

PREPRINT

Published in: Issues in Science and Technology Librarianship, Winter 2007

<http://www.istl.org/07-winter/viewpoints.html>

The Chemical Abstracts Centennial: Whither CAS?

David Flaxbart
Chemistry Librarian
University of Texas at Austin

2007 marks the 100th anniversary of the founding of Chemical Abstracts Service (CAS), and in anticipation of any hoopla being planned by CAS itself it seems appropriate to pause and take a look both back and forward at an information service that is monumental in its scope and impact on libraries.

The early history of *Chemical Abstracts* has been well documented. (Brown & Weeks 1952; Crane 1958; Maizell 1998) It was not the first major subject-based indexing service, but it rode a wave of enthusiasm for such endeavors in the early years of the 20th century. From the beginning it was an official operation of the American Chemical Society, expanding an abstracting section that had been published as a supplement to the *Journal of the American Chemical Society* since 1897. It was an academic undertaking, relying for many years on volunteer abstractors. After a year at the National Bureau of Standards in Washington DC and another at the University of Illinois, five staffers settled into a back room in the chemistry department at Ohio State University in Columbus in 1909. It moved to its present wooded campus just north of OSU in 1965.

Under the early editorships of W.A. Noyes Sr., Austin Patterson, and especially E.J. Crane (1914-58), CA grew rapidly in size, scope, and quality. It eventually overwhelmed its established European competitors and by mid-century stood alone as the key tool for locating the world's chemical literature. (1) What began as two little volumes in 1907 (free to members and carrying a subsidized price of \$6.00 for non-members) had by 2000 swollen into a behemoth that swallowed library budgets and shelves at an alarming pace. Its colossal 1997-2001 Collective Index occupies 35 linear feet and lists for \$43,000.

CAS did much more than just abstract the journal literature. CA's patent coverage has always been thorough and continues to expand. CAS established and refined a highly sophisticated system of chemical nomenclature to aid in indexing and retrieval of chemical substance information. (Giles & Metanowski 1993) In the early 1960s CAS began an extraordinarily precocious program to use computers to aid indexers in their work. It pioneered the computerized keyword-in-context (KWIC) concept with its new *Chemical Titles* publication in 1961. But it was the Registry system (1965) that revolutionized the chemical information field. It used computer algorithms to analyze and store the 3D description of a molecular structure and make those structures searchable. (Schulz 1994; Weisgerber 1997) Now indexers could build a database of chemical compounds described in the literature. The CAS Registry Number became the lingua franca of chemical identification everywhere.

Recognizing a business opportunity when it saw one, CAS made their computerized databases available to librarians and specialists elsewhere via their own network, STN, in 1984. An entire generation of chemistry librarians learned their skills pecking away at command-based searches while hearing the clock ticking in their heads. Drawing chemical structures under duress using only ASCII commands and displays was a daunting task, but it freed us from the strictures of keywords and unpronounceable chemical names.

From the 1980s on, CAS became an ever more lucrative cash cow for the American Chemical Society as a result of its database marketing activities, which were primarily aimed at corporate and special libraries in the chemical and pharmaceutical industries. Academic institutions did benefit from substantial online discounts (first 90, then 80 percent off) if one was willing to search after 5pm local time. But STN was first and foremost a tool for companies with high-volume needs and deep pockets. This orientation would shape much of CAS' marketing and policies for many years. The hiring of Robert Massie, a business executive, as Director in 1992 helped cement the corporate outlook at CAS.

The advent of the Web in the early 1990s took electronic information out of the hands of trained mediators and put it onto users' desktops. This drove CAS to develop the SciFinder product, a client software interface to its core databases that end users could access themselves without having to make an appointment or pay a time- and results-based fee. Predictably, SciFinder was rolled out to corporate customers well before an academic version was launched. But its capabilities and innovations thrilled all who saw it and raised academic interest to a fever pitch. SciFinder Scholar was made available for academic site-licensing in 1997.

And it was then that the longstanding symbiotic relationship between Chemical Abstracts Service and the academic community began to fray. The initial licensing cost of SciFinder Scholar was so high, and so out of line with what libraries had been paying for other database tools in the sciences, that outrage and accusations of price gouging quickly took root. Restrictions on access to SciFinder Scholar based on an inadequate number of "seats," the time of day, and a user's personal affiliation further rankled librarians. CAS came to be regarded by some librarians not so much as a partner in the scientific enterprise, but as a profit-driven business. CAS defended these measures as central to protecting its market position and preventing for-profit uses that it claimed would erode its business. After some lengthy negotiations with groups of librarians, CAS eventually moderated its pricing structure somewhat. Nevertheless, SciFinder Scholar is one of the most expensive resources on the market and, despite its ease of use and usefulness as a teaching tool, is still not a realistic option for many small schools with limited budgets. And many users chafe at the proprietary client-server model and lack of a more flexible web interface, and the cost of adding necessary seats remains very high.

As it looks back on its first hundred years, Chemical Abstracts Service is at a crossroads. It has much to be proud of. It is the premier indexing service in the world, in any subject. Its coverage and scope are unmatched and stretch well beyond pure chemistry. It has registered over 30 million chemical compounds and indexed over one million documents in 2006. The full backfile of CA, back to 1907, has been digitized and re-indexed and added to the CAPLUS file available via SciFinder; and CAS is now retroactively adding some pre-1907 records to the database. (2) Calculated and experimental property data have been added to thousands of Registry compound records. These enhancements have come at no additional charge to customers. The

SciFinder client has undergone steady improvement in the decade since it appeared, and although there are a few annoying gaps in its capability, it is an excellent product overall, which librarians can blame (or thank, depending) for abruptly ending the era of mediated chemical literature searches.

The investment required for this level of innovation, development and expansion is not trivial, and it helps to justify to some extent the high prices libraries pay for CAS products. Services such as customer support, training materials, and documentation also add to the price. Librarians should acknowledge that on many levels CAS offers a product far superior to, say, *Engineering Index*, and factor that into their assessments.

But 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. Following on its legal battles with Dialog in the early 1990s, CAS successfully lobbied in 2002 to shut down PubScience, a free index service operated by the Department of Energy (Foster 2002). Perhaps encouraged by this victory against what it considered unwarranted government competition, CAS and ACS in 2005 launched a savage frontal assault on NIH's PubChem database. Unlike DOE, the NIH stood its ground and PubChem has so far survived (Drake 2005; Marris 2005; Univ. of California). Most quixotic of all, the ACS sued Google in 2004 over – of all things -- the use of the word "Scholar," claiming rather absurdly that the new Google Scholar database would somehow confuse scientists who allegedly assume "Scholar" could mean only SciFinder. CAS settled (and lost) amid much derision. In spite of these public relations fiascoes, the Society remains implacably opposed to any kind of open-access initiative that might impact its monopoly on chemical information, and this begs the question of whether CAS and its parent organization exist to further the scientific enterprise, or merely to make money.

Ultimately the fate of all information providers – including libraries – rests on the fickle tastes of end users, whose outlook has changed significantly in recent years. CAS' claims of high quality, comprehensiveness, and expert curation, however valid, may fall on deaf ears as users shift from a "no stone unturned" approach to a "good enough" model of discovering information that Google enables so well. CAS' insularity and resistance to change may eventually cause customers to go elsewhere for their information needs, following paths of less resistance and lower cost. Their take-no-prisoners reaction to perceived threats alienates customers and rank and file society members alike.

Librarians are the natural allies and unpaid sales force for high quality information providers like CAS. Despite CAS' occasional over-aggressive tendencies and high prices, most of us remain convinced that real science cannot get done without the intensive use of databases like theirs. So far, most chemists seem to agree. But nothing is forever. CAS faces a true challenge in their second century, to convince an upcoming generation of Google-ized users, OA advocates, downsizing corporate libraries, and institutional subscribers that their approach is still worth the cost of paying for these products and persuading people to use them.

There's no shortage of librarians willing to give CAS advice, but here's a start: CAS should grow a new generation of users by allowing more undergraduate institutions to take advantage of

SciFinder by significantly reducing its bottom-tier price structure. It should stop worrying so much about inappropriate use and eliminate the much-criticized prohibition on walk-up users in libraries. It should drop its seat-based pricing requirement and allow research institutions to have sufficient access for the price of a full subscription. It should explore ways to move away from proprietary software interfaces and towards a web-based interface, though this will take time. And finally, CAS and the ACS should take a page from other major STM publishers and embrace some aspects of Open Access and interconnectivity as potential boons to their business, rather than trying to litigate and lobby them away. It should stop attacking legitimate government efforts to assist scientists and get taxpayer-funded information into the hands of people.

Chemical Abstracts Service has a long, admirable record of being a pioneer and a trend-setter. It deserves acclaim for the database system it has built. Let us hope that it can continue to adapt to changing times and stay in the forefront of scientific information where it belongs.

NOTES

1. CA's primary competitors were *Chemisches Zentralblatt* (1830-1969) and *Journal of the Chemical Society - Abstracts* (1871-1925, later *British Abstracts*, 1926-53).
2. Added retrospective coverage includes pre-1907 U.S. patent records, selected landmark papers, 1890-1906 abstracts from *Journal of the Chemical Society*, and articles from the *Journal of the American Chemical Society* and *Journal of Physical Chemistry* back to their first volumes.

REFERENCES

- Brown, Charles A. and Weeks, Mary E. 1952. *A History of the American Chemical Society, seventy-five eventful years*. (Washington: American Chemical Society)
- Crane, E. J. 1958. *CA today: the production of Chemical Abstracts*. (Washington: American Chemical Society)
- Drake, Miriam. 2005. "A Cauldron bubbles: PubChem and the American Chemical Society." *Info Today*, June 6, 2005. <http://www.infotoday.com/newsbreaks/nb050606-1.shtml>
- Foster, Andrea L. 2002. "Energy Department Seeks to Close Web Site That Searches Scientific Journals." *Chronicle of Higher Education*, 49(2) Sept. 6, 2002, A46. <http://chronicle.com/weekly/v49/i02/02a04604.htm>
- Giles, P.M. and Metanomski, W.V. 1993. "The History of chemical substance nomenclature at Chemical Abstracts Service." in *Organic chemistry: Its language and its state of the art*. (Basel: Verlag Helvetica Chimica Acta), 173-196.
- Maizell, Robert E. 1998. "Chemical Abstracts Service: History and development." in *How to find chemical information*. 3rd ed. (New York: Wiley), 60-106.

Marris, Emma. 2005. "American Chemical Society: Chemical reaction." *Nature*, 437 (Oct. 6, 2005), 807-809.

Schulz, Hedda. 1994. *From CA to CAS online: databases in chemistry*. 2nd ed. (Berlin: Springer)

University of California. Office of Scholarly Communication. The American Chemical Society and NIH's PubChem. http://osc.universityofcalifornia.edu/news/acs_pubchem.html (Accessed 1/22/2007).

Weisgerber, David W. 1997. "Chemical Abstracts Service Chemical Registry System: history, scope, and impacts." *Journal of the American Society for Information Science* 48(4), 349-360.