The ADONIS Experience

James Andrew Braid
Head of Service Development, The British Library Document Supply Centre
Paper presented at the CD-ROM Conference, York, 18th and 19th September 1989

Abstract

The fundamental concept of Adonis is that new technology can be used to reduce the cost of supplying articles from printed journals. The idea is not a new one, it dates back some 10 years. Early projections showed that the technology was not ready. In the mid-1980's further advances in technology made possible a limited trial. As a result of that trial the publishers have decided to go ahead with a commercial service from 1991, in experimental form. At about that time the technology was becoming available to put the theory into practice.

The Reasons for Adonis

The underlying reason for Adonis arose in the mid to late 1970's. At that time publishers of scientific and technical serial literature became increasingly concerned about possible loss of subscriptions to their journals. There were several reasons for this:

1. The decline in library budgets. The beginning of cuts in library budgets began in the mid-1970's. This also coincided with the time when libraries were trying to find additional money for automation projects.

2. The decline in journal subscriptions. Acquisition costs are by far the largest part of a library's non-staff cost and so had to bear the lion's share of any cuts.

3. The growth of other new media. There was a rapid increase in online database searching during the 1970's. This led to an increased demand for more esoteric articles and thus to a demand for full text from non-mainstream publications.

4. The growth of document supply. In spite of the above reasons, publishers viewed the tremendous increase in traffic at document supply centres as the main cause of decline in journal subscriptions.

The first attempt to remedy the situation that the publishers took was to try to change the Copyright Law. Changing laws is not a quick process. In the United Kingdom it took over 10 years. Neither is it guaranteed to succeed. In the United Kingdom, as far as the "fair use" conditions are concerned there is little changed in the new legislation. It is true that the publishers had more success in the United States where the Copyright Clearance Centre was set up in the early 1980's.

The publishers looked for other ways to overcome the problem of controlling the use of material that they had published at an article level. They turned to technology. The basis of this approach was that if technology could deliver articles more cheaply than the traditional, manual methods then the cost saving could be shared between publisher, as a royalty payment, and the supplying organization. Elsevier were the first publisher to examine this possibility. They quickly discovered that the costs were very high; too high for a single publisher. They grouped together to form a consortium to share the costs. Membership of the consortium has changed over time but there are four principal members; they are Blackwell Scientific Publications, Elsevier Science Publishers, Pergamon Press and Springer Verlag.

From the outset the Adonis Consortium decided that facsimile encoding was essential for the project. The Consortium decided that it was important that a true replica of pages from conventionally printed journals should be produced by the system. In 1980, and even today, it is not possible to do this by any means other than facsimile encoding. This had important consequences in the volume of data generated for each page that is scanned. They planned to use optical disk storage techniques which were just becoming available in 1980.

As originally conceived the project failed. This was mainly because the Consortium had got their sums wrong. They decided to play a waiting game until technology could give them what they wanted. It proved to be a long wait. It was not until 1986 that the Consortium decided to go ahead, albeit with a project much reduced in size to that originally proposed.

Facsimile Encoding

Traditional electronic storage, processing and communication of text uses character encoded text. That is, each individual character is represented by a number of bits, typically 8 bits or one byte. There are various conventions for encoding text in this way. One of the more universally accepted standards is the American Standard Code for Information Interchange (ASCII) format. A typical page in a scientific journal contains anything up to 10,000 characters, equivalent to 10,000 bytes or 10 kilobytes in ASCII encoding. Character encoding in this form has been used ever since computers were invented. It works well and is an efficient means of storing data. It cannot cope with non-textual information, ie. diagrams, pictures, graphs etc., that may appear in scientific and technical journals. Often, particularly with scientific publications, illustrations form an integral part of the information.

There are several ways of converting non-textual material into machine-readable format. The method that is most widely used, and one of the few methods that can deal with all forms of images printed on a page, is facsimile encoding. The principle involved is very simple. A page is scanned by a light sensitive device which monitors areas of blackness and whiteness on the page. This information is converted into an electrical impulse which is used to set the value of a bit to either "0" for white areas or "1" for black areas.

The finer the original scanning the more detail that will be captured. However, the finer the scanning the greater the number of bits generated. The bit stream that is generated is often referred to as a "bit-map". A typical resolution of 200 discrete areas per inch scanned in both directions produces about 4 million bits per A4 page. (200 lines per inch equates to about 8 lines per millimetre) If the resolution is increased to 400 lines per inch then nearly 16 million bits are generated. 400 lines per inch is the minimum resolution at which the human eye cannot detect any breakdown of image in any direction. Data compression techniques can be used to reduce the volume of data to be stored or transmitted. Compression ratios vary greatly, those used commercially in telefacsimile machines compress by a factor varying between 10 and 20 to 1 for text but are less efficient for illustrations. There are proprietary compression algorithms for illustrations and text which can achieve factors of up to 80 to 1. With standard algorithms encoding typical pages from scientific journals a factor of about 15 to 1 can be achieved. This means that, for a resolution of 400 lines per inch, a typical page generates about 120 kilobytes of data. This is a magnitude greater than character encoding.

Use Studies

The introduction of high volume optical storage media gave the Adonis Consortium the opportunity to explore their original concept. They quickly realized that, even with the high storage capacity of optical storage, because of the vast volume of data generated by facsimile encoding they would have to find a method to limit the volume of data. Initially they decided to make a compromise on the scanning resolution used. 300 lines per inch was chosen as resolution which was adequate for the sub- and super-scripts in scientific journals and yet did not generate an excessive amount of data. The Consortium also decided to limit the number of journals to be stored on the system. The obvious criteria was to store only those titles which were in greatest demand.

In 1980 the Consortium did not know which journals were in highest demand for inter-library loan and document supply purposes. They approached the (then) British Library Lending Division who carried out a survey monitoring use over a period of time. It was discovered that the use of journals was concentrated on a limited number of titles and that use was highest for journals less than 3 years old. The results of a single survey were insufficient so far as the Adonis Consortium was concerned. They approached other document supply centres to determine if similar request patterns applied. Further surveys were carried out including one international comparison of five centres. The results of these studies showed that 20% of all demand at the five centres could be satisfied by 514 titles and 80% of demand by 20% of all titles (see figure 1).

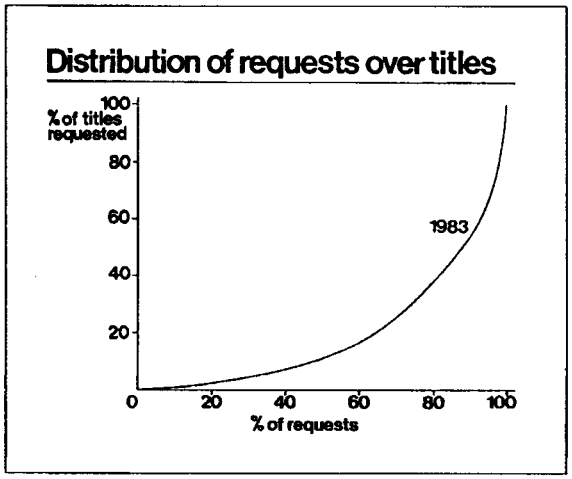

Figure 1

In 1980 the figure of around 3,000 titles seemed to be an economic number. Much research had taken place using 12 inch diameter optical disks and various types of advanced technology for scanning and printing the images. (In fact much of the pioneering work done for Adonis can still be seen today in such systems as that installed at the Library of Congress). These proposals failed for financial reasons. The cost of the scanning equipment and the cost of the printing terminal were high - at least $250,000 each. The Consortium had also over-estimated the costs of the supply centres. The project as originally conceived was not economic.

The Interim Phase

Although the Consortium realized that the project was not viable economically, their original premise was still valid. If a cheaper method of scanning and a cheaper output terminal could be found then the project may become viable.

In the early 1980's various studies were carried out on behalf of the Adonis Consortium. One suggested the use of a scanning bureau, but no other use for the bureau could be identified at that time. Another study reduced the number of titles to be scanned. Although this reduced the scanning costs it did not reduce the costs of the supply centres, in fact it increased them. At this time the idea of limiting the project to one subject area was discussed. The earlier studies had shown that use of biomedical material was concentrated on a limited number of titles and that much of this use was for current material.

CD-ROM and scanning Bureaux

The introduction of CD-ROM in about 1985 gave the possibility of low-cost, easy to use workstations for the printing of items stored on CD-ROM. The cost was about 10% that of a workstation for larger WORM disks. This meant that it should be possible for more libraries to be able to afford to use the system. There were two drawbacks to the use of CD-ROM. First, they are relatively expensive to master and second their capacity is somewhat limited for facsimile encoded pages. The high mastering costs are compensated by the fact that the cost of making multiple copies is much less than WORM disks. The capacity was not thought to be an insurmountable problem if a jukebox could be used.

In 1985 the European, Japanese and American Patent Offices signed a tripartite agreement to exchange patent specifications in facsimile encoded format. The European Patent Office, which had in excess of 60 million pages to scan decided to tender for the work to be done. As a result three scanning bureaux were set up, in England, France and West Germany. The Adonis Consortium approached one of these bureaux to scan the journals. The standard to be used was exactly the same as that for patent scanning. An agreement was reached for the scanning to be done at an economic price.

The Trial

In July 1986 the Adonis Consortium decided to go ahead with a limited, two year trial. This was on the understanding that the Consortium would also receive some partial funding from the Commission of European Communities. Initially, 300 journals in the field of biomedicine were chosen. This included publishers who were not members of the Consortium. The following additional publishers have some titles included in the trial - Butterworth, Churchill-Livingstone, Mosby, Munksgaard, Thieme and Wiley. For a variety of reasons the total number of journals was later reduced to 219. All issues of the chosen journals for the years 1987 and 1988 were to be included. Use of the system at the trial site libraries was to be during 1988 and 1989.

Various contracts were negotiated for the production of the discs. It had been known for some time that the only way to obtain a true facsimile image of a page was to scan the page after it had been published. In order to do this it is necessary to prepare printed journals for scanning. This was done by the Adonis Office in Amsterdam. The indexing of the journals was also done in Amsterdam by Excerpta Medica. The scanning was done in England by Scanmedia. The tapes of the scanned images and the index to those pages were sent to the CD-ROM production facility, Philips-du Pont Optical in Hannover. A diagram of the production cycle is shown in figure 2.

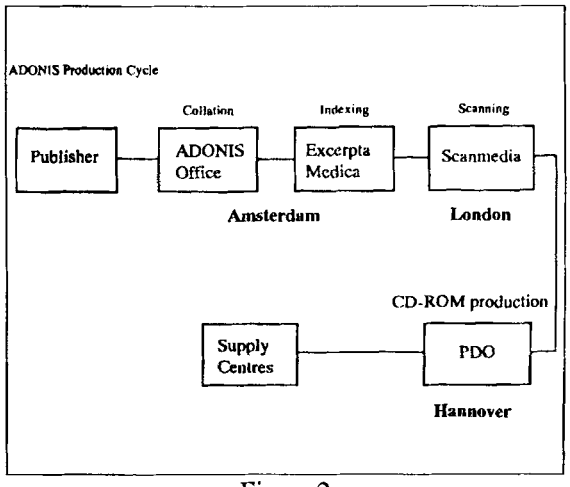

Figure 2

The Adonis Article Identifier

In any machine-readable database it is far easier to refer to individual entries by number. Adonis is no exception to this. The problem facing the Consortium was that there was no standard method of referring to articles by number. There are two standards for article identification systems. One is BIBLID (BIBLiographic IDentification), ISO 9115. The other, the Serial Issue and Article Identifier (SAID), is an American National Standard (ANSI Z39.56). Both BIBLID and SAID were developed mainly as library tools and are page dependent. They can be generated independently but are fairly cumbersome to use (up to 38 characters long).

Adonis adopted a simpler system which is page independent and as a result can be allocated by publishers in advance of publication. The format of the identifier is:

1234-5678 88 99999 X
ISSN Yr No Check digit

A major disadvantage of the Adonis Identifier was that it was allocated at the time of indexing. This means that the number is not known in advance and articles cannot be requested by this means. In the full operational system this number may be allocated in advance of publication and appear on the first page of an article and then be quoted in abstracts, databases, etc. The British Library Document Supply Centre (BLDSC) is carrying out some comparative trials of the Adonis Identifier, BIBLID and SAID in the DOCMATCH project.

The Adonis Workstation

The BLDSC has had a long association with the Adonis Consortium. When the Consortium decided to go ahead with the trial BLDSC offered to develop a workstation which would retrieve and print the articles stored on the CD-ROMs. To carry out this work the BLDSC also received partial funding from the Commission of European Communities. It was decided at the outset that, if possible, an off-the-shelf workstation should be used. BLDSC awarded a contract to the French company MC2 for the development of the proprietary workstation developed by Laserdata. This comprises an IBM PC AT or compatible machine with 640 kbyte memory and a 40 Mbyte hard disk, a Hitachi CD-ROM drive, a Ricoh laser printer of 300 line per inch resolution and a high resolution screen capable of displaying a full A4 page at 150 lines per inch.

As each new CD-ROM was received, at a rate of about one every 10 days, the disc was loaded into the workstation and the index information was downloaded. The new index information was then merged with the existing index. In this way a cumulated index of articles was generated on the hard disk of the workstation.

The level of indexing used on Adonis is only sufficient to uniquely identify articles. The indexed fields are:

Journal title	Issue number
Page number	ISSN
Year of publication	Article title
Author - up to four	Adonis Identifier
Volume number	

Once an article has been identified it can either be read on the high resolution screen or printed out. This can be done immediately or else in batch mode. The screen and printer are both of high quality and give very good results. One drawback of the trial system, which uses 4 year old technology, is that it is impossible to search and print simultaneously. Also these processes are not as fast as is desirable in a normal working environment.

The Users

There are several other centres in the world using the ADONIS system in addition to BLDSC. Four centres are known as core libraries. They were all actively involved with BLDSC in the initial specification of the system. The other three core centres are -

Centre de Documentation de Science et Technologie, Paris
Central Medical Library, Cologne
Royal Academy of Science, Amsterdam

There are other centres throughout the world -

Europe:
Technical Information Library, Hannover
Institute of Science and Technology, Madrid
Karolinska Institute, Stockholm

Australia:
National Library of Australia, Canberra

U.S.A.:
Information on Demand, Washington
University Microfilms, Ann Arbour

Japan:
Kinokuniya, Tokyo

Mexico:
Universidad Autonoma de Nuevo Leon, Monterrey

All the centres are interested in determining how new technology can be used to improve their

internal processes. All are making a financial contribution of about $30,000 as a token subscription for the discs plus the cost of the workstation. In addition to the subscription fee there is also a use fee paid by some of the centres. They are contracted to make all copies for requests for Adonis material from the system. Results to date are somewhat mixed. All centres report usage as being less than they had predicted. For most centres use has been less than 5% of total demand, for BLDSC it has been less than 1%. This has posed severe operational problems but once requests arrive at the workstation they are processed more quickly and slightly more cheaply [in the majority of cases] than by using conventional manual methods. As with all high technology systems there have been problems - the only lasting one concerns the laser printer which is working beyond its design capacity.

There are other problems which require solution before the system becomes fully operational. The handling of large numbers of discs has been resolved by the addition of a CD-ROM jukebox. During the two years of scanning 84 CD-ROM's were produced. This is a quite unmanageable number if there is a high demand on the system. BLDSC looked for a CD-ROM jukebox. There were none in existence so it was necessary to specify and procure one. This was done during late 1988 and it became operational [probably the first in the world in everyday use] in January 1989. As a spin-off in the development of the jukebox the company responsible, Next Technology of Cambridge, have totally redesigned the jukebox into a design that is more compatible for networking. This design, called Voyager, is now operational with Adonis at BLDSC.

A further problem concerns the matching of requests against the index. At present this is a manual and fairly slow operation. The BLDSC has received a grant from the European Commission to develop a system to automate this matching process. Obviously this can only be done for those requests that are in electronic format. The system, called DOCMATCH, has been developed by the University of Bradford and is based on work that they had done previously in developing the Universal Standard Book Code [USBC] matching algorithm. Although not in daily use DOCMATCH is being used experimentally and successfully at BLDSC.

One further area for improved efficiency would be in the direct electronic transmission of articles from the workstation. Theoretically, this is a simple matter. As the pages are stored on the discs in facsimile format it should be possible to transmit those pages to remote facsimile machines. In practice this has proved to be a difficult process to achieve but a way of doing it has now been found. Once this is operational and linked with the DOCMATCH system as input it will be possible for the whole process of requesting, processing and delivery of documents to be carried out automatically.

An interesting by-product of the Adonis system is the production of use data at an article level. This is the first time that such data has been collected on such a large scale at this level. Early indications are that use of articles, like that of journals, is highly concentrated on a limited number.

The Future

Early in 1989 the four original members of Adonis Consortium decided to examine the possibility of launching Adonis as a commercial service. It has always been realized that there must be a wider market than just document supply organizations. The commercial service will be aimed in the first instance at pharmaceutical companies as well as document supply centres. A list of over 400 target journals for inclusion in the system has been produced although the agreement with the publishers to be involved has not yet been reached. It is planned to begin the service in 1991. There will be improvements, mainly in speed of access to the workstation and a greater number of pages per disc, the aim is 15,000 pages per CD-ROM. As yet the price has not been fixed but Adonis will probably operate as a not-for-profit organization dealing with the production of discs and indexes. Each publisher will set a use charge which will be in addition to the annual subscription paid to Adonis. As yet there is no indication of what these prices will be. The publishers face a further dilemma in attempting to maintain the existing subscriptions to proposed Adonis titles. There will be some sort of financial incentive for customers of the Adonis service not to cancel subscriptions.

Conclusion

The Adonis project has had a long history. After a very long period of gestation a practical trial has now taken place. The success of the trial is difficult to assess. From the Consortium's view it must be judged a success as they have taken the decision to launch Adonis as a commercial service. The view of individual publishers is less easy to gauge. Time will show with the number of publishers who agree to participate in the commercial phase of Adonis [or

perhaps with the number who do not]. From a library's standpoint the trial has provided conflicting evidence. It can be shown that the Adonis method of supplying documents is both quicker and cheaper than conventional methods. Likewise, it can be shown that Adonis is slower and more expensive! The major potential customers for using Adonis, large industrial libraries, have not yet had the chance to try it. Time will be the judge in this instance. However, it should not be libraries who are the main beneficiary of the Adonis service. The aim of all technological innovation should be to give a better service to the end-user. In the case of Adonis this is the research worker. In the trial service it is unlikely if the end-user noticed any difference. This may well also be the case with the first stages of the commercial service. It is only when further developments take place, such as information arriving in electronic format directly into a researchers desk top computer that a real change will take place.

Acknowledgement

Much of the information on the history of ADONIS is taken, with permission, from:
"ADONIS - The story so far"
by R Campbell and B T Stern
in the book "CD-ROM Fundamentals to applications" edited by C Oppenheim (Butterworths, 1988 0-408-00746-X)