CHARLES H. DAVIS

Professor
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

How We Got Where We Are: A Brief Chronology

The creation of machine-readable databases and computer-based services has always been predicated on the availability of appropriate hardware and software. During the first generation (roughly 1949 to the late 1950s), very little happened because the machines were slow, had relatively little storage capacity, and were extremely expensive. In addition, most programming was done at the machine level — a tedious process. Second generation computers (from the late 1950s through the early 1960s) used transistors instead of vacuum tubes, which meant they were faster, more reliable, held more data, and could be afforded by institutions smaller than the federal government.

To facilitate programming, a number of higher-level languages were developed during the early years. FORTRAN was designed primarily for scientific and engineering applications; ALGOL, the first of the so-called procedure-oriented languages, provided an internationally recognized structure for program documentation; LISP eventually proved valuable in studying artificial intelligence; and COMIT, the first language designed specifically for text processing, was used in computational linguistics and early studies in information retrieval.

Higher-level languages greatly facilitated software development, because programs using them were shorter, easier to understand, and could be used on a variety of computers, unlike programs written at the machine level. It was also during this period, in 1958, that Hans Peter Luhn of IBM described the mechanized production of keyword indexes as well as an automated current-awareness service called SDI (Selective Dissemination of Information).

The language COBOL was introduced in 1960 and had special importance for libraries. Unlike earlier languages, it facilitated the

handling of large alphanumeric records and files. It was the first language well-suited for use with MARC records, which were developed by the Library of Congress during the mid- to late 1960s. Magnetic tapes of these records were distributed to selected libraries so that the use of cataloging information in this format could be studied. At this time, considerable work was done on the automation of library technical services such as acquisitions, circulation, and cataloging. Another programming language also appeared in the early 1960s: SNOBOL, which might best be described as a successor to COMIT; it was very popular for text-processing applications but saw only limited use in the automation of library technical services.

The early to mid-1960s saw the transition, especially in scientific information handling, from labor-intensive, error-prone tasks to automated processing—often to expedite the efficient production of printed products (e.g., *Index Medicus, Chemical Abstracts*). Keyword indexing, SDI, and other batch-mode processes became popular. This period also ushered in the third generation of computers, which featured integrated circuitry, greater emphasis on direct-access storage (especially magnetic disks), and improved facilities for telecommunication.

New programming languages included PL/I and BASIC. PL/I incorporated the numerical capabilities of FORTRAN, the file-handling of COBOL, and the most crucial text-processing features of COMIT and SNOBOL—all in a structure that looked like ALGOL. PL/I has been used extensively in library automation; BASIC was originally designed to help students learn programming while online to mainframe computers.

Who was creating databases? Institutions and agencies of the federal government, e.g., the Library of Congress (LC), the National Library of Medicine (NLM), the National Aeronautics and Space Administration (NASA), the Atomic Energy Commission (AEC), and the Commerce Department; large professional societies, e.g., the American Chemical Society (through the Chemical Abstracts Service) and the American Psychological Association; some large universities through grants, e.g., SPIRES and BALLOTS at Stanford University; and private enterprise, e.g., the Institute for Scientific Information (ISI) with *Index Chemicus* and *Science Citation Index*.

The late 1960s saw the start of large bibliographic utilities such as OCLC, RLIN (originally BALLOTS), WLN, UTLAS—all of which required third generation hardware and software as well as substantial improvements in telecommunications technology. Probably the most popular and certainly the largest venture of its kind, OCLC was not regarded initially as the source of a database for online searching, but rather as a means of producing and distributing catalog cards for individual libraries.

The early 1970s were distinguished by the advent of minicomputers and the growth of online search services designed for efficient searching via telecommunications (e.g., System Development Corporation's ORBIT; Lockheed's DIALOG; the New York Times Information Bank; and later, BRS [Bibliographic Retrieval Service]). The online search services featured Boolean search logic for interrogating large files. Minicomputers permitted the development of local and regional online circulation systems, usually with truncated bibliographic records. In addition, the fourth generation of computers appeared, characterized by even higher processing speeds and greater storage capacity. Another new programming language, Pascal, was also introduced. Named for the French philosopher and mathematician, it is similar to PL/I but is more streamlined and has superior implementations that are now available for all sizes of computer.

The late 1970s brought Altair, Radio Shack, and Apple microcomputers—toys at first, with libraries purchasing a few, mostly for entertainment and as an inducement to use other library services. Because it was a simple language with few hardware requirements, BASIC became the most popular programming language for these smaller machines.

During the 1980s, microcomputers grew from 8- to 16- and even 32-bit word machines, meaning that they quickly assumed the power formerly associated only with minis and mainframes. The internal "clock" speed of these computers has also increased dramatically, from 4.7 to 33 MHz and even higher. In the early 1980s, IBM entered the microcomputer business in a big way with its PC (Personal Computer) using Microsoft Corporation's operating system, MS-DOS. Although not "state-of-the-art," the IBM-PC became an industrial standard because of IBM's enormous marketing capabilities. In 1984, Apple introduced the first of its Macintosh microcomputers, which, although not compatible with the IBM machines, offered different capabilities including graphic user interfaces and a "mouse" for quick placement of the cursor on the computer's monitor. Software manufacturers, inspired initially by the Macintosh series, have begun to explore options using graphic user interfaces as alternatives to the traditional command- and menu-driven systems. Most of the newer interfaces and software packages are designed for the larger and faster microcomputers made available only recently. Also of considerable importance has been the introduction of CD-ROM (Compact Disc-Read Only Memory). It permits the storage of about 550 million characters on one 5 1/4 inch disc, can store graphics as well as text, cannot be disrupted by magnetic fields (and therefore has archival potential), has made significantly more data available to individual users, and may replace magnetic tape as the distribution medium of choice.

Individuals, businesses, and libraries have purchased microcomputers for such operations as word processing, spreadsheet analysis, database management, information retrieval, and electronic mail. Widely available off-the-shelf software packages (too numerous to mention) perform these and other tasks. Optical storage technology, especially CD-ROM, has been linked with micros, creating local workstations for bibliographic database searching without the cost of telecommunications. Hardware and software developments make possible local and regional online catalogs of full bibliographic (MARC) records.

While there is talk of a fifth generation of computers, it is generally considered to be in progress, and the differences between generations have become more subtle. Newer languages include Ada, which is similar to Pascal; PROLOG, which is used primarily for work in artificial intelligence and expert systems; and both Microsoft and Turbo Pascal, which are microcomputer-based supersets of standard Pascal that contain numerous useful string manipulation functions. The newest versions of Pascal also feature object-oriented programming, which deliberately blurs the distinctions between programs and their data; it is meant to go beyond conventional procedures to simplify computer programming and make it accessible to a wider audience. Intended primarily for programming professionals, the language C and its object-oriented extension C++ occupy a level somewhere between machine- and higher-level programming.

Current areas of interest include, but are certainly not limited to: studies of the concept of user friendliness, experiments with expert systems, the use of microcomputers as intelligent terminals, system interface design, a reexamination of the roles of batch-mode and online services, studies of the dichotomy between end users and intermediaries, large databases, full-text systems, and the library as an access point for community databases.