HISTORY OF ON-LINE DEVELOPMENT

Joseph Becker*

My aim in this talk is to present a personal perspective concerning the development of on line technology in the field of information science. I believe the best way to understand the impact of new information technology is to know the people who were its prime-movers — the experimenters and the pioneers. For after all, it was their courage to innovate and their personal spirit that made things happen during the past forty years.

Therefore, as I elaborate on developments and trends, I will wherever possible interleave the commentary with mention of the work of selected people whom I had the privilege of knowing along the way. Many of them were early ASIS members. I wish there was room to mention them all. The sum of these historical vignettes will, I hope, provide you with insights and perspectives you might not glean any other way.

The 40s

My own interest in information technology began in 1944. As a young officer in the Army, I used punched cards to store index terms and retrieve technical documents. At the time, it was considered daring to apply punched card technology to non-numerical work and I was pleased to learn later that librarians like Majorie Quigly and Ralph Parker were just as reckless. Today punched card work is considered trivial. In 1944, people viewed it as pure magic.

* Joseph Becker, president of Becker and Hayes, Inc. U.S.A.
The following year Vannevar Bush authored his seminal article entitled *As We May Think*. In it, Bush described a machine he called "Memex" that continually processed facts and figures and retrieved them swiftly on command. During World War II, Bush had been Head of the U.S. Office of Scientific Research and Development and, most likely, was familiar with early computers like WHIRLWIND and ENIAC. Whether he foresaw their application to information is not known, but he was in a favorable position to appreciate the computer's potential.

Circumstances brought Bush and Ralph Shaw together in 1945. Ralph Shaw was the Librarian of the U.S. Department of Agriculture. He persuaded Bush to build the Memex machine and to apply it to the Bibliography of Agriculture. With help from engineers at the U.S. National Bureau of Standards, a prototype device was designed and a year later it was operational.

I first saw the Bush-Shaw Rapid Selector in Ralph's office on Independence Avenue in Washington. It's data base consisted of 35mm film stock on which were recorded copies of the abstracts found in the Bibliography of Agriculture. Codes appeared alongside each abstract in the form of black and white dot patterns. These were binary coded representations of the index terms associated with each document. Inserting the inverse of these binary codes into the selector, and running the film, caused the machine to "blackout" each time it encountered a target index term. The secondary result of each "hit" was the photographic transfer of the corresponding abstract onto a piece of undeveloped film. Thus, by running the entire film spool, all of the abstracts satisfying the search strategy were recorded incrementally on one strip of undeveloped film. The film strip containing the accumulated abstracts was then chemically treated like any other film, and the developed strip given to the user as the search printout.

Neither Bush nor Shaw referred to their work as data base searching and extraction but the genealogy of today's on-line information utilities leads us directly back to the roots of the
Memex experiment.

Shortly thereafter, the implications of the "information explosion" in science and technology were aired and discussed at the first international conference on the subject, held in London in 1948 by the Royal Society. At that time it was already clear that the publishing rate in science and technology was increasing exponentially, and that specialization in the individual sciences and interdisciplinary research were generating multiple uses for the same information. Although interest in information storage and retrieval thus received its start in the world of science, it soon spread to other areas, particularly business, industry, government, and education.

This widespread recognition of the information explosion and of the urgent need to employ new means to cope with it was a major factor in the movement to introduce technology into information work.

Another key factor responsible for the development of information technology was the impact of the development of four major technologies: computers, communications, photography, and television. The fruits of research and application in these four areas provided highly significant data which implied new methods and techniques for storing and retrieving information. Unforeseen at the time, however, was the convergence of these four technologies and the effects this would have on changing the information environment. This convergence eventually provided capabilities in summation that were more powerful than the individual parts.

The 50s

During this era, several specialists devoted themselves to research into the problems of information organization and they used punched card, punched paper tape, and jury rigged equipment in the process.

Among them was Mortimer Taube, who is identified as
the founder of coordinate indexing, which provided a method for coordinating index terms as combinations rather than permutations. Taube called his index terms Uniterms and his coordinate index consisted of a set of Uniterm cards on which appeared the identification numbers of the documents relevant to each Uniterm. Searching was accomplished by selecting those Uniterm cards pertinent to a request and correlating their document numbers. Fred Jonker built a machine to facilitate the searching process. He called it the Termatrix. It looked like an egg candler and functioned somewhat as one. The selected Uniterm cards were placed in the device in front of a light source. This immediately identified the document that were relevant to the search. It was called the “peek-a-boo” system then, today we use the phrase “optical coincidence” to describe the same process.

This was also the era of knitting needle technology. Calvin Mooers proposed a concept for superimposing the codes for the subjects in a document one atop the other. This technique was particularly applicable in situations where coding space was at a premium on edge notched cards. In its day, it represented the most sophisticated way of using technology to perform correlative searches.

It was at this juncture that first generation computers made their debut and a spact of new technological ideas emerged: P.R. Bagley wrote a doctoral thesis at MIT in which he proposed that a special purpose “information searching machine” be engineered to compare large blocks of data simultaneously, thus reducing overall search time. Hans Peter Luhn of IBM invented a special searching device. Luhn’s machine “read” IBM cards longitudinally rather than cross-wise. When IBM punched cards were laid end-to-end this way it was like reading a ribbon of codes and this was the beginning of magnetic tape reading principles which came later.

James W. Perry, Allen Kent, and Jesse H. Shera at Western Reserve University built an experimental machine with limited tape capabilities albeit punched paper tape. The device used relays
as switches, punched paper tape as the medium for storing abstracts, index terms, and search requests. The search logic was programmed into wiring boards similar to those found in punched card equipment. The WRU search machine was able to search an abstract a minute and handle 10 questions simultaneously. It was the forerunner of today's high speed text search machines sold by the General Electric Co. and Logicon, Inc.

The 60s

Once second generation computers appeared — such as the IBM 701, 704, and 709 — researchers like H.E. Tillet, P. Baxendale, A. Barton, L.N. Caplan, M.E. Maron, L.C. Ray, H.P. Edmundson, Don Swanson, T.R. Savage, D. Climenson, and many others used this equipment to exercise the computer's ability to perform automatic indexing, automatic abstracting, machine translation, automatic searching, automatic dissemination, automatic question-answering, etc. This was the heyday of language data processing and it was this research, performed on second generation computer technology, that provided the basis for much of today's on-line and other text-related information systems.

The story is not complete, however, unless we go back to one other area of research and experimentation, performed before third generation computers and time-sharing were commonplace, to test and demonstrate, the concept of on-line bibliographic storage and retrieval. Hal Borko was responsible for such research when he worked at the System Development Corporation (SDC) back in 1964. At that time SDC had a Q32 military computer which possessed attributes of third generation computer technology long before they were available publically. The military used the Q32 to process early warning air defense data but Borko managed to get time on the machine to test "BOLD", an acronym which stood for Bibliographic On-Line Display. Many information scientists, however, to this day refer to it as
Borko’s Own Little Demonstrator.

In retrospect, it is interesting to note how many of the features found in today’s on-line bibliographic search services were also incorporated in BOLD. As its data base, the program used 4000 bibliographic records and abstracts taken from technical reports. These were indexed by subject, and were also accessible by author, title, document number, and so forth. The terminals used in the demonstration were connected by cable, rather than by telephone line, to the Q32, and the display consisted of a TV console with a light pen attachment. The user submitted his request using standard descriptors. In turn, the program displayed a matrix on the TV screen, showing the document numbers selected by the machine in response to the search strategy. It also displayed the number of relevant records in the file that met the search criteria, and gave the user the option to change his or her mind with respect to search logic at any point in the program. The light pen was used to select individual pieces of data or for printing out the entire bibliography on a teletype printer or on a magnetic tape.

BOLD was the precursor of SDC’s Orbit system and Lockheed’s Dialog system and undoubtedly affected the development of other data base services which exist today. As important, it was a test bed for applying the most advanced technology available during that period to the problems of information retrieval. It was the emergence of third generation computer technology, like the Q32, and its marriage to telecommunications technology that was to have a most significant impact on the information environment.

Retrieval at a distance of digital information from bibliographic and other files presupposes the existence of an interconnected communications network. The standard commercial telephone network was used in 1964 to demonstrate an on-line bibliographic searching system operated by the American Library Association (ALA) at the New York World’s Fair in “Library USA”, an exhibition of the U.S. Pavilion.
The technology used by Library USA consisted of a UNIVAC 490 with a “fastrand” direct access drum memory. Six UNIVAC Unisets were part of the configuration too. A uniset can best be described as a first generation computer terminal. These terminals were pre-wired to perform a restricted set of request functions. Six reference librarians used them to request a variety of printouts. Essays and reading lists at the elementary and adult educational level in English, German, and Spanish were available demonstrating the machines ability to handle romanized language texts. 800,000 such printouts were generated on highspeed printers (1200 lines/min) during a 12 month period.

This computer system was also used to communicate with teletype machines throughout the country over the standard dial-up telephone system. The public was thus able to make direct inquiries to the UNIVAC 490 in New York over standard teletypewriters and receive identical printouts to those provided visitors at the Fair.

In addition to the essays, the Fastrand drum also stored selected entries from the Reader's Guide to Periodical Literature. Each week the publisher of the Guide, H.W. Wilson Company, sent ALA duplicate copies of its raw-form index entries on 75 subjects covered by eighteen different periodicals which they regularly indexed. The 75 subjects reflected topics contained in the U.S. Exhibit. On command, from in on-site UNIVAC Uniset or from any teletypewriter, the 490 would respond to subject inquiries by printing out the latest 10-15 index entries pertaining to the requested subject.

Library USA demonstrated three potentials of on-line technology: (1) the computer's ability to dynamically update a file; (2) its ability to interface smoothly with a communications network; and, (3) its ability to serve many users simultaneously from remote locations.

Third generation computers like the UNIVAC 490, the GE 635, and the IBM 360 became available in 1967. During this period, work by J.C. R. Licklider and Mike Kessler at MIT, in a
History of on-line Development

project called MAC (Multiple Access to Computers), stimulated development of the "time-sharing" and "remote access" concepts. As these on-line concepts took hold, former geographic barriers to information came tumbling down.

The 1970s and the Future

For information scientists, the years 1970 to 1980 will probably be remembered as the Information Technology Decade. It was during this period that experiments and demonstrations gave way to practice and operational programs. Information science and technology had come of age.

Also, during this period, the pace of technological change quickened and individual contributions to the field became more diffuse and harder to identify than in previous years. Since 1970, people of every professional stripe have become acquainted with information technology in every conceivable context.

Without a doubt, the spectacular spread of on-line computers has been the most significant influence of all. In prior years, computers were mainly for institutions. After 1970, however, they were manufactured for the mass market. The millions of personal computers available today are destined to become smaller, cheaper, and faster. At the same time, they will become smarter — having more memory, more storage, and more stand-alone capabilities.

This means, for example, that a computer terminal will someday be able to accept instructions from a user, perform an information search at a designated time, use limited judgement in contacting other sources of information, revising the strategy of a search as circumstances dictate, and hopefully learn from past mistakes. In time we many even see a pocket size information terminal, much like a hand held calculator, which will bring us individual messages on command.

Technological developments in the computer industry are certain to continue. Manufacturers are already employing magical
silicon chips called microprocessors that condense data and programs into a single, miniaturized form. A form, that in the future, will enable us to swap applications with greater facility than before.

Supercomputers are also on the horizon. They are expected to incorporate parallel and associative processing principles to a significant degree and provide a springboard for further development of artificial intelligence. Another technological upgrade likely to occur is in the field of mass direct storage. In this context, electro-optical devices such as videodiscs are expected to play an increasing role for mass storage of potentially vast proportions; and, new versatile machines called "file processors" are destined to supersede today's disk file controllers.

New developments in communications technology are likewise to impact tomorrow's information environment. The signals emitted by tomorrow's information machines — computers, facsimile equipment, TV cameras, videodisc players, and so forth — all require broad band-width channels for their efficient communication. Our narrow band telephone wires, originally designed to carry voice signals, are simply not up to the job.

Once broad-band-width channels are in widespread use, a totally new, integrated telecommunications infrastructure will be available for information exchange. A broad-band-width national telecommunications network will make it possible to mix signals bidirectionally — voice, digital, and video signals — and thus to accommodate different information appliances on the same line at the same time. It will enable us to tap into a single cable in our home or office and, through gateways in our local area network, reach out to external information networks in our own country and throughout the world. It will vastly increase the density of information message traffic between and among human beings.

In 1979, the Xerox Corporation introduced the Ethernet network concept, which incorporates many of the local features of this new development. Ethernet makes it possible to link
different office machines into a single network over a common coaxial cable. It meets several objectives. It allows many users access to the same data — say one or more electronic data bases; it enables a set of users to share the same central computer; it permits the information network to grow regardless of the changes which may occur in the individual devices — such as model upgrades, design changes, etc. The latter benefit is particularly significant: it means that an organization can plug-in new machines gradually, as its needs dictate, or as technology develops new and better information machines. In other words, it provides a hedge against technological obsolescence.

Of course, it will be a decade or more before a broad-bandwidth telecommunications upgrading takes place. But, the trend is certainly in that direction. Unmistakable signs are already evident. Facsimile and electronic mail companies are now serving several hundred cities across the world. Also, beginning next year after deregulation, AT & T intends to offer its subscribers many new network services. National telephone and telex networks in other countries are developing similar integrated services digital networks capable of carrying voice, text, data, and video signals. As communication links are used more to create geographically and functionally decentralized information networks, the data processing industry and the communications industry are more likely to compete rather than co-exist. Whether these two giants eventually blend into a superindustry is presently unclear, but, regardless of organization, one general outcome of the trend is clear. Namely, that technology will soon provide us with new and expanded capabilities for moving and distributing electronic information in all forms, from place to place.

Additional new capabilities can also be expected to emerge from the intersection of the two major technologies with video technology. Information systems already exist in which a data base of audio and visual information is stored on a videodisc, connected to a microprocessor, and its use time-shared over remote TV displays. A microprocessor enables the user to interact
with a video data base in the same conversational way he now interacts with an on-line computer data base.

Unhappily, there is a serious drawback in bringing video text to the home and that is inadequate line definition. American TV displays, cameras, and related TV equipment are geared to a 525 line standard. This line definition is neither an international standard nor is it sufficient for reading a page of normal text on a TV screen. A sharp, more vivid screen image, requires a TV system with almost double this definition. In an era where electronic and visual media are assuming ever greater importance as sources of information and entertainment, technical advances should certainly be sought and welcomed. But, the issues surrounding high definition television are as much political as they are technical. The technical information advantages to be gained from video technology will not be realized until matters concerning compatibility and economic benefit are thought out and negotiated at the international level.

Conclusion

The conclusion to this essay can be briefly summarized. Over the past forty years, we information scientists have been instrumental in shaping the development of an information processing revolution rivaled in importance only by the agricultural and industrial revolutions. Although we helped instigate this revolution, and once influenced the development of its dominant technologies, this latter role is now largely out of our hands. Our mission today is to perfect information systems, to produce information software, poresprove hohare psodoctivity, and, in the process, to help people realize the societal benefits which the new information technology can provide.

The world's ability to cope with the future will, to an important degree, depend on how well and how rapidly we are able to integrate new information technologies into the mainstream of the country's economic and social life. There is no technology that can do this for us. What's needed is the continuing human touch of the next generation of information scientists.