10

DATA RETRIEVAL AND USE OF ICT IN RESEARCH

O. A. Fakolujo

PREAMBLE

A wide range of developments in ICT, covering hardware, software and networking technologies, bring about ongoing changes in the science system and research methodology. They include significant improvements in computing power and storage capacity and better networking and search technologies. Such developments have allowed scientists and researchers to make rapidly increasing use of Internet and ICT tools. However, there are concerns that the Internet is becoming inadequate for certain scientific purposes. Several governments and universities have recently taken initiatives to develop faster networking technologies to meet the needs of science. ICT-related changes underlying the evolving science system have three main sources:

- Technological change in the ICT industry (mostly driven by needs unrelated to science)
- Scientists' and Researchers' efforts to develop their own tools; and
- Government programmes specifically designed to foster developments in ICT and apply them to scientific needs, e.g. the US High-Performance Computing and Communications (HPCC) programme.

The application areas of ICT include:

- Research activities
- Job search
  a. You can reach deeper into your local area as well as take your search far beyond your regular boundaries
  b. You can access current information at all hours of the day or night
  c. Using the Internet in your search demonstrates leading-edge skills
  d. The Internet lets you meet new people and initiate new relationships with others in your profession or region
  e. The Internet can help you explore career alternatives and options that you maybe haven't considered
- Fund Raising: securing grant
- For marketing

MAIN TECHNOLOGICAL DEVELOPMENTS

Conventional computers solve problems by performing programme instructions one at a time in a strict sequence. Electronics manufacturers have provided users with increasing computing power at decreasing cost for many years, essentially by squeezing a greater number of ever smaller transistors and other components onto chips, thanks to continuous advances in lithographic techniques (*Science*, 1996). They have also found alternative ways of getting more processing power from computers, for example by using reduced instruction set computer chips (RISC) or special-purpose chips to perform designated tasks faster than a general-purpose processor (*New Scientist*, 1996).

The supercomputers used for research generally have custom-made, expensive processors which provide better performance. Up to the late 1980s, vector supercomputers were the most powerful computers. They were the only option for researchers with truly large problems, who used them to perform calculations simultaneously on long strings of numbers, i.e. vectors (Pool, 1995). The research potential of parallel computers has been demonstrated more recently. These multiprocessor machines break major programming tasks down into smaller problems which they solve simultaneously. This method remains quite difficult, however, and cannot be used to solve all problems.

Cheaper off-the-shelf components and software have generally contributed to increased use of information technology. A new generation of extremely powerful off-the-shelf commodity chips is also at the heart of an emerging standard parallel architecture (Matthews, 1996). Even working on their own, these chips attain speeds of up to 200, 300 or 600 million flops (floating point operations per second). Many off-the-shelf components are also available for certain scientific instruments. Plug-in circuit cards allow new features to be added to personal computers (PCs) without much adjustment. Complex software, increasingly available for Windows Operating System (OS), contributes to the use of technology by non-specialists at lower cost. Various storage and information delivery technologies continue to co-exist.

Traditional storage systems such as the CD-ROM (compact disk – read only memory) are still being used by publishers, particularly where current Internet access limitations would result in very slow access, e.g. when the package contains great quantities of data. New products that combine CD-ROM data with information on the World Wide Web (WWW) or online services give publishers the opportunity to deliver huge amounts of data on CD-ROM and then use the Internet to offer updates or transactions. The Digital Video Disk (DVD) can store seven times as much data as a CD-ROM and deliver a moving picture quality that outshines laser disks. It is particularly useful for multimedia publishing and will enable educational software, in particular, to incorporate more video. The mass-storage industry continues to develop technologies that can handle increasing quantities of data, thereby satisfying the needs of scientists carrying out large-scale simulations, experiments and observation projects.

Electronic networks constitute the infrastructure which provides scientists with new means of communication that give them access to data, information and software in cyberspace, allow them to share and control remote instruments, and that link distant learners to virtual classrooms and campuses. Scientists currently have access to various types of networks, including campus, national and international research networks, which are increasingly interconnected. The main network, the Internet, began in the late 1960s as a network providing a limited number of researchers with shared interactive communication among computing systems at different locations. It has become a network of networks that can be accessed by anyone with a computer and a modem. Since 1991, the WWW has been a very powerful and convenient way to navigate through the world's collection of networked computers. Through hypertext links, it connects information on the network to other sites. Special graphical interfaces known as Web browsers, such as Netscape Navigator, Microsoft Internet Explorer and Netcom NetCruiser, allow users to read hypertext.

Rapid advances in computing power and the explosive growth in network connectivity have generally enhanced the use of distributed systems. The potential of networking several computers to perform tasks similar to those performed by supercomputers is also being tested. In addition, systems capable of co-ordinating different types of computers, including traditional supercomputers, parallel computers, workstations and PCs, are emerging (Economist, 1996). Hardware and software for using a network of workstations as a distributed computing system on a building-wide scale are being developed (National Science and Technology Council, 1995).

The development and use of digital data and information rely on a broad range of technologies. Non-digital data requires data acquisition technology, such as optical character recognition, while direct use of data collections requires database management systems. Text analysis and information retrieval techniques (including text, index and image compression, indexing, routing, filtering and visualization techniques) sometimes enhanced by artificial intelligence, are needed to index, search, retrieve and present information. Furthermore, data mining technologies can be used to filter large amounts of data for useful patterns. Methods for handling information help users more effectively search, learn about, organize and use data and information. Search tools, for example, can go through millions of articles from current and back issues of electronic journals in almost any discipline. They help users navigate online services and save time. Search engines such as Altavista, Excite, Infoseek, Lycos, Web 12 Crawler and Yahoo constantly tunnel through and catalogue Internet documents. There are also limited area search engines that index only Internet resources relevant to a specific subject and thus raise the speed and efficiency of searches. Internet search technology is still, however, in its infancy.

Many ICT applications used by scientists, such as access to databases, information services and e-mail, were originally based on narrowband technologies; broadband technologies were only needed for video applications. However, the growth of the Internet and new interactive – often multimedia – applications has led to a rapidly growing demand for high bandwidth technology, which may also be needed to process large amounts of data.

COMMUNICATION AMONG RESEARCHERS

Researchers have used ICT-based communications – or the Internet – mostly as a natural extension of other communications tools. Apart from greatly enhancing the quantity, quality and speed of communication among researchers, ICT use has also had various effects on the organization of research work. Collaboration patterns have changed, research work base has widened as more researchers are able to participate, and the hierarchies have sometimes been affected.

The Growth of Collaborative Arrangements

Improved communication due to ICT may contribute to an increase in the size of professional networks. For example, among oceanographers, intensive e-mail users report larger professional networks. In biology, chemistry, mathematics and physics, collaborations have also increased in size, apparently in association with the use of ICT. In experimental particle physics, the Internet has facilitated experiments in which a large number of people collaborate effectively. A more significant change in the organization of research has been the increase in remote collaboration, particularly at international level. Computer networks have reduced the need for co-workers to be at a single location. Consequently, a new form of research work has emerged, the "extended research group". This is typically a large, unified, cohesive, co-operative research group that is geographically dispersed, yet co-ordinated as if it were at one location and under the guidance of a single director. It provides access to colleagues and to equipment, software and databases that are traditionally part of laboratory organization, without regard to geography. These "collaboratories" rely heavily on ICT for coordinating their work.

E-mail over the Internet enables researchers to overcome many barriers to communication due to geographic distance, such as time, costs and language. The main requirement is that all members of the group have Internet addresses. E-mail was preferred to the telephone because scientists who travel may be hard to reach by phone, but can be contacted at their virtual address, because written messages allow time for formulating answers before responding, and because colleagues whose native language is not English preferred written communication. Therefore, E-mail is considered next best to face-to-face interaction and a good medium for facilitating collaboration among researchers. However, many researchers emphasize the importance of establishing common understanding of the research problem through intensive, face-to-face interaction before engaging in computer-mediated collaboration. For example, the modiago scholarship presently has two students from the Department of Electrical / Electronic Engineering, University of Ibadan (Nigeria) and two others from another country working on the same project yet not debarred by geographical distances.

With closer links among geographically dispersed researchers, the international community of scholars is becoming denser. For a given research topic, ICT allows the creation of more complex work groups with more fluid structures. Virtual research teams can be formed and link a variety of researchers, each of whom contributes his or her skills to the project. Projects take advantage of networks to obtain access to the precise skills needed, and researchers gain access to projects that demand their skills. As a result, the research topic, rather than geographical proximity, determines collaboration decisions.

Effects on Status and Hierarchy

ICT-based communication can lead to greater decentralization or less difference in status, because interaction over the Internet provides fewer clues to status, rank, and gender than face-to-face or even mail or phone communication (Walsh, 1997). Group decisions are consequently less influenced by the status of those proposing particular solutions. Moreover, by its informal nature, e-mail reduces lower-level researchers' caginess about contacting higher-level ones. It may facilitate the creation of new ties among remote collaborators and give researchers with lower status easier access to their more eminent colleagues with whom they may eventually publish results jointly. On the other hand, it may create even greater disparity in publication rates as top researchers become attached to a greater number of research projects via e-mail contacts. E-mail communication may also allow researchers who previously lacked the access necessary to keep up to date to become active participants and possibly core researchers. So far, no significant correlation has been found between age or institutional prestige and ICT use as a predictor of productivity (Cohen, 1995).

To the extent that status distinctions remain however, individuals with high status will continue to exert more influence on group decisions. As the technology has been developed, more status cues are being inserted into the communication. E-mail addresses, for example, are evolving from a nondescript assembly of letters and numbers to a combination of family name, institution or company, and country of registration. Also, other mechanisms for introducing the status-reinforcing procedures of earlier communication technologies (mail, telephone) are beginning to appear. For example, high-level researchers increasingly use gatekeepers to screen their e-mail just as they screen letters and calls. Similarly, if ICT violates existing work norms or status distinctions, it may not be used. New technology can also change part of the basis for existing status distinctions. ICT can, for example, enhance the status of younger colleagues who are more familiar with the latest technology. It may also provide peripheral researchers with wider access to crucial resources – such as computing facilities, software or databases – which have traditionally been unequally distributed. Improved access could reduce the gap between more and less eminent researchers.

In general, ICT has allowed more researchers to have access to the latest information and thus remain up to date. This has been particularly meaningful for those at less prestigious institutions. However, there is a significant difference between having access and being present. Researchers at top institutions have access to oral information and seminars as well as research papers. They also have access to specialists who know which information and papers are important. The filtering provided by local and informal communication is an important part of the process of finding research information. Researchers at large institutions usually also have better access to funding and equipment.

Generally, while ICT helps improve access to information, it does not overcome disadvantages due to a lack of direct contact with top scholars in the field. ICT use may thus lead more to a broadening of the resource base than to a change in the hierarchy of institutions. While ICT can be used to provide broad access to resources, it can also be used to limit access. Netnews bulletin boards are generally open to many users and are used to announce new findings, discuss substantive issues, and receive answers to questions from unknown colleagues. More field-specific distribution lists may be announced through direct contact with existing research ties, thus enabling a more specialized exchange of information. Some fields seem to have more potential to benefit from technology than others.

ACCESS TO INFORMATION

Rapid advances in ICT have made it possible to handle digital data and information in large volumes at ever-increasing speeds and have resulted in sharp reductions in the cost of storing, filtering, processing, compressing, and retrieving data for interpretation and retransmission. ICT has increased researchers' ability to access information by supplying them with increasingly powerful tools at decreasing cost, thus enabling new ways of working. Researchers have frequently been the first to use ICT in a new or comprehensive way, as in the case of the Internet. On the whole, this has significantly improved the efficiency of information -based work.

Digital Resources for Researchers

In the past, traditional libraries held the keys to research and knowledge. Today, "digital libraries" store and manipulate large collections of material in electronic form. The development of digital libraries is closely linked to that of network information systems, which increasingly allow access to resources when and where users desire it. Prodigious quantities of general and sector-specific information are now available off-line on CD-ROMs and online, increasingly over the Internet. With ICT, access to this information can already be obtained at low incremental cost. As systems become more sophisticated, users will benefit from a growing capacity to navigate among information resources at low cost.

Databases

The value of scientific and technical databases to research organizations continues to increase. Estimates suggest that both the amount of data they contain and their total number expand by about 10 percent a year. Internet tools, in particular, have made information more readily available to a growing base of scientists and engineers, as database service providers have started moving to Web-based systems. Web browsers such as Netscape are excellent database interfaces; their broad acceptance has extended the potential user base to the research community (R& D Magazine, 1997).

Several factors help make data sets collected by scientific projects available to broad communities of users. Since the tools used to collect, transmit, and analyse data generate or require digital signals, the data are already in digital form and are therefore easily communicated over digital networks in a timely way to researchers world-wide. Furthermore, when researchers have public support for major research projects, they are encouraged to disseminate data widely so as to maintain that support.

Digital Library Initiatives

There are vast numbers of projects for developing digital libraries. They currently focus on issues of access costs and digitization technology. The key technological issues, however, are how to search and display desired selections from large collections on the Internet. Research on digital libraries concentrates on how to develop the necessary infrastructure to manipulate effectively the massive amounts of internet information (Computer, 1996).

Many traditional libraries which are not yet involved in large-scale digitization of publications are nevertheless increasing their holdings of electronic documents. These can be powerful tools for research and may reduce subscriptions to printed publications by enabling electronic access to other libraries' holdings. A "free-rider" problem may arise, however, if all libraries follow this policy. Access to unpublished student research stored at universities is generally limited, thereby reducing the transfer of knowledge contained in unpublished scholarly work. It is estimated that over 10 percent of all research performed in the hard sciences each year had already been done. Providing electronic access to this data source might improve scientists' productivity by enabling them to focus on the appropriate issues.

The Digital Library Initiative (DLI) is the flagship research effort of the US National Information Infrastructure Initiative. Ultimately, the digital library would involve an entire network of distributed repositories where objects of any type can be searched within and across indexed collections. With its many partners and large testbeds, the DLI is structured to encourage technology transfer. Once the DLI has stimulated basic research in various enabling technologies and enabled several digital library testbeds, it is expected that IT companies, traditional libraries, publishers, organizations, and users will join forces to develop the knowledge repositories that will play an essential role for all of society in the 21st century.

EDUCATION AND TRAINING OF RESEARCHERS

ICT contributes directly to teaching, learning, and research and provides a support function to researchers by enabling access to digital libraries, archives, databases and information services. ICT can have positive effects on learning by opening up access to educational resources, by supporting the learning process, and by supporting skill development. However, this requires efficient planning, and learners, teachers, and institutions that are willing and able to adapt.

The enhanced use of ICT in teaching may also help to improve academic productivity, thus enabling researchers to spend more time on research. Researchers may also need better education and training to use ICT efficiently for research work. ICT opens up access to education by removing many of the temporal and spatial constraints to information and knowledge. Furthermore, the availability of learning materials based on ICT can greatly improve learning resources. Computers support the learning process by helping to create a student-centred rather than a teacher-centred environment, one which is more flexible and adaptable to individual needs. Working groups formed around the computer can also help prepare learners for a world in which many problems are addressed by teams. Nevertheless, this potential can only be realized with high-quality software and significant efforts by all those involved (OECD, 1997).

Efficient use of ICT allows students to develop the kinds of skills and competencies that many educational reform panels have viewed as essential (OECD, 1997). Basic skills such as arithmetic can be mastered with computer-aided drill and practice, while writing skills can be developed with word processing, which makes writing and revising easier. A deeper understanding of complex scientific concepts in mathematics and science - particularly where experiments are not feasible or are dangerous - can be gained through computer simulations. Last but not the least, the use of these technologies for learning may establish familiarity with technologies that are increasingly needed by individuals in a technology-driven society.

BIBLIOGRAPHY

COHEN, J.A. 1995. *Computer Mediated Communications and Publication Productivity among Faculty in Association of Jesuit Colleges and Universities (ACJU) Institutions*. Doctoral dissertation, State University of New York at Buffalo. 78

COMPUTER. 1996. Digital Library Initiative, May, Los Alamitos, California.

ECONOMIST. 1996. Supercomputers: Divide and Conquer, 14 December, p. 94.

MATTHEWS, R. 1996. Truly, Madly, Deeply Parallel, *New Scientist*, 24 February.

NATIONAL SCIENCE AND TECHNOLOGY COUNCIL 1995. *HPCC: Foundation for America's Information Future*, Washington, DC.

NEW SCIENTIST 1996. Simply Faster, 9 March.

OECD 1997. *Information Technology Outlook 1997*, OECD, Paris.

POOL, R. 1995. Supercomputers: Off-the-shelf Chips Conquer the Heights of Computing, *Science*, 8 September.

R&D MAGAZINE 1997. Database Service Providers Move to Web-Based Systems, June, pp. 69-72.

SCIENCE 1996. Can Chip Devices Keep Shrinking, Vol. 274, 13 December.

WALSH, J.P. 1997. Telescience: The Effects of Computer Networks on Scientific Work. *Consultancy paper for OECD*, University of Illinois at Chicago, April, mimeo.