USING TEXT PROCESSING, COMPUTER NETWORKING AND SATELLITE TELECOMMUNICATION TECHNOLOGIES TO PUBLISH PRIMARY SCIENTIFIC AND TECHNICAL INFORMATION

John R.U. Page
*International Institute for Applied Systems Analysis, Laxenburg, Austria*

RR-82-41
November 1982

Research Reports, which record research conducted at IIASA, are independently reviewed before publication. However, the views and opinions they express are not necessarily those of the Institute or the National Member Organizations that support it.


Copyright © 1980 ECSC, EEC, EAEC, Brussels and Luxembourg.

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage or retrieval system, without permission in writing from the copyright holder.
In the past year there has been an upsurge of interest in the problems of scientific and technical communication at the personal level within and between less developed countries and between them and developed countries. Concern over these problems was expressed in the so-called Declaration of Mexico of 1981. The Canadian International Development Research Centre called together a small working party to consider this document, clarify the issues and, at least, define further actions to be taken to reduce the severity of the problems. The Information Technology Task of the Management and Technology Area of the International Institute for Applied Systems Analysis (IIASA) was represented by two participants in these discussions.

This was appropriate, as IIASA's Information Technology Task is concerned with the impact of this very rapidly developing technology on societal systems. One such system is the production, structuring, and ordered distribution of scientific technical and other related types of information, a key element in research, development, and planning. Furthermore, John Page, one member of the Information Technology Task’s group had considerable experience and expertise of such systems and, indeed, had written about them.

It is timely and appropriate, therefore, to issue this paper in which Mr. Page discusses the possible impacts of a systematic combination of fully electronic publication technology, communicating text processors and satellite telecommunications on the information generation and distribution system.

ALEC M. LEE
Chairman
Management and Technology Area
Using text processing, computer networking and satellite telecommunication technologies to publish primary scientific and technical information.

John R.U. PAGE

Computer Science Department,
International Institute for Applied Systems Analysis (IIASA),
Laxenburg, Austria

Abstract

The concept discussed in this paper rests on the premise that the conventional system fails to meet the needs for rapid exchange of the results of research and development in certain areas of applied science and technology: conventional publication involves several handlings of the material, is labour-intensive and is therefore slow and expensive. The paper suggests that significant improvements could be made by the use of compatible text-editing systems by authors, referees and publishers, interconnected over computer networks. By creating a master digital record at the outset, many new output options are possible in addition to the conventional journal or report. Examples are new forms of current awareness publications, in the shape of personalized journals and the like, and immediate information retrieval techniques leading to delivery of full texts via the network. To pass full texts through the network will require the use of high-capacity satellite telecommunications, employing small Earth stations. Finally, it is suggested that a programme of studies should be carried out, in association with actual experiments using the European Space Agency’s Orbital Test Satellite.

Background: the general concept

The aim of this paper is to present some ideas on the marriage of text processing, computer networks and satellite telecommunications to provide a system for publishing and disseminating scientific and technical information. The basis of the concept is that, where appropriate, texts should be prepared and edited using a standard word processing system. Collaborative publications by several authors not resident in the same locality could be prepared using computer networks (either terrestrial or
satellite based) to exchange texts and comments. Refereeing, whether formal or informal, could be dealt with in the same way. The master of the final text would be made available for "publication" in various forms according to need; the bandwidth of the satellite communications system would enable copies of this master to be transmitted in original digital form to different locations for local processing.

Such a concept is admittedly revolutionary, and it must be emphasized at the outset that it is not proposed as a total alternative to traditional forms of scientific and technical publication, i.e. journal articles, monographs, proceedings, and report series and other forms of "grey" literature. The extent to which an all-electronic publishing system could become a complementary or alternative system to present-day methods cannot as yet be forecast: this would require an in-depth economic and operational analysis, and many of the factors cannot yet be quantified.

There are however some pointers which begin to make such a system look attractive for certain classes of scientific and technical information. In 1974, D.B. McCarn of the National Library of Medicine (1) pointed out that the costs of normal publication and distribution were increasing by at least 6% a year, while the cost of data handling and storage of the same material on a computer was falling by perhaps 30% a year. If these trends continued, McCarn concluded, at some point in the relatively near future a cross-over point would be reached at which the routine production and distribution of a formal printed version would not be cost-effective compared with its computer-stored counterpart.

McCarn was referring to the relative cost-effectiveness of consulting the printed Index Medicus as compared with the use of a terminal to retrieve references from the MEDLINE system, but the same kind of relationship would in principle exist in the case of certain types of primary literature. By and large, these trends have continued since 1974, although technological innovation in printing and publishing may have had some effect in holding down the rate of cost increase in that area.

The need for change

The rising costs of conventional publication and distribution appear to be having undesirable effects on the process as a whole, viewed as a means of dissemination of the results of research and development in science and technology. Rising costs have led to increasing subscription fees for journals; library budgets have not increased to the same extent, and in consequence, circulation figures, and therefore revenue, have also tended to fall. Some publishers' organizations claim that the increase in interlibrary loan services, particularly photocopying, has also contributed to this result.

Nevertheless, the volume of material for publication, in one form or another, is apparently still growing. Outside the needs met by monographs and report series, the volume of scientific and technical publication is spread over a very large number of journals and periodicals: one major document supply organization subscribed to over 48,000 journal titles in 1976/77, from which it was able to satisfy over 95% of requests(2).
In spite of the existence of several major serials in many of the main scientific disciplines, a very large number of journals have relatively small circulations. Consequently, it does not seem that the large volume of demand for publication can result in economies of scale.

Moreover, conventional methods of drafting, revising, reviewing, editing and preparing texts for publication are highly intensive in skilled labour and this adds to costs and increases the delay between the point at which a piece of research is completed and the results appearing in print. Even with the most modern methods of handling the final text (computer-aided composition and printing), the time which elapses between completion of the work and its dissemination is rarely less than one year, and in many cases is over two years. Although several interesting new developments have been introduced, such as synoptic journals, with the complete article available as back-up in print, typescript or microfiche, there appears to be no means by which really significant improvements in timeliness can be made within the conventional publication system.

The case for all electronic systems rests on the premise that the number of "handlings" of the material in conventional systems results in high costs and a lack of timeliness which limits their usefulness as a means of information exchange. But how important is timeliness, and what value would the client, in his twin capacity as author and user, place on major improvements in this area?

Hard data, enabling answers to such questions to be given, is not available. It seems clear though that the requirement for timeliness differs between disciplines: in those areas of pure or theoretical science having as yet no immediate industrial or economic applications, timeliness may be considerably less important than in areas of applied science and technology in which the frontier of knowledge is advancing very rapidly and the economic or commercial impact of research and development is more immediate. In certain highly specialized or esoteric areas of pure science, practitioners will be few in number and therefore will probably know each other, so that an effective current-awareness service may be built up on the basis of personal contact and correspondence; a gain in timeliness of formal publication may therefore by irrelevant in the context of information exchange.

Finally, another aspect which must not be forgotten in any discussion of the need for reform of the publication process is that information exchange is only one of the objects of formal scientific publication. Other factors, such as publication being the traditional means of claiming priority for new knowledge, its role as a means by which peer judgement can operate, and (at least in some areas) as a necessary element in career advancement, act as a brake to reassessment of the process as a means of information exchange.

Nevertheless, most users/producers of scientific and technical information would agree that the information exchange role is predominant; the high growth rate of on-line services, and the increasing volume of demand for document copies, are pointers to the growing concern for timely and complete information. Some quantitative details on the relationship of these two factors are given in reference (3).
The technologies: state-of-the-art

In this section we shall very briefly review the current state of development of the three relevant technologies required in realizing an all-electronic publishing concept, together with current moves towards integration to provide a total system.

Text Processing

Text processing constitutes a major growth area in computer applications and the technique has achieved a major breakthrough in commerce and industry during the past few years. Many industrial and commercial organizations find it more cost-effective to buy a mini-computer system with several terminal work stations than to rely on normal typing facilities for the production of reports and memos of all kinds. At the other end of the spectrum, text processing is now replacing other more conventional methods of composing printed matter of all kinds including newspapers.

Aside from the easy manipulation and editing of texts, contributing greatly to cost reduction (no re-keyboarding) and the rapid production of a fully edited master, a feature highly important for scientific and technical information handling is that a digital record is produced automatically during the initial keyboarding operation. This digital record provides a convenient stepping-stone to subsequent retrieval and delivery parts of the dissemination system.

The output of a text-processing terminal or its associated minicomputer/microprocessor may be interfaced with a telecommunications facility in exactly the same way as any other computer or computer terminal, e.g. they may be connected to a telephone-type circuit (leased or dialled connexion) by modems. Thus, a text drafted in one location may be edited in near real-time in another, either through the public switch telephone network or a computer network. Further, text-processing machines can send and receive texts in this fashion without interrupting the performance of local work at the remote locations. A major manufacturer of text-processing hardware and software has provided quantitative evidence of the explosive growth of interest in communicating systems of this kind (4). Within one year (1978-79), the proportion of systems ordered with communications interfaces rose from less than 10% to over 30%.

While almost every mini-computer manufacturer offers a text-processing system with full telecommunications capability, the systems from one manufacturer are not necessarily compatible with those from another. The need to agree on telecommunications standards has given some impetus to a move towards general compatibility, but there is still a long way to go before texts produced on one manufacturer's system can be edited on another system over a computer network and the revised version displayed on the originating terminal: technically there seems to be no insuperable barrier to such a possibility.

Computer Networks

The expansion and elaboration of computer networking facilities throughout the industrialized world during the last decade is well-known, and needs no underlining
here. The major US networks, e.g. Tymnet and Telenet, are now nearly global in coverage: Tymnet extends from the Americas to Europe, the Far East and Japan. After a very slow start, it now seems that thanks to the Euronet initiative, Europe will have its own international public general-purpose data network, interconnecting with national networks within the near future. We are now entering an era in which the emphasis is on network interconnexion, and with the development of the X.75 protocol system, global interconnexion of data networks can only be hindered by administrative and political problems.

Present-day computer networks provide facilities for many types of traffic between computers and input/output devices such as terminals, printers, etc., but most have been designed to meet the special requirements of multiplexing interactive traffic between large numbers of terminals and relatively small numbers of host computers. This is admirably served by the packet-switching technique, which may not however be optimal for the file-transfer type of operation, which requires that a large quantity of data, perhaps of the order of many millions of bits, is transferred rapidly at a steady rate between two fixed locations in the network. Transfer of whole text is, from the computer point of view, more in the nature of a file-transfer operation than interacting with a remote data-base.

Broadly, there are two types of difficulties which could arise in using a packet-switched network for this kind of operation. First, speed of the operation is limited by the maximum speed at which data can be accepted by the circuits entering the main spine of the network (for all practical purposes 9.6 kb/s). Second, a fast steady bit stream between two fixed locations may inhibit the effectiveness of routing procedures, thus degrading the network performance as a whole. The extent to which a given network, such as Euronet, can handle file transfer in addition to interactive traffic is a matter requiring further study and experiment.

Satellite Telecommunications

Two basic properties of satellite communications, as compared with those of terrestrial networks, are particularly relevant to electronic publishing: they are the much greater bandwidth (and therefore bit-rate achievable) as compared with terrestrial data circuits, and the broadcasting, or multi-destination, possibility inherent in transmissions through a satellite link. Geo-stationary communication satellites of increasing sophistication, in the shape of the Intelsat series, possessing these attributes have been in regular operational use for several years, but with little impact in the shape of new or cheaper data services.

This is partly due to common carrier and PTT tariff policies, but is also a function of the technical characteristics of the Intelsat services. In the frequency range used (4 to 6 GHz), and with the on-board power available, very large and costly Earth stations are needed to receive and transmit signals to the satellite. The system is designed for global communications, i.e. for transmission and reception at equal strength over the total area covered by the satellite, a further reason for employing large and powerful Earth stations. Thus, data traffic to and from the satellite must use the terrestrial net-
work to access the single ground station provided within a country or region, and consequently the theoretical benefits of satellite telecommunications are invisible to the data processing user.

However, a new generation of regional communications satellites is now emerging: these are also geo-stationary but concentrate their signals within a particular region using a beaming technique. New parts of the frequency spectrum are employed: the present European prototype communication satellite now in orbit (Orbital Test Satellite - OTS) operates at 11-14 GHz, and this and other design features make it possible to use relatively small and therefore cheap Earth stations with a dish antenna of no more than 3 metres diameter. The cost of such transmit - and - receive Earth stations is at present of the order of 100 000 European Accounting Units, and it may readily be appreciated that with a multiplicity of such stations satellite data communications could become much less dependent on ground-based facilities for local distribution. OTS and an associated Earth station are illustrated in Figures 1 and 2.

Figure 1: The Orbital Test Satellite (OTS)

Since satellite communications are all digital, there are no special limitations on the rate at which data can be sent and received: rates of a megabit per second may easily be achieved, making massive file-transfer operations much more practical than in the case of terrestrial computer networks. Experimental file transfers using OTS have already been made to test the hardware and software required to interface the computers and Earth stations, with surprisingly low error rates.
The multiple-destination property is also near the point at which practical applications can be tested. The TDMA (Time-Division Multiple-Access) reservation system seems to provide the most economic use of the satellite in this respect; in this, the satellite can accept simultaneous traffic between pairs of Earth stations together with traffic "broadcast" from one Earth station to several others. This possibility is clearly relevant to the requirement for simultaneous text distribution to a number of locations.

Reference (5) discusses communications satellite technology in terms of the various informatics applications now under active study or development.

Figure 2: Typical Earth station for use with the Orbital Test Satellite

Integration of the Technologies

The impact of satellite communications on informatics is receiving a great deal of attention in both Europe and the USA; in both regions, the current emphasis is on the need for integrated satellite and terrestrial networks, each portion undertaking the tasks best suited to it.

In Europe, however, there is not yet total agreement on how satellite systems can best be integrated into the existing networks and services, although there is increasing realization that the advent of the small Earth station opens up entirely new possibilities.

In the USA, the DOMSAT programme, together with the move away from telecommunications monopolies in data transmission, has resulted in proposals by organizations outside the telecommunications field to offer a whole spectrum of new services including electronic mail, video-teleconferencing, high-speed facsimile, text processing
and distributed data processing of all kinds. These employ a mix of satellite and terrestrial communications technology and cover electronic data processing applications of all kinds. The proposals of the Satellite Business Systems corporation is a well-known example: in this IBM, in association with COMSAT Laboratories, is offering a large capacity, digital, across-the-board data service using satellites, and Earth stations located on the customer’s premises; clients will be large corporations, able to use the capacity offered. The Xerox corporation has recently made a proposal on rather different lines, reviewed in reference (6), which involves a new terrestrial-satellite network to be called XTEN (Xerox Telecommunications Network), composed of roof-top microwave transceivers operating in the 10 GHz band and linked to city network nodes at which frequency conversion to the 4 to 6 GHz satellite band will be made. This mode will constitute the interface with a satellite Earth station. The local microwave radio links will use Frequency Division Multiplexing techniques, and the frequency chosen will permit frequency reuse. It is proposed to offer all types of distributed data processing services on XTEN including teleconferencing backed up by still-frame video and high-speed facsimile. The proposal is currently under consideration by the Federal Communications Commission.

An outline design for an all-electronic publications system

The design of an electronic publishing system can be described in terms of the three main functional elements constituting the publication process in general. These are:

(i) **Origination**, i.e. preparation of material for publication, including any necessary refereeing or approval process.

(ii) **Processing**, i.e. the transformation of the material to a form suitable for widespread dissemination (usually involving text composition or other forms of re-keyboarding).

(iii) **Dissemination**, i.e. the distribution process itself. With present technology, distribution means the printing or reproduction of the text in a journal or as an issue in a report series, and its physical transmission to a point at which it becomes accessible to potential users.

In practice, a fourth process has been found necessary to enable a user to become aware of publications relevant to his particular problem, and therefore the primary literature is abstracted, catalogued, and indexed for inclusion in a secondary publication, designed as a retrieval tool. Almost all such retrieval publications are now fully computerized, and the data bases are available for real-time interactive interrogation using computer networks. The main features and requirements for electronic publishing in each of these areas are summarized in the following paragraphs.

**Text Preparation**

The main requirement is that compatible word processing systems are used at all stages in the chain, from the production of the first draft by an author up to final amend-
Figure 3: Flowchart for electronic publication

ELECTRONIC PUBLICATION – FLOWCHART

ORIGINATION

AUTHOR(S) TP

PUBLISHER TP

PROCESSING

MERGE/ SUB-EDIT

CONVENTIONAL JOURNAL

PERSONALISED JOURNAL

ALERTING SERVICES

MICRO-FICHE

SECONDARY DATABASE

DOCUMENT DELIVERY

RETRIEVAL

MASTER TEXT

ORDERING ON-LINE

CONVENTIONAL OR OVER NETWORK AS APPROPRIATE
ments as a result of refereeing or approval procedures. This part of the process could work without sophisticated computer telecommunications between the parties involved, relying on the postal transmission of floppy discs or other forms of temporary storage. But, if the costs are reasonable, transmission of the full text using computer networks would be much preferable, bearing in mind that the text is in digital form from the start. Taking into account the rapid growth of present-day terrestrial computer networks, it is not unreasonable to expect that within the next few years authors, editors, referees, and publishers will all have such facilities available, whether using terrestrial or satellite links.

Composition and Reproduction

In the normal case, where reproduction is by some form of printing process, the author's final text will have to be re-keyboarded to provide the necessary master for reproduction, whether this be an offset plate or type. At this stage, layout is usually standardized and pagination added. Although efforts to spread the re-keyboarding load are becoming increasingly common, e.g. the requirement that authors provide camera-ready copy, such measures are not altogether satisfactory. Where the author's final text is available in digital form, however, no re-keyboarding is necessary, since a master for offset printing can be produced directly from the digital record, using the editing facility to add pagination and other material necessary if the text is to be combined with others in one publication. Modern commercial text-processing systems employ clear and attractive type faces which result in good quality masters, far removed from the impermanent appearance of normal computer printout. It may be noted that prestigious journals such as the Harvard Business Review are now produced on a text processor. The reproduction of drawings and photographs using a digital system is at present unsatisfactory, except where highly sophisticated, high resolution systems are employed, and material of this kind may therefore require separate treatment.

Although this discussion has been concerned with using electronic methods to produce a master for printing at a remote location, it may be observed that the existence of a digital, machine-readable form of the text from the outset, together with the capacity of satellite links to deliver this text rapidly to single or multiple destinations, may make the production of an analogue master for printing unnecessary, except possibly for archival or other purposes not specifically concerned with rapid information exchange. If adequate systems for retrieval of information in a particular text are available to potential readers, and if new forms of delivery systems, making use of broadband communications can be introduced as a normal service, then an analogue (print) version could be produced only when and where necessary.

Distribution Systems

Where multiple copies of a printed version are required, normal distribution systems would be employed, with two additional options: microfiche direct from the digital version by a COM process, or local reproduction in a number of remote centres by multi-destination transmission of the text. The most interesting case is, however, that in which transmission and conversion into analogue form is undertaken on
demand, either as a result of the particular article or report being retrieved as relevant to a particular user's problem, or as part of a new-style current awareness service.

In the latter concept, users would be able to request new articles matching a particular interest profile at regular intervals (often referred to as the personalized journal concept). In this case too, the output would either be in the form of COM-produced microfiche, or text transmitted to a local printer via a satellite link. In all cases in which full text transmission is required, there would need to be a local delivery system as an interface between the high-speed satellite system and the low-speed local printer, i.e. the production of a low-speed record on tape or floppy disc, for example, which would then be physically transported to the user's location.

Retrieval

If a demand publication system of this type is to be satisfactory, highly efficient retrieval methods must be employed. This would imply the construction of a data-base in parallel with the processing of the text at the publishers. Assuming that an abstract will always form part of the text, it would not be difficult to construct a data-base for free-text-retrieval based on the title and abstract, plus author and corporate source, at the time the texts are being processed for "publication"; this could be done with little or no additional intellectual or manual work as a routine computer task. If indexing were required, additional intellectual work and therefore increased costs and time would be involved. Thus a thesaurus or vocabulary-index data-base could not be made available for searching until the full texts of the articles or reports were available for distribution.

It may be observed that this implies a merging of primary and secondary publication functions, and retrieval data-bases constructed on the basis of what might be described as journal or report series coverage, rather than in terms of material drawn from a number of sources in the same field. It would of course be possible to produce major discipline or mission-oriented data-bases by aggregation of "primary" data-bases, in the same way as is done today, but by speeding up the process using digital records rather than by reworking printed documents. However, such aggregated data-bases could not be produced until after the original documents were available for delivery, in conformity with present practice. Thus, it is probable that, should this form of demand publication be further developed, the optimum design of retrieval systems will require considerable further attention, perhaps in the direction of providing a two-tier facility, the lower for quick retrieval of relevant items from readily identifiable sources, and the second for in-depth retrieval from a large number of sources.

Integration with Other Services

From a telecommunications point of view, the design for an all-electronic publications system sketched here could only become practicable if it were part of a whole new range of high-capacity information and data services offered on a public service basis. The flow chart in Figure 3 and the network layout sketched in Figure 4 should therefore be interpreted in this sense.
Figure 4: Integrated network layout

INTEGRATED NETWORK

DATA RATES

- \(10^6\) bls
- \(10^3 - 10^4\) bls
- \(10^2 - 10^3\) bls
Conclusions

The Requirement

The scheme outlined in this paper is based on the premise that changes are needed in the conventional methods of publishing the results of research and development activity in certain areas of science and technology, in order to improve timeliness and to provide direct access to relevant information. The quantity of material for publication, the labour-intensive nature of conventional systems, the number of "handlings" required between the author's first draft and actual publication, and the time-lag between primary publication and the appearance of tools for retrieving the information, are all factors which increase the total cost of information exchange, and reduce its effectiveness as a total system.

The Possibility of Technical Solutions

The key element in the solution proposed is the creation of a digital machine record at the very beginning of the process as a whole, i.e. the author's draft; this is then electronically manipulated to avoid re-keyboarding and is transported by means of computer networks using, where necessary, the very high-capacity and multiple-destination characteristics of satellite data communications. From a purely technical point of view, text processing, computer networking and satellite telecommunications have all reached a state of development in which the required features are all state-of-the-art, or nearly so. Text-processing systems which can be interfaced with computer networks are already commercially available, although compatibility problems require to be tackled; international computer networks on a public service basis are now at the point at which their universal availability within industrialized regions of the world is beginning to be taken for granted, and the interconnexion of networks is becoming less of a technical than a political and organizational problem. The characteristics of current prototype regional communication satellite systems permit the use of small ground stations and multi-destination transmissions at very high data rates, thus making it possible to transmit whole texts very rapidly.

Operational Considerations

An important feature is the almost total flexibility in the type of outputs which become possible as a result of the digital nature of master texts and their electronic distribution. The conventional options of physically distributed printed versions still remain, but are supplemented by a wide variety of other products. From a systems point of view, the scheme has the advantage that a functional integration of the various parts of the information dissemination and acquisition process becomes possible, each function being tuned to its neighbours in the chain. Under the present system, primary and secondary publication processes are often largely independent, involving costly additional intellectual and clerical effort, with heavy additional resources being required for retrieval based on intellectual and machine reprocessing of material already processed at earlier stages.
Organizational Factors

It seems very unlikely that the technical infrastructure required to operate an all-electronic publications system will be created solely for this application of tele-informatics. However, the new possibilities of integrated ultra-high capacity data networks as vehicles for a whole spectrum of new information and data services of all kinds are now receiving serious consideration, particularly in the USA: a publication facility fits naturally into this more general concept. Indeed, the publication system described could be viewed as a special case of a more general concept for information creation, management and distribution, implicit in the rapid development of tele-informatics.

There are problems in realizing this kind of concept in Europe, but there are grounds for guarded optimism that these may be solved: the collaboration between the Commission of the European Communities and the PTTs in establishing Euronet and between the European Space Agency and CEPT in the OTS programme and its operational successor, the European Communications Satellite system (ECS), together with intense activity in the information and data networking areas at the national level, indicate a rapidly changing climate in which this type of innovation becomes possible.

In the meantime, there are many unknowns, including the whole question of economics. A programme of studies is required to quantify some of the operational and economic factors involved in electronic publishing; this study programme might well be carried out in association with actual experiments using OTS, which is available for applications-oriented experiments in 1980. An experimental programme would enable some of the organizational questions to be resolved without disturbing existing structures. One might, for example, use an existing technical report series in a high-technology area as a guinea pig to assess the overall improvement of timeliness against user reaction to an integrated system involving computer-stored master texts, printed outputs, simultaneous information retrieval and electronic document delivery.
References


