FROM LIBRARY STACKS TO COMPLEX ONLINE SEARCHES
THE EVOLUTION OF CHEMISTRY INFORMATION MANAGEMENT
Information grows and changes over time. New facts, concepts and ideas continually augment, mold and even overturn previous knowledge and beliefs. In the 21st century, this expansion has been particularly palpable because globalization and the interdisciplinary nature of research have expanded the scope and diversity of information sources. Keeping pace with knowledge in any scientific discipline has become a challenge. Chemistry is no exception. Successfully capturing what is relevant to your research requires the right tools.

Information management systems were born from the explosion in scientific research of the early 19th century. The accumulated data was collected, organized, and recorded in published books. With time however, the drive to process rapidly growing amounts of information and make it directly accessible to users prompted the transformation of printed sources into the sophisticated electronic information systems of today. Each new generation of information management tools leveraged advances in technology to address deficiencies of previous generations and empower the researcher to better find, gather, integrate and synthesize information.
FINDING THE RIGHT INFORMATION

Just forty years ago, an initial examination of the current state of information relevant to a research idea required weeks of work. Finding physicochemical properties of compounds, experimental data on relevant reactions, and identifying colleagues with the latest work in the field would have required sifting through card catalogs and microforms, checking library stacks for particular journals or textbooks, and requesting unavailable items from other libraries. Searching filed patents to clear the intellectual property space required contacting patent offices. Each aspect of the search could take days or weeks and most likely necessitated support from librarians or other information specialists.

The widespread introduction of microforms in the first half of the 20th century greatly improved the storage and availability of information at libraries. Microforms were inexpensive and compact, saving costs and precious space. They required no specialized equipment to view (just a magnifying glass) and under the right conditions, microforms were resistant to tears and stable for long-term storage. This meant that libraries could place larger collections of bibliographic and periodical items in the hands of users; however, although they brought more information to the user they did little to save time in the actual search process. Ultimately, it was the systematic indexing of the information in library collections that made it possible to find the right information quickly.

An often overlooked development that emerged from the growing number of books and microforms of library collections was the creation of indexing systems to lead the user to the right book volume or microfilm reel where they could find the information needed. Examples of such systems recorded each item in a library – book, journal, graphic, map, musical record, microform, etc. – on an information card and copies of that card were filed under relevant keywords and bibliographic data (e.g., author name, title) in card catalogs. A user could then find all items under a given keyword and locate them in the library stacks. These indexing systems were limited and static, but they were precursors to the useful indexing powering the first electronic databases of the 1980s.

Rapid advances in computer hardware and programming made it possible to generate comprehensive databases that could be accessed from a computer terminal. The digitization of information contained in printed forms eliminated physical space issues and entries in these databases could be extensively indexed since this electronic "tagging" had no space restrictions like the physical card catalogues. Finally, the development of highly organized indexing systems and algorithms to extract entries from the database assumed much of the search work, dramatically reducing the amount of time needed to find relevant information to only hours.

Ultimately, it was the systematic indexing of the information in library collections that made it possible to find the right information quickly.
The advent of the commercial internet and graphical user interfaces catalyzed a new approach to information access. The focus of development shifted to bringing information directly to the user and enabling the exploration of the content of databases. On one level, this meant developing mechanisms to extract selected information from the main text and even the footnotes of publications and feed it into a database. Beyond delivering citations relevant to a particular search topic, these new information management systems reveal the data contained in each citation that is actually critical to the question at hand. On another level, this meant connecting entries in a database in such a way that a search could be easily expanded and even lead down novel information paths. The connections generated captured synonymy, dialect and discipline-specific terminology, and related topics.

These two driving forces have and continue to elicit the creation of increasingly sophisticated organizational frameworks that accommodate a broad spectrum of relationships among database entries. Beyond the classic classification structures of equivalence (x is the same as y) and hierarchy (x is a subset of y), new frameworks describe relationships between database entries that can power, for example, computation or automated inferencing. A striking example of this expansion is the linking of physicochemical, bibliographic and reaction data to chemical structure. While classifying data by compound structure was already practiced in printed chemistry handbooks, the new electronic version of this organization translates compound names (and synonyms) into corresponding chemical structures, expands searches to include structurally similar hits, and automatically generates synthesis plans. In this way, intelligent algorithms capitalize on these sophisticated frameworks to take over some of the arduous work of integrating and synthesizing information that was previously done by hand.

Nevertheless, these first databases were restricted in access. Complex cost structures for their use made it prohibitive to simply explore the content. Searches needed to be targeted and efficient, which required the user to have an understanding of the database structure and underlying programming language, and this conversely required significant training. Thus, although more information could be accessed more rapidly, for the most part the output still had to be brokered by a specialist. Furthermore, developments to this point addressed the collection of information but did little for its integration and synthesis. The output from electronic databases still needed to be manually cross-checked with existing knowledge and then processed.
The interaction of user and database was also revolutionized by these two driving forces. Incorporating a comprehensive description of the interconnectedness of information in a database opens that content to a broad range of search approaches. The upshot is that relevant chemistry information can be accessed by any user, regardless of their specialty or the nature of the question they wish to answer.

Furthermore, with a few mouse clicks, the relevance of a search can be expanded to include tangential but connected information, which can in turn lead to novel insights otherwise inaccessible through strictly defined search criteria. Ultimately, new chemistry information systems are designed to generate reliable and insightful results starting from an intuitive query.

**EVOLVING HAND-IN-HAND WITH INFORMATION NEEDS**

The evolution of information management systems has resulted in comprehensive databases embedded in descriptive organizational structures. The next challenge in remaining abreast with evolving information needs is to make access to this information as intuitive as asking a question. To this end, an effective information system must understand not only what information is relevant to a search, but also the context of the question and how the information will be used. Given advances in database structuring, the ever-increasing sophistication of algorithms, and open feedback from users, meeting this challenge is in the foreseeable future.
CONTACT INFORMATION

Please visit www.elsevier.com/reaxys

AMERICAS
(8am – 8pm CET – St. Louis)
Tel: US toll-free: +1 888 615 4500
Tel: Non toll-free: +1 314 523 4900
Email: usinfo@elsevier.com
Email Brazil: brinfo@elsevier.com

ASIA AND PACIFIC
(9.00am – 6pm SST – Singapore office)
Tel: +65 6349 0222
Email: sginfo@elsevier.com

EUROPE AND ALL OTHER REGIONS
(9.30am – 5.30pm JST – Tokyo office)
Tel: +31 20 485 3767
Email: nlinfo@elsevier.com

JAPAN
(9.30am – 5.30pm JST – Tokyo office)
Tel: +81 3 5561 5035
Email: jpinfo@elsevier.com
www.elsevier.com/jp