated at the University of California, Riverside, in 1986 for the construction of a database of digitized ftir spectra. The team involved also developed algorithms for spectra evaluation (75). Other sources of spectral libraries include Sprouse Scientific, Aston Scientific, and the American Society for Testing and Materials (ASTM).

3.6.1.3. Mass Spectroscopy. A collection of 125,000 spectra is maintained at Cornell University and is available from John Wiley & Sons, Inc. (New York) on CD-ROM or magnetic tape. The spectra can be evaluated using a quality index algorithm (63, 76). Software for use with the magnetic tape version to match unknowns is distributed by Cornell (77). The collection contains all available spectral information, including isotopically labeled derivatives, partial spectra, and multiple spectra of a single compound.

A second important database is the NIST/EPA/MSCD (formerly NBS/EPA/ MSCD), jointly administered by the NIST, EPA, and the Mass Spectrometry Data Center (MSDC) of Nottingham, U.K. Many of the almost 55,000 spectra, especially those of environmental interest, are generated directly from the compounds at one of the sponsoring laboratories (67). A quality index algorithm, based on the one developed for the Wiley database, is available to evaluate the reliability of each spectra (64, 78). Multiple spectra for a single compound are not included. Both the Wiley and NIST/EPA/MSDC databases incorporate older noncomputerized data collections as well as spectra from the literature (65). A merged version of these two databases is available from Wiley on CD-ROM or tape. SpecInfo offers a portion of the Wiley spectra as well as spectra from BASF, Max Plank Institute in Mulheim, ETH in Zurich, and the geo/petrochemical collections from Kernsforschungsanlage (KFA) Julich and Delft University. Fein-Marquart markets software for the PC called MASCOT, which includes a collection of about 1500 spectra of environmental compounds and forensic drugs (79). The NCLI integrated system, SDBS, has over 10,000 mass spectra (80, 81).

3.6.1.4. Miscellaneous. NIST has a reference database of critically evaluated x-ray photoelectron and Auger spectral data, which is designed to run on PCs. It is searchable by spectral lines as well as by element, line energy, and chemical data (82). The Nuclear Quadrapole Resonance Spectra Database at Osaka University

of over 10,000 records is available in an MS-DOS version (83). The NCLI system, SDBS, has esr and Raman spectra, along with nmr, ir, and ms data, as described.

3.6.2. Public On-Line Databases

These databases are accessible through on-line commercial services.

3.6.2.1. Chemical Information System. CIS was the first public on-line system to make a collection of spectra readily available. The initial work was done by the NIH, EPA, and other U.S. government agencies. Spectra data can be displayed in tabular or graphic format. Chemical substructures can be searched on all CIS databases using the Structure and Nomenclature Search System (SANSS). The structure query entry can be facilitated if the user has the PC front-end package, SuperStructure (71). A database of over 11,000 ^{13}C-nmr spectra, collected by the Royal Dutch Chemical Society, can be searched interactively or by batch for shift, multiplicity, and intensity. The Infrared Search System (IRSS) allows retrieval spectra for known compounds as well as for unknown comparisons. IRSS contains over 4500 spectra for the EPA and the Boris Kidric Institute in the former Yugoslavia. CIS has two ways to search libraries of mass spectral data. The Mass Spectral Search System (MSSS) is used to search the NIST/EPA/MSDC database and a small collection of chemical ionization spectra. Besides chemical and formula information, MSSS can search by individual peaks or by full spectrum using a probability-based match. The Wiley Mass Spectral Search System (WMSSS) is similar to the MSSS but searches against the Wiley database.

3.6.2.2. SpecInfo/STN. A large proportion of the SpecInfo nmr, ir, ms databases described, along with additional data taken from the literature, are available through STN (Scientific and Technical Network International). Graphical access is possible through several PC emulation packages such as STN EXPRESS. MESSENGER command language is used along with several specialized commands, which were developed at BASF. CHESS is a chemical structure similarity search query. Coupling constants and nmr information can be searched using SPECAL, and spectra similarity searches are run with GETSPEC. Peak values can also be searched with all these techniques.

3.6.2.3. JICST/JOIS. The Japan Information Center for Science and Technology (JICST) Mass Spectral Database is accessible to users in Japan through the JICST Factual Database System (JOIS-F). The database uses the NIST/EPA/ MSCD data collection supplemented by spectra from the Mass Spectrometry Society of Japan (84).

3.6.2.4. CAN/SND. The Canadian Scientific Numeric Database Service (CAN/ SND) is provided by the Canada Institute for Scientific and Technical Information (CISTI), a division of the National Research Council of Canada. It contains 140,000 ir spectra of 96,000 compounds. Entries consist of peak locations and some intensities. This system is searchable on-line using the SPIR (Search Program for Infrared Spectra) (85). Table 9 summarizes the available databases in the area of spectra.

3.7. Reactions

CASREACT File is a chemical reaction database containing over 118,500 records with reaction information derived from documents covered in the Organic Section of *Chemical Abstracts*. The file is available from STN. Coverage includes journals from 1985 to the present and patents from January 1991 to the present. The records contain structure diagrams for reactants and products, CAS Registry Numbers for all reactants, products, reagents, solvents, and catalysts, yields for many products, and textual reaction information. The reactants, reagents, and products are structure searchable with a single reaction query. Roles, reaction sites, and mapping of atoms between reactants and products are also structure searchable (86).

The database is updated biweekly with new records. The primary search terms in CASREACT are CAS Registry Numbers. Registry numbers may be input directly or may be contained in L-numbered answer sets and crossed over from the Registry file. Typically, all substances that are in CASREACT are flagged in the Registry file. In addition to the Registry numbers and structures to search the database, the feature Functional

Table 9. Summary of Spectra Databases

Producer	Type	Number of spectra	Availability	Notes
Aldrich-Nicholet	ftir	10,600	PC-based	compounds found in Aldrich catalogue
ASTM	ir	145,000		
Aston Scientific	ir			Quick-Search software
Boris Kidric Institute	ir	4,500	STN (vendor)	
Bruker	^{13}C-nmr ^{1}H-nmr	19,000 900	spectrometer-based	
Canadian Institute for Scientific and Tech-nical Information	ir	140,000	CAN/SND (vendor)	
Coblentz Society-JCAMP	ir	4,400	PC-based	
EPA	ir	3,300	STN (vendor) and spectro-photo meter	vapor phase
Fein-Marquart	ms	1,500	PC-based	environmental and forensic; MASCOT software
Fraser Williams (Scientific Systems Ltd.)	^{19}F-nmr		PC-based	PC-SABRE search software
Infrared Data Com-mittee of Japan	ir	19,000	PC-based	peak wavenumbers and intensities
Mass Spectroscopy Society of Japan	ms	6,000	JICST-JOIS	
NCLI, Japan	^{13}C-nmr ^{1}H-nmr ir ms Raman esr	6,000 6,000 25,000 10,000	PC, CD-ROM, tape	integrated system available; Spectral Database System (SDBS) software
NIST	xps		PC-based	
NIST/EPA/MSDC	ms	55,000	STN (vendor) PC-based, CD-ROM, tape	Wiley Quality Index algorithm
Novosibirsk Institute for Organic Chemistry	^{1}H-nmr	50,000		
Osaka University	nqr	10,000	PC-based, CD-ROM, tape	
Royal Dutch Chemical Society	^{13}C-nmr	11,000	STN (vendor)	
Sadtler Laboratories	^{13}C-nmr, ir, ftir	30,000 60,000	PC-based	PPC search software; substructure searchable; MACCS, ISIS link
Sigma-Nicolet	ftir	10,400	PC-based	compounds found in Sigma catalog
SpecInfo Chemical Concepts	^{13}C-nmr ^{1}H-nmr ^{15}N-nmr ^{17}O-nmr ^{19}F-nmr ^{31}P-nmr ^{11}B-nmr ir ms	100,000 13,000 1,000 900 1,900 2,200 9,000 50,000 160,000	STN (vendor)	Bremser-BASF collection Nöth, Munich University collection BASF; Hummel Standards BASF; portion of Wiley, Max Planck Inst.; ETH; geo/petrochemical
Sprouse Scientific	ir			
Toyohashi University	^{1}H-nmr	4,000	PC-based	full spectrum searchable
Tsukuba University	^{13}C-nmr		CD-ROM	polymers
University of California, Riverside	ftir			
University of Vienna	^{13}C-nmr	16,000	PPC-based	CSEARCH; MACCS, ISIS searchable
University of Wisconsin	^{1}H-nmr			proteins
Varian	nmr		PC-based	
John Wiley & Sons, Cornell University	ms	125,000	CD-ROM, tape	isotopic labeled compound; Quality Index algorithm

Groups allows use of different functional groups either as starting materials or products. An alphabetical listing of the CAS functional groups is as follows.

Acetal
acid halide
acylmetal
aldehyde
alkene
alkyne
amide
amidine
anhydride
azide
azine
aziriding
azo
benzenoid
carbamate
carbonate
carboxylate
carboxylic
chloramine
cyclic alc
cyclic alkene
cyclicdiene
cyclicketone
cyclopropyl
diazo
diene
dihalide
diimide
disulfide
dithioacetal
dithiocarboxylate
dithiocarboxylic
enamine
enol
enol ether
epoxide
ether
gem-dihalide

glycol
guanidine
halide
haloformate
halohydrin
hemiacetal
heterocycle
hydrazide
hydrazine
hydrazone
hydroperoxide
hydroxylamine
imide
imine
iminoether
isocyanate
isonitrile
ketal
ketone
lactam
lactone
metal halide
metal hydride
metal metal single
metal phosphine
metal sulfide
metalarene
metalcarbonyl
metalcyclopentadienyl
metallocarbocycle
metalnitrogen
metalnitrosyl
μ-carbonyl
nitrile
nitro
nitrone
nitroso
nitroxide
organometal
orthoester
oxime

peroxide
peroxyacid
phenol
phosphate
phosphite
phosphonate
phosphonium
phosphorus ylide
π-alkene
π-alkyne
π-allyl
primary alc
primary amine
quaternary ammonium
quinone
secondary alc
secondary amine
selenide
selenol
silyl
silyl enol ether
sulfenyl halide
sulfide
sulfinate
sulfinic
sulfinyl halide
sulfonamide
sulfonate
sulfone
sulfonic
sulfonyl halide
sulfoxide
tertiary alc
tertiary amine
thioamide
thiocarboxylate1
thiocarboxylate2
thiol
thiolactam
thiophenol
triazene

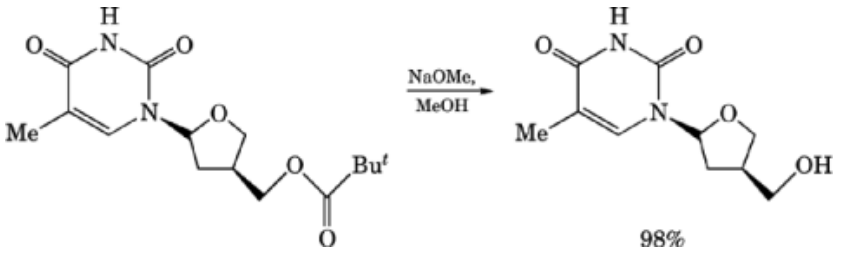

Fig. 1. Sample reduction of an ester group to the corresponding primary alcohol (CASREACT) (87): s(carboxylic or carboxylate)/fg.rct (s) (primary alc)/fg.pro.

trihalide
unsatdacid
unsatdaldehyde
unsatdamide
unsatdester
unsatdketone
unsatdnitrile
urea
and vinylhalide.

A sample of reduction of carboxylic acids to the corresponding primary alcohol is demonstrated in Figure 1.

CHEMINFORMRX from FIZ-Chemie Berlin and *CHEMREACT* from Springer-Verlag Infochem GmbH are other databases covering reaction information area. Reaction information is also available from Beilstein, Gmelin, and CA.

REACCS(reaction access system) is a specialized database management system for chemical reaction information. It is designed to store, search, retrieve, and display molecules, reactions, and the data associated with them (88). REACCS allows a variety of databases to be searched either individually or globally. The databases contain information such as bibliographic references, reaction conditions, yields, reagents, catalysts, and reaction class. Input may be either graphic or alphanumeric. In a REACCS database, a molecule is represented by a Stereochemical Extension of Mongan Algorithm (SEMA) name, two-dimensional coordinates, and structure keys. The SEMA name uniquely defines a molecule in terms of its atoms, their connectivity, stereochemistry, charge, and isotopic properties (89).

REACCS is organized into six operational modes: MAIN and BUILD, which are used to draw molecular structures, build reactions, and construct graphic queries; and SEARCH, VIEWLIST, PLOT, and FORMS, which are used to create custom forms to display data associated with reactions and molecules. Each of the modes provides a characteristic menu and a set of options, which normally perform tasks that relate to the general function of that mode.

With its flexible and logical search language, REACCS can retrieve molecular structures, the atoms and bonds that are transformed in a reaction, relative and absolute stereochemistry, the role (reactant, product, solvent, or catalyst) of a molecule in a reaction, reaction data (eg, temperature and yield), literature references, and keyword descriptions of reaction types.

Data storage in REACCS is hierarchical: related data are stored separately but also are grouped under a single descriptive category. For example, in the Theilheimer database, the treename is the complete hierarchical

name of a piece of data and is composed of three components: entity, parent datatypes or category of data, and field datatype. All REACCS databases include VARIATION. VARIATION is usually the highest parent datatype in the reaction hierarchy. VARIATION can be used to store more than one complete set of reaction data with a reaction. To keep track of the data associated with different variations or multiple reactants and products in the same reaction, line numbers are appended to some of the datatypes in a treename.

The four Build Menus in REACCS are Structure, Query, Top, and HighlightRxn. Structure menu contains the basic drawing commands used to construct the backbone of the structure. Query menu contains the commands used to add flexible structural parameters to the query. Top menu contains commands used to build reactions and to store and retrieve reactions, molecules, and graphic queries. HighlightRxn menu contains commands that apply atom/atom mapping and reaction centers to the current reaction. Atom/atom mapping is used to identify the reaction centers and increase accuracy and efficiency by letting the searcher specify that a particular atom in a reactant must correspond to a particular atom in the product.

3.8. Integrated Systems

Until recently, each of the numerous databases and sources of information available to chemists and technologists had to be searched individually, and selected results either printed for file storage or downloaded to an in-house or private computer system for easy future access.

Molecular Design Limited (MDL) (90) has marketed an Integrated Scientific Information System (ISIS), which provides the capability to query multiple systems, including binary, text, proprietry, and relational databases across global networks, thereby providing transparent desktop access to multiple autonomous data sources. ISIS links databases, networking protocols, and computer platforms, providing a consistent environment across personal computer formats, by the use of three interrelated components: ISIS/Draw, a state-of-the-art chemical drawing package which works on multiple computing platforms; ISIS/Base, a database management program which allows storage of data and provides customized information that can be constructed on the desktop and stored or printed; and ISIS/Host, which operates the server and interacts with clients and databases.

4. Patent Information and Search Methods/Services

Patents are unique as primary source documents because of their stylized format, specialized language, presence of legally significant claims, descriptive drawings, and frequent disclosure of chemical compositions as generic (Markush) structures. The term patent is commonly used to identify a broad family of patent publications at various stages of prosecution by national or regional issuing authorities, eg, unexamined applications, grants, and reissues. A comprehensive review of electronic databases that contain patent information, including legal aspects, is available (91) (see also Patents and trade secrets).

4.1. Bibliographic Databases

A number of bibliographic and textual patent database producers have enhanced indexing to facilitate accurate retrieval with the use of controlled vocabulary, specific compound and substructure search capability, and additional explanatory text. Some database producers provide access via full text or inclusion of only some or all bibliographic details, title, and abstract, and others provide patent text or databases on CD-ROM for PC or network use. Newer resources allow users to input full or partial chemical structures graphically; the structures are automatically converted to appropriate functional group or topological coding for subsequent searching.

4.1.1. World Patents Index

WPI, produced by Derwent Publications, Ltd., contains records of patent publications from 32 issuing authorities around the world. Since 1970, all chemical technologies are included. Prior to then, content varies as follows: polymers from 1966, agricultural chemicals from 1965, and pharmaceuticals from 1963. Records contain bibliographic data and an abstract, which describes the novel features of the patent. The patent titles and abstracts in WPI are created by Derwent and generally give a more definitive description of the patent's content than the title and abstract, which appear on the document itself. Records also list equivalent or family member patents that have been identified as covering the same or related invention(s) (92).

WPI can be searched by bibliographic data, such as inventor, assignee (including standardized assignee codes), patent or application numbers or dates, and International Patent Classification, or by the in-depth indexing unique to the database. This indexing includes Derwent classification codes based on chemical composition, utility, or processes; chemical fragmentation codes (93), which define compounds by atoms, functional groups, ring systems, and carbon chains present; and polymer codes, which represent polymer type, monomers, processing, properties, and utility. About 2100 specific chemical compounds, which are important to the chemical and pharmaceutical industries, are searchable by Derwent Registry Numbers.

4.1.2. U.S. Patents

This file, produced by Derwent, Inc., covers U.S. patents from 1971 to the present. The database includes all bibliographic and front page information and the text of all claims. (From 1971 to 1974 the claims from many patents were not available from the United States Patent and Trademark Office (USPTO) source tapes, and therefore are not included.) The complete claim text can be searched from 1971 but can be printed only from 1982. Titles and patentee names are present in their original form, neither expanded nor standardized. There is no enhanced indexing. Examiner citations are directly searchable, and USPTO classification is updated when the tapes are received from the Patent Office.

4.1.3. Claims

The CLAIMS databases are produced by IFI/Plenum Data Corp. They cover only U.S. patents and include both bibliographic and auxiliary files.

CLAIMS BIBLIO includes an abstract and claim in addition to basic bibliographic information for chemical and chemically related U.S. patents from 1950 and for all patents from 1963. All claims are searchable and printable from 1971; claims for many patents are not available from 1971 to 1974. From 1972, many titles have been enhanced with additional keywords to describe the invention more clearly and to indicate the presence of a drawing; chemical structures have been converted so that they display in linear format. Many company names have been standardized, and USPTO classification is updated annually to reflect reclassification projects.

CLAIMS UNITERM adds enhanced indexing to the chemical and chemically related patent records; general terms to describe processes, properties, end products, etc; specific compound terms (over 15,000); and chemical fragment terms to describe generic compounds.

CLAIMS COMPREHENSIVE (subscriber access only) further enhances the UNITERM indexing from 1964 with roles for all indexed compounds and polymer class terms and with links and negation codes for the fragment terms describing chemical compounds (94).

CLAIMS CITATION includes examiner citations (prior references cited during prosecution of the patent application) against all W.S. patents from 1947. Citations are not directly searchable in the three bibliographic files.

CLAIMS REASSIGNMENT AND REEXAMINATION gives information on post-issue actions: reissues from 1950, reassignments from 1980, reexaminations from 1981, extensions from 1986, expiration for nonpayment of fees from 1985, and reinstatements.

CLAIMS CLASS contains the titles of the classes and subclasses of the *USPTO Manual of Classification*. It can be searched by title words to locate pertinent classes to use in the bibliographic files and by class/subclass numbers to identify the classification assigned to a known patent. It is updated annually.

4.1.4. APIPAT

This is the patent database produced by the American Petroleum Institute and covers patents from 1964 of interest to the petrochemical industry, including petroleum refining, pollution control, uses of petrochemicals, and catalysts. Enhanced indexing includes terms applied from a hierarchical thesaurus with automatic posting to the broader terms in the hierarchy. Fragments called chemical aspects are linked to describe each compound, and the compounds are further linked to roles (eg, reactant or product) and use (eg, antioxidant or lubricant). ORBIT provides access to a merged APIPAT/WPI file, which allows searchers to draw on the strengths of both databases without the need to search them separately (95).

4.1.5. EPAT

The European Patents Register is produced by the European Patent Office (EPO) and the Institut National de la Propriete Industrielle (INPI). The database provides bibliographic, including the first claim of granted (BI) patents, and legal status information on all European patents and published applications. Coverage is from June 1978, the beginning of EP publication, and now includes over 450,000 records. Titles on all records and claims of granted patents are in French, English, and German. Abstracts, which have been included from 1988, are in the original language of the patent publication. The database is updated weekly on the day of publication.

4.1.6. JAPIO

This database is produced by the Japan Patent Information Organization and is based on the Patent Abstracts of Japan provided by the Japanese Patent Office. The database is updated monthly and contains all Kokai Tokyo Koho (published unexamined patent applications) published as of October 1976. Records appear in JAPIO approximately six months after publication of the unexamined patent application. English language abstracts are provided for the majority of applications filed by Japanese applicants. Applications by non-Japanese applications do not have abstracts, but bibliographic information is included. Searchable fields include the International Patent Office Classification and JAPIO classification (96).

4.1.7. INPADOC

This database has been produced by the European Patent Office since January 1991. It presently provides patent family and legal status information for over 16 million patent publications. Legal status, including oppositions, lapses or expiration for nonpayment of fees, and patent grants are provided for 11 issuing authorities. Patent family and bibliographic information is available for 56 national and regional patent offices, many providing coverage from as early as 1968. Titles appear in their original language and abstracts are not available. The database is updated weekly.

4.1.8. PHARM

This file contains bibliographic records of patents in the fields of pharmaceutical chemistry and biology. Coverage includes European, French, and U.S. patents from 1986, German and British from the 36th week of 1992, PCT from 1993, and French Special Patents for Medical Compositions (BMS) from 1961. Records emphasize the pharmaceutical aspects of the invention. PHARM is produced by INPI (97).

4.2. Full-Text Databases

The bibliographic databases discussed contain only a portion of the total information in the patent documents. Although several patent issuing authorities have, or are developing, full-text databases (98), remote on-line access to such databases has been limited. Two vendor-provided full-text patent databases are LEXPAT, produced by Mead Data Central, and PATFULL, produced by Dialog Information Services. The textual information for both of these is obtained from tapes available through the USPTO. These databases contain the full text of U.S. patents issued from 1974, but graphics, drawings, and chemical structures are not included. Searching can be based on words from the full document, USPTO Classifications, or International Patent Classifications. In 1994, Dialog announced the European Patents Fulltext database which contains the complete text of European published applications and patents, and the European PCT published applications. The database is produced by the European Patent Office and the text is in the original language of the application.

4.3. Chemical Substructure Databases

Several patent databases are searchable by chemical substructure (99). These are designed to give higher relevance of retrieval when searching chemical compounds than the bibliographic or full-text databases.

MPHARM is a companion database to the bibliographic PHARM. It contains the specific and generic structure records for compounds disclosed in patents included in the bibliographic database. Compound numbers located in MPHARM can be searched in PHARM to retrieve the corresponding bibliographic records (100).

WPIM (World Patents Index Markush), produced by Derwent Publications, Ltd., contains the specific and generic structure records for compounds in the patents included in Derwent Sections B (Farmdoc), C (Agdoc), and E (Chemdoc) since 1987. Sources include patents from 29 industrialized countries as well as European and PCT patents and also items from Research Disclosure and International Technology Disclosures. The compound numbers of relevant references found in WPIM can be searched in Derwent's WPI database to retrieve the corresponding bibliographic information.

MARPAT, produced by Chemical Abstracts Service, contains the generic structure records for patent publications since 1988, which are included in the CA file. Sources include patents from 26 countries plus EPO and PCT publications. Bibliographic records for retrieved references can be directly accessed in this database (101).

4.4. CD-ROM Databases

Since about 1989, CD-ROM format bibliographic and image patent databases have become available as current awareness, reference, or image storage and retrieval tools. The databases are designed for stand-alone PC or local network use, and in some cases they may be alternatives to retrieving patent information on-line. Image files dramatically reduce storage requirements compared to paper or microfilm. Records on each disk are searchable by the bibliographic information from patent documents or from weekly official patent office gazettes or bulletins. The USPTO and vendors are investigating user interest in full image technology subsets, eg, genetic engineering, biotechnology, and acid rain, in the CD-ROM format. Derwent Publications Ltd. provides *Documentation Abstracts* from 1992 as CD-ROM, with full-text abstracts and bibliographic information, including figures or structures. JAPIO provides image and index data as well as full-text for unexamined Japanese applications and utility models from 1987 to the present. Table 10 summarizes U.S. patent information resources available on CD-ROM.

The European Patent office ESPACE series of CD-ROM products are summarized in Table 11.

Table 10. CD-ROM Databases

Database	Producer	Vendor	Description
APS (Automated Patent Searching)	U.S. Patent and Trademark Office (USPTO)	MicroPatent	covers U.S. granted patents 1975–present; first page and exemplary claim; updated monthly within two weeks of final issue date each month; cumulated to one disk/three years
CASSIS	USPTO	USPTO Office of Electronic Data Conversion and Dissemination	CASSIS, the Classification and Search Support Information System of the USPTO, comprises three subfiles: CASSIS/BIB, bibliographic information for utility patents from 1969 and for others from 1977; CASSIS/CLASS, USPTO classification by patent number of class/subclass; CASSIS/ASSIST, index to U.S. Manual of Classification: U.S. Manual of Classification, Class Definitions; IPC, U.S. Classification Concordance; Manual of Patent Examining Procedure; Attorneys/Agents Roster, etc
FullText	USPTO	MicroPatent	covers U.S. granted patents 1991–present; full text including examples, tables, but excluding drawings, structures; updated monthly, not cumulated; 12 disks/year
OG/PLUS	USPTO	Research Publications	CD-ROM version of the USPTO Official Gazette; covers 1990–present; includes searchable subfiles: PATENTS ISSUED, images of O.G. pages searchable by bibliographic fields and first page abstract; PATENT STATUS File, track-ing post-issuance actions, eg, reexaminations, corrections; and LITALERT, containing records of patent suits filed by U.S. District Courts with the USPTO; updated monthly; six disks/year
Patent-Images	USPTO	MicroPatent	covers U.S. granted patents 1990–present; backfile to 1975 available; uses the same Patsoft software as the ESPACE products
PatentView	USPTO	MicroPatent	covers U.S. granted patents from 1986–present; backfile to 1974 to be available in 1994; customized subsets, eg, by company or technology, may be ordered

Table 11. ESPACE Series of CD-ROM Databases

Database	Producer	Description
ESPACE-EP	MicroPatent, Research Publications, EPO, WIPO	contains full text and images of EP applications (EP-A) or granted (EP-B) as separate collections; disks are up-dated weekly (85–90/year); covers 1989–present
ESPACE-FIRST	Micropatent, Research Publications EPO, WIPO	contains scanned images of first-page information of all EPO and PCT inter-national applications; bibliographic fields and patent titles are searchable in English, French, and German; disks are updated bimonthly (five disks/year); covers 1989–present
ESPACE-UK	MicroPatent, Research Publications, EPO	contains full images of U.K. A documents; updated monthly; covers 1991–present
ESPACE-WORLD	MicroPatent, Research Publications, EPO	contains full text and images of PCT applications; titles are searchable in English, French, or German; updated bimonthly (35–40 disks/year); covers 1981–present
ESPACE-ACCESS	MicroPatent, EPO	bibliographic data with abstracts for European (EPO) patent applications; disks are updated quarterly and are cumulative; covers 1978–present

4.5. MARKUSH TOPFRAG

Searching chemical compounds indexed in Derwent's World Patent Index database using fragmentation codes is a complex operation. It requires selecting appropriate codes from the coding manual or coding sheet and linking the codes in time- and subject-dependent search statements. In 1987 Derwent began to index chemical compounds based on an algorithm that allows for a topological search in file WPIM.

Derwent's TOPFRAG family of products is PC-based software that automates the selection of search codes and strategies. A chemical structure is input graphically using a drawing program, and the software generates the appropri-ate codes that define the structure. MARKUSH TOPFRAG, introduced in 1993, is the third generation of TOPFRAG and is the first designed to run under the WINDOWS environment. This software generates the chemical fragmentation codes for searching in WPI as well as the atoms, bonds, connections, etc, to be used in searching WPIM. The codes are generated in a line-by-line search strategy format. This strategy can be entered manually or, with some word processing, can be formatted to be uploaded via a communications software package.

5. Environmental and Safety Information and Search Methods and Services

Since the 1970s environmental and safety information and awareness have been characterized by legislation and regulation at both the state and federal levels. These actions have spurred a need to collect, organize, and retrieve information to aid in compliance with these laws. In an attempt to find a balance between public health, ecological balance and amenities, and industrial development, information supporting government, industry, and public actions has grown rapidly. The need to find information quickly has increased the demand for developing and using environmental databases.

Numerous reviews of environmental, safety, and health information sources have been published since 1981. A comprehensive review entitled "Environmental Information" was published in the *Annual Review of Information Science and Technology* (ARIST) in 1986 (102) and was updated in 1992. A three-part series entitled "Environment Online: the Greening of Databases" was published in *Database* magazine in August 1991, October 1991, and August 1992 (99, 101, 102). Part 1 covers general interest databases, Part 2, scientific and technical databases, and Part 3, business and regulatory databases; "Environment Online: Update 1993" appeared in the December 1993 issue of *Database* (103). This issue also reported a future trend for accessing electronically stored data for business purposes, eg, the use of the Internet telecommunications network to contact sources of information within the government directly (104). The article provides Internet connection numbers to such information sources as the Environmental Protection Agency (EPA), Online Library System, and the Department of Agriculture's National Agricultural Library, as well as toll-free numbers to EPA- and Department of Labor-sponsored bulletin boards. The government is considering providing a public on-line information system containing information sources and services within the government and indicating where to obtain them (ca 1994).

Comprehensive reviews of medical databases (105) and health and toxicological information systems (106), including search aids in each field, appeared in ARIST publications in 1983 and 1990. Toxicology information was reviewed in 1983 (103) and medical and health information in 1990 (100). Reviews of electronic government information (107) and engineering information systems (108) have also been published and provide an expansion of database knowledge for readers who require crossover information in these fields.

There are public and private databases. Public databases are produced by the government and private enterprise and are commercially available through database vendors, such as STN, DIALOG, BRS, ORBIT, and NLM, and through various universities. Private databases are produced by government agencies, corporations, or other organizations for in-house use by their employees or others affiliated with them. The information in these databases may be made available on a need-to-know basis to individuals or corporations in the public or

private sectors. Knowledge of the existence of private databases is usually obtained by personal contact within an organization or thorough disclosure in published literature. Private industrial databases are not usually accessible by the public, although information contained in them on hazardous materials must be reported to the EPA under provisions of TSCA (Toxic Substances Control Act) Section 8e.

5.1. Public Databases

The most comprehensive list of publicly available databases is the two-volume *Gale Directory of Databases*, hereafter referred to as The Directory, published in 1993. Its index lists 140 databases under the subject heading Environment, 18 under Environmental Engineering, 17 under Environmental Health, 36 under Environmental Law, 50 under Waste Management, 61 under Toxicology, 39 under Industrial Hygiene, 63 under Hazardous Substances, 12 under Health Law, and 27 under Safety. Many entries are repeated from category to category, indicating information crossover within technologies and reinforcing the multidisciplinary nature of these technologies.

The environmental and safety databases are listed in Table 12. They are grouped by category, which is useful when searching the literature: Bibliographic/Directory, Current Research Projects, Legal/Regulatory (Table 13), Numeric/Data, and Newsletters. The Directory may be referenced for details of specific subject entries and for database availability. Databases, such as Chemical Abstracts, Biological Abstracts, and Engineering Index, are not included; for their details, see The Directory.

Table 12. Environmental and Safety Databases

Database	Producer	Subject
	Bibliographic directory	
Acid Rain	Reed Reference Publishing Group Bowker A&I Publishing	acid rain
ACIDOC	Quebec Ministere de l'Environment	acid rain
AGRIS	United Nations Food and Agricultural Organization (FAO)	agriculture
Applied Science & Technology Index	H. W. Wilson Co.	environment
AQUALINE	WRc plc	water resources
AQUAREF	Environment Canada	water resources
BALTIC	Sweden Statens Naturvardsverk (SNV)	Baltic Sea
BIOLIS	Informationszentrum für Biologie	environmental biology
Biological & Agricultural Index	H. W. Wilson Co.	environmental science
BNA Books	Bureau of National Affairs (BNA)	employment law
Chemical Activity Status Report (CASR)	Chemical Information Systems, Inc. (BNA)	chemicals under EPA review
Coal Database	IEA Coal Research	coal science
Current Awareness in Biological Sciences	Pergamon Press Plc.	toxicology, ecology
Current Contents Search	Institute for Scientific Information (ISI)	environmental sciences
DECHEMA/DETEQ	Dechema	German environmental equipment technology
Directory of Occupational Safety & Health (OSH)	Labour Canada	Canadian OSH-related legislation
Eastern European Energy Report	Strategic Marketing	Eastern European environment industries
ECOCERVED	Ecocerved	Italian environment
Ecology Abstracts	Cambridge Scientific Abstracts (CSA)	ecology, environment
EDF-DOC	Electricite de France (EDF)	electric power
ELIAS	Environment Canada	Canadian environment
EMBASE	Elsevier Science Publishers	biomedical, OSH, toxicology

Table 12. ***Continued***

Database	Producer	Subject
ENERGIE	FIZ Karlsruhe	energy-related aspects of environmental and biomedical sciences
Energy Science & Tech-nology (EST)	U.S. Department of Energy (DOE)	energy conservation
Enviroline	Reed Reference Pub-lishing Group	natural resources
Environment Protection	VINITI (Vsesoyuznyi Insitiut Nauchnoy i Tecnicheskoy Infor-matsii)	environmental protection
Environmental Biblio-graphy	International Academy at Santa Barbara	environment
Environmental Mutagen Information Center Data Base	Oak Ridge National Laboratory	physical agents tested for mutagenic activity
Environmental Resources Technology	Petroleum Abstracts	petroleum exploration, production, transport
Environmental Teratology Information Center Database	Oak Ridge National Laboratory	teratology
Facilities Index System	U.S. Environmental Protection Agency (EPA)	EPA-regulated sites
GEOBASE	Elsevier/Geo Abstracts	earth sciences
Hazardous Communi-cation Standard Compliance Manual Database	BNA	OSHA hazards
Health and Safety Science Abstracts	CSA	hazards control
ISTP&B Search	ISI	environmental science
MSDS/FTSS	Canadian Centre for Occupational Health and Safety (CCOHS)	chemicals
National Environmental Data Referral Service Database (NEDRES)	U.S. National Environ-mental, Satellite, Data and Information Service (NESDIS)	environmental data
NATUR	SNV	Swedish environmental issues
Occupational Health & Safety: Seven Critical Issues	BNA	OSHA concerns
Oceanic Abstracts	CSA	marine sciences
OIL	Oljedirektoratet	Scandinavian oil industry
Pollution Abstracts	CSA	pollution
POWER	DOE	energy production
Report on Defense Plant Wastes	Business Publishers, Inc. (BPI)	waste management
REPROTOX	Columbia Hospital for Women	chemical effects on human reproduction
SIREN	Portugal Centro de Estudosem Economia Energia dos Trans-portese e Ambiente	Portuguese environment
Standards & Direct-ories/Normes et Reportoires	Canadian Centre for Occupational Health	Canadian occupational health and safety
Superfund	Pasha Publications, Inc.	hazardous waste cleanup
Tanker	Institut Francais du Petrole (IFP)	shipping accidents with oil released
Toxicological Aspects of Environmental Health	Biosis	pollution effects on environment
Toxicology Abstracts	CSA	toxicology
TOXLINE	U.S. National Library of Medicine (NLM)	toxicology, pesticides, mutagens
UMEDIA	Institut der Deutschen Wirschaft	environmental issues
Umwelt-Datenbank	Online Gesellschaft für Informationsvermitt-lung mbH	German environmental protection products
Umweltliteraturdatenbank (ULIDAT)	Deutshes Umweltbundesamt	German environmental topics
VROMDOC	Netherlands Ministry of Housing	Dutch land registry

Table 12. ***Continued***

Database	Producer	Subject
Waste Management & Resource Recovery	International Research and Evaluation (IRE)	waste management
WATERNET	American Water Works Association	wastewater treatment
	Current research projects	
Coal Research Projects Data Base	IEA Coal Research	coal technology
Energy Research in Progress (ERIP)	Energy Research Development Corp.	Australian energy and conservation
Environmental Research Projects (ENREP)	Commission of the European Com-munities	European Community (EC) environment
EUREKA	Eureka Secretariat	EC's Eureka program
Umweltforschungs-datenbank (UFORDAT)	Deutsches Umweltbun-desamt	Austrian environment
	Numeric data	
AQUA II	Infochem Computer Services Ltd	thermodynamic properties of aqueous solutions
BAKER	J. T. Baker, Inc.	material safety data sheets (MSDS) for 1500 chemicals
CERCLIS Database of Hazardous Waste Sites	EPA and CIS	hazardous substances releases reported to the EPA
CHEMEST	Technical Database Services, Inc. (TDS)	properties of pharmaceuticals and chemicals
Chemical Evaluation Search & Retrieval System (CESARS)	Michigan State Depart-ment of Natural Resources	toxicological data on 370 toxic chemicals
Chemical Hazards Re-sponse Information System (CHRIS)	U.S. Coast Guard	water transport of hazardous chemicals
Chemical Identification of Medicine File (ChemID)	NLM	nomenclature and structure of 200,000 chemicals in NLM file
Environmental Chemicals Data and Information Network (ECDIN)	Commission of the Euro-pean Communities	toxicity of 122,400 chemicals in the environment
Environmental Fate (ENVIROFATE)	EPA	fate of 800 chemicals released to the environment
Environmental Fate Databases	Syracuse Research Corp.	DATALOG, CHEMFATE, BIOLOG, BIODEC files on fate of organic chemicals released into the environment
Hazardline	Occupational Health Services (OHS)	regulatory data on 90,000 hazardous chemicals
Hazardous Chemicals Information and Disposal Database (HAZINF)	University of Alberta	handling instructions for hazardous substances
National Analysis of Trends in Emergencies System (NATES)	Environment Canada	hazardous spill incidents in Canada
OHS Material Safety Data Sheets (OHS MSDS)	OHS	information on 85,000 OSHA-documented chemicals
OHS MSDS Summary Sheet	OHS	information on 10,000 chemicals
Oil and Hazardous Materials Technical Assistance Data System (OHM-TADS)	EPA	technical support for dealing with dangers from oil or hazardous substances
Radioactivity Environ-mental Monitoring (REM)	EC	radioactivity data from the EC relating to food chain contamination
Toxic Chemical Release Inventory (TRI)	EPA	estimated releases of toxic chemicals from 20,000 industrial sites
TSCA Plant and Production Data (TSCAPP)	EPA	production data for TSCA Chemicals
TSCA Test Submissions (TSCATS)	EPA	4200 TSCA chemicals
	Newsletters	
BNA Chemical Regu-lation Daily	BNA	legislation on pesticides, chemicals
BNA Daily News	BNA	U.S. government legal issues

Table 12. ***Continued***

Database	Producer	Subject
BNA Environmental Law Update	BNA	legal actions on EPA rules, Super-fund cleanup, wetlands, etc
BNA International Environment Reporter	BNA	global pollution control legis-lation, conferences, treaties
BNA Occupational Safety & Health Daily	BNA	legal issues affecting occupational health and safety
Brazil Watch	Orbis Publications, Inc.	Brazilian politics, economics, business
Business and the Environment	Cutter Information Corp.	global environmental policies
California Planning & Development Report	Torf Fulton Associates	California land-use regulations and growth control
Environment Watch: Latin America	Cutter Information Corp.	Latin American environmental initiatives
Environment Week	King Communication Group, Inc.	environmental issues, acid rain, recycling, greenhouse effect
Environmental Business Journal	Environmental Business Publishing, Inc.	environmental industry business
Global Environmental Change Report	Cutter Information Corp.	global warming, ozone depletion, acid rain, deforestation
Golob's Oil Pollution Bulletin	World Information Systems	global oil spills: control, prevention, pollution
Greenhouse Effect Report	Business Publishers, Inc. (BPI)	global climatic warming
Ground Water Monitor	BPI	groundwater legal issues, hazardous waste disposal
Industrial Environment	Worldwide Videotex	industrial environment improvement measures
Industrial Health & Hazards Update (IH&HU)	Merton Allen Associates	industrial health and hazards regulations
Louisiana Industry Environmental Alert	Environmental Compliance Reporter, Inc.	industrial regulations and actions in Louisiana
New Jersey Industry Environmental Alert	Environmental Compliance Reporter, Inc.	DEP and EPA policies and effects on New Jersey environmental issues
Nuclear Waste News	BPI	nuclear waste management, safety, security
Occupational Health & Safety Letter	BPI	health and safety in the workplace
Occupational Safety & Health Reporter	BNA	worker safety and health issues
Ozone Depletion Network Online TODAY	Environmental Infor-mation Networks	stratospheric ozone depletion
PressNet Environmental Reports	PressNet Systems, Inc.	state and local environmental issues
Texas Industry Environmental Alert	Environmental Compliance Reporter, Inc.	regulations and EPA actions affecting Texas industries
Toxic Materials News	BPI	legislation related to TSCA
Toxics News	Capitol Reports	hazardous wastes, air and water quality
U.N. Conference on Environment and Development	United Nations Conference on Environment and Development	press releases, documents, speeches from this conference
Waste Information Digests	International Academy at Santa Barbara	waste management, recycling
Waste Treatment Technology News	Business Communications Company (BCC)	handling and management of hazardous waste
World Environment Outlook	BPI	global environmental issues

Table 13. Legal/Regulatory Environmental Databases

Producer	Subject	Coverage
BNA California Environment Daily	Bureau of National Affairs (BNA)	California environmental law
BNA Chemical Regulation Daily	BNA	pesticides, chemicals, biotechnology
BNA Chemical Regulation Reporter	BNA	chemicals, pesticides, hazardous wastes
BNA Daily Environment Report	BNA	state international environmental issues
BNA Environmental Law Database	BNA	chemistry, pesticides, environment
BNA Environmental Law Update	BNA	environmental policy
BNA International Environment Daily	BNA	pollution, waste management
CELDS Environmental Regulations	University of Illinois at Champagne-Urbana	U.S. environmental regulations
Daily Report for Executives	BNA	government regulations
Environment Reporter	BNA	pollution, hazardous waste, environment
Environmental Compliance Update	High Tech Publishing Co.	business compliance with environmental standards
Environmental Health News (EHN)	Occupational Health Services (OHS)	environmental and occupational health
Environmental Law Reporter	Environmental Law Institute	U.S. environmental law
Guide to Federal Environmental Laws	BNA	federal environmental laws for human resource professionals
Statens Naturvardsverk	DAFA Data AB	Swedish National Environmental Protection Board statutes
Toxic Law Reporter	BNA	hazardous waste laws and news
WESTLAW Environmental Administrative Law Database	West Publishing Co.	state environmental regulations in U.S.
WESTLAW Environmental Law Library	West Publishing Co.	U.S. environmental law
WESTLAW Topical Highlights	West Publishing Co.	federal and state court decisions on environmental law

5.2. Private (EPA) Databases

The U.S. EPA maintains a list of approximately 600 current information systems, as well as some of the models and databases used within the organization. The list is published in *Information Systems Inventory* (ISI) which is updated yearly and maintained by the Information Management and Services Division of the Office of Information Resources Management (109). ISI lists the system name and acronym, system level, responsible organization, contact person, legislative authorities, database descriptors, access information, hardware and software, system abstract, and keywords.

ISI is available in hard copy and electronically at EPA's headquarters and regional libraries, and through the National Technical Information Service (NTIS). The electronic form may be installed on IBM PC-compatible computers or placed on local area networks, and run under Microsoft WINDOWS or WordPerfect's Library program. The Macintosh version is no longer available. The 1993 update will include the ISI hardcopy, PC disks, and the PC system user manual. EPA also publishes ACCESS EPA, which provides sources of information, databases, and publications within the EPA. Chapter 5 of that publication includes important environmental databases in air and solid waste, pesticides and toxic substances, water, and cross-program (110). EPA also provides databases accessible through EPA libraries, which describe the private EPA and commercial databases available to library users (111).

6. On-Line Search Aids

6.1. MACCS-II

The Molecular Access System is a chemical information management system from Molecular Design Limited (MDL), San Leandro, California. It offers menu-driven graphical input for building, maintaining, and accessing chemical structures and any associated data, eg, chemical and physical properties, biological activity, toxicity data, pricing, safety, and supplier information. Substructure searches are done by drawing the compound with a mouse or light pen and activating the search process. The software interprets the drawn atoms, bonds, and stereochemistry of a chemical structure; retrieves appropriate compounds; and graphically displays them with their corresponding data. MACCS-II allows for customization of the environment to suit various applications. Because it is designed for both mini- and main-frame computers, this system is suited for developing proprietary personal or corporate databases in which unique or new compounds can be registered together with the data of interest (112). Several companies such as Sandoz and Zeneca have taken advantage of this feature.

The following chemical databases are available for searching in MACCS-II. *Chemical Directory Database* contains a combined catalogue of 66 commercial suppliers of more than 77,000 organic chemicals. *MACCS-II Drug Data Report*, based on the Prous Drug Data Report, includes 39,000 compounds with information on therapeutic indication, biological action, phase of development, related patents, and literature references. *MUSE Database*, the tutorial database for MACCS-II, contains over 100 compounds.

MDL has also developed a special version of MACCS-II called MACCS3D. This is used for storing and searching three-dimensional chemical structures and related data. MACCS3D provides convenient graphical input for building three-dimensional queries based on atom distances, bond angles, torsions, planes, centroids, and excluded volumes. Special facilities are also provided for viewing three-dimensional images of retrieved compounds and related data.

The MACCS3D databases available for searching are *MUSE3D Database* which contains over 700 compounds and is useful in learning to use MACCS3D; *FCD3D* is the Fine Chemicals Directory Database of commercial chemicals; and *CMC3D* is the Comprehensive Medical Chemistry Drug Compendium containing about 6000 compounds. *MDDR3D* is the Drug Data Report, Volume 1, which includes biological data.

MACCS-II enables direct interface with other database management systems, such as the Relational Database Management System (RDBMS) and Oracle, so that databases which contain text and numeric data for which special interfaces are normally needed can be constructed. For example, an Oracle MACCS-II linked system is currently being used by the National Institute on Drug Abuse (113) to develop a database that will allow scientists to determine the molecular structures of cocaine and other controlled substances as well as designer drugs.

7. Optical Disk-Based Information and Document Image Systems

Optical-based storage technology has joined paper, microfilm, and electronic/ magnetic technologies as another medium for the storage, retrieval, and management of information (114–116). Optical media differ from magnetic media in that the information is encoded and read by means of laser optics. Information stored on optical disk may be either in a searchable text format (ASCII) or in a format containing only bit-mapped images, usually obtained as output from a scanner. Through the scanning and digitization process, pages that consist of printed text, graphics, photographs, drawings, handwriting, tables, etc, are converted to their binary representation and are stored as bit-mapped images on the optical media. The information stored on optical disk often has counterparts in other formats, such as printed publications or on-line databases or files. Advantages of document imaging systems for complementing, enhancing, or replacing traditional paper- or microfilm-based systems include increased storage capacity, ease of access via automated retrieval, simultaneous searching

and viewing at multiple workstations, speed of access and delivery resulting in productivity gains, improved customer service, document security, document integrity via preservation and elimination of lost or misfiled documents, and networking and integration capabilities for these systems (114). Among the types of optical media that have been developed, the two most common for information storage and retrieval systems are both optical disk-based systems, namely CD-ROM and WORM (write once read many).

Searchable text information on optical disk can range from bibliographic databases, abstracts, and indexes, to full texts of books, journals, and reference publications, to numeric data. Printed text, graphics, drawings, photographs, spectra, and tabular information may also be stored on optical disk as images. Applications of document-image systems are numerous and span all types of industries and organizations. The potential for integrating, linking, and networking document image systems with other existing electronic databases and systems is being explored in efforts to bring the virtual library to reality.

Several projects such as CORE (Chemistry On-Line Retrieval Experiment) at Cornell University, Project Mercury at Carnegie-Mellon University, RightPages at AT&T, and Red Sage at the University of California, San Francisco are in progress and illustrate the issues arising from implementation of on-line information systems that combine text and image (6, 116).

The objectives of these projects are to investigate how new desktop interfaces to scientific information affect the way scientists access and use the information from databases, journals, and other proprietary and published literature and to address the technical aspects of assembling these seamless networks and user-friendly interfaces. These and other challenges that lie ahead for users, publishers, and suppliers of CD-ROM, optical disk, and document image storage systems encompass some of the same issues that are affecting other areas of information science and technology. The issues of standardization, copyright, and the admissibility of proprietary electronic and optical records as legal documents are not being resolved as quickly as the underlying storage and retrieval technologies are advancing.

Optical media have the capacity to handle much larger amounts of digitized information than equivalent sizes of magnetic media. These capacities are increasing each year. A standard 4.75-inch CD-ROM can hold more than 680 MB (megabytes) of data, roughly the equivalent of 250,000 pages of text. Storage capacities of WORM disks range from 250 MB for 5.25-inch disks to more than 14 GB (gigabytes) for 14-inch ones; this translates into 20,000 and 130,000 imaged pages, respectively. Whereas CD-ROM has emerged as a suitable medium for the information publishing industry, WORM applications are used mainly for archival document image storage systems.

7.1. Publications on CD-ROM

The *Gale Directory of Databases*, Volume 2, contains a comprehensive listing of available CD-ROM based products (117). *Analytical Abstracts*, *Chemical Abstracts 12th Collective Abstracts and Index*, *MEDLINE*, *Chemistry Citation Index*, the *CASurveyor* series (separate disks on specific areas of chemistry), and the *PolTox* series (pollution and toxicology) are among some of the bibliographic databases and indexes available in CD-ROM versions. Full-text sources on CD-ROM include catalogues: *Aldrichem Data Search;* encyclopedias: *Kirk-Othmer Encyclopedia of Chemical Technology* and *Polymer Encyclopedia* (complete text of the *Encyclopedia of Polymer Science and Technology*); dictionaries: *CHCD Dictionary of Organic Chemistry* and *CHCD Dictionary of Inorganic Chemistry*; and directories: *Chem Sources*. Examples of numeric data sources available on CD-ROM are toxic and hazardous chemicals information in *Chem-Bank*, x-ray powder diffraction patterns in *Powder Diffraction File*, and properties of plastic materials in *CenBASE/Materials*. *Beilstein Current Facts in Chemistry* on CD-ROM allows searching of current bibliographic citations on organic chemistry by chemical structure and other physical and chemical properties. Some journals on CD-ROM are available in full-text searchable format; for example, *The Lancet* (1989–1994). The *ADONIS* collection of CD-ROMs contains full-text images of all articles, letters, and abstracts from more than 490 biomedical journals. Other commercial document image collections on CD-ROM include U.S. patents on *PatentImages* and *PatentView*, European

patent applications on *ESPACE*, and Department of Defense Standards on *DoD Standardization Service*. The *Worldwide Standards Service* on CD-ROM contains a comprehensive index to standards and specifications from more than 375 international organizations and associations; the full standards as images on CD-ROM are also available as part of this service.

7.2. CORE

The CORE Electronic Chemistry Library is a joint project of Cornell University, OCLC (On-line Computer Library Center), Bell Communications Research (Bellcore), and the American Chemical Society. The CORE database will contain the full text of American Chemical Society Journals from 1980, associated information from Chemical Abstracts Service, and selected reference texts. It will provide machine-readable text that can be searched and displayed, graphical representations of equations and figures, and full-page document images. The project will examine the performance obtained by the use of a traditional printed index as compared with a hypertext system (SUPERBOOK) and a document retrieval system (Pixlook) (6, 116).

7.3. Project Mercury

This system, at Carnegie-Mellon University, implements electronic library architecture through distributed rather than centralized computing. Elsevier and the Institute of Electronic and Electrical Engineers are also helping to distribute full-page images to users' desktop computers across the campus network. In this project a search is performed in the INSPEC database, citations are selected on the screen, and the full article appears for viewing. Detailed statistics are gathered to provide the publishers with information that helps them devise marketing, pricing, and delivery strategies (118).

7.4. RightPages

The RightPages image-based electronic library, developed for users at AT&T Bell Laboratories, is an on-line version of the periodical shelves of a conventional library. An added feature is that RightPages alerts users to the arrival of new journal articles matching a specified profile and enables them to examine pages in the article and browse other articles in the database (119). AT&T is also collaborating with Springer-Verlag in another electronic library project at the University of California, called Red Sage (6). Forty Springer-Verlag journals in the areas of molecular biology and radiology will be scanned into a bit-mapped image system and transformed into searchable text via OCR (optical character recognition). The user interface for this project will be AT&T's RightPages and will contain a hyper-paper feature that will allow users to view pages containing figures immediately by pointing to the citations for those figures.

8. Private Bibliographic and Text Databases

Personal computers have introduced new ways to handle private biblographic and text files. The most important factors to consider to achieve satisfactory results in building a bibliographic or text database are the type of information to be stored and the needs of the user. Types of information include correspondence, research results and documentation, meeting notes, and bibliographic references. Needs of the user to be considered should include the potential number of users of the database, restrictions for the access and display of the information because of privacy or proprietary reasons, and the retrieval mechanisms (eg, by keyword, authority list, controlled vocabulary, author, title, date, or other document or information attributes). In addition, criteria for selecting and encoding information for the database need to be established.

The potential size of the database and the number of users will influence the decision to use free text, keywords, or natural language vs a thesaurus-controlled vocabulary for subject description or analysis. In free text or natural language, individuals may select keywords or they may be automatically generated by the software used. Controlled vocabulary assignment is done using an authority list or thesaurus, which is subsequently used for retrieval. For example, the controlled term "preparation" could be used to retrieve documents on synthesis, manufacturing, preparing, formation, or reactions of. In free-text searching, all of the above words would need to be searched. Controlled vocabulary assignment requires more effort to input but less effort to search; free-text keyword assignment requires less effort to input and more to search. Additional information is available on indexing or term assignment (120–125).

If the database is to contain published literature, many software packages have the capability to download or copy from on-line databases. Also, most terminal emulation software packages have a copy or capture feature. When downloading from commercial databases, license agreements and copyright requirements must be honored (check with the database vendors as to their specific licensing agreements). In addition to published literature, proprietary information can be added to the database by copying or downloading from an in-house computer system (125).

The type of hardware or computer system to be used and the potential size of the database should also be considered. For some databases, a personal computer may be adequate. For others, especially if there will be multiple users, a mainframe computer or a network of personal computers may be required. Storage capacity and response time are parameters that should be considered. However, computer technology changes so rapidly that vendors and computer experts should be consulted when building any personal databases.

Another factor frequently overlooked in private database creation is commitment to the support and maintenance of the database. Support involves training users, solving software and hardware problems, and upgrading the software when new features become available or are needed by the end user. Maintenance of the database includes adding new information, deleting information no longer wanted, and correcting information in the database when errors are detected. Once the needs and requirements are documented, available software products should be surveyed to determine which one is most suitable and whether customization to fit the user's needs is required. Customization may include defining the format of the field or area in the database in which the information is to be stored. Some examples are the format for the name of an author, eg, last name and initials, or last name and full first name; how many characters should be allowed in a field for a title or abstract; and the format for a date, eg, year/month/day or day/month/year. Depending on the software selected and the skills of the database user, customizing can be done by the user or someone trained in computer technology. Libraries, research groups, and individuals can use a variety of software products to develop private databases in their specific areas of interest (126, 127). Software packages available for building databases range from generic personal computer database management systems, which require customizing to software designed specifically for bibliographic files. Most of the software is for Macintosh or IBM compatible PCs. Some examples are PROCITE (128), NOTEBOOK II (129), ENDNOTE (130), LIBRARY MASTER (131), PERSONAL FILE SYSTEM (132), ASKSAM (133), BASISPLUS (134), and PERSONAL LIBRARIAN (135). A buyer's guide to software products for management of information by industrial scientists is available (128). Vendors and their product literature and product reviews in library, information science, and personal computing journals are good sources of information. Private databases offer a fast, convenient, and economic tool to aid researchers in planning, interpreting, and reporting their results.

BIBLIOGRAPHY

"Literature, Mechanized Searching" in *ECT* 1st ed., Vol. 8, pp. 449–467, by J. W. Perry, Bjorksten Research Laboratories, and R. S. Casey, W. A. Sheaffer Pen Co.; "Literature of Chemistry and Chemical Technology" in *ECT* 2nd ed., Vol. 12, pp. 500–529, by T. J. Devlin, Esso Production Research Co., and B. H. Weil, Esso Research and Engineering Co.; "Information Retrieval Services and Methods" in *ECT* 2nd ed., Suppl. Vol., pp. 510–535, by E. Garfield and C. E. Granito, Institute for Scientific Information, and A. E. Petrarca, Ohio State University; "Information Retrieval " in *ECT* 3rd ed., Vol. 13, pp. 278–336, by M. H. Graham and L. Y. Stroumtsos, Exxon Research and Engineering Co., A. B. Lamy, Essochem Europe, Inc., and B. Lawrence, Exxon Corp.

Cited Publications

1. H. Skolnik, *The Literature Matrix of Chemistry*, John Wiley & Sons, Inc., New York, 1982, p. vi.
2. G. Wiggins, *Chemical Information Sources*, McGraw-Hill Book Co., Inc., New York, 1991.
3. R. Maizell, *How to Find Chemical Information*, John Wiley & Sons, Inc., New York, 1987.
4. J. Branin, *Collection Building* **9**(3–4), 20 (1989).
5. M. Khalil, *Library J.* **118**(2), 43 (1993).
6. S. Borman, *C. & E. News* **71**(24), 10 (1993).
7. *Advanced Technology Libraries* **20**(12), 3 (1991).
8. *The High-Performance Computing Act of 1991*, PL 102–194, 105 *STAT*. 1594, Washington, D.C.
9. R. Tennant, J. Ober, and A. G. Lipow, *Crossing the Internet Threshold: An Instructional Handbook*, Library Solutions Press, San Carlos, Calif., 1993.
10. E. Krol, *The Whole Internet User's Guide & Catalog*, O'Reilly & Associates, Inc., Sebastopol, Calif., 1992.
11. S. Fishman, *The Copyright Handbook*, Nolo Press, Berkeley, Calif., 1992.
12. L. L. Schaper, *Proceedings of the 12th National Online Meeting, 1991*, Learned Information, Medford, N.J., 1991, 349–352.
13. J. Garrett, *Online* **15**(2), 22 (1991).
14. R. K. Bose, *Copyright Issues in Multimedia*, SRI International Business Intelligence Program D92-1692, Menlo Park, Calif., 1992, p. 12.
15. W. Peryman, *J. Library Admin.* **15**(1–2), 81 (1991).
16. M. Fleming, ed., *Online Factbook*, Digital Information Group, Stamford, Conn., 1992.
17. C. T. Meadow, *Database* **11**(5), 14 (1988).
18. *Ibid*., p. 23.
19. R. K. Summit, *Online* **11**(1), 61 (1987).
20. M. E. Williams, in K. Y. Marcaccio, ed., *Gale Directory of Databases*, Gale Research Institute, Detroit, 1993, pp. xvii–xxvii.
21. S. E. Arnold, *Online* **13**(2), 6 (1989).
22. P. P. Massa, *Bull. Am. Soc. Info. Sci.* **17**(6), 8 (1991).
23. R. Brody, *Online* **17**(3), 66 (1993).
24. T. B. Chadwick, *Online* **13**(1), 26 (1989).
25. N. Garman, *Online* **13**(1), 6 (1989).
26. R. Basch, *Database* **14**(5), 13 (1991).
27. M. O'Leary, *Database* **13**(2), 15 (1990).
28. J. Thompson, *Online* **13**(6), 11 (1989).
29. H. Pemberton, *Online* **16**(2), 12 (1992).
30. S. N. Bjorner, *Online* **14**(4), 90 (1990).
31. M. O'Leary, *Online* **16**(3), 29 (1992).
32. R. J. Massie, *The American Chemical Society and Dialog Information Services Settle Litigation*, press release, Chemical Abstract Service, Columbus, Ohio, Oct. 29, 1993.
33. D. T. Hawkins, *Online* **17**(4), 98 (1993).
34. M. O'Leary, *Online* **17**(1), 34 (1993).

35. P. Lane, *Info. Today* **9**(11), 1 (1992).
36. A. J. Van Camp, *Online* **16**(2), 102 (1992).
37. A. Novallo, ed., *1994 Information Industry Directory*, 14th ed., Vol. **1**, Gale Research Inc., Detroit, 1994.
38. *The CIS (Chemical Information System): An Overview*, Chemical Information Systems, Towson, Md., 1993, p. 1.
39. L. Eichler and J. Newland, *Database* **16**(3), 48 (1993).
40. C. Wilson, ed., *The Chronolog* **21**(4), 91 (1993).
41. *DIALOG Database Catalogue*, Dialog Information Services, Inc., Palo Alto, Calif., 1993, p. 3.
42. *ORBIT Database Catalogue*, Info Pro Technologies, 1993, McLean, Va., p. 1.
43. *Questel Database Catalog-1993*, Questel, Inc., Arlington, Va., 1993, p. i.
44. *July 1993 STN Database Catalog*, Chemical Abstracts Service, Columbus, Ohio, 1993, p. 4.
45. B. D. Christie, B. A. Leland, and J. G. Nourse, *J. Chem. Inf. Comput. Sci.* **33**(4), 545 (1993).
46. A. J. Gushurst, and co-workers, *J. Chem. Inf. Comput. Sci.* **31**(2), 447 (1991).
47. M. G. Hicks and C. Jochum, *Anal. Chim. Acta* **235**, 87 (1990).
48. R. Attias and J. -E. Bubois, *J. Chem. Inf. Comput. Sci.* **30**(1), 2 (1990).
49. J. M. Barnard, *J. Chem. Inf. Comput. Sci.* **33**(4), 532 (1993).
50. M. G. Hicks and C. Jochum, *J. Chem. Inf. Comput. Sci.* **30**(2), 191 (1990).
51. A. Barth, in S. R. Heller, ed., *ACS Symposium Series 436*, American Chemical Society, Washington, D.C., 1990, 24–41.
52. *Beilstein and Gmelin on STN*, Springer-Verlag, Heidelberg, Germany, 1993.
53. P. G. Dittmar, *J. Chem. Inf. Comput. Sci.* **23**(3), 93 (1983).
54. J. R. McDaniel and A. E. Fein, in W. A. Warr, ed., *ACS Symposium Series 341*, American Chemical Society, Washington, D.C., 1987, p. 62.
55. Beilstein-Institüt für Literatur der Organischen Chemie, Germany.
56. Gmelin-Institüt für Anorganische Chemie ünd Grenzgegiet, Germany.
57. MPD Network, Columbus, Ohio.
58. Technical data, Technical Databases Services, Inc. (TDS)-Numerica, New York.
59. S. E. Jakes, and co-workers, *J. Molec. Graphics*, **5**(1), 41 (1987).
60. F. H. Allen, M. J. Doyle, and R. Taylor, *Acta Crystallog., Section B*, **B47**(1), 50 (1991).
61. R. A. Engh and R. Huber, *Acta Crystallog., Section A*, **A47**(4), 392 (1991).
62. C. Craver, ed., *The Coblentz Society Desk Book of Infrared Spectra*, The Coblentz Society, Inc., Kirkwood, Mo., 1977.
63. A. N. Davies, H. Hillig, and M. Linshceid, *Proceedings of the Scientific Computing and Automation (Europe) Conference*, 1990, p. 445.
64. E. Pretsch, M. Farkas, and A. First, *Proc. Int. CODATA Conf.*, 11th ed., 1988, p. 176.
65. W. A. Warr, *Chemometrics and Intell. Lab. Sys.* **10**, 279 (1991).
66. S. G. Lias, *J. Res. Natl. Inst. Stand. Technol.* **94**(1), 25 (1989).
67. E. Kono, A. Hidetsugu, and S. Sasaki, *Joho Kagaku Toronkai, Kozo Kassei Sokan Shinpojiumu Koen Yoshishu* **13–18**, 149 (1990).
68. M. G. Weller, in H. Collier, ed., *Recent Advances in Chemical Information*, Royal Society of Chemistry, London, 1992, 197–208.
69. S. R. Heller, *Chem. Int.* **13**(6), 235 (1991).
70. O. Yamamoto, K. Hayamizu, and M. Yanagisawa, *Anal. Sci.* **5**(2), 141 (1989).
71. O. Yamamoto, K. Hayamizu, and M. Yanagisawa, *Joho Kagaku Toronkai, Kozo Kassei Sokan Shinpojimu Koen Yoshishu* **13–18**, 53 (1990).
72. K. S. Lebedev, E. A. Otmakhova, and I. V. Gritsenko, *Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim. Nauk* **5**, 73 (1990).
73. D. E. Meyer, W. A. Warr, and R. A. Love, eds., *Chemical Structure Software for Personal Computers*, American Chemical Society, Washington, D.C., 1988.
74. J. L. Markley, and co-workers, in M. Ikehara, ed., *Protein Eng. Proc. Int. Conf. Protein Eng., 2nd*, 1989, Japan Scientific Society Press, Tokyo, 1990, p. 285.
75. P. R. Griffiths and C. L. Wilkins, *Appl. Spectros.* **42**, 538 (1988).
76. D. D. Speck, R. Venkataraghavan, and F. W. McLafferty, *Org. Mass Spectrom.* **13**, 209 (1978).
77. F. W. McLafferty, and co-workers, *Int. J. Mass Spectrom. Ion Phys.* **47**, 317 (1983).
78. D. T. Terwilliger, and co-workers, *Biomed. Environ. Mass Spectrom.* **14**, 263 (1987).
79. *CIS (Chemical Information System): An Overview*, Chemical Information Systems, Inc., Towson, Md., 1993.

80. K. Hayamizu, *Kagaku to Sofutowea* **14**(3), 191 (1992).
81. K. Tanabe, K. Hayamizu, and S. Ono, *Anal. Sci.* **7**, 711 (1991; Suppl.), *Proc. Int. Congr. Anal. Sci.*, Pt. 1.
82. J. R. Rumble, Jr., D. M. Bickham, and C. J. Powell, *Surf. Interface Anal.* **19**(1–12), 241 (1992).
83. H. Chihara and K. Mano, *Z. Narurforsch., A: Phys. Sci.* **47**(1–2), 446 (1992).
84. K. Okano and A. Abe, *Joho Kanri* **31**(1), 56 (1988).
85. G. H. Wood, J. R. Rogers, and S. R. Gough, *J. Chem. Inf. Comput. Sci.* **29**, 118 (1989).
86. P. E. Blower, in P. Willett, ed., *Modern Approaches to Chemical Reaction Searching*, Gower, Aldershot, U.K., 1986, p. 146.
87. *Chem. Pharm. Bull.* **41**(11), 1906–1909 (1993).
88. *REACCS Information Management Reference Manual*, revision 8.0, Molecular Design Ltd., San Leandro, Calif.
89. W. T. Wipke and T. M. Dyott, *J. Am. Chem. Soc.* **96**, 4825 (1974).
90. E. Aberdeen, *Group Market Viewpoint* **4**(11) (Nov. 12, 1991).
91. J. T. Butler, *LawLibrary J.* **84**, 121 (1992).
92. E. S. Simmons, *Database* **8**(1), 49 (1985).
93. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **24**(1), 10 (1984).
94. N. Lambert, *Database* **10**(6), 46 (1987).
95. S. M. Kaback, *World Pat. Info.* **11**(2), 95 (1989).
96. E. S. Simmons, *Online* **10**(4), 51 (1986).
97. M. P. O'Hara and C. Pagis, *J. Chem. Inf. Comput. Sci.* **31**(1), 59 (1991).
98. J. Lucas, *World Pat. Info.* **14**(3), 167 (1992).
99. K. A. Cloutier, *J. Chem. Inf. Comput. Sci.* **31**(1), 40 (1991).
100. P. G. Alston and F. W. Stoss, *Database* **15**(4), 17 (1992).
101. T. Ebe, K. A. Sanderson, and P. S. Wilson, *J. Chem. Inf. Comput. Sci.* **31**(1), 31 (1991).
102. R. R. Freeman and M. F. Smith, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARTIST)*, Vol. **21**, Elsevier Science Publishers, New York, 1986.
103. R. Basch, *Database* **16**(6), 47 (1993).
104. M. J. Cronin, *Database* **16**(6), 47 (1993).
105. C. B. Tilly, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. **25**, Elsevier Science Publishers, New York, 1990.
106. H. M. Kissman and P. Wexler, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. **18**, Elsevier Science Publishers, New York, 1983.
107. P. Hernon and C. R. McClure, in, M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. **28**, Elsevier Science Publishers, New York, 1993.
108. E. N. Mailloux, in M. E. Williams, ed., *Annual Review of Information Science and Technology (ARIST)*, Vol. **24**, Elsevier Science Publishers, New York, 1989.
109. EPA, *Information Systems Inventory*, NTIS kit order no. PB-94-500501, Environmental Protection Agency, Washington, D.C., 1993.
110. EPA, Access EPA, EPA/IMSD-91-100, GPO order no. 055-000-00378-5 or NTIS PB91-151563, Environmental Protection Agency, Washington, D.C., 1993.
111. EPA, *Databases Accessible Through EPA Libraries*, Environmental Protection Agency, Washington, D.C. (in press) (1994).
112. V. E. Hample, D. P. Grubb, and A. Moulik, in J. R. Williams, N. A. Gocken, and J. E. Morral, eds., *Computerized Metallurgical Databases*, Metallurgical Society, Warrendale, Pa., 1988, p. 19.
113. K. D. Schwarz, *Gov. Comp, News* **11**(2), 41 (1992).
114. L. F. Lunin, in M. E. Williams, ed., *Annual Review of Information Science and Technology*, Vol. **22**, Elsevier Science Publishers, Amsterdam, 1987, p. 179.
115. E. K. Brumm, in M. E. Williams, ed., *Annual Review of Information Science and Technology*, Vol. **26**, Learned Information, Inc., Medford, N.J., 1991, p. 197.
116. L. Krumenaker, *Science* **260**, 1066 (1993).
117. K. Y. Marcaccio, ed., *Gale Directory of Databases*, Vol. **2**, Gale Research Inc., Detroit, Mich., 1993.
118. K. M. Ginther-Webster, *AI Rev. Prod., Serv., Res.* **3**, 25 (1990).
119. G. A. Story, *Computer* **25**, 17 (1992).

120. R. Fidel, *Special Libraries* **83**(1), 1 (1992).
121. B. R. Boyce and J. P. McLain, *J. Am. Soc. Info. Sci.* **40**(4), 273 (1989).
122. F. W. Lancaster, C. Elliker, and C. T. Harkness, *Ann. Rev. Info. Sci. Technol.* **24**, 35 (1989).
123. S. C. Biswas and F. Smith, *Library & Info. Sci. Res. II*, (2), 109 (1989).
124. P. Willett, *Document Retrieval Systems*, (1988).
125. A. Hopkinson, *9th International Online Information Meeting*, 295–304 (1985).
126. T. E. Wolff, *Database* **15**(2), 34 (1992).
127. C. Moon, *Int. J. Info. Manage.* **8**(4), 265 (1988).
128. D. Cheney and G. Jenks, *Library Software Rev.* **7**(6), 411 (1988).
129. G. Lundeen, *Database* **12**(3), 36 (1989).
130. S. Stigleman, *Database* **15**(5), 50 (1992).
131. A. Raeder, *Database* **14**(5), 67 (1991).
132. T. Hanson, *Library Micromation News*, (33), 16 (1991).
133. K. Watterson, *Data Based Advisor II*, (6), 45 (1993).
134. E. Perez, *Database* **15**(6), 45 (1992).
135. G. W. Lundeen and C. Tenopir, *Database* **15**(4), 51 (1992).

General References

136. G. Anderson, *Acquisitions Librarian* **6**, 3 (1991).
137. A. R. Haygarth-Jackson, *Serials '85. Proceedings of the U. K. Serials Group Conference*, 29–47 (1985).
138. S. James, *Library Rev.* **39**(4), 21 (1990).
139. H. Martin, *Law Library J.* **82**(1), 129 (1990).
140. D. Tyckson, *Reference Librarian* **34**, 37 (1991).
141. Du Pont Co., Scientific Computing Division, Experimental Station, Wilmington, Del., 1988.
142. L. Katz, *Rolling Stone* **77** 25–36 (Apr. 15, 1993).
143. E. Krol, *The Whole Internet*, O'Reilly & Associates, Inc., Sebastopol, Calif., 1992.
144. C. A. Lynch and C. M. Preston, in C. A. Lynch and C. M. Preston, *Annual Review of Information Science and Technology*, Vol. **24**, Learned Information, Medford, N.J., 1989, 263–312, 140 references.
145. Online files from Sura.Net.
146. M. Rapaport, *Online* **15**(3), 33 (1991).
147. H. N. Tillman, *SpeciaList* **16**(3), 1, 3 (1993).
148. B. Brain, *Australian Academic and Research Libraries* **22**(3), 167 (1991).
149. M. K. Duggan, *Online* **15**(3), 20 (1991).
150. J. Garrett, *Proceedings of the 12th National Online Meeting, 1991*, Learned Information, Medford, N.J., 1991, 99–107.
151. L. Garson, *J. Chem. Inf. Comput. Sci.* **24**(3), 119 (1984).
152. M. Hattery, *Info. Retrieval & Library Automation* **28**(10), 1 (1993).
153. M. Jensen, *Serials Rev.* **18**(1–2), 62 (1992).
154. J. Ogburn, *Library Acquisitions, Practice and Theory* **14**(3), 257 (1990).
155. M. K. Booker, *Encycl. Materials Sci. and Engineering* **2**, 796–800 (1986).
156. N. Herrlich and J. Wierer, *Kunstst. Ger. Plast.* **79**(10), 106–109 (1989).
157. F. C. Allan and W. R. Ferrell, *Database* **12**, 50–58 (1989).
158. J. G. Kaufman, *Engineering with Computers* **4**, 75–85 (1988).
159. Third Chemical Congress of North America, American Chemical Society, June, 1988, Toronto, Canada, Herman Skolnik Award Symposum, Scientific Numerical Databases, Present and Future.
160. Association for Information and Image Management (AIIM), *Information & Image Management: The State of the Industry 1993*, Association for Information and Image Management, Silver Spring, Md., 1993.
161. P. Jasco, *CD-ROM Software, Dataware and Hardware: Evaluation, Selection and Installation*, Libraries Unlimited, Englewood Cliffs, Colo., 1992.
162. J. Langlois, *CD-ROM 1992L An Annotated Bibliography of Resources*, Meckler Corp., Westport, Conn., 1992.
163. G. Tapper and K. Tombs, *Admissibility of Document Imaging Systems*, Meckler Corp., Westport, Conn., 1992.

164. NERAC, bibliography, *Full Text Database Searching*, pub. no. PB92-860469, Tolland, Conn., 1992.
165. NERAC, bibliography, *Database Design*, pub. no. PB92-861665, Tolland, Conn., 1992.

CYNTHIA S. BARCELON-YANG
EVELYN L. BROWNLEE
EMMETT D. CALHOUN
BRUNO A. CAPUTO
CHARLES C. CUMBO
JOSEPH P. DANISZEWSKI
DOUGLAS A. ECKEL
KENNETH H. GLASPEY
DARLYN C. GREEN-KOCHER
MARIANNE B. GRUBER
MARGARET M. ISSELMANN
THOMAS C. JOHNS
ALICIA P. KING
DAVID M. KRENTZ
FLORENCE H. KVALNES
LURAY M. MINKIEWICZ
BEHROOZ NAZER
ANGELA K.G PARSONS
CAROL R. PERROTTO
RITA D. RATLIFF
JEANETTE C. SIKES
AMIE H. WEBSTER
Du Pont Company

Related Articles

Copyrights, Trademarks; Databases; Patents; Information storage methods