Office of the Future — Starting Today

Mark S. Trasko
Dynamic Information Systems Corporation
Denver, Colorado

INTRODUCTION

Through the past several decades, computer hardware and software have evolved and expanded to meet our various needs. Scientific data processing, process and manufacturing control, database management are three domains where computers are predominant. In the 80's, a fourth area of application will rapidly emerge: the realm of inter- and intra-company communication, often referred to simply as "office automation." Electronic mail, word processing, automated message systems and time management tools are some of the many functions of office automation systems that will increase productivity.

This paper surveys current activities in the office automation field. It then explores some future directions for this technology, and implementation potential on the HP3000. Finally, the paper focuses on a product, DATADEX/3000, intended to serve as a cornerstone of office automation systems.

THE PRESENT

Currently, office automation tools are available for most computer systems. These products provide a variety of functions, with emphasis in the following areas:

- Word processing, often with access to an on-line dictionary for spelling error detection and correction and word hyphenation.
- Electronic mail facilities to distribute documents that exist on the system (potentially a computer network). Most mail systems provide for verification that a document has been received, replies, the inclusion of comments prior to routing to further destinations, etc.
- Electronic memo systems which are similar in capabilities to the mail systems just described, except that text is generally brief, and is entered interactively by the sender. As with mail systems, most electronic memo facilities allow routing to multiple destinations, adding of comments, etc.
- Electronic scheduling. This includes personal time management tools and the scheduling of meetings and shared resources such as conference rooms and equipment.

The above office automation facilities are available on computer systems provided by several manufacturers, including IBM, DEC, Prime, Wang and others. As of January of 1982, Hewlett Packard had introduced word processing hardware and software. Given HP's publicly announced commitment to the "interactive office," it is likely that they will soon release additional tools in the office automation area. In addition, independent HP3000 software vendors are likely developing products aimed at this market which will soon be available.

Beyond the four categories of commonplace office automation tools just discussed, some more advanced products are available from two vendors that indicate possible future directions for office automation.

Xerox has the Star Work Station, an expensive but powerful system that allows the manipulation of documents and other textual information through the use of icons. Icons are graphic representations of either actual physical devices or physical equivalents of logical entities such as files. For example, a disc file might be represented as a file folder or letter, depending on whether the file contained one or several documents. A Star terminal includes a "mouse" to facilitate easy movement of the cursor among the several icons typically displayed on a screen. Placing the cursor on an icon and pushing a button on the mouse designates the device or entity associated with that icon as the source or destination in an operation.

One icon can father several other icons, similar to a menu screen. Thus, an operator can start with a view of the whole office, then focus all the way down to an individual character in one document. Since common visual associations are used, little training is required before an operator can command a wide variety of operations. However, the high cost of the system may have to be reduced before the savings in training costs justify its widespread usage.

A second advanced office automation facility has been announced by Wang. Called "DVX" (Digital Voice Exchange), it is an audio version of an electronic memo system. DVX is faster and for most people easier to use than textual memos, and may well be more effective since faithful reproduction of the speaker's voice is preserved. Wang is also developing a system to process digitized speech, allowing direct editing of speech with a fair degree of flexibility. However, word processors have far more flexible editing capabilities, so individuals that use a dictaphone almost exclusively and wish to

edit their own speech are the most likely candidates for this system.

The facilities just described are predominant areas of concentration in the office automation field today. Their use will grow rapidly throughout the next several years, and it is likely that HP or third party vendors will provide products in these areas to meet the needs of HP3000 users. The current state of the art in office automation emphasizes productivity and efficiency gains at two levels:

- Making clerical staff more effective.
- Reducing the time spent by management on communications "overhead."

For example, word processing systems provide significant time savings and productivity gains for clerical help, particularly when all or part of a document is used on more than one occasion. These systems also reduce the time spent by management and staff in the interaction required to finalize a new document, because editing, formatting, and printing operations can all be accomplished more quickly.

Similarly, electronic message systems and scheduling facilities reduce the time spent by management and staff in numerous non-productive but otherwise necessary activities. Message systems allow efficient communications within a company, helping to eliminate much of the "telephone tag game" that is usually prevalent. Scheduling systems can greatly reduce the time required for the iterative process of scheduling a meeting, especially if attendees maintain an on line schedule of times they are available.

Generalizations are almost always dangerous. However, viewing office automation as the implementation of various tools to increase the efficiency and reduce the cost of communication seems appropriate. Nearly all the office automation tools currently available deal with the generation, manipulation, or distribution of textual information. (Speech is grouped with written text, since they share a common purpose in office systems.) The one exception, electronic scheduling, still falls in the category of communication in the broad sense. Scheduling systems reduce the communication time required to arrange meetings, and meetings themselves are interactive communication!

THE NEED

Close examination of the current state of the art in office automation reveals a significant deficiency. Available systems deal primarily with "outbound" communications — textual information that will be disseminated. This certainly is a crucial function. All businesses must communicate information to other businesses. Large companies must also communicate information from office to office. And individuals in all but the smallest organizations often must communicate on other than a "live" basis.

However, these tools do not address the equal, or potentially greater, need to manage "inbound" communications. Most communication is bidirectional, with a high likelihood that as much or more information is received as is transmitted. Correspondence, legal documents, periodicals, books, data sheets, brochures, catalogs, all are examples of textual information that usually originates from outside the office area. This information is not only very important, but is received in large and ever increasing quantities. Yet current office automation tools can do little to support the management of this information. They are addressing only half of the need.

What is meant by "manage information?" The implications are the same as in the phrase "database management system." A DBMS organizes data so that a desired subset of the data can be retrieved quickly when required. The larger the database, the greater the need for efficient retrieval capabilities. Similarly, the more textual information a company must manage, the more important rapid access to that information becomes.

Even if the information has been internally generated, or has been received by electronic mail and thus can be stored on-line, the difficulty in managing it is nearly as great as with hard copy documents received from outside sources. Whether the information resides in computer files or file cabinets makes little difference. The key issue that must be addressed is: Can needed information be retrieved quickly and easily, or must exhaustive searches of all documents be performed to retrieve the ones desired?

Unfortunately, database managers such as IMAGE cannot by themselves meet this need. Textual information is excellent for communication, but poorly suited for storage and retrieval using conventional database managers. Text is unstructured and free format, whereas database managers are optimized for storage and retrieval of structured, formatted data. A new strategy is required if information of a textual nature is to be managed.

Before considering such a strategy, a pertinent question should be answered: How great is the need to manage such information? The ability to harness the wealth of information that a company receives on a daily basis, to eliminate duplication of effort in creating, maintaining, and retrieving such information, will provide a company with a significant competitive edge.

Information is important at all levels of a company. Executives and managers must have access to a wide range of information so that they can base decisions on the most accurate, up to date information that exists. The efficiency and productivity of staff personnel depends on fast, flexible access to information pertinent to their activities and responsibilities. Technical professionals function in a constantly changing technological environment. The quality and competitive standing of products that they design are impacted by their ability to access state of the art information in their fields. Successful support of delivered products is heavily de-

pendent on the availability of up to date product information to personnel in the field.

THE INFORMATION BASE CONCEPT

Meeting the needs of those who must have access to information can be termed Information Base Management. Facilities that support this function will be a key cornerstone to future office systems. The ideal information base provides fast, flexible access to the wide range of information vital to a company's operation.

As noted earlier, the scope of information that a company must deal with is very broad. Most of it comes from outside the immediate office area. Periodicals, letters, brochures, data sheets, catalogs, many legal documents, all are examples of such information. Documents that were not created internally or received through electronic mail cannot be maintained on-line, but instead must be physically filed. Again, whether a document is physically or electronically filed matters very little. The crucial issue is whether or not documents containing the required information can quickly be located among the huge store of textual information that a company must maintain.

Thus, a powerful method to index or "key" documents is required in order to ensure retrieval. The viability of the information base concept hinges on the strength of its keyed retrieval capabilities. Specific requirements are discussed in the Implementation section later in this paper. However, assuming that sufficiently powerful keyed retrieval is available, a serious question arises. How should these documents be keyed? By company name? Person name? Subject? Date? Filing by only one key provides little assurance that desired documents can always be found. For example, a file organized by company name is of little value in instances where only a company representative's name is known. One option is to maintain multiple physical copies of the documents, each filed under a different key. But the cost of maintaining multiple copies can be prohibitive, and the likelihood of inconsistencies between the files grows quickly. How does an information base solve this dilemma?

An information base frees a company from inflexible physical filing strategies. A document is physically (or electronically) filed using a unique document number (or name) that is assigned to it. The document is then electronically catalogued in an information base by those keys that provide optimum retrieval flexibility for the type of information being filed. Thus, the document number, several keys, and typically a one or two line document summary might be stored in the information base.

To retrieve a document, it is first located in the information base catalog via the key that is most appropriate. Document summaries would aid in the screening process, especially if several documents qualify based on the key and key value used. The document's unique identification number is then used to retrieve it from the physical file in which it resides.

Documents can be filed by any mix of keys desired; the mix can be varied from file to file. Company name, person name, date, subject, contract, phone number, zip code, author, publisher, product — all are examples of potential keys. No matter how many keys are used, only one physical copy of the document is needed.

Document management is an excellent information base application for several reasons:

- The majority of documents that a company has in its possession are hard copy. Little has been done to date to apply computer hardware and software solutions to the problem of hard copy document management.
- The term "document," as used throughout this paper, is a very broad one. It includes correspondence, periodicals, data sheets, brochures, books, catalogs, contracts, etc. Documents that must be managed originate both from within and without the company.
- The wealth of information contained on hard copy documents that a company customarily receives is virtually limitless. By employing information base management techniques, far more of this information can now be exploited than ever has in the past.

Although document management is huge in scope, it is only one of many potential information base applications. Some examples of other applications include:

- On-line Rolodex-type files.
- Corporate directories, including phone "books."
- Human resource or component-product information bases.

These applications would likely maintain all data on-line, unlike the sample document cataloguing system, which typically maintains the bulk of the data in physical files. Regardless of whether physical files are used or not, there is no logical distinction between the two strategies. An information base is not constrained to be totally resident on one computer system, but can be the integration of several information stores into one logical entity.

INFORMATION BASE IMPLEMENTATION

To implement an information base as conceptualized in this paper, the following conditions must be satisfied:

- Keyed sequential access must be available to entries in the information base. Keyed sequential access includes generic retrieval, ascending sequential retrieval, and if possible, descending sequential retrieval.
- The ability to key (index) information by multiple keys is required. Typically, 3-5 keys are sufficient, but an upper limit of 8 or possibly more might be required in some situations.
- A catalog approach to managing textual data is

needed to manage hard copy documents, and optionally, on-line documents. Use of this technique makes an information base invariant to whether documents are stored on or off line.

- Data security, high performance in multi-user environments, and high reliability must be ensured.
- A convenient query facility must exist for user interface to the information base.

By definition, flexible and powerful retrieval capabilities are the essential element of the information base concept. An information entry must be accessible even when the exact value of its key is not known. For example, a search for a document by the author's name must be ensured of success even if the name is not fully specified. There are several reasons why retrieval by partial key, known as "generic" retrieval, is extremely important:

- The correct, full key value may not be known — frequently the case with names of all types.
- The correct key value may be known, but the information may have been originally entered with the key value incorrectly spelled.
- Lengthy keys are time consuming to enter fully and precisely. Often, the first 4 or 5 characters of a key are sufficient to select the desired entry. For example, "HEWL" is sufficient to uniquely select "HEWLETT PACKARD" if no other company in the information base has a name beginning with those four letters.

Keyed sequential access, which includes both generic and sequential retrieval, is an absolute necessity when information is keyed by names of any type. Without it, information can easily be lost if key values are incorrectly spelled during entry. Correct spelling is a severe restriction with names, because it implies exact punctuation, use of spaces, etc., not just correct spelling of each component of a name. For example, "TRASKO,MARK S," "TRASKO,MARK S.," and "TRASKO MARK S" are worlds apart if generic access is not available. Thus, an exhaustive search of the information base is required to locate entries with misspelled keys unless the incorrect spelling can be exactly guessed. With keyed sequential access, a reasonable guess can be made (generic retrieval) to get close, in alphabetic order, to the desired entry. Then, names can be scanned in forward or backward order until the desired name is found, similar to searching for a name in a phone directory.

The need for keyed sequential access is met by IMSAM/3000, the IMAGE Sequential Access Method. IMSAM, an enhancement to Hewlett Packard IMAGE, provides keyed sequential access, including generic retrieval and ascending and descending sequential access, to entries in IMAGE data sets. An IMAGE database enhanced with IMSAM/3000 serves as an excellent facility for an information base, meeting the first four implementation criteria outlined at the beginning of this section. Entries in an IMAGE database may be keyed by up to 16 items, any number of which can be designated to have IMSAM access. Since IMSAM is totally implemented under the IMAGE umbrella, all of IMAGE's data security, performance, and reliability features are maintained. In addition, existing IMAGE application programs and tools may be used to access the information base.

The fifth requirement for an information base approach to information management is met by Datadex/3000. Datadex, a specialized query facility employing IMSAM, provides a powerful and convenient way to access information in an IMAGE database enhanced with IMSAM. Datadex provides a full set of commands that allow information to be added, deleted, modified and retrieved in several different ways.

Five commands are available to exploit IMSAM access capabilities. The Datadex Find command allows retrieval by partial key, while using any of the five relational conditions (=, >, >=, <, <=) to control the retrieval. For example:

F COMPANY = HEW finds the first company whose name starts with "HEW."

F COMPANY > HEW finds the first company whose name starts with "HEX" or higher.

F COMPANY <= IN finds the first name in descending order (the highest) starting with "IN" or less.

F COMPANY-REP < SN finds the first name in descending order (the highest) starting with "SM" or less.

Unlike Query, which allows similar operations, an entry is retrieved immediately by Datadex, because IMSAM supports keyed sequential access. The capability to do relational Find operations is thus built into the structure of the database. With Query, a Find that does not specify a key value fully or does not use the "=" relational condition requires an exhaustive serial read of the data set. This can take hours to complete on a large database.

The Datadex Next and Previous commands are often used following a Find command. Next and Previous allow the user to browse forward or backward in sequence of any key desired, examining entries one at a time. The List command allows listing of a range of entries, in ascending or descending key sequence, on the line printer. For example, all companies starting with the letters I through M could be listed by:

L COMPANY = I / M for a listing in ascending key order.

L COMPANY = M / I for a listing in descending key order.

The Xfer command works much like the List command, except that entries may be transferred to an MPE file, another database, or the terminal screen, and data is not formatted. This command allows a range of en-

tries to be transferred to a holding file or data set, then reported on using Query or a vendor supplied or user written program.

Datadex provides a turn-key facility that can be used to implement numerous information base applications. All functions required to maintain and access on-line Rolodex-type files, corporate directories, document cataloguing, and many other applications are provided by Datadex. Design an IMAGE database to fit your needs, and Datadex will do the rest.

CONCLUSION

The information base concept will play a crucial role in future office environments. It may well surpass conventional, outbound communications office automation tools in importance. Information bases provide needed information to individuals at all levels of a company. The resultant increases in productivity and efficiency, and the ability to base decisions on the best information that exists, will provide a company with a competitive edge.

Using IMAGE as a foundation, adding the power of keyed sequential access with IMSAM/3000, and employing Datadex as a query facility, an information base can be implemented with capabilities that equal or surpass any facility on any computer system available today.