

## Experiences with ExtraMED, an electronic full-text biomedical journal collection on CDROM

Dirk Schoonbaert

Assistant Librarian, Institute of Tropical Medicine Prince Leopold,  
Nationaalestraat 155, 2000 Antwerpen, Belgium

**Abstract:** *ExtraMED is an electronic full-text library distributed on CDROM, incorporating over 200 biomedical and health-related journals. Unlike Adonis, for example, ExtraMED does not specialise in high-impact journals, but focuses on local or regional biomedical journals from developing countries, most of which are not indexed by the major indexing services. Especially for specialties such as tropical or traditional medicine, this constitutes a valuable source of additional information. Also, as a bibliographic database ExtraMED is a worthwhile supplement to MEDLINE, but obviously not an equivalent alternative. The annual subscription fee is £2000 and includes 12 monthly CDROMs. Each disc contains about 8000 pages, so ExtraMED provides an estimated 100 000 pages a year. The practical advantages of this type of electronic document delivery are discussed, as well as the actual coverage and relevance. Though ExtraMED is certainly a laudable initiative, there are ample indications that the producers may have underestimated the problems of its practical realisation. The project suffers considerable delay, and as only six CDROMs were published in over 20 months, the contents still present a somewhat chaotic picture with lots of gaps. Fortunately things are getting gradually better. The retrieval software is generally adequate and includes a few remarkable features such as powerful truncation, proximity searching and automatic synonym look-up. Display and printing capabilities are good, though somewhat cumbersome in the latter case.*

### 1. Introduction

ExtraMED is an electronic full-text library distributed on CDROM, incorporating over 200 biomedical and health-related journals. Unlike Adonis, for example, ExtraMED does not specialise in high-impact journals, but focuses on local or regional biomedical journals from developing countries, most of which are not indexed by the major indexing services. Especially for specialties such as tropical or traditional medicine, this constitutes an interesting source of additional information. The practical

advantages of this type of electronic document delivery are obvious: a vast number of references can easily be retrieved (bibliographic database) and the articles themselves, which in their printed form are generally not easy to come by, are now immediately available (full-text images). Because of the CDROM format, the information is far less bulky than the paper versions and the management overhead is minimal, compared to subscribing to each individual journal separately (and claiming missing issues). ExtraMED is produced by the London based com-

pany Informania Ltd. It originated as a project of the World Health Organization (WHO), which for many years has been active producing regional supplements to MEDLINE (WHO 1994). ExtraMED is also supported by other international bodies such as UNICEF. The annual subscription fee is £2000, for which you receive 12 monthly CDROMs. Each disc contains about 8000 pages, so ExtraMED provides an estimated 100 000 pages a year. ExtraSCI and AgROM Extra, comparable series in cooperation with UNESCO and FAO, have already been announced.

### 2. Coverage

As librarians at an institute of tropical medicine, my colleagues and I were obviously interested in ExtraMED, and subscribed to it when it was first announced. Although it is essentially a great initiative, I have mixed feelings concerning its practical and technical realisation.

After initial discussions, ExtraMED started processing its first CDROM early in 1994. More than 20 months after subscribing, I had received only six discs. Publication pace has recently started to improve, so ExtraMED may soon reach cruising speed.

A survey of the number of megabytes and articles on each disc is included in Appendix 1. One obvious conclusion is that the CDROM discs are not filled to their potential. Undoubtedly this is partially due to the delay in the reception of original material and the pressure to publish more or less regularly. As each year 12 CDROMs are scheduled to be added, and disc swapping should be avoided as much as possible, I hope this filling ratio will soon be optimised.

At the start of the project, some 200 journals had agreed to join the ExtraMED Consortium, with many others to follow, including 90 newsletters re-

**"ExtraMED can only aim to supplement the bibliographic MEDLINE information and can certainly not replace it"**

cently announced. The list of 200 journals is dominated by Asian and African titles. China is especially well represented with over 60 journals (though for most, only the titles and abstracts are in English, the full-text being in Chinese characters). Remarkably, the list does not include any Latin American titles. This is probably because ExtraMED originated at the WHO Eastern Mediterranean Regional Office. Fortunately, many Latin American journals are covered by the Latin American and Caribbean Literature on Health Sciences (LILACS) database, an initiative of the Pan American Health Organization (PAHO), produced by BIREME in Sao Paulo. LILACS has been available online and on CDROM for many years, but it is a bibliographic, not a full-text database. Possibly individual journals may later join the ExtraMED Consortium or the PAHO may one day publish a comparable CDROM collection of its own. Less striking at first sight is the unbalanced coverage of African journals: only about 15 titles published in sub-Saharan Africa are included.

ExtraMED is not just an electronic stack of journal issues but can also be used as a bibliographic database. You can either search for a specific article (by an author, in a journal, limited to a year, etc.) or look for articles on a certain topic. Generally, ExtraMED's contribution on this level tends to be stressed. However, the claim that ExtraMED could be considered as the database of choice for tropical medicine should be put into perspective. For one thing, ExtraMED can only aim to supplement the bibliographic MEDLINE information and can certainly not replace it. It features some 10 000 articles per year, from 200 source items, compared to MEDLINE's almost 400 000 articles and over 3400 sources. It is true that only a small proportion of MEDLINE deals with tropical medicine, but it is equally true that the most influential literature in this specialty tends to be published in sources covered by MEDLINE or other international databases (Roelants 1987, 1995). I shall not now join the discussion as to whether this partially egg and hen problem is justifiable (Zielinski 1995). Neither do I want to belittle the merits of these local journals, which actually do contain lots of valuable information (though 'tropical medicine' and 'medicine in developing countries' are not necessarily synonymous). Still, it would be a mistake to consider the present selection as the core literature for tropical medicine. Choosing a limited number of journals is always a somewhat arbitrary procedure. This is true for high-impact journals, but even more so for local journals with less obvious international acclaim. Whatever the selection,

**"ExtraMED's major value is to be found in its electronic document delivery aspect"**

there can always be found many others of equal merit. The philosophy of ExtraMED's selection criteria is to present 'the best of the rest', i.e. those journals that MEDLINE does not give a 'fair chance'. Within this context, it is curious that more than 30 of the 200 ExtraMED journals are actually indexed by MEDLINE.

ExtraMED's major value is to be found in its electronic document delivery aspect (see above) as it offers full-text access to local literature that would otherwise be hard to get at and facilitates the local and international distribution of this literature. As a collection of over 200 journals with an average price of £10 for each title, ExtraMED certainly constitutes value for money for libraries interested in tropical medicine and international health. Unfortunately, there is a practical shortcoming, probably resulting from the database point of view: as retrieval capabilities are quite good, there may appear to be no need to list what particular journals and issues are included on each disc. But the librarian's viewpoint is somewhat different — and libraries probably constitute the majority of the ExtraMED customer potential. It certainly is necessary that specific articles can be found easily in the system, but it is also important to know what exactly you can provide both internal and external customers with. A clear listing of journal issues included would certainly enhance ExtraMED's attractiveness to librarians. Informania now seems to acknowledge this demand, as the sixth disc actually did feature an ASCII file listing the issues included. It would be even more helpful if this list were in alphabetical order and available (preferably in a cumulative form) from within the program.

One major aspect which the ExtraMED producers have probably overestimated is the speed of delivery of the journal issues provided by their local publishers. The six CDROMs so far contain 359 issues: 177 of the 204 announced titles are included, of which 123 appear with one single issue (21 journals have four or more issues). This is understandable as this represents only half a year of information (but actually produced in over 20 months!). The actual contents are not a good example of a harmonious journal collection. This becomes clear when looking at the distribution of publication years (see Appendix 2). I have enough experience with journals from developing countries not to expect to receive my copies with the same regularity and punctuality as the international top journals, and I know that a number of journals have publication delays of three years or more, but I do not see the relevance of including articles from 1986, for example, in a 1994-present database. Also, many issues are still missing; sometimes there are gaps of several years; often titles start with mid-volume issues; and so on. Informania has promised to fill in the current gaps gradually. The contents of the first few discs were quite haphazard, but judging from discs 5 and 6, Informania is now making good on its promise.

Another problem is that some of the graphical full-text images were obviously scanned carelessly. For example, on some pages the first millimetres of each line or the bottom line(s) are missing. Also disturbing is the fact that the third disc contains nearly 300 records lacking a publication year.

This obviously does not improve the reliability of searches, nor does it inspire a great deal of confidence.

### 3. Software

ExtraMED uses Idealist retrieval software, developed by Blackwell Scientific Publishers. I am using the Windows version; a Windows NT version also exists and a Macintosh version was announced. In order to run smoothly, a 32-bit Windows expansion, which is provided with the program, needs to be installed first. During the actual software installation, four different icons are created: ExtraMED (the database program proper), CD-Manager (to install new CDROM discs), User registration (to allow the system administrator to define a number of user names, which may be linked to a password and a 'cost centre'), and User statistics (to log the actual use of each journal, user and cost centre). This valuable log information is a fine feature that would not be out of place in any sophisticated database program.

During installation of new discs, all bibliographic information, indexes and abstracts are copied to the hard disk and as such this can be used as a database by itself. The actual page images remain on the CDROMs, and only need to be retrieved for actually viewing the full-text articles. This splitting up of bibliographic and full-text information is a good option, allowing fast retrieval, but it should also be borne in mind that as the indexes copied to hard disk for the first six CDROMs take up some 12 Mb of hard disk space, probably over 20 Mb will be needed for each year of ExtraMED.

Though the system is now working fine, installation of the ExtraMED software was not without problems. Originally, most of the full-text images could not be displayed on screen, resulting in an error message. Contrary to earlier specifications, the 4 Mb RAM of my installation proved insufficient. This appeared to be the fault of Microsoft, which had underestimated the memory needed to run its own 32 bit Windows expansion. Adonis is reported to have encountered the same problem in the beginning. Upgrading to 8 Mb RAM solved this problem. For one thing, this clearly implies that ExtraMED needs a relatively powerful PC to run smoothly, which may partially compromise its suitability for efficient widespread distribution of biomedical information in developing countries. Of course, yesterday's high-end PCs are today's entry-level machines, and most present-day software gets ever more memory hungry, so this may not be an objection after all, provided that developing countries can and will follow this dazzling pace.

After installing a program upgrade, the CD-Manager failed to recognise the ExtraMED discs in the 'D:' drive. Trying to remedy this problem led to many hours of frustration, at the end of which I was ready to give up. Eventually, Informania found out the cause: in my Windows installation 'English (International)' was not installed as language in the international settings of the control panel. Changing this single (external) parameter solved the problem.

"The actual page images remain on the CDROMs, and only need to be retrieved for actually viewing the full-text articles"

### 4. Retrieval

As ExtraMED is a Windows application, most functions are available via buttons, ALT menus and shortcut (function) key combinations. Online help is available from any screen.

Retrieval is fast and can be highly differentiated: practically all fields can be searched, including volume number, page ranges, number of pages and so on. The ExtraMED interface offers two basic templates: 'searching all fields' and 'searching by field'. In practice, both allow global and field-specific searching, including a number of fields in groups, such as 'text fields' or 'numeric fields'. The major difference is that the first incorporates natural language phrases (for example 'containing the word . . .'), while the second uses the more abstract standard Boolean operators AND, OR and NOT in a step by step approach. Straightforward retrieval using a command mode is also available in both (for example 'YPB=1993' to limit the search to 1993 as publication year), but the Boolean operators are recognised as such only if they are entered between brackets (for example 'africa [AND] aids'). Comparative searching with operators such as 'greater than', 'greater than or equal to', 'less than' and 'less than or equal to' is also available. Interestingly, fields can be selected on whether they actually contain data or are empty. Various degrees of proximity searching can be specified: search terms appearing within the same paragraph, within the same sentence, or a specific (maximum) number of words apart. Ranges can be both numerical and alphabetical. Searching is case independent and truncation capabilities are impressive: left hand and right hand truncation pose no problem, and even masking within a word is executed very fast.

Different retrieval templates can be defined with the 'Search-U-Like' feature. The two standard templates feature automatic index browsing with the possibility to select index items as search terms. This index itself is not field-specific (though the 'searching by field' template highlights the index terms if they are present in the specific field demanded) and includes only individual words. For some fields, such as 'journal name', an index of full field contents would have been more helpful than just the constituent words.

The software is generally adequate to find the information 'as-it-is' in the database. Finding a specific article is fast and easy. Searching for subject information, however, is basically limited. Since no keywords are included, only terms explicitly included in the title or abstract field can be retrieved. This lack of keywords also precludes a good thesaurus with powerful

"Probably over 20 Mb will be needed for each year of ExtraMED"

**"The 'synonym' list can be customised and the feature can be activated as a system default or turned off"**

'explode' facilities (cf. MEDLINE's MESH), and obviously reduces precision and recall. This lack of structured keyword searching is partly overcome by a few interesting additional techniques: a synonym list activates supplementary search terms. For example, 'malaria\*' yields additional hits which do not contain the term 'malaria', but are triggered off by terms like 'mosquito', 'falciparum' or 'chloroquine'. The 'synonym' list can be customised and the feature can be activated as a system default or turned off.

A comparable retrieval option is called 'lookalikes' or 'soundalikes' and is meant to neutralise spelling errors. In practice, this implies that 'today' and 'toothpaste', for example, are closely related. It is not clear if and how results from this function can be modified. This feature may prove helpful at times, but it is a good thing it is optional and not a system default.

The system uses a stopword list. This list can be modified but only from outside the program, for example using Windows Notepad. This is not a very user-friendly procedure and, more seriously, the changes made are not reflected in subsequent sessions. Just as before, the shortlisted words keep popping up in the retrieval results. Obviously, the index files themselves cannot be modified, as they are copied from the CDROM discs. Yet it might still be possible that the software ignored stoplisted items from this index, but this does not appear to be the case.

The active hit list can be modified (using the Boolean operators OR, AND, NOT or their natural language equivalents 'widen', 'narrow', 'exclude'), but the subsequent search formulations are not kept as separate sets. So it is not possible to get a survey of all search formulations used in the present session, or to combine a set 3 with a set 6. On the other hand, the most recent search formulation can always be recalled, even if it was executed weeks before. It is possible to save search histories for use in later sessions. Individual hit lists can also be saved.

## 5. Display

ExtraMED features several display levels. First, the 'overview hit list' shows one line per record, consisting of author(s), year of publication and title. For the titles, upper and lower case is used indiscriminately, which gives a rather sloppy appearance. The actual combination of fields can be

**"ExtraMED works on a flat fee principle: no royalties or other extra costs are involved for printing and downloading"**

customised, but only from outside the system, changing an initialisation file in the Windows directory. According to the comprehensive manual, the order between individual records can be self-defined, but I could not find how to achieve this. It appears that this Idealist standard feature was deliberately omitted from the ExtraMED package, but may return in a future software upgrade (Trimmer 1995). At the bottom of the 'overview hit list', the search expression is listed in command mode. The screen layout can be customised: for example you can define what bars are included on the screen. Records can be marked (or dropped) while viewing the 'overview hit list'. This subset can then be selected as an active 'hit list'.

Full-record display features the complete bibliographic information of each record, including an abstract. The fields included and their relative order can be changed by defining one or more alternative display sheets. Hit terms are highlighted automatically. Unlike the overview hit list and the full-record display, full-text display does not use machine readable text but displays images (in TIFF format) of the articles themselves, which are retrieved individually from the CDROM. This works very well, although the result is ultimately dependent on the quality of the scanning of the original. At first a full-page view is displayed in the window (which itself can be sized up to fill the whole screen). From this full-page image, at least two subsequent enlargements can be activated. If the text required is not on the active disc, in most cases a message prompts you to load the right disc (e.g. 'cannot open d:\ad940001.img'), but this message could certainly be formulated in a more user-friendly fashion.

## 6. Printing and exporting

Hit lists and bibliographic references can be printed during retrieval or display. Printing the full-text articles is somewhat more complicated, as the articles selected are first stored in a print queue from which actual printing is generated afterwards. I experienced some problems with this. When I did not specify individual printer settings for the headers, the hit lists and the text itself and relied on the (correct) default values, the system invariably claimed it could not print because of 'insufficient printer information'. The headers — which include bibliographic information, credits and copyright indications, but also management information such as to which centre costs should be accounted — are certainly helpful in larger libraries or network environments. But it should be equally possible to ignore them, as in standalone situations they can be a waste of time and paper. Hit lists, full records and even TIFF images of full-text articles can also be downloaded to a file on disc. For libraries, it is interesting that ExtraMED works on a flat fee principle: no royalties or other extra costs are involved for printing and downloading.

## 7. Conclusion

ExtraMED features the full-text images of over 200 biomedical journals from developing countries at a average price of

£10 per journal. As such, it is certainly an interesting electronic library for the study of tropical medicine, though not necessarily an essential one, as the major international journals are not included and the actual selection is debatable. At this time ExtraMED still suffers substantial delay, and as only six CDROMs were published by the end of 1995, the contents still present a somewhat chaotic picture. The producers may have underestimated the problems of its practical realisation. As things are gradually getting better, these are probably growing pains which will soon be overcome. Also, as a bibliographic database ExtraMED is a worthwhile supplement to MEDLINE, but obviously not an equivalent alternative. The retrieval software could still be improved upon, and a few things that were promised have failed to materialise. Yet it is generally adequate for retrieval and includes a few remarkable features such as powerful truncation, proximity searching, and automatic synonym activation. Display and printing capabilities are good, though somewhat cumbersome in the latter case.

### Post Scriptum

By the end of March 1996, a total of 10 discs, featuring nearly 10 000 articles, had been published.

### References

- ROELANTS, G. (1987) Citation analysis in the field of tropical medicine, *Annales de la Société Belge de Médecine Tropicale*, **67**, 315–318.
- ROELANTS, G. (1995) Which health journals are most useful? *World Health Forum*, **16**, 69–73.
- TRIMMER, D.V. (1995) Personal communication from Informania Limited.
- WHO (1994) Health literature services, in *The Work of WHO 1992–1993: Biennial Report of the Director-General*, World Health Organisation, Geneva, 125–126.

ZIELINSKI, C. (1995) New equities of information in an electronic age: the Third World needs First World information — how about the other way round? *Br Med J*, **310**, 1480.

### Appendix 1: analysis of contents in ExtraMED discs 1–6

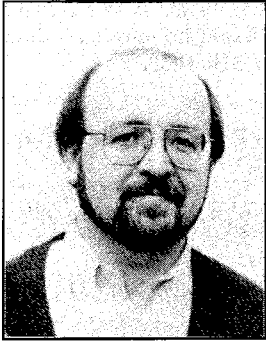
Disc	94/1	94/2	95/3	95/4	95/5	95/6
CDROM memory actually used:	237 Mb	430 Mb	363 Mb	370 Mb	279 Mb	314 Mb
Articles (cumulative)	1062	1683	2765	3685	4504	5455

### Appendix 2: Distribution of publication years

Disc	94/1	94/2	95/3	95/4	95/5	95/6
1995	0	0	0	0	0	0
1994	0	608	490	679	795	816
1993	405	420	826	1256	1660	1964
1992	497	495	785	988	1188	1667
1991	54	54	94	117	180	221
1990	12	12	97	121	121	167
1989	15	15	46	69	78	91
1988	42	42	56	69	86	86
1987	29	29	43	58	68	115
1986	8	8	34	34	34	34
No publication year	0	0	294	0	0	0

### Appendix 3: Informania Ltd

ExtraMED  
Informania Limited  
PO Box 1359  
London W5 3ZP  
UK  
Fax: +44 (0)1730 265398



## The author

### **Dirk Schoonbaert**

Dirk Schoonbaert graduated from the University of Antwerp (Germanic Philology 1980, and Library and Documentation Science 1985), and has worked in several scientific libraries. Since 1986 he has been Assistant Librarian at the Prince Leopold Institute of Tropical Medicine in Antwerp. His professional interests include the electronic aspects of information management and retrieval.