

Research Information Take Away
or
How to serve Research Information Fast and Friendly on the Web

Peter Linde, Librarian, and *Leif Lagebrand*, IT Specialist,
University of Karlskrona/Ronneby, Sweden
Database: <http://www5.hk-r.se/fou/forskinfo.nsf>
e-mail: (peter.linde@bib.hk-r.se)

Abstract

In 1997 the library department at the University of Karlskrona/Ronneby was asked to develop a database which could be used to collate and present all the research material and ongoing research projects at the University in electronic form. If possible, full text versions of all the material should be made available.

In this paper we will discuss the experiences and problems we have gathered during 6 months of developing an electronic research archive.

Using Lotus Notes software we have, with some assistance from consultants, developed a database into which the researchers themselves feed data using web-forms.

Since the project commenced in the summer of 1997 we have created:

- An electronic archive of research documents which uses pdf-files for full text presentation.
- Procedures for cataloguing. In addition to being stored electronically, paper copies of each document are stored.
- Material which is generally available via WWW. They are searchable by subject, type of document, name of institution, or text search within the database, and one can print out references or full text versions of them. The material is also accessible from search engines such as Alta Vista.
- A database which complies with the guidelines stipulated by the Swedish government¹, the Education Department² and the Swedish University Board³.

Background

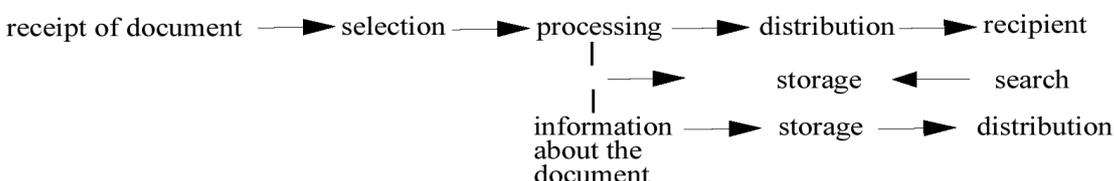
During the spring term of 1997 an interim research editorial committee was formed at the University of Karlskrona/Ronneby to streamline the publishing, distribution and storage of the department's research material. One of the tasks for this committee was to seek funds for developing a database as an electronic catalogue of research material. The project was named DELFIN (Direkt Elektronisk Lagring av Forsknings INFORMATION = *direct electronic storage of research information*). The first thing the

committee had to do was to specify guidelines for processing research material at the University.

For the system to work it was essential to ensure the reliable delivery of material from the various institutions.

The committee agreed that the written word was the best way to present research at the University. In this way a profile of the University could be presented, contacts formed and maintained with institutions and sponsors, while the University would become part of the scientific community. As the volume of research and number of completed papers increases, it is worth considering what happens when we process and distribute written documents.

The figure below illustrates the general form of an information system



Using the above diagram to illustrate how our processing will be organised with the aid of the research information database DELFIN, the stages shown can be described in more detail as follows:

RECEIPT OF DOCUMENT: Researchers complete a web-form, adapted to suit the type of document in hand. This contains fields for bibliographical details and the option of attaching an original file (required for database indexing) and a PDF-file, or a PS-file which can then be converted to PDF-form by the database administrator. The contents of the web-form are sent directly to the database.

SELECTION: This takes place at department level, where each contribution is examined before being entered into the database. The document is checked to see that it is in fact research and that it is of satisfactory quality.

PROCESSING: The administrator checks that the necessary fields have been completed and that the necessary files have been attached. A PDF file is created to enable users to read or print out the material.

DISTRIBUTION: The PDF file is dispatched electronically to the printer's for "Print on Demand". In this way circulation can easily be adapted to requirements. Research reports must be sent to the University library at the same time as they are catalogued and made available on the Swedish research library's collective catalogue LIBRIS. The department which has produced the document is responsible for its distribution to other organisations, university departments and authorities. Since the library catalogues and stores the physical documents, it is also responsible for distributing them on request to users, either in electronic form or, as a loan, in hard copy.

INFORMATION ABOUT THE DOCUMENT: Notice of new material added to the database appears automatically on the database home page. New material is announced on a continual basis through the University's internal channels.

STORAGE: The material is stored electronically in the database and as hard copy in the library.

RECIPIENT: The recipient of the document, or information about the document, should be able to search the database in Swedish or English using free text, subject, institution, author, title, etc. In addition to the research community, industry and the general public are important users of the information. As far as possible the user should be able to access the full text version of the document electronically.

Target Group

Possible users of the research database cover a broad spectrum. We aim primarily at the following categories:

- Researchers here and at other established research institutions, and also university students.
- Teachers and students at other educational institutions, especially at secondary schools (high schools), where students are preparing for university.
- Companies operating in the region needing research and development.
- Media
- Organisations awarding grants and sponsorships

What can we provide?

As can be seen, we offer a service both to the expert and to the layman, within our region and beyond. This has meant meeting certain conditions when we created the database. The user should be able to expect the following of our database:

- The information it contains should be of high quality and presented clearly
- It should be easy to search for information (browsing)
- The database should be available on the WWW
- Its structure should have as few levels as possible
- Relevant information should be included to enable one to find persons/organisations with the skills/knowledge one is looking for
- On-line help
- Links between documents and to other databases / organisations
- If the full text version is unavailable on-line then it should be possible to borrow or copy the document from the University library.
- Generation of reports
- Accessible from search engines such as Alta Vista

What can the database do?

This question is obviously important since the project depends on the researchers themselves providing the contents of the database. For the administrators, the merits of a central database are obvious, but in our experience, if we wish to motivate researchers, we have to be very specific and preferably be able to demonstrate major benefits to convince them that it is worth their while taking on any extra work of this nature. For this reason we have tried to be extremely sensitive to their needs and wishes.

In other words we want the database to be seen as a useful tool for institutions and researchers in their work, as a tool for project planning; and as a means of attaining the standards prescribed by the government and authorities.

On top of this the goal of the project has been to provide:

- A more efficient means of disseminating research information, partly by being connected to national projects
- The creation of a central source of information
- Simple and methodical processing without duplication or extra work on the part of the author or institution. They should not need their own home page because they can link directly to the database.
- Access for everyone, via the Web
- Access to full text versions of the material as far as possible.
- Simplify the identification of a certain document.
- Procedures for cataloguing documents in electronic form and hard copy.
- Enable the generation of reports and printouts of the contents of the database for specific purposes at the institutions
- By extension create a basis which may be used for publishing
- Simplify communication between the University and local industry.
- Meet government requests to co-operate with and supply information to industry and the community.
- Provide administrative procedures to safeguard integrity when processing research material at the University.
- Promote the research activities of the University both internally and externally.

The contents of the database

From the start we wished to create a database which contained all the research material written by or in conjunction with staff at the University.

The University was established in 1989 and it was not until more recent years that research activity really got going. There are now 6 departments with about 250 staff and 2000 students:

Department of Telecommunications and Mathematics
Department of Spatial Planning and Civil Engineering
Department of the Humanities
Department of Mechanical Engineering

Department of Signal Processing

Department of Computer Science and Business Administration

This means that the number of research documents is still easy to take in stock. For the purposes of the database, the documents can be classified as follows:

1. Research Reports
2. Articles
3. Conference Proceedings
4. Monographs
5. Licentiate dissertations
6. Theses
7. Book chapters
8. Research projects
- 9*. (Master's Dissertations).
- 10*. (Information for the industry)

- *These are not as yet included in the database but will be during the spring 1998

-

1. Research reports of the University of Karlskrona/Ronneby

This series is common to the whole university, and contains articles, proceedings and licentiate dissertations. At the moment there are about 20 new issues a year. The level of activity varies considerably between the different departments, but an increase can be expected.

2-7. Newspaper articles, conference proceedings, monographs, licentiate dissertations, theses and chapters of books

There is at present no common register for documents of these types. Our ambition, copyright permitting, is to scan in these documents and publish them in full text form (PDF or HTML).

8. Information on current research

To keep users up to date on current research we use a web-form similar to that used for already existing documents. The idea is that the database and its contents will be technically compatible with specifications from national bodies such as the Swedish University Board or supranational bodies such as the EU project, Cordis.

9. Masters Dissertations

These often form part of a research project. It is also in the interests of local industry if attention is brought to these projects as many of them are the result of collaboration between the University and individual firms.

During the spring project work from Master's degrees will be added to the database and work from Bachelor's degrees will be incorporated in a parallel database. We are working together with Teknopol (one of the University's channels of contact with industry in the region), which is setting up an agency for liaison on project work

between students/departments and companies/authorities. By integrating these two projects we facilitate the successful search for personal contacts, knowledge and material, making the exchange of information between the University and local industry easier.

10. Information for the Industry

One of the aims of the University is close co-operation with industry in the region. The research database ought to be a possible source of information for companies wishing to know what research is going on and which projects researchers and students at the University are involved in. The database should provide an interactive means of communicating about research work and the possibilities for related practical work to be carried out together with one of the companies.

Design

Our aim was to create a database which could be maintained and developed within the existing organisation of our library. Technology itself is not our strong point. For this reason we wished to avoid having to write our own software or order custom-made software from outside consultants. So we appraised software from major reliable firms who could provide ready-made well-tested database systems compatible with WWW and who could guarantee up-grades. Our final choice was Lotus Notes produced by Lotus Inc., a subsidiary of IBM.

Several factors influenced our choice, but most of all Notes offered a ready-made solution which on initial inspection seemed to offer the facilities we required. The server program was reliable and cheap. With just a little support we were able to develop precisely the WWW-interface we wanted, and in doing so were able to retain and even enhance our own ability to administer the system. We were soon able to get started and create a system which could be further developed.

The starting point for the design was that the researchers themselves, using a web-form, enter the data into the system. This means that they do not need to learn to use any new software and that all information received comes directly from the original source, ensuring utmost integrity.

When designing the input form we tried to keep it simple, with as few fields as possible. All the forms were to be similar. The input process would take advantage of the fact that the database could fetch information from tables and insert these in the appropriate fields depending on what the researcher chooses. Since the researcher's original document is already stored in HTML or in programs such as Latex, FrameMaker or Word, we wanted to facilitate cutting and pasting from the original document – anything to minimise the work involved in creating new records.

We had to change the design of the user interface quite radically from its original form, but even at the beginning we aimed to make it easy to browse through records and through subjects/types of document/institution. Most users find the information they seek by browsing rather than using sophisticated search techniques^{4,5}

Technical Specifications

The research database was developed during the autumn using software from Lotus Notes Domino 4.6 . For the server we used a PC: Dell System 2200/266MHz Poweredge 2200/266 Pentium II with battery back-up.

Client: Netscape 3.02 – 4.x or Microsoft Explorer 3.02 – 4.x

Classification and Indexing

One of the hardest tasks in setting up the database was to decide on procedures for classification and indexing. Since the material covers such a broad spectrum – from humanities and social sciences to computer science – and because there is no established system which covers all these areas, we have chosen to use a simple keyword classification system tailored to those subject areas where research is currently taking place at the University. We have taken the Swedish General Bibliographical Standard as a basis and adjusted it to suit our needs in consultation with those responsible for research in each department.

This means that when inputting a new record, the researcher chooses a subject word from a list on the screen. We realised that in our case it would be unrealistic to ask that the researcher use a large complicated classification system such as UDK, Dewey or LC.

We have also given the researchers the freedom to choose their own keywords for indexing. We work on the assumption that the researchers are themselves capable of defining these for their particular field.

Six of the input form fields are generated automatically as metatags within a field which, when the document is converted to HTML, are included in the header section.

These fields are Title, Author, E-mail address, Keyword, URL and language. The metastandard used is DublinCore. Here is an example from the database:

```
<META NAME="DC.date.current" CONTENT="(SCHEME=ISO31) 1997-10-31">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#date">
<META NAME="DC.title" CONTENT="Quility Monitoring in Robotised Short Circuiting GMA
Welding">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#title">
<META NAME="DC.creator.name" CONTENT="Stefan Adolfsson">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.creator.email" CONTENT="stefan.adolfsson@isb.hk-r.se">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#creator">
<META NAME="DC.subject.keyword" CONTENT="Quality monitoring, GMA Welding,
Detection">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#subject">
<META NAME="DC.identifier.url" CONTENT="http://www5.hk-
r.se/fou/forskinfinfo.nsf/b8396790ab81883ec125647600412b54/931EA3D3E02C61F6C125654E00457E
81?OpenDocument">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#identifier">
<META NAME="DC.language" CONTENT="(SCHEME=Z39.53) ENG">
<LINK REL=SCHEMA.dc HREF="http://purl.org/metadata/dublin_core_elements#language">
<SCRIPT LANGUAGE="JavaScript">
```

We realise there is a weakness in not adopting a major well-tested system of classification. For this reason we will in the spring be including an additional element in the above metatag. In each document the administrator will enter a CERIF (Common European Research Information Format)-classification. For common market countries this system is becoming a standard for subject classification and exchange of research information.⁶

Another reason for doing this is so that we can successfully link our research database to national and international projects through which, using *gateways* – computers which link two or more networks, one will be able to search different databases using the Z39.50 standard.

Searching and navigation

We have put a lot of effort into making it possible to navigate the database to a large extent by *browsing*, i.e. where the user arrives at the relevant section without having to formulate any search questions. We feel that this approach is important because the target group is so wide.

We have tried to keep the search page itself as simple as possible. For the most part the user searches by author or title. The headings Author or Title do not cause any problems, even for the most inexperienced database user. If the required document cannot be found using these fields the option is given of doing a full text search – i.e. a search not only of all the fields but also through the full text versions of all the documents

The database does not offer facilities for saving search results or for searching through fields other than those predefined.

The presentation of the search results still leaves much to be desired. We have not made so much progress here, but are aware of the problem, as we are of the problem of generating reports both on the screen and to file.

One of the strongest wishes of the researchers at the beginning of the project was that their material should be searchable from Alta Vista and other major search engines. This is a major problem, both for search services and database suppliers such as Lotus and Microsoft. Dynamic databases, which generate HTML pages “on the fly” are simply not searchable from these engines, which just do not understand the URL-addresses the databases generate for their documents.

Since this was a fundamental requirement from our researchers we had to find a solution. To this end we used an older Lotus program – Internotes Webpublisher, which creates HTML documents from a Notes database. Thus using this program we can create HTML documents containing fields of interest for search engines such as Alta Vista. We believe that in the near future search services offered by AltaVista and others will improve. They will be under pressure from users and database operators to find ways to index database documents too. For this reason we have prepared ourselves by creating a header-field in each document which includes metatags such as DublinCore. See also under “Classification and Indexing”.

Document Input

Documents are input as follows:

When a document has been selected for publication in the University research database, it is first examined at department level to confirm that it is in fact research, and that it is of a standard acceptable to the department and the University. Each department is responsible for this. Thus the receipt and selection of material is entirely decentralised and delegated to the individual departments.

We, the research editorial committee, have emphasised how important it is also for the departments to make their research accessible through reports in our own periodicals. We have tried to make it easy and cut down on any additional work, and wish to point out the advantages:

Through “pre-publishing” in such a form as the periodical, the publication becomes to some extent an official article which can then be distributed without inhibition at seminars. Conference and periodical publishing is well known to take months or years. Registration means that the results cannot be stolen after they have been printed in the research reports; activities are publicised immediately at the University and on the Internet for the benefit of the researchers themselves and their departments. This helps create a good atmosphere at the University.

Note that international periodicals do not normally have any objections to prior publishing of results in local research reports.

We encourage the authors to be aware of copyright laws and of the agreements signed. We hope that as many documents as possible can be presented in full text form as PDF-files. This presents no problem for our own reports but for other types of document we are obliged to restrict ourselves to bibliographical details. By encouraging “pre-publishing” in the research reports we hope to be able to present some full text versions of the documents which would otherwise only be available in print.

When the department has given the go ahead the researcher opens the following URL using their web-browser:

<http://www5.hk-r.se/fou/forskininfo.nsf>

Then they click on the link To the Archive, then Submit Document, fill in the User ID and password which they obtain from the database administrator at the library.

Next they choose an appropriate form.

There is a help-link for those entering a document for the first time.

The author must inform the administrator via e-mail how many copies of the research report are to be printed, to whom they should be sent and who is to be invoiced for the printing costs. The latter, usually the department secretary, should also be sent a copy of this message.

Validation of data

After input an e-mail is sent automatically to the administrator informing that a new record has been added to the database. The administrator edits the record and, if the document is one of our own research reports, gives it a number and a fly-leaf using a PageMaker template, then creates a PDF-file. This is sent to the printers who print the required number of copies and an additional 5 copies for the library. By law a further 7 copies are sent to the Swedish University Library.

Before paper version is printed the electronic version is already in place in the database.

The administrator also submits the research report to the national research library database LIBRIS, where it is searchable and is linked to the relevant document in our database. The report is also included in our own library catalogue.

By request we have included certain options for researchers whose documents are converted to PDF-files. They can choose whether the PDF-document can be printed out or not and whether it will be possible to copy sections of text from it.

Similarly there is a button where they can indicate whether or not the original document being added to the database can be made available on the Web.

An original file (Word, Latex, FrameMaker) is attached by clicking on Browse, whereby it is then possible to locate the original file on their own hard disk. When the file has been found they click on "Send to database". A PS-file of the original file, needed when the administrator creates a PDF-file, is attached in a similar way.

The author can only make changes to the record during the interval between when it has been sent and when the administrator has entered the PDF-file. After that access to the file is denied to all but the administrator. If the author wishes to make further changes then the administrator must be contacted.

Publication

Today it is natural to choose the WWW as a medium of publication for our research documents. There is no other medium in the research world which has made such an impact and which is so effective in relation to publication costs.

We use PDF-files as a simple way to publish documents in full text form. It is a cheap and simple way to create files which can be read over the WWW. The researcher who feeds in the document information does not need to be familiar with or use Adobe Acrobat software. Although there are researchers who do send completed PDF-files directly, most simply attach the original file and a PS-file, which of course can be created by anyone who has a postscript-printer drive installed on their computer.

In the research database we pick out all documents which have been submitted during the last month and present them in a news bulletin, from which they can easily be forwarded to the University's own news page on WWW. We have also discussed publishing summaries of new research documents in the local press and in the University's own printed information channels on a regular basis. So far, however this remains merely a possibility and has been given low priority.

We have also discussed developing a more interactive interface for staff at the University whose job it is to make contact with companies in the region. It would be conceivable that companies / product developers could look for expertise using our database and they could also make suggestions for project work for students using a web-form.

Maintenance

Through our control group and through ordinary e-mail we strive constantly to listen to the needs and wishes of our researchers in terms of new ideas for development and new features for the database. Lotus Notes is such a simple system that we have so far been able to adapt and maintain the database almost entirely ourselves. We have had

about 40 hours of consultant support. It is our experience that this type of support and close contact with expert developers is essential for the development and maintenance of a database of this type.

The day-to-day administration is currently dealt with by one person working part-time. It is difficult to assess how much time will be needed for development, editing and input of new records. Roughly speaking we can say that it takes between half an hour and an hour to edit a record and create the PDF-file.

Comments

In a project of this type some problems on the way are inevitable. We have solved many of them but some remain. Even in a small university like ours it is sometimes difficult to inform everybody about the project and what it could mean for our organisation. The research committee and the University board have not yet decided on the status of the research database. Submission of research material is still voluntary. In academic spheres departments usually treat their publications in their own way without instructions from above. We have therefore constantly had to justify the database, outlining its advantages for researchers and the departments, and have consequently understood the importance of promotion and sensitivity to researchers' needs. To be able to offer researchers a viable and advantageous system is important, but equally so is that there is a system for handling and cataloguing research documents which applies to the whole organisation. This is planned for the spring.

The question of copyright is another issue which spreads uncertainty within electronic publishing. Clearly we would like to publish everything in the database in full text, but there are problems.

In practice we get round the problem of copyright by only publishing our own research reports in full text in PDF-format, together with those where the author still owns the publishing rights. We warn authors to be aware of copyright when they sign a contract with a publisher prior to the publication of an article, so that they know whether or not they have the right to publish electronically on the organisation's own web pages.

Subject classification of records has not been totally clear. We have used our own system, adapted to the research actually taking place, and have doubts as to whether this is the right answer. For this reason we are considering using CERIF, which is the EU proposal for a subject classification standard for research. We shall be appraising this system for possible implementation during the spring.

External search engines such as AltaVista and HotBot and so on the WWW have difficulty in indexing dynamic databases such as ours and cannot as yet search for meta-information such as DublinCore. This means that our documents have limited access from these engines.

We hope that the major search services will develop engines which can reach individual documents in dynamic databases. We have prepared for this but for the time being we use automatically generated HTML-pages as a parallel to the research database on a web-server, to make the contents of the database searchable from the outside.

References

1. Government proposal 1997/98:1 budget proposal for 1998 Vol. 8 chap. 5.5.9 Doctorate theses. "The theses can be distributed to a broader public by, for example, being put on the university's electronic research information system"

2. Government authorisation for the budget year 1998 concerning grants to universities and colleges of higher education. page 7.

"The universities shall to a greater extent co-operate with the surrounding community, for example industry, authorities, organisations, cultural organisations and educational institutions. This means that the universities shall

- Be more receptive to experiences and problems from outside
- Keep the public informed about its courses and research
- Make it easy for the local community to gain access to relevant information about research results (cf. higher education act 1992:1434 chap 1 § 2)

By 10 May 1999 at the latest the universities must submit to the Department of Education their plans for co-operating with the community, and what measures have been taken during 1998 in connection with this plan. The same report should include details of steps taken to develop a research information system, especially with reference to Internet. Such a system should be in operation by the beginning of 1998, in accordance with the government commission to universities board 21 November 1996.

3. National Agency for Higher Education. Publishing information on Swedish research on Internet. URL:
<http://www.hsv.se/verksamhet/utredningar/safari/index.html> (1997-02-03)

"The National Agency for Higher Education has been commissioned by the government to develop a system for publishing research information on Internet. For the background to this decision see the proposal "Research and society 1997". The project, named SAFARI, aims at making it easy for the general public to gain an insight into current research, enable companies to find people with the required knowledge, and researchers a fast, cheap, high-integrity means of passing on information about their own research to colleagues and others around the world and of seeking researchers and information connected with their own current research.

4. Yuan, Weijing. End-user Searching Behavior in Information Retrieval: a Longitudinal Study. *Journal of the American Society for Information Science*. 48(3):218-234, 1997.

5. Chang, Shan-Ju; Rice, Ronald E. Browsing: A Multidimensional Framework. *Annual Review of Information Science and Technology*. 28: 231-276, 1992.

6. European Commission. Community Research and Development Information Service. URL: <http://www.cordis.lu/> (1997-02-03)

"CERIF is the Common European Research Information Format. Developed in 1988 by a European Working Group on Research Databases, CERIF provides a common usable format for databases containing information on current research projects. The CERIF research documentation standard was published in the Official Journal of 13 July 1991 (OJ No L 189 of 13.7.1991), following a Commission."