Some Observations on Piecemeal Electronic Publishing Solutions in the Pharmaceutical Industry†

WENDY A. WARR

ICI Pharmaceuticals, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, England

Received January 18, 1991

The scientist in a pharmaceutical company prepares his compound documents, using a variety of text and graphics systems, in a laborious cut-and-paste procedure, and transmits a hard copy to a records center, where photocopies and microforms are made and indexing takes place. When the end-user wants to retrieve his report, he asks an intermediary to do a search. The result will be a report number or access address. The hard-copy report then has to be manually retrieved, perhaps copied or printed, and delivered by the mailman to the end-user. Assuming that the technical difficulties in preparation and transmission of electronic compound documents can be overcome, what are the management implications for records centers and can end-users cope with full-text retrieval? It is also evident that research end-users are not accessing full-text online databases to any large extent. This paper considers progress toward a more ideal system.

INTRODUCTION

Text information management systems (TIMS) aim to manage not only text but also information in general, i.e., numbers, images, drawings, and photographs. Despite some sweeping claims made for certain TIMS, no one product meets all the information needs of ICI Pharmaceuticals; nor can the variety of products we use be successfully integrated yet to meet *all* our needs. This theme and the related topics of compound documents and electronic publishing are the subject matter of this paper. The term "electronic publishing" is used in its widest sense. The paper discusses both the production of electronic documents in-house and the related techniques used to produce hard-copy books and journal articles which appear also in full-text online databases. The end-users under consideration are primarily those in research departments. I shall not be considering lawyers (who successfully use online full-text databases) or others who need to retrieve full-text, but my comments on office automation do apply to a wider group of end-users than those in research departments.

EXISTING SYSTEMS

The existing "system" is in fact not singular: it consists of piecemeal solutions with a lack of integration. This is not an indictment of ICI but a reflection of the unsatisfactory state-of-the-art across the whole industry.

Compound Document Preparation. The scientist in a pharmaceutical company prepares his compound documents using a variety of text and graphics systems, in a laborious cut-and-paste procedure, and transmits a hard copy to a records center, where photocopies and microforms are made and indexing takes place. When end-users want to retrieve a report, they ask an intermediary to do a search. The result will be a report number or access address. The hard-copy report then has to be manually retrieved, perhaps copied or printed, and delivered by the mailman to the end-user.

We need first of all to look at the various systems used to create and retrieve the information in the compound document.

Word Processing. DEC's All-in-1 office automation system is widely used by managers, research workers, and secretaries at ICI Pharmaceuticals. Broadly speaking, research scientists are not encouraged to do their own typing, so the text is prepared by a secretary, and graphics prepared by a scientist (or patent agent) are pasted in manually. The All-in-1 word processor is WPS+.

† This paper is dedicated to Professor Michael F. Lynch, scholar, teacher, researcher, man of learning and integrity, on the occasion of his Silver Jubilee at the University of Sheffield, Department of Information Studies, 1990.

However, Microsoft Word, WordPerfect, MacWrite, WriteNow, Wordstar, and other PC-based word processors are all in use in various parts of ICI.

All-in-1 is not the only ICI standard. In many ICI businesses, IBM's Professional Office System PROFS is used. This does at the moment make communication among the various ICI businesses more difficult, but there is a long-term strategy, which I will mention later.

Many ICI staff have become accustomed to retrieving documents by using keywords on All-in-1. However the concept of registries, as opposed to old-fashioned filing systems, is not well established. It has been stated[1] that "far more company information is held and retrieved in office filing cabinets than via text retrieval systems". This still holds true in 1990.

Text Retrieval. Filing and text retrieval are often confused as concepts.[2] While filing allows users to find things if they can remember where they put them, text retrieval allows them to look for things that they may not have considered important when the data was first entered.

While the research worker retrieves on keywords in All-in-1 and the manager relies on a secretary who always knows where they filed it, some information professionals in ICI Pharmaceuticals are now using BASISPlus to pull the Reports Collection out of the dark ages of KWIK lists and card indexes. This exercise has, as yet, impinged little on the end-user. Other parts of ICI are still devotees of STAIRS and ASSASSIN, which presumably will not ease the electronic transmission of information throughout the company.

Chemical Structures. The usual imaging technologies are not adequate for chemical structure information systems. In such systems the structure is stored as a connection table, often with two-dimensional coordinates for display purposes. The connection table permits structure and substructure searching and is sometimes referred to as a "computable image". The connection table can be converted into another form for image or vector representation. Some de facto standards for this are discussed later.

Within ICI a wide variety of systems is used for handling chemical structures. The in-house database is handled by Molecular Design Limited's (MDL's) Molecular Access System, MACCS,[3–5] on a VAX cluster. Reaction indexes are handled by ORAC (Organic Reaction Access by Computer).[6,7] Several other software packages are used in computer-aided synthesis design. Scientists working on Quantitative Structure Activity Relationships (QSAR) use the system marketed by Daylight Chemical Information Systems.[8]

The spreadsheeting of generic structures with "R groups"

0095-2338/91/1631-0181$02.50/0

plus data in tables for QSAR is highly desirable but not yet possible in ICI, although Molecular Design Limited, Fraser Williams Computing Services, and others are proposing solutions.

Many software packages have appeared for structure drawing on microcomputers.[9] Of these, ChemDraw, for use on Macintoshes, has perhaps proved the most popular. Although ICI end-users are very enthusiastic about ChemDraw, managers, until recently, had some concerns about it. When ChemDraw was first acquired Macintoshes were considered "renegade" hardware. The situation has now changed to the extent that, in 1991, Macintoshes may well be subject to the same end-user support model as IBM-PC's are. ChemDraw was also originally criticized because it was a stand-alone application. Nowadays ChemDraw files can be imported into DAYLIGHT[8] and STN Express[5] and other links are possible. The fact that chemistry management is somewhat suspicious of time spent on drawing in ChemDraw for internal documentation where high quality is not required is a snag that may well pertain to other structure drawing packages.

Although ChemDraw is no longer "stand-alone" it is still not a full-functioned chemical structure management package because it does not have facilities for building chemical databases and substructure searching them. A few structure management packages have appeared over the last few years.[9] (In fact, one or two have even disappeared.) Only two need mentioning here: PSIDOM[9-11] from Hampden Data Services (HDS) and the Chemists Personal Software Series (CPSS)[5,9,10,12] from MDL. PSIDOM's structure drawing module is known as PsiGen and its chemical database module as PsiBase. The CPSS database module is known as ChemBase and the chemical word processor is ChemText.

ICI Agrochemicals has used ChemBase in conjunction with WordPerfect on a MicroVAX in a prototype system that allows compound documents, containing structures and text, to be transmitted using VAXMAIL.

Neither CPSS nor PSIDOM (in its "native" form) has made much impact on ICI Pharmaceuticals. However, PsiGen is actually in daily use in that it is the structure drawing module of software packages from other vendors, e.g., STN Express[5] from STN International and TOPFRAG[11] from Derwent Publications.

STN Express is the front end of choice for accessing CAS ONLINE[13,14] throughout ICI, although not all of the 200 or 300 end-users[15] yet have the right hardware and training. Unfortunately, it is not possible to download structures from the CA Registry File or Beilstein Online[16] using STN Express. We have published a preliminary evaluation[17] of an alternative, ChemTalk Plus (from MDL), and looked at its successor, the Chemists Access System. Unfortunately these products also do not permit downloading of structures as computable images from databases on STN, although they can produce high-quality compound documents. It is possible to make a ChemBase database of structures downloaded (slowly and expensively) from Beilstein Online on DIALOG, but the quality of the structures is poor compared with the structures on the STN version. (It should be noted that the online database quality and downloading problems are largely outside the control of MDL.) Since STN Express has advantages such as scrolling graphics, and since Chemical Abstracts Service does not cooperate in the development of the Chemists Access System, we continue to recommend STN Express within ICI.

Electronic Document Transmission and Online Databases. ICI as a whole runs in-house current awareness services based on tapes supplied by CAS, ISI, and others. The searches are run on a corporate ICI mainframe, but hits can be transferred to end-users' IBM PC's at ICI Pharmaceuticals. Here they are used to update individual literature reference "card-boxes" in the PC-based Journal Reference Retrieval system. Unfortunately at the heart of the system is the low-technology solution of carrying a tape manually from an IBM corporate mainframe to a VAX at the ICI Pharmaceuticals site.

Most of the searches on online databases run by intermediary information scientists in the Pharmaceuticals business are postprocessed and then transmitted by All-in-1 to the client.

Library and reports collection bulletins are also distributed on All-in-1, but the rather complex New Drugs Bulletin, a compound document of chemical structures and data on competitors' drugs, cannot be produced electronically, let alone linked to the MACCS database which holds the structures. It might be possible to use a product such as ChemText as the publishing tool, but the real problems are the lack of relational database management tools and the need for software to link the structure and property databases to the publishing software. In the meantime, the MACCS database has a very low usage since it contains virtually no data.

Some bioscience end-users have been trained to search a variety of databases on DataStar. Chemistry end-users are fortunate in that just one major database (Chemical Abstracts) meets most of their needs, and this is available in its fullest form on STN. However, chemistry end-users are *not* using the full-text database CJO (Chemical Journals Online) on STN. The reasons for this will be discussed later.

We are only just beginning to use databases on CD-ROM. CD-ROM is almost an irrelevancy in this paper; as a medium it can hold compound documents but the medium in itself is no use without some integrated software solutions. Nor is fancy hypertext software of use to us unless it is compatible with other systems we are using, or so completely user-friendly and vital to our business that we are willing to support yet another system.

THE IDEAL SYSTEM

This section deals with a "realistic ideal" system rather than a blue-sky ideal. Compound documents should incorporate sound (e.g., music and telephone messages) and motion video as well as chemical structures, images, and text. I shall ignore sound and motion video. I shall also not talk about CANDA's, or computerized New Drug Applications, despite the fact that it was ICI who filed the first optical disc NDA with the Food and Drug Administration in 1988.

One description[18] of the post-1995 Records Management (RM) system foresees functions such as:

- Outgoing correspondence automatically transferred to the RM system from wordprocessing.
- Outgoing electronic mail automatically transferred to RM, via wordprocessing.
- Incoming paper correspondence scanned for storage.
- Incoming electronic mail and faxes automatically transferred to RM, via wordprocessing.
- Material from teleconferencing, video, voice mail, etc. captured in a multi-media RM system.
- Automatic indexing and abstracting by intelligent software.

It is a moot point whether all these facilities are "ideals". Certainly I would agree with Ward[1] that:

> "Text retrieval must be recognized as being part of a much larger process of text management. It is simply not enough to be able to identify the existence of a document if one cannot display it, edit it, and transmit it, securely and accurately".

In the ideal system, documents retrieved will be easily revisable, with text and chemical structure editors both available in one system. The chemical structure editor will be compatible with other systems used for substructure searching of internal and external databases. It will also cope with the more exacting requirements of patent preparation, for example:

An ability to cope with very inventive representations of a chemical structure to encompass the full scope of the claims, e.g., circles to represent variable size rings with substituents attached to the ring but not from any defined atoms.

The ability to show all atoms, i.e., all hydrogens and all carbons.

The ability to override the chemical intelligence of a program (for example, valency checking).

The ability to cope with generic substituents, such as R, X, Y, Z, aryl, heterocycle, etc.

The ability to include fragments of a structure in text and in tables.

The ability to draw reaction schemes.

The availability of suitable fonts for subsequent optical character recognition (OCR).

It will be possible to download structures from CAS ONLINE and Beilstein Online into revisable documents or inhouse databases.

The New Drugs database (under MACCS or perhaps its successor) and the related data (in a relational database management system as yet unspecified) will be in an integrated system that allows the end-user to do searches on his desktop workstation and the information scientist to publish a professional-looking hard-copy and electronic bulletin.

Users of both PROFS and All-in-1 will have equal facilities within the ICI Strategic Office System, STRATOS, will be able to transmit documents to each other regardless of location and hardware, and will all have access to an electronic directory.

BENEFITS TO THE BUSINESS OF NEW TECHNOLOGY

A Dataquest survey has shown that scientists spend 75% of their time performing administrative tasks (e.g., on the telephone, attending meetings, writing reports and memos). U.S. companies spend $100 billion a year managing paper, and 75% of all the information collected is never used again.

A survey in a U.S. pharmaceutical company showed that nearly 30% of the R&D personnel budget, or 10 million dollars a year was attributable to publishing activities. A $500 000 investment in electronic publishing technology could be recovered in just one year if only 5% of this time could be saved.

Digital Equipment Corporation, in *Digital CD News*, claims that companies spend 6–10% of their budgets on creating, revising, maintaining, and distributing documents. Of the costs which go to make up the publishing process, printing only represents 10% of the total. Research and creation accounts for 55%; review and approval 20%; support and coordination 15%. Given these figures, too much attention is paid to alternative output media (paper, film, CD-ROM, WORM, etc.) and not enough to all the other elements of the publishing process, they claim.

It is also claimed[1] that the records in a four-drawer filing cabinet cost £15000–£20000 to create and about £700 a year to maintain. A misfiled document can waste time valued at £60, and 2% of all filing cabinet transactions result in misfiling.

Whichever figures one believes, it is obvious that there would be large cost benefits if significant amounts of professional time (scientists, patent agents, information professionals) could be saved.

Research and Intellectual Property Departments (which include Information Services) at ICI Pharmaceuticals produce tens of thousands of unique pages a year. This volume of paper increases to over a million pages once photocopying is included. Unfortunately perhaps 30% of the unique pages are in Information Services bulletins, which have to be produced by cut-and-paste and photocopying techniques because the information is not available electronically.

TECHNICAL CONSIDERATIONS

Changes in the Computing Environment. Many of our problems in the past have resulted from the inability to integrate software, network hardware, and exchange files. Since software applications always lag behind hardware developments, in discussing progress toward the ideal system, we need first to consider changes in the total operating environment.

There is a current trend toward network computing, whereby the local area networks (LANs) of work groups are broadened and integrated into an enterprise-wide, multivendor network. Host and desktop systems reside on a single network. Distributed computing, cooperative processing, and client/server computing (three terms increasingly being used interchangeably[19]) are the computing styles used within enterprise-wide networks. Client/server computer splits the processing of an application between a front-end portion executing on a PC or workstation and a back-end portion running on a server. The front end typically handles local data manipulation and maintains the user interface and the back end handles database and other number-intensive processing.

If the client/server model is to work within a heterogeneous computer environment, standards for networking and database interrogation must be widely adopted. The reader is referred to another paper[19] in this issue, and references therein, for more detail about standards.

The well-known Open Systems Interconnection initiative begun some 13 years ago by the International Standards Organisation (ISO) started by defining a seven-layer model. Much of the preliminary standard-making work was then carried out by bodies such as The Comitē Consultatif International Tēlēgraphique et Tēlēphonique (CCITT), The European Computer Manufacturers Association (ECMA), and the Institute of Electronic and Electrical Engineers (IEEE). Standards for all seven layers now exist. However, the exercise is still not complete after 13 years, and both vendors and users are reluctant to ignore other "standard" architectures such as Transmission Control/Internet Protocol (TCP/IP) or IBM's Systems Network Architecture (SNA). For example, DEC had promised to provide full OSI implementation in its DECnet architecture by September 1990 but failed to do so. In the interim it has embraced TCP/IP. The U.S. Government has mandated the Government Open Systems Interconnection Profile (GOSIP), which has increased emphasis on OSI, but, as with many standards, there is a lengthy battle among manufacturers (who need to keep a competitive edge and advertise "better" or exclusive products), users (who want better services), and governments (who press for deregulation).[20]

It must be remembered that open systems and Open Systems Interconnection are not the same. OSI is one of several possible routes to open systems.[20] It is also not yet clear which standard will emerge as the winner for document exchange. ISO 8613, entitled "Information Processing—Text and Office Systems—Office Document Architecture (ODA) and Interchange Format (ODIF)", is endorsed by the ECMA but is in competition with manufacturers' own standards such as DEC's Compound Document Architecture (CDA), and Digital Document Interchange Format (DDIF), and IBM's Document Content Architecture (DCA). ODA fits into the seventh layer of OSI.

An international initiative started by the U.S. Department of Defense, Computer-aided Acquisition and Logistics Support (CALS) incorporates various standards such as SQL for querying databases and Standard Generalized Markup Language (SGML) for electronic publishing. SGML is also an

ISO standard and has a related Standard Document Interchange Format (SDIF).

The CCITT Group 3 and Group 4 standards for image compression are now part of most document content standards.

Chemical structures can be represented as bit-maps, metafiles, or connection tables. Bit-mapped images require large amounts of computer storage space, and the images cannot be moved, deleted, or restyled.[21] However, for information available only in hard copy, a bit-mapped image is the only option. For electronically available information, a second option is a format known as a graphics metafile. Metafile images can be rotated and scaled-up or scaled-down, but they are not truly computable.

Connection tables permit structure and substructure searching and may be converted to another form for importing into word processing or other applications. Some de facto standards are the Hewlett Packard Graphics Language (HPGL), Encapsulated PostScript (EPS), and Computer Graphics Metafile (CGM).

Standards are also emerging for connection tables. One of the important ones is the Standard Molecular Data Format (SMD).[22]

Generation of 2D coordinates for structure display is a nontrivial problem, but coordinates are often stored in the connection table. SMD format incorporates a standard for storage of coordinates.

Software Developments. There are as yet few applications available for the new client/server hardware.

One operating environment development of particular interest to the field of electronic publishing is the emergence of Graphical User Interfaces (GUI's) such as Windows Version 3 and Presentation Manager, which introduce IBM microcomputer users to the world of the WIMP (Windows, Icons, Mice, Pointers) interface already familiar to Macintosh users.

In the Windows-orientated environment, two or more applications may be open concurrently and files can be exchanged between applications either via the clipboard or directly through Dynamic Data Exchange (DDE). In this new environment the integration of solutions that were previously piecemeal becomes a real possibility.

Town has published an interesting article[19] on chemical information systems in the new "mix-and-match" or "plug-and-play" open systems computing environment.

STN Express is now available in a Windows version.

MDL has embarked on a major distributed systems project formerly known as the Scientists Workstation (SWS) and now called ISIS (Integrated Scientific Information System) and already has some software at the β-test stage. Ultimately, the new system will replace the CPSS/MACCS product line. How successfully MDL can integrate access to online public databases remains to be seen.

MISCELLANEOUS RESEARCH PROJECTS

The British Library's Quartet project[23] has addressed a number of issues including document delivery. They were able to make use of the ADONIS CD-ROM based database of articles in biomedical journals. (ADONIS stands for Article Delivery Over Network Information Systems.[24])

Tuck[23] describes a just-in-time concept for information. There should be no more stockpile of photocopies waiting until needed; no three-week delays in document ordering. You should pay only for what is needed, and research productivity would be increased. The user finds an article he wants, or gets the bibliography checked, by online searching, requests it from the British Library by electronic mail, and receives the document by Group 4 FAX over British Telecom's Integrated Services Digital Network, ISDN. (At the time only an Integrated Digital Access, IDA, prototype of ISDN was available.) Certain parts of ICI are already experimenting with this technology. This is significant because the Group 4 FAX standard has been incorporated into the new international standards that allow compound documents to be transmitted between various devices. Unfortunately, the Group 4 FAX trials at ICI Pharmaceuticals's libraries showed that the technology and associated support services are not yet reliable enough.

There are already some public online systems which include graphics images. At present the transmission of bit-mapped images in DIALOG's TRADEMARKSCAN-FEDERAL and HEILBRON databases is rather slow. There is also the problem that the images can be retrieved but not searched[25,26] (although "design code" searching has recently been implemented). Graphic data for the German patent system, PATDPA (on STN) is scanned in CCITT Group 4 format and converted to outline vector representation using a system called SCORE (Scan Conversion for Outline Representation of Images).[27-29] Text and vectorized graphics are transmitted along the telecommunication network as ASCII data. Transmission of the vectorized data is much more effective than transmission of the original digitized graphics: the compression factor is more than 60%. STN Express can be used to display the graphics on a wide range of screens and printers.[29]

Questel is working on image retrieval for the INPI (Institut National de la Propriêtê Industrielle) database of French trademarks.

A division of The American Chemical Society has studied the challenges associated with the submission of machine-readable manuscripts.[30] The issues involved are much more complex than most authors imagine. Although the ACS Books Department receives 95% of its manuscripts on diskette and has to keyboard very little material, the situation is very different for the ACS Journals Department. Retaining the high quality of the products and reducing turnround time cannot necessarily be done at the same cost. One important factor to be considered is the integration of systems for the production of ACS journals both in hard copy and as full-text files online.[31,32]

Usage of Chemical Journals Online (CJO) has been disappointingly low for the following reasons.[33] A maximum of four years' worth of primary journal issues is available; the database needs to go back further than that. The breadth of information online is not enough: too few journals are mounted (even though publishers other than ACS are involved). Graphics (tables, chemical structures, etc.) are not included, and this greatly reduces the usefulness of CJO to an end-user. CJO is not linked to secondary data files online. The database is aimed at end-users but the marketing has not been directly targeted at them and they are not the biggest group of users of STN. Solving these problems involves a number of technical and economic issues which have been detailed in the literature.[33]

The same developments that ACS claim[32] will make it easier to include graphics in full-text journals online are also of interest to those of us in internal electronic publishing, namely:

- Newer composition systems allowing scanning of graphics at the compositor workstation.
- Higher baud rates.
- Advances in compression of data.
- Enhancement of vectorization techniques.
- Decrease in storage costs.

A new project Chemistry Online Retrieval Experiment (CORE) is currently being run by ACS's R&D group, Cornell University, Bellcommunications Research (Bellcore), OCLC, and Chemical Abstracts Service.[34] ACS is providing CORE with computer tapes of the text of ACS journals. Bellcore is scanning the microfilm issues of the journals and detecting and storing the graphics parts. A complete text plus graphics database will be stored on magnetic media and optical discs. Seven years' worth of ACS journals will be available to 150 Cornell University faculty members and students. The CORE project will evaluate the usefulness of the full-text service to these scientists and will investigate other issues as follows:

How difficult is it to build text/graphic files?

What are the acceptability, advantages, and disadvantages of paper versus electronic delivery of information; bitmapped versus ASCII/graphic-composed page images; Boolean versus coordinate index searching; free text versus indexing for searching?

What will be the network loading?

What are the workstation requirements?

What are the advantages and disadvantages of magnetic or optical storage?

Finally, another research area worth mentioning is that of Optical Character Recognition of chemical graphics. Fraser Williams Computing Services have reported work in this field.[35] Fifty-five thousand chemical structure diagrams in Chapman and Hall's *Dictionary of Organic Compounds* have been scanned by a new program, Recog, to produce parameterized diagrams that can be converted by a further program, Constr, into a connection table based database. Concurrently, a team at IBM's Almaden Research Center have developed a more sophisticated program OROCG for optical recognition of chemical graphics.[36] The program allows documents containing chemical structures to be optically scanned so that both the text and the chemical structures are recognized. The structures are converted directly into MOLFILES (a type of connection table proprietary to MDL) suitable for direct input into chemical databases, molecular modeling programs, image rendering programs, and so on. At least two other teams are known to be working on the digitization of molecular structures.[37,38] This sort of technology opens up new possibilities for producing structure searchable databases from hard copy—rather different from the usual concept of drawing structures and doing publishing with the same structure drawing package.

MANAGEMENT ISSUES

All the discussion so far has assumed that the ideal systems speculated upon are totally desirable. In practice, there are a number of management issues to consider.

First there is the question of who owns the data. Scientists often assume that *they* own the data, but in reality it is likely to be the intellectual property of their employers. Certainly someone needs to be in control of classification, retention policy, security, standardization, and the establishment of house style. The consultancy Strategic Information Management claimed in a recent newsletter that "only 5% of U.K. companies employ records managers and most of these are kept well away from any important work".

The concept of the paperless office raises another issue that records managers understand well. There is still a need for hard copy (or microfilm), and not just for the old-fashioned user or the user who reads on trains and in bed. Certain regulatory authorities and lawyers are not happy to accept optical discs and electronic-only documents.

Information professionals (at least in Europe) tend to assume that they have a monopoly on specialized expertise in the text retrieval area that cannot be found within user departments and Information Systems departments. Mahon[39] claims that data processing staff have, almost with no justification, made themselves indispensable to management, while the text retrieval specialists have not. One is tempted to ask *why* the text retrieval specialists have not made themselves indispensable to management. Are they really indispensable?

I would certainly assume that the information science manager has a better feel for the external world (including external information sources, and competitors's activities) than the average DP/MIS manager, whose department tends to be much more introverted. However, the DP/MIS people have control of that vital asset, the enterprise-wide communication network.

DP/MIS people are not convinced that information professionals can design or build databases under BASISplus or like systems. One of the appeals of Paralog's TRIP has been the ease with which end-users can build applications without the need for many man-years of support and maintenance from MIS professionals. Certainly problems such as our New Drugs system are constantly delayed by the need for justification of the resource required in terms of MIS expertise.

The move away from PC-based systems with users seemingly in control (but corporations not in control) toward distributed computing, will in fact give *more* power to MIS departments,[40] although the end-user will be given more computer power at his desk or workbench.[19] Jeffries[40] claims that application generation will require more computing expertise than was needed previously and the complexity of the operations will mean that senior management insist on the involvement of personnel with the data processing skills appropriate for running distributed systems.

Surely all this, together with the current economic climate in the pharmaceutical industry, signals the time when information science departments, end-users, and MIS departments must strive above all to *cooperate* with each other in the strategic interests of the company. Information professionals cannot rest upon their laurels or console themselves with the mystique of text retrieval expertise. They have to make a positive acceptance of their change of role and work toward a new future. This is, of course, not easy in the current climate of change, where management is striving to achieve the appropriate organizational structures to handle the eternal triangle of the increasingly computer-literate end-users, the MIS departments who supply them with systems, and the information department caught as the intermediaries who often take the blame.

CONCLUSION

In this paper I have described many problems and not many solutions. I have outlined some features of an ideal system and mentioned some standards and some changes in the computing environment which herald an era of "mix-and-match" or "plug-and-play". The launch of Windows Version 3 has been a major step toward this era but it does not offer true multitasking as OS/2 eventually will.

In the brave new world of chemical information at the desktop, integrated solutions will replace piecemeal solutions. Cut-and-paste techniques and computing department mystique will be a thing of the past. The user will be in control. Chemical structure editors will handle a greater variety of structures and will link much more transparently to text editors. Fast and accurate transmission of compound documents will be possible and high-quality hard copy will be generated when needed. There will be much better links between in-house chemical information systems and online ones, allowing downloading from public databases and easy preparation of compound documents in-house. (There will also be trend toward *storage* of commercially available databases in-house as an alternative to online systems.)

The possibilities sound exciting and should be ultimately achievable. However one should be very cautious in thinking about time scales. We have waited a long time for document architecture standards; for true multitasking under OS/2; for all the possibilities of open systems. Promises have abounded for at least two years, but I suspect it could be at least another two years before real integrated solutions can be implemented.

What I have not covered is the issue of major vendors who will not cooperate with each other to produce the solutions the end-user wants. To what extent will we be able to achieve what we want in the brave new world of open systems, despite the unwillingness of certain major players to work together? This could be the subject for another paper.

ACKNOWLEDGMENT

The ICI systems mentioned in this paper have involved so many people that they cannot all be named here. It must be recognized that the work was a team effort and I am merely the mouthpiece. I would however like to single out Mr. Frank Loftus for his work on STN Express, Dr. Martyn Wilkins for his work on CPSS and the Chemists' Access System, Mrs. Denise Ledgerwood for her list of requirements for chemical structure drawing in patents, and Mrs. Margaret Hayter for advice on records management and information bulletins. The opinions expressed in this paper are, however, my own. Finally, I would like to thank Dr. William Town of Hampden Data Services for much advice on network computing and for reading and criticizing an early draft of this paper.

REFERENCES AND NOTES

(1) Ward, S. E. Text Retrieval—Its Place in Information Technology. In *Text Retrieval the State of the Art*; Gillman, P., Ed.; Taylor Graham: London, 1990; p 5.
(2) Clough, R. Text Retrieval: Technology and Marketplace, Reality and Hype. In *Text Retrieval the State of the Art*; Gillman, P., Ed.; Taylor Graham: London, 1990; p 93.
(3) Anderson, S. Graphical Representation of Molecules and Substructure Search Queries in MACCS. *J. Mol. Graphics* **1984**, *2*, 83–90.
(4) Chemical Structure Search Systems and Services. In *Communication, Storage and Retrieval of Chemical Information*; Ash, J. E., Chubb, P. A., Ward, S. E., Welford, S. M., Willett, P., Eds.; Ellis Horwood: Chichester, 1985; Chapter 7, p 182.
(5) Kasparek, S. V. *Computer Graphics and Chemical Structures*; Wiley-Interscience: New York, 1990.
(6) Johnson, A. P.; Cook, A. P. Automatic Keyword Generation for Reaction Searching. In *Modern Approaches to Chemical Reaction Searching*; Willett, P., Ed.; Gower: Aldershot, 1986; p 184.
(7) Johnson, A. P. Reaction Indexing: an Overview of Current Approaches. In *Chemical Structures: The International Language of Chemistry*; Warr, W. A., Ed.; Springer-Verlag: Berlin, 1988; p 297.
(8) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: an Integrated Tool for Computer-Assisted Molecular Design and Pharmacophore Recognition from Geometric, Steric and Substructure Searching of Three-Dimensional Molecular Structures. *J. Comput.-Aided Mol. Des.* **1989**, *3* (3), 225–251.
(9) *Chemical Structure Software for Personal Computers*; Meyer, D. E., Warr, W. A., Love, R. A., Eds.; American Chemical Society: Washington, 1988.
(10) Meyer, D. E. User of Microcomputer Software to Access and Handle Chemical Data. In *Chemical Structures: The International Language of Chemistry*; Warr, W. A., Ed.; Springer-Verlag: Berlin, 1988; p 251.
(11) Town, W. G. Microcomputers and Information Systems. *Chem. Br.* **1989**, *25* (11), 1118–1120.
(12) del Rey, D. Applications of Personal Computer Products for Chemical Data Management in the Chemist's Workstation. In *Graphics for Chemical Structures Integration with Text and Data*; Warr, W. A., Ed.; ACS Symposium Series 341; American Chemical Society: Washington, DC, 1987; p 48.
(13) Farmer, N. A.; O'Hara, M. P. CAS ONLINE: A New Source of Substance Information from Chemical Abstracts Service. *Database* **1980**, *3* (4), 10–25.
(14) Rich, C. Searching STN International. *Database* **1986**, *9* (5), 116–120.
(15) Warr, W. A.; Haygarth Jackson, A. H. End-User Searching of CAS ONLINE: Results of a Cooperative Experiment Between Imperial Chemical Industries and Chemical Abstracts Service. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 68–72.
(16) *The Beilstein Online Database*; Heller, S. R., Ed.; ACS Symposium Series 436; American Chemical Society: Washington, DC, 1990.
(17) Warr, W. A.; Wilkins, M. P. Graphics Front Ends for Chemical Searching and a Look at ChemTalk Plus. *Online* **1990**, *14* (3), 50–54.
(18) Freeman, G. Some Likely Developments in Information Technology and Records Management. *Rec. Manage. J.* **1990**, *2* (2), 44–49.
(19) Town, W. G. Integration of Microcomputer and Mainframe Information Systems. *J. Chem. Inf. Comput. Sci.* **1991**, *31* (2), 176–180.
(20) Tuck, B. OSI and Library Services. *Library and Information Briefings 17*. British Library R&D Department, 1990.
(21) Barstow, J. F.; del Rey, D.; Laufer, J. S. Problems and Solutions in Generating Scientific Manuscripts for Publication. *Am. Lab. (Fairfield, Conn.)* **1988**, *20* (7), 82–85.
(22) Barnard, J. M. Draft Specification for Revised Version of the Standard Molecular Data (SMD) Format. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 81–96.
(23) *Project Quartet*; Tuck, B., Archer, D., Hayet, M-C., McKnight, C., Eds.; British Library R&D Report LIR 76, 1990.
(24) Campbell, R. M.; Stern, B. T. ADONIS—a New Approach to Document Delivery. *Microcomput. Inf. Manage.* **1987**, *4* (2), 87–107.
(25) Thompson, N. J. DIALOGLINK and TRADEMARKSCAN-FEDERAL: Pioneers in Online Images. *Online* **1989**, *13* (3), 15–26.
(26) Chadwick, T. R. DIALOG Comments on Imaging Capabilities: an Interview with Fred Zappert. *Online* **1989**, *13* (3), 28–30.
(27) Tittlbach, G. Electronic Publishing and Document Delivery of German Patent Information. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 13–17.
(28) Niedermeyr, W. Integration of Technical Drawings in a Databank System. In *Graphics for Chemical Structures: Integration with Text and Data*: Warr, W. A., Ed.; ACS Symposium Series 341; American Chemical Society: Washington, DC, 1987; p 143.
(29) Detemple, W. Future Enhancements for Full Text Databases. *Online Rev.* **1989**, *13* (2), 155–160.
(30) Brogan, M.; Garson, L. R. Requirements for and Challenges Associated with Submission of Machine-Readable Manuscripts. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 271–277.
(31) Martinsen, D. P.; Love, R. A.; Garson, L. R. Multiple Use of Primary Full Text Information—a Publisher's Perspective. *Online Rev.* **1989**, *13* (2) 121–133.
(32) Hearty, J. A.; Rohrbaugh, V. K. Current State of Full Text Primary Information Online with Recommendations for the Future. *Online Rev.* **1989**, *13* (2) 135–140.
(33) Hearty, J. A. Primary Information Online: Today's Problems, Tomorrow's Solutions. *Inf. Serv. Use* **1988**, *8*, 93–105.
(34) Garson, L. R. Private communication, 1990.
(35) Hyams, P. OCRing Chemical Graphics. *Inf. World Rev.* **1988**, *July*, 18.
(36) Boyer, S.; Casey, R.; Miller, A.; Zilles, K. Optical Character Recognition of Chemical Graphics. In *Chemical Structures 2* (Proceedings of the Second International Conference on Chemical Structures) Springer-Verlag: Berlin, 1991; (in press).
(37) Johnson, A. P. Private communication concerning work at the Maxwell Research Institute, Leeds, 1990.
(38) Contreras, M. L.; Allendes, C.; Elvarez, L. T.; Rozas, R. Computational Perception and Recognition of Digitized Molecular Structures. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 302–307.
(39) Mahon, B. Information Technology—IT's View of Text Retrieval. In *Text Retrieval The State of The Art*; Gillman, P., Ed.; Taylor Graham: London, 1990; p 14.
(40) Jeffries, A. Distributed Text Databases: Issues and Potential Applications. In *Text Retrieval The State of The Art*; Gillman, P., Ed.; Taylor Graham: London, 1990; p 112.