Repository Profile

Chemical Abstract Service

(Version: May 28, 2009)

By Marie Waltz, Center for Research Libraries, with technical analysis by James A. Jacobs.

About the Long-Lived Digital Collections Case Studies

With funding from the National Science Foundation, CRL is engaged in a two-year project to analyze eight established, "long-lived" collections of digital data and content. These case studies build upon the *Trustworthy Repositories Audit & Certification Checklist* (TRAC) criteria and the audits of the Portico, LOCKSS and ICPSR repositories conducted by CRL in 2006-2007 to test and refine those criteria.

The CRL case studies serve a different purpose than the aforementioned audits. While the audits probed the soundness of repository organizational and technical infrastructure, the case studies identify practices, strategies and mechanisms that have enabled repositories to sustain massive digital collections over substantial periods of time. As such the case studies contribute to the understanding of repositories and other sources of scholarly information in digital form, to inform decisions about investing in those resources.

The American Chemical Society's Chemical Abstracts Service is the subject of the present case study. As producers and custodians of major databases and indexes for the field of chemistry, CAS and its operations are a matter of interest to libraries, researchers, scientists and others.

This profile of CAS will be updated periodically, as CRL further examines CAS archiving practices and strategies, past and present. CRL welcomes comments and corrections from readers.

A Note about Sources:

This report is based upon information obtained from users of Chemical Abstracts Service (CAS) databases, from published sources, and from the American Chemical Society (ACS) and CAS itself. CAS's 2007 centennial resulted in *CAS Surveys Its First Century*, a detailed history of the service written by Eric Shively and first published in a special issue of the journal *Chemical and Engineering News*. Section Two of the present report draws heavily upon this key source. Beyond this, CAS declined to respond to most of CRL's request for information about their technical processes and procedures.

In addition, CRL is indebted to Andrea Twiss-Brooks, Science Librarian at the University of Chicago and Patricia Kirkwood, Engineering, Mathematics, & Computer Science Librarian at the University of Arkansas for sharing their knowledge and expertise on the products produced by the American Chemical Society and CAS.

1) *Overview*

CAS is a division of the American Chemical Society. CAS and its partner organizations index and abstract a large portion of the world's chemical information and literature. In doing this work CAS has created and maintains key collections, databases and indexes; produces and distributes tools and interfaces for managing and analyzing chemical related information; and provides an array of services to those in the field of chemistry. CAS database products serve a variety of scientific disciplines, including biomedical, chemical, materials, and agricultural sciences and engineering. Most users of those products are in academia, the chemical industries and government agencies.

2) *Mission and History*

The mission of CAS is "to be the world leader in disseminating chemical and related scientific information."[1] While the service's mission statement does not specifically refer to managing or maintaining digital information per se, it does speak to providing access to information that enables scientific discovery.

[1] "CAS Information Use Policies." CAS Website. <http://www.cas.org/legal/infopolicy.html>

Since 2007, the American Chemical Society mission has been "to advance the broader chemistry enterprise and its practitioners for the benefit of Earth and its people[2]." With this statement the ACS identified its goal of becoming a global chemical association. CAS, which has always sold its products around the world, continues to develop and market products to the global chemical community. CAS's business purpose is to provide subscribers with worldwide chemical and chemistry-related science information.

CAS is the larger of two revenue-generating divisions within the ACS. The other is the ACS Print Publications Division, which is responsible for producing the Society's journals and other publications that are issued in traditional published media.

History

The ACS began abstracting scientific journals in 1907. The journal *Chemical Abstracts* (often referred to as *CA*) was created at that time by and for scientists in the United States, as an expansion of the abstracting section that had been published as a supplement to the *Journal of the American Chemical Society* since 1897. At the time European chemistry abstracting services neglected much of the work published in American journals. The first editor of *Chemical Abstracts* was William A. Noyes, Sr. Noyes worked from his office at the National Bureau of Standards, along with two part-time editors and a secretary. Many of the abstracts were produced by 129 volunteers, who during the service's first year generated approximately 12,000 abstracts.

In 1909 *Chemical Abstracts'* editorial office moved to the campus of Ohio State University in Columbus, Ohio. At that time Austin M. Patterson, a professor at OSU, became the journal's editor. In 1915, editorship passed to Evan J. Crane, another OSU professor, who would head the operation until 1958. *CA* editorial offices, and later the CAS Division, were based on the OSU campus for 56 years. The offices remain near the OSU campus today. Faculty and students of the OSU chemistry department continue to benefit in a number of small ways from being near CAS (job opportunities, scholarships, etc.), although OSU's Chemistry Department no longer contributes significantly to CAS operations.

CA's third editor, Evan J. Crane, contributed substantially to the development of the CAS organization and greatly expanded its worldwide network of abstractors. To communicate with his army of volunteers Crane devised an innovative newsletter called *The Little CA*, which helped

---

[2] "About ACS." American Chemical Society Web site <http://portal.acs.org/portal/acs/corg/content?_nfpb=true&_pageLabel=PP_TRANSITIONMAIN&node_id=225&use_sec=false&sec_url_var=region1&__uuid=e2647d31-4e76-48d6-ab91-c63a7f6d07d4>

keep volunteers informed about the guidelines and policies governing the writing of abstracts for *CA*.

Crane introduced a number of new tools for chemical research, including indexes that made it possible for chemists to easily locate published information about specific chemical names and processes. In 1917 Crane published the *CA Decennial Index (1907-16*), which introduced a new system for naming and indexing compounds that allowed related compounds to be grouped together for easy reference. Crane's tenure saw a number of innovations designed to facilitate research. In particular, Crane was concerned that various branches of chemistry would create their own nomenclatures and that this would eventually lead to "nomenclatures out of step with one another." To prevent this he developed a new unified system of chemical nomenclature, based on classified subject entries rather than simple word indexing.

Crane also supported the idea of building a new tool for recording research results, which would "systematically record the new information obtained and for keying it to serve further progress effectively."[3] Crane moved CAS to a new building, which would be owned and operated by CAS rather than renting space on the Ohio State University campus.

It was Crane who presided over CAS's most significant budgetary crisis. A business recession was the likely cause of disappointing 1955 financial results, where CAS revenue fell well short of meeting its operating expenses. The crisis led the ACS Board of Directors to establish a new "break even policy." The new policy required CAS to become completely self-supporting and to price its products as necessary to make the service less reliant upon Society membership revenues.[4] Up until that time CAS was subsidized by the ACS. The new policy also changed the status of CAS within the ACS organization, making it an operating division of the Society. As an operating division, CAS was led by a director, and the key decisions were thenceforth made in large part by a committee, rather than directly by the ACS board.

In 1958, Evan J. Crane, who had led CAS for 43 years, retired. During his long tenure he had established CAS as a stable organization. With Crane's departure, Dale B. Baker was charged with modernizing the CAS publishing process. Baker hired G. Malcolm Dyson to lead the

[3] "1951: Evan J. Crane (1889–1966)," *Chemical and Engineering News* 2007, 2 Feb 2009 <http://pubs.acs.org/cen/priestley/recipients/1951crane.html>

[4] Shively, Eric, "CAS Surveys its First 100 Years," *Chemical and Engineering News* 2007, 2 Feb 2009 <http://pubs.acs.org/cen/coverstory/85/8524cover2.html>.

automation efforts. G. Malcolm Dyson had gained some recognition for devising a notation system that represented chemical structures uniquely and unambiguously in a linear sequence of letters and numbers. Dyson began his work at CAS by combining his ideas for a notation system with new database software developed by IBM.

From 1959 to 1960, CAS ran a pilot project to publish test issues of an electronic chemistry periodical. To support the work they obtained a grant of $112,000 from the National Science Foundation and leased an IBM 1401, one of the first electronic date processing systems. IBM's software included key-word-in-context (KWIC) indexing, a tool familiar to today's searchers although it was a revolutionary concept in the early 1960s. KWIC enabled CAS to produce complete indexes by computer. CAS used KWIC to create indexes by title, author and Journal Coden.

CAS's application for federal funding from the NSF was a departure from previous ACS practice. Until this point the Society had accepted no government funding. However, without outside financial assistance the expense of developing automated indexing systems would have been prohibitive for CAS. Over the next two decades the National Science Foundation contributed approximately $30 million to the creation of CAS systems.

CAS's first electronic products were distributed through third-party commercial vendors on magnetic tapes, which customers used to load the CAS database onto large mainframe computers. The first electronic publication was entitled *Chemical Titles* (CT). Because *CT* was not easy to search, searches were conducted by trained librarians and other information professionals. This became a new service area for science librarians, who scheduled and often mediated complex online searches in the CAS database. This role for science librarians has been largely phased out but lasted almost thirty years, until the technology became easy enough for less skilled researchers to navigate.

CAS continued to distribute its databases on magnetic tape until the 1980s when it began to experiment with its own online service. The new service, *CAS ONLINE*® was introduced in November 1980.[5] This new online database was accessible to users who owned an HP 2647A, a microcomputer introduced in the mid1970s.[6]

---

[5] Shively, Eric. "CAS Surveys Its First 100 Years" *Chemical and Engineering News* special issue: http://pubs.acs.org/cen/coverstory/85/8524cover2.html

[6] Brnyko, Barbara. "CAS turns 100," *Information Today* Feb 2007 <http://www.cas.org/ASSETS/72B436BFBE3542B1BFB6E5FA4545FA53/Information%20Today%20-%20CAS%20Turns%20100.pdf>

In 1983, CAS concluded an agreement with Fachinformationzentrum (FIZ), a German producer of scientific and technical information for chemists, to create a new on-line service called *STN*® *International.* *STN*® was introduced to the European community in May 1984. *STN*® was distributed around the world, though data was kept at its two nodes, Karlsruhe Germany and at CAS's Ohio facility.

As the Internet began to become the preferred channel for disseminating scientific information, CAS began to put more resources into online delivery and services. Databases such as *CAplusSM*®, a database of global chemistry literature from 1967 to the present, were introduced. *SciFinder*, an easy-to-use desktop search tool was launched in 1997 for use by corporate subscribers, and was later introduced for the academic community. Today it is the primary platform for CAS subscribers.

In 2005, CAS introduced handheld options for their databases. Today, Blackberrys and other hand-held devices can be set up to access information from the *CAplus*® and *CAS Registry*® databases.[7]

In 2006, ACS CEO Madeleine Jacobs reported that the Chemical Abstracts Service had achieved a record year in revenues and net contributions, and that it was experiencing increases in both volume and usage worldwide, with particularly strong growth in Asian markets.

3) CAS Governance

Until 1955 CAS was governed directly by the American Chemical Society Board of Directors. However, this changed in 1956 when CAS was made an operating division within ACS. The change in governance occurred because ACS wanted CAS to cover its own costs and be less of a financial burden to the Society.[8]

---

[7] "World-Leading Scientific Databases Now Accessible via Handhelds: CAS puts millions of chemical substance records and literature references in scientists' hands," *Access*, June 2005, 2 Feb 2009 http://www.aardvarknet.info/access/number53/monthnews.cfm?monthnews=03.

[8] Shively, Eric. "CAS Surveys Its First 100 Years" *Chemical and Engineering News* Special Issue, June 11, 2007: <http://pubs.acs.org/cen/coverstory/85/8524cover2.html>

Leadership for CAS was then vested in the newly created position of director, who reported to the ACS board. The director position remained until 2003, when the title of the incumbent, Robert Massie, was changed to President. The CAS director/president position reports directly to the ACS Executive Director & CEO. CAS has had five directors since its inception, with three serving significant amounts of time. Evan Crane served for 43 years, Dale B. Barker 30 years, and the current president, Robert Massie, to date had served 19 years. Continuity of its leadership may be among the reasons CAS has been so successful.

To ensure that CAS remained accountable to the ACS, a committee was created to serve as liaison between the ACS Board of Directors and CAS. This committee, originally called the Standing Committee on CAS, was formed in 1965. The Standing Committee was superseded in 1978 by the ACS Society Committee on CAS (CCAS) and has since been renamed the Joint Board-Council Committee on CAS (JBCCP).[9] The JBCCP reports to the ACS Governing Board for Publishing and the ACS Board of Directors. Today the JBCCP largely acts as a liaison between the membership and the Board on "affairs of Chemical Abstracts Service, including the reporting of each group's activities to the other."[10]

A financial crisis in the 1990's led to a second reorganization of CAS along more commercial lines. The ACS Governing Board for Publishing was created in 1996 and meets biannually. It includes four compensated publishing experts.[11] These experts have executive experience in both industry and academia and this arrangement provides ACS insight on the competitive and financial aspects of the publishing environment. In addition to the publishing experts, the chairman of the Society's Board of Directors, members of the ACS Executive Board and the CAS director are also members of the Governing Board of Publishing. The Governing Board holds general responsibility and authority for the operations and performance of Chemical Abstracts Service and the ACS Publications Division except for the appointment of editors and the editorial content of journals and magazines. The ACS Board also continues to be responsible for appointments, and for setting prices for CAS services and subscriptions.

In addition to the aforementioned governing bodies, CAS began to create user councils for North America, Europe, and Japan in 1983. The councils were originally formed in conjunction with

---

[9] Ibid.
[10] ACS. "Supplementary Information," *Charter, Constitution, Bylaws, and Regulations of the American Chemical Society as Revised through January 1, 2009* <http://portal.acs.org/portal/fileFetch/C/CTP_004181/pdf/CTP_004181.pdf>
[11] Marris, Emma. "Chemical Reaction," *Nature* 6 Oct 2005: 809.

CAS's introduction of online services, to provide a way of communicating with information users and obtaining feedback from same on CAS services. Similar functions had been fulfilled in the past by the CAS Advisory Board, established in 1964, and the CAS Editorial Advisory Board, created in 1975.

4) Content and Data Collections

CAS produces a variety of databases that are combined in different distribution models at different pricing levels. The databases contain chemical information, including abstracts of articles published in chemistry-related publications and patents, a registry of chemical substances, and indexes to a wide range of chemical literature. CAS also abstracts materials in non-chemical journals, because scientific findings in other fields, such as physics and biology, are often of interest to chemists.

At present CAS reviews and indexes more than 10,000 major scientific journals and patents from 57 patent authorities. At the end of 2007 over 26 million abstracts for books, journal articles and patents were available in CAS databases, and over 93 million substances had been indexed.[12]

The main CAS databases are: *Chemical Abstracts (CA)*, *CASRegistry*, *CASREACT*, *CHEMCATS*, *CHEMLIST*, *CIN*, *MARPAT*, and *TOXCENTER*. CAS products are more than simple journal databases or indexes, but provide added value through highly sophisticated systems of chemical nomenclature that aid in the organization, indexing and retrieval of chemical substance information.[13] CAS continually augments its products with new tools, such as *AnaVista* for modeling the data in *STN*. ® (A CAS database created in partnership with a number of European scientific organizations) and improving existing ones, such as adding the feature of being able to remove duplicates in the *SciFinder* interface.

Chemical Abstracts

Chemical Abstracts (CA) was CAS's original publication, begun in 1907, and its successor, the *CA* database, continues to be a key product. (A weekly edition of *Chemical Abstracts* is still

---

[12] CAS. *Statistical Summary 1907-2007*. Feb 2008. <http://www.cas.org/ASSETS/836E3804111B49BFA28B95BD1B40CD0F/casstats.pdf>

[13] Giles, P.M. and Metanomski, W.V., "The History of Chemical Substance Nomenclature at Chemical Abstracts Service in Organic Chemistry: Its Language and its State of the Art," *Helvetica Chimica Acta*, 1993: 173-196

distributed in print, CD-ROM, microform, and through a number of online database services.) Over 55% of the material abstracted in *CA* is from English language sources, with the rest in any of 66 languages. Materials abstracted include articles in print and electronic journals, patents, book chapters, dissertations, reviews, proceedings, and Web preprints. Other products have resulted from *CA*, among them indexes and a numbering system for new chemicals called CAS Registry numbers. CAS is no longer directly responsible for the print version of *CA*, which is a product of the Publications Division of the ACS.

CA was the first chemistry print journal to offer a number of finding tools for its users. These include semi-annual indexes, which allow users to search by controlled key words and subject terms, author names, synonyms for chemical names, and trade names. Substance indexes are by CA Index Name only, not synonyms or trade names. The latter must be translated via the Index Guide thesaurus. Abstract numbers include the CA volume number; not issue or page numbers.

The print version of *CA* is still used by some institutions and businesses, generally because it is less expensive than the electronic version. Because the indexing of *CA* began with the print, the electronic database still reflects some of these earlier indexing choices. For example, chemical names in *SciFinder Scholar* are still governed by the nomenclature rules developed for *CA* in print. Abstracts continue to be assigned numbers, which include issue and page numbers based on the *CA* print locations.

CAS Content

For *CA* and its other products, CAS creates and indexes chemical information in a variety of ways. Listed below are some of the most common types of data that CAS offers. This list is not exhaustive, but is meant to illustrate how CAS serves data to its users and why these data resources are useful for chemical researchers.

Abstracts

Abstracting was the original service CAS provided through its *CA* print publication. The database evolved from the print edition of *CA*. Currently over 10,000 journals are abstracted for *Chemical Abstracts*. CAS chooses a broad range of journals relevant to chemistry and has always included peripheral journals which are normally considered to be outside the bounds of chemical literature but are relevant for chemistry research.

Through the 1960's CAS relied primarily on volunteers to monitor scientific journals and produce abstracts, while indexing was always the responsibility of paid CAS staff. By the 1960s the number of CAS volunteers had increased to 3,300 people in 55 countries.[14] During the late 1960s CAS began hiring full time chemists to write abstracts in the service's Columbus, Ohio offices. There were several reasons for this change: volunteers were unable to keep up with the expanding chemical literature; there was a great deal of complexity involved in coordinating all the administrative aspects of the volunteer abstractors; computer processing was making it possible to create a faster, more efficient workflow; and the CAS financial situation had improved.[15] Although CAS continued to obtain some abstracts from British, German and Japanese chemical societies, these arrangements had all ended by the early 1980's.

In 2003 CAS completed digitization of its print materials back to the original 1907 *CA* issues. Prior to this, only material back to 1967 had been available online. Chemistry is a field in which most information does not become obsolete and so CAS has made the effort to digitize older material, providing full-text online to all of its legacy chemical information.

Indexes

Indexes provide tools for research which give citations to sources of chemical information in print and electronic journals through pointers such as author names, subject terms, chemical formulas, chemical names, and cited references. Some CAS indexes provide hyperlinks to e-journal articles and other sources cited, through which subscribers and others authenticated for those sources can access them directly.

Patent Information

Patent registrations are an important form of publication for industrial research. CAS indexes new patents in certain classes that are applicable to chemistry. CAS is not the only organization providing access to patents. Google, the U.S. Patent and Trademark Office, the patent offices of other countries, and other commercial database firms also provide access. However CAS provides more than merely the patent itself: CAS indexers will often expand a patent's abstract and add CAS Registry numbers for specific chemical compounds, all relevant to researchers.

---

[14] Abelson, Philip H, "Chemical Abstracts after 75 years," *Science* July 2, 1982: 7.
[15] Baker, Dale B., Jean W. Horiszny, Wladyslaw V. Metanomski, "History of Abstracting at Chemical Abstracts Service." *Journal of Chemical Information and Computer Science* 1980: 199.

CAS Databases

Besides compiling and indexing chemical information, CAS organizes its content into a number of different databases, customizing the content for different subscribers. Below are brief descriptions of seven CAS databases. Some of these databases are available through several data distribution interfaces, including CAS's *Scifinder, STN* and several other services.

CAplus

CAplus covers journal articles and patent documents in many scientific disciplines from 1907 to the present, plus more than 133,000 pre-1907 journal records. English language summaries are produced from scientific literature published in more than 50 different languages from over 185 different countries. Publications abstracted in *CAplus* include conference proceedings, technical reports, books, dissertations, reviews, meeting abstracts, electronic-only journals, and Web preprints

CASREACT®

CASREACT is a database of information about chemical reactions. The database contains information about more than 15 million single- and multi-step reactions from journal articles and patents.

CAS Registry

The *CAS Registry* was introduced in 1965. CAS registry numbers (referred to as RN's) are assigned to *CA* index names, commonly used synonyms, polymer class terms, structure diagrams and molecular formulas. These numbers are then used by CAS staff to index all chemical substances mentioned in an abstracted publication. RN's are a unique, unambiguous, machine-language representation of a two-dimensional structure of a chemical compound, together with a method for recording additional data such as stereochemistry.

In addition to enhancing search capabilities, the *CAS Registry* system gives chemists a tool for quick and reliable "identity checks" of chemicals which often have lengthy, incomprehensible names. A researcher can build their own database for a particular chemical compound by searching the literature on a particular CAS Registry number. In addition, users can also search a three-dimensional description of a molecular structure.[16].

---

[16] Schulz, Hedda. *From CA to CAS Online: Databases in Chemistry.* 2nd ed. (Berlin: Springer, 1994)

CAS has begun to make the scope of the *CASRegistry* even broader by drawing upon sources of substance information beyond the traditional patents and journal literature, including new web-based resources of value. CAS continues to improve the reach and timeliness of its patent information, with special focus on China and other Asian sources of growing technological importance.[17]

Chemcats®

Chemcats is a CAS database of information about commercially available chemicals and their worldwide suppliers. Currently, there are 24,033,991 commercially available chemicals listed in the database.

Chemical Industry Notes (CIN)

CIN is an online version of a print publication that covers events and developments in the worldwide chemical industry. It indexes approximately 80 business-oriented journals, newspapers, and related periodicals that reflect recent events in the chemical industry. The records contain bibliographic information, indexing, abstracts, chemical names, and CAS Registry Numbers®. There is an online thesaurus for the Geographic Term (/GT) field.

Chemlist®

Chemlist is a database of chemical substances that are regulated in key markets across the globe. Currently, the database contains approximately 247,332 inventoried/regulated substances

Marpat®

A Markush patent claim, (named for the inventor who first claimed a generic structure), is a patent claim made for a family of chemical compounds. A chemical family is a group of elements or, more commonly, compounds that share certain chemical characteristics and have a common name. This relationship is important for chemists because a family of chemical characteristics can be patented even if the inventor has not tested or even prepared all members of the family. *Marpat®* allows the chemistry researcher to determine whether an unidentified chemical on which she is working has been patented through its familial affiliations, something that would otherwise be exceedingly time-consuming.

---

[17] "In profile: Chemical Abstracts Service." *scientific-computing.com* 6 Jan 2009, 2 Feb 2009<http://www.scientific-computing.com/features/feature.php?feature_id=188>

Searching the *Marpat* database gives patent searchers a way to discover whether a molecule is original or has been discovered before filing a patent claim. CAS provides 780,328 searchable Markush structures in the *Marpat* database.

TOXCENTER

TOXCENTER replaced an earlier database *TOXLINE,* and combines information from several biological information services, including *BIOSIS*, the American Society of Health-System Pharmacists (ASHP), the National Library of Medicine (NLM), and CAS itself. The database was created to replace the National Library of Medicine's *TOXLINE* database already on *STN*®. According to CAS, NLM stopped delivering new content to *TOXLINE*, and then ordered the removal of the file altogether by July 2008. NLM continues to maintain a newly designed version of *TOXLINE*, available free on the *TOXNET* service.

5) Services

Because of the level of sophistication of their products CAS provides a number of services that support chemical research and development. Database training is a major feature of CAS customer service and part of their marketing appeal. CAS offers training databases, videos and other materials for users on their website. They provide a number of learning databases that contain subsets of the information available in larger CAS databases. Many of the learning databases provide users the opportunity to practice searching CAS information without incurring high fees.

Through its *Document Detective Service* CAS provides copies of documents abstracted or cited in *CA* or contained in publications and U.S. Patents monitored by CAS, from 1971 to the present. When copyright restrictions apply, CAS will lend a copy of the original instead, or will forward requests for materials not owned by CAS to the British Library Document Supply Centre.

Science IP is a search service that researches and identifies matching or related chemical compounds for a given substance, and information on potential restrictions on same. The service performs sophisticated structure and substructure searching of chemical files. The service is designed for the intellectual property community, i.e., patent searchers, attorneys, and research and development professionals.

Other CAS contract services including inventory control, substance identification and Registry number assignment.

6) *Interfaces and Tools*

In the 1970s, when CAS began to distribute its databases on magnetic tape, computerized searching became an option for CAS information and content. At first, the databases were searched mainly by librarians and information specialists who had learned the command language and had mastered the requisite information navigation skills. These intermediaries were necessary until the mid-1990's when better software and desktop computers and networks became more prevalent. Described below are a few of the more familiar distribution and discovery platforms for CAS content and databases.

SciFinder and *SciFinder Scholar*

SciFinder was introduced as a locally-installed client for professional and industrial chemists in 1995. *SciFinder* was tailored to two user groups, professional and industrial chemists. It facilitated searching of CAS databases without the assistance of a librarian or specialist. Originally the platform was installed on a user's desktop but today *SciFinder* is available through the Web. *SciFinder Scholar,* introduced to the academic community in 1999, was similar to *SciFinder* except that it offered multiple users simultaneous access.

The Web version of *SciFinder* was introduced in 2008. It offers access to content based on a user's subscription type. CAS discontinued the client version of *SciFinder* in early 2009, leaving the Web as the sole distribution platform. The Web version of *SciFinder* searches using text terms and offers tools to refine and analyze search results.

One of CAS's successful practices has been devising tools to help researchers better mine the information available in its databases. Some of the latest tools for *SciFinder* include *SubScape*. Released in October 2007, *SubScape* is a substance visualization tool that permits scientists to visualize substance answer sets in a previous version of *SciFinder*. Features include a cluster map that represents base frameworks using a clustering algorithm and structure descriptors from the *CAS Registry*. In addition, a Bioactivity Indicators Chart presents bioactivity terms for which relationships exist between *CAS Registry* substances and documents represented in *CAplus.*

Scientific and Technical Information Network (STN®)

In 1983 the *STN*® database was introduced in Europe. *STN*® was the first platform that provided online access to published research, journal literature, patents, structures, sequences, properties, and other data from a variety of producers. Adding to the global reach of CAS, the Japan Association for International Chemical Information (JAICI) became the Provisional STN® Tokyo Service Center. The STN® cooperative information conglomerate today provides access to more than 220 databases.

STN® was the first to offer researchers the ability to draw chemical structures and use these drawings to search in published literature and patents. This type of search reduces the possibility of missing a chemical structure due to variations in nomenclature. *STN*® provides structure-based searching of a number of chemistry databases including *CAS Registry*, *CAS MARPAT*, and non-CAS products such as *Beilstein*, and *Derwent World Patent Index*.

STN® *AnaVist*™ is an analysis and visualization tool for patents. In 2007, *STN*® *AnaVist*™ 2.0 was expanded to include searching a non-CAS database, *Derwent World Patents Index* (DWPI℠), along with the extensive patent content of *CAplus*.

In 2007, CAS and FIZ Karlsruhe launched *STN*® *Viewer*™, a web-based tool that helps lawyers, and other patent professionals evaluate patents by allowing the user to view major sections of the patent, its publication stages and its chemical family members.

To support data mining, CAS permits the extraction of metadata for use within CAS databases to *STN*® tools without first displaying the data.

DIALOG®

Cengage's *Dialog* database platform provides access to some CAS information in *CA Search* file numbers 204, 308-314, and 399. File 204 is a training database, while files 308-314 each contain a subset (by years) of all *CA* information. File 399 holds all of the CA information in one file. Subdividing the files gives searchers the ability to save time by reducing the amount of data searched. CAS information in *Dialog* is updated weekly. Although *Dialog* does not offer direct access to CAS Abstracts, searches conducted via *Dialog Web Guided Search* give users the option of ordering abstracts from CAS®, via remote links, integrated in the *Dialog*® search results. *Dialog*® does supply the descriptors, identifiers, and Registry numbers. *Dialog* CA files are made available on a pay-per-use basis.

OCLC FirstSearch

OCLC's FirstSearch interface has provided access to *Chemical Abstracts Student Edition*, a subset of the *Chemical Abstracts* database, since 2000. *Chemical Abstracts Student Edition* serves the needs of smaller academic institutions by providing access to CAS information for undergraduate students. In 2008 *Chemical Abstracts Student Edition* contained abstracts for approximately 316 journals and 300,000 dissertations.

The OCLC *Chemical Abstracts Student Edition* does not index books, book chapters (except for some annual review-type series), conference proceedings, patents or foreign language journals. Specific substances are indexed in CA Student edition only by *CA Index* name and *CAS Registry* number. *Chemical Abstracts Student Edition* includes records dating to 1967 for selected publications. No full text is available. Currently the database offers over 4,500,000 records. The journals covered in *Chemical Abstracts Student Edition* are selected on the basis of wide holdings in academic libraries in the U.S.

7) Technical Analysis

CAS maintains a large-scale data repository. As early as the 1970s some CAS databases had more than a million records and required storage space for hundreds of millions of bytes.[18] Today CAS databases include index information on more than 30 million documents, 42 million substances, 60 million sequences, 16 million reactions, and 26 million chemicals. As many as 3,000 references are added daily. As noted in a CAS press release, it took 30 years for CAS to publish its first million abstracts, but by 2006, it had indexed over a million records in a single year.[19] In 2007, CAS added an average of 20,924 records per week to its databases.[20]

Scale is not the only preservation challenge faced by CAS's data repository; scope also presents significant preservation challenges. Scientific data sets for chemical structures and reactions present unique challenges without a clear, established tradition. Chemical properties and the structures of molecules create a complex problem for information representation and therefore for

[18] Anzelmo, F.D., "A Data-Storage Format for Information System Files," *IEEE Transactions on Computers* 1971: 39-43.
[19] Notess, Greg R., "Chemical Abstracts Service (CAS) announced patent information improvements in CAS databases available through STN, SciFinder, and SciFinder Scholar," *Online* 2007
[20] Chemical Abstracts Service, "CAS Sets Record for New Scientific Data Recorded in one Week, Additional Database Expansions Announced. Press Release," *PR Newswire* 22 Aug 2007, 2 Feb 2009 <http://www.iptoday.com/news-archived-article.asp?id=942&type=business>

information preservation. These are not new problems; chemistry has a long history of coping with the difficulties and complexities of representing chemical information in an unambiguous way.[21,] Over the years chemists and computer scientists have tried a variety of methods. They have used names, two dimensional drawings, "connection tables," linear code, proprietary machine representations, ciphers, notation systems, and "fragmentation codes" to record this information.

Storing the information so that scientists can search for and retrieve and analyze the information with computers is one challenge. Being able to represent the information so that humans can visualize and understand the information is a different problem. Being able to do both, or convert from one form of representation to another makes the task particularly complex.[22,23,24] There are more than 40 different file formats that have been designed by academics and commercial software vendors to electronically store chemical structures and chemical reactions.[25] While there are many "standards," there is no one, universally accepted standard.

The preservation challenges CAS faces are three: the quantity of data to be preserved, the diverse nature of the data, and the complexity of much of the data. An examination of the published record of CAS's activities over the years, suggests how other data repositories might meet similar challenges.

Historical record of technological innovation

One indicator of the quality and reliability of CAS's preservation techniques is the existing historical record of its successes in maintaining and improving its collections for more than a hundred years. There are articles in the published literature that address aspects of CAS's technology, although there is less detail in articles published during the last thirty years than in

---

[21] Cf. Engel, Thomas and Johann Gasteiger, "Chemical structure representation for information exchange," *Online Information Review* 2002: 139-145; and Cooke, Helen, "A historical study of structures for communication of organic chemistry information prior to 1950," *Organic Biomolecular Chemistry* 2004: 3179–3191

[22] Engel, Thomas and Johann Gasteiger, "Chemical structure representation for information exchange," *Online Information Review* 2002: 139-145.

[23] Cooke, Helen, "A historical study of structures for communication of organic chemistry information prior to 1950," *Organic Biomolecular Chemistry* 2004: 3179–3191

[24] Powell, Evelyn Constance, "A History of Chemical Abstracts Service, 1907-1998," *Science and Technology Libraries* March 2000: 93-110.

[25] *MN.convert, Inter-conversion of Chemical Structure and Reaction File Formats, User Manual Version 1.0. ().* (Erlangen, Germany :Molecular Networks GmbH, Nov 2003) 2 Feb 2009 < http://www.molecular-networks.com/software/convert/mn_convert_manual.pdf>

earlier articles. In addition, on the occasion of CAS's one hundred year anniversary in 2007, there were several articles that outline, in a general way, its progress over a century.

We know, for example, that data input has evolved significantly, particularly over the last forty years. In the 1940s and 1950s CAS staff edited and indexed abstracts on index cards manually[26] and chemical structures were drawn by hand. This changed in the mid-1960s, when data began to be entered on punch cards, and then evolved later using chemical typewriters to create paper tape and subsequently magnetic tape. By the 1970s chemical structures were input directly to computer disk and the process was partially automated. By 1979, CAS was using "a computerized system of direct graphical input of chemical structures."[27] By 1993 CAS was using a new system that algorithmically named unique chemical substances (the "Algorithmic Name Generation System," or ANGS).[28]

Today we know little about how CAS stores data after input, but we do know about some of the work CAS has done to standardize the representation of chemical data. Certainly one of the most visible CAS innovations in this area was the creation, in 1965, of the Chemical Registry system. It was innovative in several ways. First, it created a way of accurately and unambiguously referring to each chemical compound. The *CAS Registry* number is, for a compound, somewhat like a Social Security number is for a person: a standardized, unique identifying number. This identification number is easy to store and easy for both computers and humans to use, recognize, and understand. This allows CAS to accurately index and cross reference chemicals across databases and even across different kinds of databases (e.g., bibliographic and chemical structure databases).

Second, it provides an easy and accurate way for chemists to refer to chemicals. As noted by a CAS employee, "the Registry System now serves not only as a support system for identifying substances within CAS operations, but also as an international resource for chemical substance identification for scientists, industry, and regulatory bodies. It has become the worldwide

---

[26] Brynko, Barbara, "CAS turns 100," *Information Today* 1 Feb 2007, 2 Feb 2009 <http://cas.org/ASSETS/72B436BFBE3542B1BFB6E5FA4545FA53/Information%20Today%20-%20CAS%20Turns%20100.pdf>
[27] Powell, Evelyn Constance, "A History of Chemical Abstracts Service, 1907-1998," *Science and Technology Libraries* March 2000: 93-110.
[28] Lohr, James E., "Building the CAS databases," *Database* 1998. : 52-54, 2 Feb 2009 <http://portal.acm.org/citation.cfm?id=297376>

authority for chemical substance identification information."[29] Reflecting CAS's ability to be flexible and evolve over time, the *Registry* itself has gone through at least three versions, *Registry I* (1965), *Registry II* (1968), and *Registry III* (1973).[30]

We also know a little bit about what CAS was doing in the 1960s to store data. At the same time CAS was developing the *Chemical Registry* identification numbers, they were also developing a standardized storage format that was "designed to be independent of storage medium (e.g., tape, disk, drum)." CAS's newly designed system could accommodate both fixed-length and variable-length data elements and would permit the addition of new data elements and the deletion of existing data elements with no effect on existing program code. This innovation made the preservation of the content somewhat independent of the program code that accesses the data. This can be seen as a key strategy that facilitates preservation of content without compromising functionality.

CAS also documented its procedures and definitions of data elements "rigorously" in hard copy form. Initially, the new system was designed to work in the specific operating environment of the time: the IBM System/360 using OS/360 and the OS/360 data management facilities. While conforming to the 360 system imposed some restrictions on what could be done, CAS made choices that would not limit it to the IBM environment. Specifically CAS chose the eight-bit byte as the basic unit of information and implemented the "ability to transmit binary data to or from storage devices so it would not be limited solely to System/360."[31]

In a 1971 article, F.D Anzelmo, a member of the Systems Development Department of CAS, noted the benefits of these technological choices to the functioning of the systems and to their flexibility. This flexibility is also key to preservation because it helps ensure that the content is separated from functionality. These choices allowed the data to be independent, and so better preserved and available to use with new tools and software functions over time. Although we do not have such detailed information about later CAS decisions and infrastructure, one hopes that they are similar in effect. It is worth noting in detail the benefits that Anzelmo described:

---

[29] Weisgerber, David W., "Chemical Abstracts Service Chemical Registry System: history, scope, and impacts," *Journal of the American Society for Information Science*. 1997: 349-360. 2 Feb 2009 <http://dx.doi.org/10.1002/(SICI)1097-4571(199704)48:4<349::AID-ASI8>3.0.CO;2-W>
[30] Powell, Evelyn Constance, "A History of Chemical Abstracts Service, 1907-1998," *Science and Technology Libraries* March 2000: 93-110.
[31] Anzelmo, F.D., "A Data-Storage Format for Information System Files," *IEEE Transactions on Computers* 1971: 39-43.

> With the utilization of SFF [Standard File Format], the CAS computer-based system operates with greater generality and flexibility. Several programs written for a particular application have been used in other applications with only slight modifications. The increased generality is due not only to the standardization of the file, but also to the use of table-driven programs to process the SFF files. Such programs have been written for editing, validation, creation of data files from raw input, file maintenance, and data selection based on the logical combination of specified data element values. Utility programs, such as those for file print or file creation, are also table-driven. Data elements which are to be transferred through the various application files rather than processed by application programs can be identified using table-driven programs. This capability is highly desirable at CAS since a number of organizations wish to process and coordinate their data with the information in the CAS system and to incorporate the resultant data within their own internal filing systems.
>
> The insensitivity of the programs to the addition of data elements allows integration of data files across several application areas. The standardized files become, in effect, an information reservoir from which programs extract specific information of interest to them and into which they deposit labeled information. The standardized format can also provide a stable interface for data exchange with users of the CAS information services and files. Users may write their own software knowing that CAS will change its processes and products only within the bounds of the standard.[32]

The "information reservoir" that Anzelmo describes is, in effect, a preservation format for content.

At the time CAS was developing these systems, they were still processing data on sequential-access computer tape, but they were already making plans for implementing the system on random-access computer disks. Such planning has key benefits to preservation as it helps ensure that the design does not lock the content into a particular computing environment and anticipates the next generation of computing environment.

Another salient aspect of CAS is the interfaces it provides to users. CAS has continually and repeatedly enhanced its user interfaces, adding new features and functionality that have garnered praise from users. One science librarian has said that the *SciFinder Scholar* interface to CAS "continues to be among the best database interfaces available to scientists today."[33] CAS pioneered the use of electronic databases as a publishing tool in the 1960's, enabling it to be both

[32] Anzelmo, F.D., "A Data-Storage Format for Information System Files," *IEEE Transactions on Computers*, 1971: 39-43.

[33] Wagner, A. Ben, "SciFinder Scholar 2006: An Empirical Analysis of Research Topic Query Processing," *Journal of Chemical Information Modeling* 2006: 767-774.

more efficient and more accurate.[34,35] In 1980 CAS launched *CAS ONLINE*, making it possible for users to search the *CAS Registry* database. In addition to these key advances in search capabilities for users, CAS has continually enhanced both the content that is searchable and the user interface used to search its databases. There were five versions of *SciFinder* released between 1994 and 1999, each adding new content and functionality. *SciFinder Scholar* was introduced in 1997 and new versions have been introduced almost every year since then.[36]

These changes, both to the underlying infrastructure that ingests and delivers the very large amount of data that CAS manages and to the user interface, indicate how CAS has successfully managed continuing, rapid technological innovation over a period of several decades. It is reasonable to assume that CAS has paid equal attention to the preservation of its underlying data through continuing, innovative changes in its preservation technologies. Unfortunately, we have no way of knowing what those changes have been and therefore can draw no more specific lessons from CAS successes

8) Users, Funders and Other Stakeholders

The primary stakeholders in CAS are chemists, ACS members, librarians, educators, ACS employees, and students. The relationships between these different stakeholder groups and CAS, and the benefits they derive from CAS services, vary significantly.

American Chemical Society members

The ACS has four categories of membership plus a Student Affiliate category.

> Regular Members are entitled to member-only subscription rates for ACS journals and a subscription (print and online) to *Chemical & Engineering News*.
> Individuals must have one of the following:
>
> - a degree in a chemical science

[34] Brynko, Barbara, "CAS turns 100," *Information Today* 1 Feb 2007, 2 Feb 2009 <http://cas.org/ASSETS/72B436BFBE3542B1BFB6E5FA4545FA53/Information%20Today%20-%20CAS%20Turns%20100.pdf>

[35] Powell, Evelyn Constance, “A History of Chemical Abstracts Service, 1907-1998,” *Science and Technology Libraries* March 2000: 93-110.

[36] Wagner, A. Ben, “SciFinder Scholar 2006: An Empirical Analysis of Research Topic Query Processing,” *Journal of Chemical Information Modeling* 2006: 767-774

- a certification as a teacher of a chemical science or relevant work experience

Associate Members are those who have not yet met requirements for regular Society membership. They cannot hold an elective office but are otherwise entitled to all benefits.

Graduate Student Members must be full-time graduate students majoring in a chemical science or a related academic discipline. They qualify for reduced subscription prices for most ACS journals.

Non-Scientist/Society Affiliate Members are individuals who are not eligible to become a regular or student member of the Society but whose major vocational efforts are directly concerned with the practice of a chemical science. Society affiliates are entitled to all regular member benefits excluding the ability to vote or hold an elective office.

Student Affiliates are undergraduate students majoring in or interested in chemistry, who qualify for membership at a substantially discounted rate and receive special student subscription prices for journals. Many students join their university's ACS chapter and remain active throughout their professional career.

CS membership begins for many chemists when they join their school's ACS student chapter in their undergraduate years. The ACS provides CAS a number of opportunities to interface with its members. The Society's annual meeting is a useful marketplace of users, and an opportunity to offer training on new and existing CAS products. Other vendors of chemical information attend the ACS meetings to market their products as well.

Chemists

In the United States approximately 84,000 men and women were employed as chemists in 2006. A little less than half of those jobs were in chemical manufacturing. Other areas where many chemical scientists find employment are in government agencies and the education sector. According to the 2008-2009 *Occupational Outlook Handbook*, employment growth for chemists over the next five years was expected to be around nine percent, on average. Much of this job growth was expected to occur in professional, scientific, and technical services firms, as manufacturing companies continue to outsource their R&D and testing.[37]

Outsourcing is a trend in the chemical and allied products industries. According to 1997 National Science Foundation data, companies in this sector outsourced about 22% of their R&D work. Work was outsourced to companies in the United States as well as in emerging economies.[38] This trend may have an effect upon future employment opportunities for U.S. chemists.

Librarians

Chemistry librarians and information specialists employed in universities and chemical corporations make up an important market for ACS products including CAS databases. The ACS has created a division to engage librarians, the Division of Chemical Information. This division fosters the sharing of expertise on science informatics, information technology, and librarianship to enable members to benefit from the experience of others and to improve the dissemination and use of scientific information. The Division offers programs, education, career mentoring, collaboration and outreach, and awards.

From the 1970s through the 1990s chemistry librarians provided the knowledge and skills needed to help researchers search the CAS databases successfully. Today mediation by librarians is no longer necessary because many CAS platforms are searchable without specialized training. However, librarians continue to provide early instruction for students in chemistry and related studies, and are often the decision makers as to what ACS products will be purchased for use in university and corporate libraries.

Many university library systems are phasing out chemistry libraries. In part this is an economy measure, although a trend in education toward more interdisciplinary programs in which chemistry is only one component is another major factor.[39] This trend is not likely to affect CAS sales, although as more databases are paid for from the same budget, without the advocacy of chemistry librarians the costs of CAS products may begin to seem inordinately expensive to

---

[37] *Occupational Outlook Handbook, 2008-09 Edition, Chemists and Materials Scientists*, (Washington, DC: Bureau of Labor Statistics, 2008) 2 Feb 2009 <http://www.bls.gov/oco/ocos049.htm.>

[38] Borchardt, John K., “Playing the economics game with outsourcing,” *Modern Drug Discovery* Mar 2000: 28-29, 31-32, 34.

[39] Garritano, Jeremy R., "Current and Future Status of Chemistry Collections and Chemistry Libraries at ARL Institutions," *Issues in Science and Technology Librarianship* Spring 2007. 4 Feb 2009 <http://www.istl.org/07-spring/refereed1.html>

university libraries. Conversely, if the population of chemistry librarians dwindles collective bargaining with ACS on price and features, as occurred in 1997 after the introduction of *SciFinder Scholar,* may no longer be possible.[40]

Educators

CAS's other educational stakeholders are college and university faculty teaching chemistry and chemistry related subjects. For these stakeholders CAS provides sources of information and support for their research, and critical visibility for the products of that research. There are currently 691 accredited U.S. universities with ACS-certified chemistry degrees. For Chemistry, BS, MA and PhD's degrees are offered. There are 646 schools with Bachelors programs, 309 with Master's programs and 200 with Ph.D. programs. Accroding to the American Chemical Society, 12,120 Bachelor's degrees and 2,321 PhDs were awarded in Chemistry in 2006. [41]

Students

Chemistry students need to have a demonstrable understanding of the general content and organization of *Chemical Abstracts* (among other chemistry publications) in order to become fully functioning chemists. To assist students, CAS provides learning databases in *STN*® and *Dialog.* Most chemistry students graduate from U.S. universities with an ACS-certified chemistry degree.

CAS Funders

CAS has been required by the ACS to be self-supporting since 1956. While its major source of revenue is the subscriber base for its databases, between 1960 and 1979 the National Science Foundation contributed approximately $30 million toward creation of the CAS database systems. The NSF Division of Chemistry has among its goals promoting the health of academic chemistry and enabling basic research and education in the chemical sciences. While this goal is consistent with CAS's mission and the NSF has funded ACS projects in the past, they have not directly funded a CAS-related project since 1979.[42]

---

[40] Flaxbart, David, "The Chemical Abstracts Centennial: Whither CAS?" *Issues in Science and Technology Librarianship* Winter 2007, 2 Feb 2009 <http://www.istl.org/07-winter/viewpoints.html>.
[41] Jacobs, Madeleine. "Report of the ACS Executive Director to the ACS Council," ACS Comment. Apr 2008 <http://pubs.acs.org/cen/acsnews/report.html>
[42] "Database Search" *National Science Foundation Awards Database (*Washington, DC: NSF) 1 Jan 2009 <http://www.nsf.gov/awardsearch/>

Universities

Universities are stakeholders in CAS as they purchase access to the data for their students, professors and researchers. In addition, universities are also heavily represented in many ACS management activities, as university administrators, faculty, and research chemists serve on ACS boards and committees, indirectly contributing to CAS management and decision-making. Another facet of the CAS-academia relationship is university reliance upon the ACS to provide ranking and accreditation of their chemistry programs. It is likely that some of these relationships help CAS sell their products more effectively than they might if they were simply a commercial vendor.

Partner Organizations

The American Chemical Society has been a persistent presence in the international community. CAS has long abstracted chemical journals from around the world, and today creates abstracts of materials in 66 languages. In fact, the Society was able to sustain its relationship with German scientists throughout World War II, continuing to obtain and publish their abstracts. The Society had a similar arrangement with Soviet scientists during the Cold War.[43]

During the 1960's, CAS began to establish formal relationships with a number of European organizations. The goals of this effort were:

> (a) decentralization of output operations, by offering selected machine-readable files of data to any agency, private or public, that may develop the capacity to search, select and package information from the files for its own purpose, with initial assistance from CAS;
>
> (b) decentralization of input operations by obtaining in return from national groups that they process their own literature according to norms compatible with the American system, and by developing more direct links with primary publishers and secondary services.[44]

In 1969 an agreement between CAS and the Council of the Chemical Societies (acting on behalf of a consortium of ten scientific and professional societies in the United Kingdom) forged the first formal link in an international network for handling secondary chemical information. In 1969 CAS also signed an agreement with Gesellschaft Deutscher Chemiker (German Chemical Association) which confirmed their joint cooperation, and a third agreement, similar in purpose,

---

[43] "In Memoriam Dale B. Baker 1920-2005," *Chemical Information Bulletin* Spring 2006: 58, 2 Feb 2009 < http://acscinf.org/docs/obit/BakerDale.htm>

[44] United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Study Report on the feasibility of a World Science Information System (Paris: UNESCO, 1971): 52.

was signed with the Scandinavian countries (Denmark, Sweden, Finland, and Norway).[45] The outcome of these agreements was that the German and Scandinavian groups would send computer-readable information to a central processing system in Columbus, OH, and share the output-both in the form of printed text and machine-searchable tapes.[46]

Although relationships existed between CAS and the chemical societies of the United Kingdom earlier, the 1969 formal agreement with the Council of the Chemical Societies resulted in creation of the Consortium on Chemical Information (COCI). The consortium included the ACS, British Chemical Society, Society of Chemical Industry, Society for Analytical Chemistry, Biochemical Society, Royal Society of Chemistry, Association for Information Management (ASLIB), and Chemical Industries Association. Under the agreement COCI would provide computer search services for CAS databases, undertake research and development on issues related to the distribution of chemical information, distribute CAS print publications, and prepare content for CAS publications.[47]

In 1980 the ACS decided to end the agreement. Why this occurred is explained by David Whiffen in his book, *the Royal Society of Chemistry: the First 150 Years*:

> Increased efficiency of scale in the CAS was not easily matched in the smaller output [of the RSC]. In the end the rate per abstract available from CAS did not match commercial viability in Britain, not least with uncertainty in the relative inflation in the two countries and uncertainty in the future dollar/pound exchange rate[48].

Today, the two organizations continue to sell one another's publications. The Royal Society is a distributor of *STN* in the UK and Ireland, while the ACS sells RSC print and electronic publications in the US.

FIZ Karlsruhe is an international organization based in Germany that provides scientific information publishing services to academic and industrial scientists. They provide electronic and print information and produce and provide access to databases that cover various fields of science such as agriculture, chemistry, engineering, and technology. Access to FIZ Karlsruhe

---

[45] United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. *Study Report on the feasibility of a World Science Information System* (Paris: UNESCO, 1971): 52.
[46] Kent, Allen, Harold Lancour, William Z, *Encyclopedia of Library and Information Science* (Boca Raton, FL: CRC Press, 1970): 498.
[47] Whiffen, David. *The Royal Society of Chemistry: the First 150 Years.* (London: The Royal Society of Chemistry,1991): 170
[48] Ibid. p. 175

databases is available for the most part through *STN*® *International*, a Web platform developed and maintained by the German organization and CAS. Many database publishers provide content for inclusion in *STN* and users subscribe to each database separately. The service is provided jointly by service centers in Germany (operated by FIZ Karlsruhe), in the USA (by CAS) and in Japan (by the Japan Science and Technology Agency). In 2007 CAS paid FIZ $2,941,147 for abstracting and sales services.[49]

In 1968, the Japan Association for International Chemical Information (JAICI) began working with CAS to provide Japanese chemical information to CAS. In 1977, JAICI entered into a cooperative agreement with CAS to distribute CAS products. In 1983 an agreement was signed between CAS and JAICI to participate in the development of *STN*. ® Under that agreement JAICI is responsible for marketing and user support of CAS services and does document analysis of Japanese patents and journal articles that are added to CAS databases. In 2006, CAS paid JAICI $6,935,403 for its services.[50]

9) Funding Model and Business Activity

Until 1933 ACS membership dues financed CAS. As a benefit of membership ACS members received a copy of the publication *CA*. A separate $6.00 subscription fee for members was added at a later date to help with publication costs. However, in 1955, with its expenses exceeding $1 million for the first time, CA faced a deficit of almost $500,000. The ACS Board of Directors addressed the crisis beginning in 1956 by establishing a new policy, requiring CAS to set its prices as necessary to "break even," or cover all of its costs. This new policy also led to the establishment of CAS as a separate business division of the ACS.

Today, the CAS funding model is very different. In a 2005 article in the journal *Nature* Emma Marris stated that CAS revenue then supported much of the ACS activities. In 2004 the CAS and ACS print publications divisions made about $340 million — 82% of the Society's total revenue — while accounting for only $300 million (74%) of its expenditures.[51]

---

[49] "American Chemical Society IRS 990 Tax Form" guidestar.org. 2006, 2 Feb 2009 <http://www.guidestar.org/FinDocuments/2006/530/196/2006-530196572-03c16aac-9.pdf>
[50] "American Chemical Society IRS 990 Tax Form," guidestar.org 2006, 2 Feb 2009 <http://www.guidestar.org/FinDocuments/2006/530/196/2006-530196572-03c16aac-9.pdf>
[51] Marris, Emma, "American Chemical Society," Nature 6 Oct 2005 : 809.

CAS revenues also illustrate how the transition from print to electronic distribution of chemical information has changed the business of chemical publishing. In 1975, 95% of ACS's revenue came from print services, and only 5% from electronic sources. In 2006, the situation had reversed, with 79% of ACS's revenue stream derived from CAS electronic databases and only 5% from print and other services. It is also important to note that 50% of CAS's revenue comes from outside the U.S. with developing countries like China and India bringing in a significant new subscriber base.[52]

The ACS reported 2006 combined earnings from CAS and the ACS Publications Divisions to be $361,710,666. With most of this revenue derived from CAS electronic databases, a quick comparison with the total 2006 total ACS revenue of $515,903,418, shows that CAS revenue contributes a great deal to the financial stability of the ACS.

While CAS continues to cover its costs, new issues that involve the service's relationship with the ACS have arisen. Dennis Chamot, the Director-at-Large of the ACS's Committee on Budget and Finance stated in an April 2008 article that, "the society ran deficits—more money went out than came in—in three of the past 10 years (2001, 2002, and 2003)." He also stated that the ACS has "become very dependent upon contributions from CAS and Pubs, as well as some investment income, to cover that uncomfortably large fraction of program costs that member dues cannot pay."[53] Indeed, the recent revenue situation for the ACS and CAS seems to be the exact opposite of that of 1955, when CAS was not covering costs and was required by the ACS to become self-supporting. Whether this relationship is a sustainable one is a question central to the future of CAS.

Pricing

Pricing has always been controversial for CAS subscriber libraries. The high cost of CAS databases has been widely viewed as a barrier to access. While large industrial subscribers can accommodate CAS prices, they are high enough to prohibit some schools and businesses from making important chemical information available to students and employees.

As mentioned above CAS changed from a publishing unit of the ACS, which provided scholarly information to its members, to a more autonomous business division whose priorities were

[52] Brynko, Barbara, "CAS turns 100," *Information Today* 1 Feb 2007:4, 2 Feb 2009 <http://cas.org/ASSETS/72B436BFBE3542B1BFB6E5FA4545FA53/Information%20Today%20-%20CAS%20Turns%20100.pdf>

[53] Chamot, Dennis, "Money Matters," *Chemical & Engineering News* 21 Apr 2008: 55.

squarely fixed on revenue generation. This new operating structure and emphasis led to significant financial success for both CAS and the Society.

In 1961 CAS first began selling subscription access to its *CT database* on magnetic tape for $30.00 per month (or $360.00 per year). By May 1968 CAS was offering five digital products, *Condensates, Chemical-Biological Activities, Polymer Science and Technology (POST)* and *CA Basic Journals Abstracts (CBJA)*. Each database was priced as an annual subscription and together the products would cost a library approximately $18,000, which in 2009 dollars is about $106,000.[54]

In 1987, CAS changed its pricing model from a simple per-connection charge to a charge per search term used and partial record viewed. Each full record a searcher accessed was a separate charge. CAS's rationale was that the value the database provided was in both the viewing of records and in the mapping from search terms to the record of interest. At that time CAS also decided to charge relatively high prices for search terms, especially for chemical structures, since this new feature purportedly insulated users from the significant cost of retrieving unneeded records.[55]

In the 2005 report, *Are Chemical Journals Too Expensive,* Robert Bovenschulte suggested that thirty percent of the world's potential users of ACS journals could not access them because their institutions could not absorb such high fees.[56] More recently CAS has attempted to address some of these pricing inequities with policies more in line with what can be accommodated in the developing world.

Unlike earlier CAS database pricing policies, pricing for the *SciFinder* databases has always been based on the number of users and the databases and services required. CAS has begun to offer special pricing programs for Bachelors and Masters degree-granting schools as lower cost

[54] Forrest, Kathryn S., *Center for Information Services, Phase II: Detailed System Design and Programming, Part 2 - A Study of Customized Literature Search Using "CA Condensates" and "CT" Magnetic Tape Data Bases, Phase IIA Final Report. (*Los Angeles: UCLA, Institute of Library Research, 1 Mar 1971: 11-12
[55] West, Lawrence A. , "Private Market for Public Goods: Pricing Strategies of Online Database Vendors." *Journal of Management Systems* Summer 2000:72.
[56] Bovenschulte, Robert., "Report: Are Chemical Journals Too Expensive," *Workshop Report from National Academy of Sciences* 2005, 2 Feb 2009
<http://dels.nas.edu/dels/rpt_briefs/chemical_journals.pdf>

options. *STN* and *STN Easy* are "pay-as-you-go" services with no subscription or monthly fees. Unlike *SciFinder*, *STN* and *STN Easy* price lists are available online.[57]

10) Successful CAS Strategies

The American Chemical Society early on achieved a prominent place in the world of international chemical research. Its products, including its published journals and CAS databases, have become indispensable sources of information to the chemical and pharmaceutical industries. And the interfaces and tools for accessing its data and information products are among the standard implements of the chemistry trade. As David Flaxbart noted in 2007, "The CAS Registry Number became the lingua franca of chemical identification everywhere."

In generating its large and continually growing family of products, CAS has been able to realize tremendous economies of scale by refining and exploiting its longstanding core of indexing and abstracting activities and its data management capabilities. Sophisticated pricing of those products, moreover, has enabled CAS to maximize its return on investment in its core activities.

Unlike many repositories of scientific information, CAS has been able to generate substantial support from sources outside the realm of academia and research. The chemical and pharmaceutical industries have provided CAS a large and lucrative market for its tools and data collections, from which CAS has reaped substantial financial rewards.

In a 1984 article "Opportunities for alternative suppliers of secondary chemical information," Peter Urbach identified four factors that contributed to CAS's early success:

> 1. The practical importance of chemical information relative to other scientific fields gave [CAS] an edge in competing for federal funds.
>
> 2. The U.S. National Science Foundation contributed $30 million dollars, which was matched by CAS, for the development of what would become the system we know today.
>
> 3. English gained acceptance as the universal language of science.

---

[57] STN 2009 Price List can be downloaded in PDF from: http://www.cas.org/ASSETS/9EDAD1B59C9442F6AFE411322726DEE0/stnprice.pdf The price list for STN Easy is at: http://stneasy.cas.org/html/english/pricecol.html

4. CAS management pursued a strategy emphasizing product quality and comprehensiveness.[58]

Two of these success factors are still operative today. English continues to be considered “the international language of science,” although with the growth of scientific publishing from countries such as Korea and China some believe that the language of choice may change. In a 2007 article Michael Heylin described this growth:

> Between 1988 and 2003, the number of U.S. [chemistry] papers grew by 2,600 from 13,200 to 15,800, with very little growth since the early '90s. Papers from the rest of the world grew by a much larger 23,100 from 44,200 in 1988 to 67,300 in 2003. The really big growth has been for Asia, other than Japan, with about 3,000 papers in 1988 and about 15,000 in 2003. The big contributors to this upsurge have been China, India, South Korea, and Taiwan.[59]

Product quality and comprehensiveness, moreover, are still strong customer drivers for CAS. And CAS continues to be perceived as the organization whose databases provide access to the greatest amount of chemical information.

Federal funding in many sciences including chemistry, however, has been flat. A 2005 American Association for the Advancement of Science study found that federal funding in the non-biomedical sciences had been stagnant for 15 years. This was attributed to a change in U.S. priorities. While health sciences, energy and national security had increased in importance, many other areas of science had lost federal funding.[60] The American Chemical Society has not been the recipient of a major grant from the National Science Foundation since 1979.

Technology

A key to CAS’s ability to support major data collections and resources has been its early and aggressive embrace of computer technology. Science has always led the publishing world in the adoption of new distribution technologies. Science and engineering journals were among the first to create online versions and provide hyperlinks to citations and multimedia. ACS was one of the earliest publishers to distribute its publications in electronic format, and was an early adopter of

---

[58] Urbach, Peter, “Opportunities for alternative suppliers of secondary chemical information,” *Journal of Chemical Information and Computer Science* 1984: 1-3,

[59] Heylin, Michael, "The Prognosis for Chemistry: U.S. dominance is still strong, but growth in Europe and Asia augurs the flattening of chemistry worldwide," *Chemical and Engineering News* 21 May 2007, 2 Feb 2009. http://pubs.acs.org/cen/science/85/8521sci2.html

[60] American Association for the Advancement of Science. *Report XXX: Research and Development FY 2006, Historical Trends in Federal R&D* (Washington, DC : AAAS, May 2006) 2 Feb 2009<http://www.aaas.org/spp/rd/06pch2.pdf>

technology that enabled scientists to search and obtain access to publications directly from their office desktops.

In fact, CAS has often advised other organizations on the application of automated processes and systems to the creation and maintenance of databases. During the early 1980s CAS and the Planning Research Corporation were chosen to develop and install a multi-million dollar computer system for the U.S. Patent & Trademark Office (PTO). This system streamlined patent processing and made the PTO the first automated patent office in the world.[61]

The history of technology deployment by CAS suggests other circumstances that contribute to the sustainability of repositories:

.

1. *A database or collection of databases is most likely to survive when it is continually updated and constantly in use.* Although neither updating nor use in themselves guarantee wise preservation choices, regular use does require constant attention and frequent decisions in order to keep the database functional. Even such relatively mundane matters as selection of hardware, operating systems and database management tools must, in such an environment, function both today and tomorrow. Conversely, choices made tomorrow must be reconciled with choices made yesterday and must guarantee functionality the day after tomorrow. Anticipating change is essential in a database that is being constantly updated.

2. *Active and continual adoption of new technologies that work in the day-to-day production environment contributes to the survivability of large, heterogeneous collections of data.* Again, simply adopting new technologies does not guarantee that content will be preserved. But prudent, functional choices help ensure that the data are not merely preserved at the bit-level but are also usable at the record level. Such choices can also add to the efficiency of the operation of the data archive making it affordable to preserve. One can imagine a much less successful outcome of CAS databases if input was still achieved with hand drawings and primitive chemical typewriters, for example.

3. *Using new technologies that meet changing user expectations contributes to the use and therefore the value of the database content.* In the case of commercially marketed databases like those of CAS, user satisfaction and growing use of the data are directly connected to income for basic operations – including preservation.

Competitive Stance

The ACS has taken an aggressive stance toward perceived threats to CAS markets. This tactic has worked well in terms of keeping competition minimal. However it has worked less well as a marketing tool.

---

[61] "CAS Offers Wide Range of Chemical Information Services" Chemical & Engineering News 31 Oct 1994.

In 2002 the ACS sued a company called Leadscope, founded by three former CAS employees, two of whom are chemists and ACS members. The suit asserted that the three were marketing a software program too similar to one on which they had worked for the Society while under agreements giving ACS rights to their creations.[62]

In the same year the ACS was one of several publishers that aligned with the Software & Information Industry Association (SIAA) to end free distribution of *PubScience*, an online database created by the U.S. Department of Energy. *PubScience* had been launched in October 1999, and offered searching across abstracts and citations of works issued by multiple publishers at no cost. It was forced to cease operations in November 2002. [63]

In 2005, the ACS turned its sights upon National Institutes of Health's *PubChem* database, which it considered competitive with its own products. According to the National Library of Medicine 2006 Annual Report, the *PubChem* database was:

> . . . designed to be a repository for small molecule data and the foundation for the massive amounts of bioactivity data that will be produced by NIH-sponsored chemical genomics centers. Approximately 70 organizations are contributing substances to *PubChem*. [64]

In published statements the ACS objected to the NIH duplicating CAS work and asserted that *PubChem* would harm CAS viability.[65] The Society's protestations provoked a backlash from many of its members. Richard J. Roberts, the 1993 Nobel Laureate in Physiology or Medicine resigned from the ACS, stating in a letter he sent to interested parties:

> [ACS] is no longer trying to be a scientific society striving towards the goals of its Congressional charter, which is to represent the best interests of the scientists who form its membership. Rather it seems to be a commercial enterprise whose principle objective is to accumulate money. The ACS management team might be well-advised

[62] DeMartini, Alayna, "Chemical Society loses lawsuit scientists awarded $27 million in trade dispute" *The Columbus Dispatch* 29March 2008, 2 Feb 2009 <http://www.columbusdispatch.com/live/content/business/stories/2008/03/28/LEADSCOPE.ART_ART_03-28-08_C12_HF9P1EG.html?sid=101> The Leadscope defendants filed a counterclaim against ACS, charging defamation, tortuous interference with business relations, unfair competition, and deceptive trade practices. The jury found in favor of Leadscope on all but the last of the counterclaims.
[63] Albanese, Andrew, "PubSCIENCE Dies Despite Comments : Opponent SIIA now takes aim at two other free public databases," *Library Journal* 15 Dec 2002: 17
[64] National Library of Medicine. *NLM 2006 Annual Report*. Washington, DC: (National Library of Medicine, 2006): 46. 2 Feb 2009 <http://www.nlm.nih.gov/ocpl/anreports/fy2006.pdf>
[65]Carroll, William F. Jr., "Statement to ACS Members from ACS President William F. Carroll, Jr.," *acswebcontent.acs.org* 3 Oct 2005, Accessed: 2 Feb 2009 <http://acswebcontent.acs.org/PDF/PubChem_open_letter3.pdf>

> to poll its members to discover if they are happy about the recent actions taken in their names.[66]

In reply to Roberts's letter, Madeleine Jacobs, ACS Executive Director and CEO, argued that *PubChem* threatened CAS's many information products and could lead to a loss of more than half of the Society's income. In the letter, she quoted an unnamed NIH official who had asserted to her that the "CAS business model is outmoded."

The ACS discontinued its NIH campaign, however, after a House of Representatives subcommittee in June 2005 declined to recommend revoking *PubChem*'s funding or otherwise limiting its mandate. The Senate followed suit the following week. The House subcommittee noted, however, that it expected the NIH to "work with private sector chemical information providers, with a primary goal of maximizing progress in science while avoiding unnecessary duplication and competition with private sector databases."[67]

In accordance with the directive from the Congressional subcommittee, the director of the NIH added a working group to the Federal advisory committee, the Board of Scientific Counselors of the National Center for Biotechnology Information. This group, The Working Group on Chemical Information Resource Coordination, is charged with advising the Board of Scientific Counselors on interactions with private sector information providers in the further development of the *PubChem* database.[68]

A minor controversy arose in 2008 among those building the chemistry sections of *Wikipedia*. These individuals had set up a page on the online encyclopedia to comment on part of their project, the validation of *CAS Registry* numbers used throughout Wikipedia pages.

Eric Shively, a CAS employee posted the following comment on the Wikipedia webpage about the project:

> Chemical Abstracts Service (CAS) objects to anyone encouraging the use of SciFinder and STN® to curate third-party databases or chemical substance collections, including the one found in Wikipedia. SciFinder and STN® are provided to researchers under formal license

---

[66] Peter Suber, Letter from Richard J. Roberts, forwarded to mailing list, *SPARC-OAForum@arl.org*. 1 Jun 2005, Accessed: 5 May 2009 <https://mx2.arl.org/Lists/SPARC-OAForum/Message/1977.html>

[67] United States, Senate, *Senate Report 109-103 – Departments of Labor, Health and Human Services, and Education and Related Agencies Appropriate Bill, 2006* (Washington : GPO, 14 July 2005), 2 Feb 2009 <http://thomas.loc.gov/cgi-bin/cpquery/?&db_id=cp109&sid=cp109LH9vq&r_n=sr103.109&item=&sel=TOC_509159&>

[68] United States, National Library of Medicine. "Notice of Meeting" *Federal Register Online* 8 Dec 2005, 2 Feb 2009 < http://www.thefederalregister.com/d.p/2005-12-08-05-23792 >

> agreements, under which the researchers agree to refrain from using these tools to build databases. We urge and expect those researchers to respect the explicit terms of the agreements they have entered into.[69]

Shively's comments prompted further discussion between CAS and Wikipedia participants, which is documented on the Wikipedia page. In the end CAS consented to validate the numbers for *Wikipedia*, noting that they would "help provide accurate CAS registry numbers for current substances listed in the Wikiprojects-Chemicals section of the *Wikipedia Chemistry Portal* that are of widespread general public interest." CAS's willingness to share their information with *Wikipedia* editors may indicate a change in its stance toward Open Access projects.

In December 2004, ACS filed a complaint in the US District Court of the District of Columbia against Google for alleged trademark infringement of the CAS *SciFinder Scholar* brand and for "unfair competition." CAS alleged that Google's use of the term: "Scholar" for its Google Scholar literature search engine constituted trademark infringement and unfair competition.[70] The suit was later settled by Google and ACS, although the terms of the settlement were kept confidential.[71] Google continued to employ the name *Google Scholar*.

International Cooperation

Since the 1960s CAS has aggressively cultivated international partnerships. Since its beginnings CAS has abstracted chemical journals from around the world, a fact evidenced by the 66 languages from which CAS abstracts are created. But the service has also established publishing and distribution arrangements with partners in Europe, the U.K. and Japan.

These arrangements have enabled CAS to achieve an efficient and strategic division of labor, thereby bringing to scale the aggregation and processing of chemical information worldwide. It has also enabled CAS to establish de facto standards for the management of chemical information and its expression. CAS dominance has driven the production by others of tools that interact with CAS databases and content.

---

[69] Wikipedia, "WikiProject Chemistry/CAS validation," *Wikipedia.org* 29 Oct 2008, 2 Feb 2009 <http://en.wikipedia.org/wiki/Wikipedia_talk:WikiProject_Chemistry/CAS_validation>

[70] MEHTA AALOK, "ACS Takes Legal Action Against Google: Google search service is said to infringe SciFinder Scholar trademark" *Chemical and Engineering News* 10 Dec 2004, 2 Feb 2009 <http://pubs.acs.org/cen/news/8250/8250acs.html>

[71] McCullagh Declan, "Google Scholar trademark case ends" *cnet.com* 19 July 2006, 2 Feb 2009 <http://news.cnet.com/Google-Scholar-trademark-case-ends/2100-1025_3-6096240.html>

11) Potential Vulnerabilities

Changes in the Chemical Industry

The *Occupational Outlook Handbook, 2008-2009* predicted that the demand for chemists in the United States would increase at an average rate of about nine percent in the next eight years. However, the same source also predicted that chemical and pharmaceutical companies that previously supported large in-house R&D operations will outsource much more of their work to small firms and non-U.S. companies.

> …chemical manufacturing industries are expected to employ fewer chemists as companies divest their R&D operations. To control costs, most chemical companies, including many large pharmaceutical and biotechnology companies, will increasingly turn to scientific R&D services firms to perform specialized research and other work formerly done by in-house chemists. As a result, these firms will experience healthy growth. Also, some companies are expected to conduct an increasing amount of manufacturing and research in lower-wage countries, further limiting domestic employment growth.[72]

This may produce a decline in some revenue streams for CAS. As larger manufacturers replace their R&D operations with private firms, which are smaller and sometimes overseas, the resources to pay for CAS products may decrease.

Open Access and Open Source Tools

In a recent article in the journal *Nature* Peter Murray-Rust argued that "conventional sources of chemical information are incompatible with the requirements of Web 2.0." He went on to assert that "chemists have no global data-collection projects" and that "the default model for chemical publishing is 'reader pays.' As a result non-subscribers (that's most of the world) have no access to a large percentage of chemical data." [73]

The scientific community was among the first to adopt Open Access (OA) models for distributing new information. Eventually, academically sound, low cost OA journals and databases published by government agencies, non-profit organizations and scholarly societies may compete with some CAS publications. The last few years have seen a surge in open access materials, including chemical information. Such growth has been signaled by the emergence of the NIH's *PubChem* and by *DrugBank*, an effort by David Wishart of the University of Alberta. If and when these

---

[72] *Occupational Outlook Handbook, 2008-09 Edition, Chemists and Materials Scientists* (Washington, DC: Bureau of Labor Statistics, 2008) 2 Feb 2009 <http://www.bls.gov/oco/ocos049.htm.>

[73] Murray-Rust Peter, "Chemistry for everyone" *Nature* 7 Feb 2008.

data collections reach critical mass it is likely that more and better Open Source tools will begin to emerge. This has occurred in the humanities and social sciences.

However, to date no OA publications provide the robust information searching tools that CAS offers. Such value-added features as *SciFinder*, the *CASRegistry* and the abstracting service have yet to be replicated in Open Access Web resources.

Protection of Intellectual Property

With the cost of data storage decreasing, CAS has rightly identified the need to protect its intellectual property, in particular the *CAS Registry* numbers and related information. The *Wikipedia* controversy mentioned above shows how new and unforeseen uses of Registry numbers can threaten CAS control of its information resources. One CAS defense is its user agreement, which prohibits users of CAS products from downloading and storing locally more than 10,000 *Registry* numbers. Enforcement of this agreement may prove difficult, but it serves as a deterrent and reminder to users of the restrictions.

Dissension among ACS Members and Subscribers

As noted above, the American Chemical Society has been widely criticized by members and others for its efforts to impede Open Access efforts like *PubChem*. In addition, the level of its executive compensation has been another source of concern.[74] The University of California Office of Scholarly Communications objected to the fact that the ACS provides executive compensation information only to its members. [75] A protest site called www.idontcare.com/acs has regularly posted the salaries of ACS executives as they were reported in the Society's annual tax filings. (The site has been accessed over 10,000 times since November 1998, suggesting a high level of interest in this topic.) In response to the controversy ACS executive director Madeleine Jacobs recently pledged to reduce executive compensation.

12) Preservation Concerns

Just as one can infer some positive technical lessons from what is known about CAS successes, one must also acknowledge technical concerns that arise from what is not known about CAS

[74] Jacobsen, Jennifer. “Chemical Society Draws Fire for Leader's High Pay" *Chronicle of Higher Education*, 3 Sept 2004 http://chronicle.com/weekly/v51/i02/02a01902.htm

[75] University of California, Office of Scholarly Communication "The Role of Scholarly Societies", accessed December 2007. http://osc.universityofcalifornia.edu/societies.html

policies and procedures. In terms of long-term preservation, those concerns revolve around the needs of scientists in the next generation of research. While CAS clearly meets the needs of scientists today, and in the past has kept pace with those needs, it is not clear whether its current planning will meet the needs of future scientists.

For example, it is not known what kinds of metadata or storage formats CAS currently uses.[76] CAS considers information about its metadata (although not its metadata formats themselves) proprietary. One can speculate that CAS uses one or more of the many open metadata formats (for storing chemical structures, for example). But it is also possible that CAS uses proprietary metadata formats that do not provide the kinds of information that will be needed by the next generation of researchers.

CAS reports that its "storage formats" are proprietary. This probably means that, as in 1971, it has created efficient ways of storing data in whatever computer systems it uses. Although these strategies have been successful, there is no way for chemists to know to what extent they will meet the future challenges of representing chemical information. As Thomas Engel and Johann Gasteiger of the Computer-Chemie-Centrum and Institute for Organic Chemistry, University of Erlangen-Nurnberg, Erlangen, Germany have noted:

> [It] should not be forgotten that there is a wide range of chemical species whose structure can not yet reasonably be represented. Furthermore, highly complex chemical scenes are being studied that cannot be communicated in electronic form.[77]

All systems involve trade-offs. In closed, proprietary systems, however, users of the data have no way of determining what preparations and processes are in place to ensure the continued integrity and survival of the data. The next generation of researchers will be either well or ill-served by design decisions made with regard to those systems today. If more about CAS systems and formats were disclosed, researchers would have a better idea of whether this large, valuable "reservoir" of chemical information will continue to meet the needs of scientists in the future. Researchers would also be better equipped to prepare for the eventuality of information loss if not.

---

[76] CAS declined to respond to most of CRL's requests for information about their technical processes and procedures for this case study.

[77] Engel, Thomas and Johann Gasteiger, "Chemical structure representation for information exchange," *Online Information Review* 2002: 139-145.

Some are skeptical of CAS's approach in general. In a 2007 critical analysis of *Chemical Abstracts*, David Flaxbart, Chemistry Librarian at the University of Texas at Austin, characterized CAS as a potentially obsolete model of scientific information management:

> [As] the information landscape shifts rapidly under everyone's feet, CAS typifies an increasingly outlying Old Guard. Its monopolistic approach to creating and marketing its information services is more and more at odds with the trends of the Google universe. CAS remains the ultimate information fortress, guarding its treasures with increasingly aggressive tactics.[78]

Thus the lack of available information about CAS preservation policies and techniques raises concerns worth considering for those designing new data repositories. Successful long-term preservation requires that the data preserved be usable in the future. It is difficult to consider any proprietary, for-profit system of data preservation an ideal model, since by definition no one but the owners of the system and those to whom they are directly accountable have a say in its preservation policies and strategies.

[78] Flaxbart, David, "The Chemical Abstracts Centennial: Whither CAS?" *Issues in Science and Technology Librarianship* Winter 2007, 2 Feb 2009 <http://www.istl.org/07-winter/viewpoints.html>

Appendix A – ACS Governance (2008)

American Chemical Society Directors

DIRECTOR	TITLE
James D. Burke	Chairman
Attila E. Pavlath	Director
Eli M. Pearce	Immediate Past President

American Chemical Society Officers

OFFICER	TITLE
James D. Burke	Chairman
Madeleine Jacobs	Executive Director
Flint H. Lewis	Secretary
Brian A. Bernstein	Treasurer
Anne T. O'Brien	Director District I
Diane Grob Schmidt	Director District II
Madeleine M. Joullie	Director District III
E. Ann Nalley	Immediate Past President
Judith Benham	Director District V
Denise Graveline	Director of Communications
Eric T. Bigham	Director District IV
Bonnie Charpentier	Director District VI
Catherine T. Hunt	President
Bruce E. Bursten	President-Elect

American Chemical Society Executives

NAME	TITLE	RESPONSIBILITIES
Ms. Madeleine Jacobs	CEO	CEO
Dr. Nina I. McClelland	COB	Chairman of the Board; CEO
Mr. Brian A Bernstein	Treasurer	Operations; Finance; Treasurer
Mr. Arthur M Bremner	HR Exec.	Operations; Personnel
Mr. David Daniel	COO	Operations
Ms. Denise Graveline	Corp. Comm.	Operations; Public Relations/Corp. Comm.
Mr. Robert J Massie	Executive Officer	Other

Ms. Elsa Reichmanis	President	Chief Executive Officer; President
Mr. Dennis Chamot	Board of Directors	Other
Ms. Madeline Jacobs	Executive Director	Other
Mr. Sergio Perez	Plant Manager	Manufacturing
Ms. Kathleen Cullins	Controller	Accounting; Finance
Mr. Flint H Lewis	Corporate Secretary	Legal; Company Secty.

2008 ACS JOINT BOARD-COUNCIL COMMITTEE ON CHEMICAL ABSTRACTS SERVICE

Dr. Spiro D. Alexandratos	Hunter College
Ms. Ann D. Bolek	University of Akron
Dr. Daryle H.	University of Kansas
Dr. Leon L. Combs	Retired
Ms. Judith Currano	University of Pennsylvania
Dr. Patricia L. Dedert	ExxonMobil Research and Engineering, Co. CHAIR
Ms. Andrea B. Twiss-Brooks	University of Chicago
Dr. Richard V. Williams	University of Idaho
Dr. Kenneth John Balkus, Jr.	University of Texas
Dr. Narayan G. Bhat	University of Texas
Dr. Don O. Rickter	Retired
Dr. Robert D. Sindelar	University of British Columbia
Committee Consultants	
Dr. Anne T. O'Brien	Consultant, Retired, Wyeth-Ayerst Research
Ms. Barbara J. Peterson	Peterson Global Knowledge Management
ACS Committee on Committees LIAISON	
Dr. Lissa Dulany	Consultant, UCB Chemicals
CAS Staff Liaison	
Mr. Robert J. Massie	CAS
CCAS Secretary	
Ms. Christine C. Casey	CAS

Governing Board of Publishing

Madeleine Jacobs	Executive Director & CEO - Chair, Governing Board for Publishing
Judith L. Benham	Chair, ACS Board of Directors
Robert J. Massie	President, Chemical Abstracts Service
Brian D. Crawford	President, ACS Publications Division
Jane N. Ryland	President Emerita, CAUSE
Mark M. Edmiston	Managing Director, AdMedia Partners

Gary B. Schuster	Dean, College of Sciences, Georgia Institute of Technology
John J. Hanley	Chairman Emeritus, Scientific American, Inc.
Jay Jordan	President & CEO, Online Computer Library Center, Inc. (OCLC)
Paul S. Anderson	Former Senior Vice-President of Chemical and Physical Sciences, DuPont-Merck Pharmaceuticals Co.; ACS President in 1997

Chemical Abstracting Service Executives

NAME	TITLE	RESPONSIBILITIES
Mr. Robert Massie	President	Chief Executive Officer; President
Ms. Christine Mc Cue	VP of Marketing	Marketing
Mr. Peter Roche	VP of Finance and Admin.	Finance; Administration
Mr. Eric R. Shively	Public Relations Group Leader	Public Relations/Corp. Communications

Appendix B -- ACS Committee Responsibilities

Committee on Publications

The committee shall be responsible for:

a. Assessing editorial quality and content of the publication program of the Society, including the system of primary publication of books, journals, scientific communications, magazines, the official organ of the Society, and miscellaneous publications; and reporting its findings to the Board of Directors;

b. Serving as a channel for communication among Society members, the Council, other users of the Society's publications, the Governing Board of Publishing and the Board of Directors to ensure that needs and support are recognized, researched, and addressed;

c. Consulting with the editors of the above publication concerning editorial policy;

d. Consulting with the Board of Directors upon matters of appointments concerning the above publications;

e. Making recommendations to the Board of Directors related to the needs and adequacy of the publications of the Society, including the need for curtailment of ongoing efforts and the initiation of new efforts, and recommending to the Council and the Board of Directors any changes to these publications;

f. Advising the Board of Directors, Council and staff on copyright policy and recommending actions to protect ACS copyright.

g. informing and educating members on copyright through national and regional symposia and through articles in the official organ of the Society and other ACS publications;

h. Monitoring developments on copyright issues at international and national levels; and

i. Informing the Council and the Board of Directors of the Committee's activities.

Joint Board-Council Committee on Chemical Abstracts Service

a. Developing an informed, constructive and supportive relationship with the society body established by the Board of Directors to manage and conduct the affairs of Chemical Abstracts Service, including reporting of each groups activities to the other;

b. Serving as a channel for information flow between the Governing Board of Publishing, Society members, and individual Chemical Abstracts Service users to help assure that each party's needs and support are researched, recognized and appreciated;

c. Encouraging and considering suggestions from the membership concerning CAS and transmitting these suggestions to the Governing Board for Publishing and the Society's Board of Directors;

d. Responding to requests by the Governing Board for Publishing that will assist in furthering the Chemical Abstracts Service mission; and,

e. Informing the Council and the Board of Directors of the Committee's activities.

Appendix C

CAS Directors

1956-1958	Evan J. Crane
1958-1986	Dale B. Barker
1986-1991	Ronald L. Wigington
1991	Clayton F. Callis (interim CAS director)
1991-Present	Robert Massie