Research Information Systems in the Nordic Countries - Infrastructure, Concepts, and Organization

Ingegerd Rabow

▶ To cite this version:

Ingegerd Rabow. Research Information Systems in the Nordic Countries - Infrastructure, Concepts, and Organization. 2009, pp.64. hprints-00433868

HAL Id: hprints-00433868

https://hal-hprints.archives-ouvertes.fr/hprints-00433868

Submitted on 20 Nov 2009

"Once the computer is put into the communication loop, the potentials for structuring, facilitating, and augmenting the communication and information exchanged among members of a research group are virtually unlimited."

. Hilz & Turoff, 1978

Research Information Systems in the Nordic Countries

Infrastructure, concepts and organization

A Report Commissioned by the Nordbib programme

June 2009/ Rev. October 2009

Table of Contents

Index of Tables

Index of Figures

1 Introduction

This report is commissioned by the Nordbib programme, and is based on a web survey of the current status of CRIS (Current Research Information Systems) and IR (Institutional Repositories) in the Nordic countries.

The survey has been conducted to investigate how Nordic higher education institutions collect and present their research output. Do they use Institutional Repositories and/or Current Research Information Systems, are these systems separate or integrated, what software is used, and how are they staffed and financed? An important part of the survey was to analyse the perceived needs for national and Nordic coordination and support regarding such specific issues as rights management, central search services, educational and promotional materials etc. The survey results are presented against international developments in Open Access, both historical and current.

The main purpose of the survey is to help Nordbib gather relevant information for promoting greater visibility to Nordic research and offer a background for creating a joint Nordic approach to further developments in Open Access.

A presentation of the survey project was given at the Nordbib Workshop on "Research Visibility – managing quality for better evaluation" 27-28 October, 2008.

We would like to thank the members of the Nordbib workshop group for valuable viewpoints regarding the final formulations in the questionnaire. We also wish to thank all our contacts in the Nordic countries for their invaluable help in identifying respondents and providing us with email addresses, and other valuable advice and support,

2 Methodology

We used an earlier survey commissioned in 2007 by the *SUHF* (The Association of Swedish Higher Education) as a template for a somewhat more inclusive survey to be sent to all Nordic countries except Sweden.[1] All questions were written in English. This initial questionnaire was presented to the Nordbib *Workshop 08* group for feedback. They suggested a number of

---

1   See Danelid & al. (2008)

changes, the most significant being a number of new questions relating to *Current Research Information Systems*.

In order to distribute the questionnaire and process the answers, we used the open source *LimeSurvey* application. LimeSurvey is written in PHP and was easily installed on a fairly standard *Fedora Linux* system running the *Apache* web server (though we had to fix some minor issues).

The web interface provided through LimeSurvey was also evaluated by the Workshop group, in order to make the design as user-friendly as possible and avoid any ambiguities. We modified the style sheets for increased readability, and put in various answer constraints to decrease the risk of typing mistakes. Many questions were also made dependent on the answers to previous questions, in order to facilitate the response and make respondent errors less likely.

HE-institution library directors in the targeted countries were selected as the primary respondents or intermediaries. A list of names with corresponding institutions, email addresses and phone numbers was compiled with the help of the *Nordbib workshop group*, various programme groups and other experts directly contacted. The initial list contained 117 names: 14 from Denmark, 51 from Finland, 9 from Iceland and 43 from Norway. (The numbers varied greatly between different countries, depending on the respective organization of higher education.) The respondents were selected in order to provide a complete map of the situation at all HE-institutions. We did not think it was possible to use representative samples.

An email template was created that provided a brief background for the survey and emphasized its importance. There were also instructions on how to proceed to complete the survey and detailed contact information, including email and phone numbers. LimeSurvey uses a simple token system which assigns a unique token to each respondent. Thus it was possible to supply a unique URL (web link) in each mail. By clicking on this link in the email message, the respondent would be taken directly to the survey without having to log in. In case the recipient of the email wished to delegate the task to someone else, it was very easy to simply forward the message to that person.

We decided to use a personal university email account as sender of the initial mail invitations in order to decrease the risk of interception by spam filters.

Unfortunately we discovered that some old email software corrupted the mail messages and thus destroyed the links to the survey. We had to manually resend invitations to those recipients

with instructions on how to copy the URL from the email message to the browser.

There were also a number of other technical problems that need not be discussed here.

We also received some responses, saying that they were not the appropriate persons to answer our survey, and that they did not know where to forward the invitation. We thus had to manually adjust the initial list, adding and removing a few names.

We obviously also received a number of automatic out-of-office replies, but this was not a problem, as that would be handled by the normal reminder system. The LimeSurvey application keeps track of which tokens have been used and thus makes it easy to handle reminders. A special reminder email was crafted and sent after a reasonable interval. Altogether nine differently worded email messages were sent as reminders to different respondent groups.

At the end of this process there was still a large number of recipients from whom we had not heard anything. We began to contact some of these recipients via the phone, but this was a very time consuming process with a low yield. (Many proposed respondents promised that they would complete the survey but the survey was never completed according to the LimeSurvey application.)

We also constructed a smaller survey on the same LimeSurvey platform that was sent to Swedish recipients in order to complement the results from the SUHF survey and make them more compatible with the results from the other Nordic countries. This survey was sent to 42 recipients.

Once we decided that we could not get any more answers, we migrated the database to a different system, and proceeded to evaluate the data.

We have also retrieved all data for the Nordic countries from *OpenDOAR*, *ROAR*, and the *DRIVER* inventories to complement and cross-check data from our surveys.

In order to aid the interpretation of survey results and propose recommendations, we have of course also made use of a large body of literature relating to different aspects of research information systems.

3 Background and History

This chapter attempts to give a general overview of the main issues relating to research

information systems. Usually the topics and concepts have been presented in their historical context to better convey their relative importance and significance. The main headings are "Current Research Information Systems" and "Institutional Repositories", corresponding to the main divisions of our survey. Research information systems were widely implemented before most institutional repositories, so they will be discussed first. Many full-text repositories were also initially created as auxiliaries to research information systems. Given that advanced institutional repository systems also provide *CRIS* functionality, the detailed discussion of standards and software solutions has been reserved for the section on *IR*.

3.1   Current Research Information Systems (CRIS)

At least since around 1970, serious efforts have been underway to develop international standards for research information systems. The main focus was on bibliographic data interchange and machine readability. These efforts were primarily motivated by the exponential growth of scientific literature and the rapid development of automated systems for scientific and technical information. [2]

Much of the early work was done within the frameworks of *UNESCO* and the *Smithsonian Institution*

Within *UNESCO* and *ICSU* (the *International Council of Scientific Unions*), a project began in the late 1960s to coordinate scientific information throughout the world and integrate it into, what was called, a "World Science Information System". The system concerned the basic sciences and was primarily intended for scientists themselves. It also concentrated on published literature, primarily journal articles and, where applicable, monographs.[3] The work within this project focused on standards for bibliographic descriptions (i.e. metadata standards), procedures for creating abstracts, the formats of scientific documents, and standards for indexing and classification.

 The *Smithsonian Science Information Exchange* (*SSIE*) was set up to facilitate planning and management of federally funded R&D. It had existed in some form since 1949, and contained information about research projects in the "life" and "physical" sciences. (The social sciences were included under life sciences.) Around 1970, close to 100 000 projects were added

2   Wysocki (1976) p 401
3   See UNESCO (1971)

annually. At the outset, the *Exchange* was used solely by government agencies to keep track of their funding, but gradually access was given to other types of clients, and more information services were added.[4] Besides a brief abstract describing the project, the database contained funder, budget, principal researchers, institutions and dates.

When discussing research information, it is customary to distinguish between the user context and the researcher context.[5] Frequently the user context has also been further divided into such categories as "practitioner", "educational", "managerial", "policy making", or "general public".[6] Beside various ways of classifying the *consumers*, research information has also been analysed in terms of the type of *use* or *utilization*. Donald Pelz and others have distinguished between the instrumental, symbolic and conceptual use of knowledge, for example.

Although, as we have seen, the early discussions about research informations systems took into account many different user categories and types of use, the actual technical work tended to be almost entirely focused on the researchers as consumers, as in the UNESCO/ICSU project. The second most important user group were the main funders, such as research councils, who wanted to keep track of their funded projects in project databases such as the *SSIE*. The problem of adopting the systems to different uses of information was often ignored in the early technical reports, and much of that discussion simply assumed that previously existing information systems (i.e. various paper-based bibliographies and library catalogues) should be adopted to electronic form. A number of limitations were posed by the high cost of memory and storage in the computer systems of the time.

The work to integrate research information systems progressed very slowly, and when, in 1992, the *Netherlands Agency for Research Information* (*NBOI**) began to create a world-wide directory of research information systems, they found about 100 different systems with different content, formats and means of access. In 1996 they made the *Directory of Research Information Systems* available on the web, and by then there were 70 different systems listed, of which 85 % were available online. Only 19% of the systems had worldwide coverage. 59% covered one or more European country, and 11% covered North America. See Table 1 below for an overview of the types of data included:

---

4   See Kreysa (1972) for a brief overview
5   Eg Landry & al (2001)
6   Eg Beyer & Harrison (1982), Closs & Cheater (1994) The UNESCO report cited above mentions, besides the producers themselves (researchers and engineers), "administrators, managers, policy-making groups, defence officials", and "educators". (p. 31)
*   *Nederlands Bureau voor Onderzoek Informatie*

Institutional	Researcher	Project	Publication	Count
No	No	Yes	No	26
No	No	Yes	Yes	14
Yes	No	Yes	No	7
Yes	No	No	No	7
Yes	Yes	Yes	No	5
Yes	No	Yes	Yes	4
Yes	Yes	Yes	Yes	7

Table 1: The types of data in research information systems 1996. Adopted from Lavieter (1997).

Gradually, however, there was more emphasis on, what was often called "innovation systems", and the need to make research results accessible for economic, health, social and political benefits. Thus, when , in the late 1980s, work came under way to integrate the European research information databases, there was more focus on other users besides researchers and project managers. A committee was set up under the EC commission for this purpose, and its first task was to integrate three major research project databases : the Flemish *IWETO,* the Dutch *CILO* (formerly *NBOI*), and the British *BEST*. When the first workshop on *European Research Databases* was held in Brussels in 1987, the emphasis was on the use of databases for policy making and evaluation, and with the long-term goal of making the databases accessible to the general public. The *European Working Group on Research Databases* was formed, and they recommended the use of the *Common European Research Information Format* (*CERIF*), as the basis for integration.

In 1994, the European Commission created the *European Research Gateways Organisation* (*ERGO*), under the "Specific Programme for the Dissemination and Optimisation of the Results of Activities in the field of Research, and Technological development, including demonstration". In this work, a number of user categories were targeted besides the researchers themselves. They were industry inventors and entrepreneurs, Innovation Relay Centres, policy makers and R&D managers, information professionals (e.g. librarians), and , to some extent, the general media.

At about the same time the *European Platform for Current Research Database Producers*, (*EuroCRIS*) was set up, and its secretariat rotated biannually among its members. In the second

period it was located in Norway at the *University of Bergen*, which had also been the location of the first biannual European CRIS conference.

The *Norwegian Computing Centre for the Humanities* at the *University of Bergen* had extensive experience in this area, and they maintained a database of research projects funded by the Norwegian research councils. By 1990, this database contained about 25000 documents, and five years later, when EuroCRIS had been created, that number had more than doubled.

Norway played a leading role in the work with Current Research Information Systems in the Nordic countries, and another important early system was also located there: *FORSKDOK*, maintained by the Norwegian *BIBSYS*. It contained information about research projects, publications and other results, research institutions, and a systematic subject classification.

Towards the end of the 1990s, mainly because of the growth of Internet, various initiatives were launched in the Nordic countries (and elsewhere) to develop more advanced research information systems in order to integrate large numbers of data sources, and present them according to universal standards.

In Sweden, the *National Agency for Higher Education* (*Högskoleverket*) was entrusted with the task of coordinating the first system for disseminating research information on the Internet. The system was presented in 1998 under the acronym *SAFARI*. In SAFARI, each HE-institution catalogued its own research and assigned metadata according to a predefined standard based on *Dublin Core*.

In 1988, the *Ministry of Research and Information Technology* in Denmark established the *DANDOK* database, which consisted of three parts: a normal reference database with published research, a project database and profiles of major research institutions. "DANDOK-basen" was maintained by the *Information Service Department* at Risø. (The DANDOK was originally created to coordinate the information and documentation services in Denmark for "research, higher education, industrial and commercial enterprises". [7]) After 10 years, the database contained more than 150 000 research references.

DANDOK-basen later became the "Danish Research Database" and was integrated in Denmark's Electronic Research Library by the year 2000. The attempt to integrate bibliographic data with project and institutional data was abandoned and replaced with links to external information sources.

---

7   Hansen & al (1983)

In Norway, the *Norwegian Current Research Database* was set up as a searchable database of projects funded by the Norwegian Research Council. In 1998, under instructions from the *Ministry of Education, Research, and Church affairs*, this was extended to a national database for "research documentation". The main stated purpose of the Norwegian database was to link "research activity" with "research results". The database was supposed to interface with external data sources containing additional data about projects, publications, institutions and more. The main motivation appears to have been to aid the general application of research findings, a "user-focused knowledge and technology transfer".[8]

3.2   Institutional Repositories

This section will begin with an outline of the development of institutional repositories in relationship to electronic publishing in general. The second subsection will briefly describe the relationship between research information systems and document repositories. The final subsection will discuss the crucial role of institutional repositories in relation to *open access*. The term *open access* will be understood in this report in the conventional sense. It is a system of publication that makes research results freely available via the Internet without any barriers whatsoever. Users may download and distribute the data without any restrictions, as long as the integrity of the data and its creator/author attributions are not compromised.[9]

3.2.1   Electronic publishing and the role of repositories

Most of the early information retrieval systems all tended to have a fairly similar structure. The *DIALOG* system, developed at *Lockheed Corporation* 1964-66, could be used as a model. The user would begin by selecting a file to be searched, then enter a boolean combination of search terms, and finally select items from the result set to be printed. It also had an "expand" command, which let users choose terms by entering a few letters.

At the same time, already the NLS system developed at Stanford Research Institute in the

---

8   See Stortingsmelding. nr. 36, 1991-92
9   This definition derives from the major *BBB* declarations on OA. (Budapest, Bethesda and Berlin)

1960s had facilities for creating, discussing, revising, formating, reviewing, and publishing scientific papers. In 1963, Chaitin and Bourne had constructed one of the first institutional repositories at SRI. It contained memos, bibliographies, technical reports, software, and documentation. It allowed both keyword and full-text searches.[10] Thus it made sense to incorporate and expand these features in the larger system.

It is interesting to note that one of the first mission critical information retrieval systems, the EMISARI system in Washington D.C, was, despite its simplicity, also an integrated communication system, with not only facilities for data retrieval, but also modules for computer conferencing, electronic mail, file sharing, etc.[11] The experiences with EMISARI were used for the later EIES system, developed in the 1970s at the *New Jersey Institute of Technology*. EIES stood for *Electronic Information Exchange System,* and was designed in order to facilitate and improve scientific communication. A number of studies had shown how most of the communication within a scientific discipline took place before publication. The scientific journals thus mainly served the need of informing researchers in other fields about important findings. EIES was set up to improve pre-publication communication among scientists. It included facilities for messaging, conferencing, shared notebooks, newsletters and journals.[12]

Because of the focus on the prepublication phase, an "electronic journal" was often seen as more akin to an electronic bulletin board than a paper journal. The only major distinguishing characteristic was that the journal only contained articles that had passed *peer review*.

 As John Senders put it in 1976:

"The 'journal' would once again exist in its etymological sense: i.e., publication would be carried on every day. It would, in fact, be a continuous process, and the present form of separated subsections of the scientific literature now called journals would no longer be necessary."[13]

There were also new systems for making research data directly available, such as the NIH-EPA Chemical Information System from 1973. This also led to renewed practices of publishing, where the data corresponding to a research article was published separately in a database.

Thus early electronic journals tended to have fluid boundaries and were seen as part of larger

---

10 Bourne & Hahn (2003) pp 14-17
11 For details of the EMISARI system see e.g. Kupperman & Wilcox (1972).
12 Hiltz & Turoff (1978) pp 18 ff
13 Senders (1976)

information and communication systems.

At the same time publishers of scientific journals had many reasons to be apprehensive about these developments. They were concerned about how to protect their brands in a world where information became more diffuse, and the existence of data in electronic form gave rise to worries about copyright protection.

When early digital versions of well known scientific journals appeared, they tended to mimic the paper versions. The electronic format was seen essentially as a new mode of distribution, rather than a new way of publishing. Instead of distributing the articles in paper form, they were distributed in electronic form, and they were meant to be printed and read on paper. (In this sense they were seen as mainly a development of earlier distribution projects, such as e.g. Elsevier's *ADONIS*, which began in 1979 and provided access to electronic versions of paper journals on optical disks.)

On the other hand, as we have seen, many information specialists and scientists had long seen a revolutionary potential in electronic publishing and worldwide network access. Visionaries, such as Bruce Schatz at the *National Center for Supercomputing Applications* and Stevan Harnad at Prince ton *University*, wanted electronic journals and archives that would improve scientific communication in general, and not just make the distribution of articles more efficient.

Harnad and others gradually developed a new electronic journal from the combination of a discussion group and an FTP archive. He wished to reform the entire process of submission, peer review, and publishing, and argued that electronic publishing would profoundly affect all phases of scientific enquiry.[14]

There were also other influential projects with (at least initially) less ambitious goals. Paul Ginsparg at the *Los Alamos National Laboratory* wanted to provide a way of distributing preprints which was more efficient than email.[15] His work resulted in one of the most successful document repositories, *ArXiv.org*, which gradually came to introduce many of the new information sharing technologies envisioned by Schatz, Harnad, and others.

When peer reviewed electronic journals first became available in the late 1980s, they were met with much scepticism. As a former editor of the *New England Journal of Medicine* put it

---

14 Harnad (1990)
15 Ginsparg (2008)

succinctly in 1991: "The hard-copy journal is hard to beat".[16] Yet, information specialists had, at least since the 1970s, pointed out a number of reasons why they saw the future in electronic publishing. The main advantages were lower cost per delivered article, decreased time interval between submission and publication, and less limitations of space (storage).[17] Early electronic journals also quickly began to exploit some of the possibilities offered by the new medium. *Current Clinical Trials*, for example, which first appeared in 1991, and required special software for full functionality, linked to the abstracts of referenced articles and had a subject-based alert service. It also functioned as a kind of preprint archive for *Lancet*.

In 1990 William Gardner found three principal advantages in using *archives* instead of *journals* as the main form of electronic publishing.[18] A: The structure of an archive is more easily adopted to users with different computer skills and tools. B: The archive may be *personalized* to fit the research profiles of individual researchers. (In essence, each user creates her own journal.) C: The archive can allow more efficient *discovery* and *retrieval* of information in an article. This may be done through user profiles and the creation of links between documents and document segments.

3.2.2   Research information systems and repositories

Despite considerable opposition from some groups, as outlined above, an increasing proportion of scientific publications became available electronically during the 1990s. More and more electronic journals began to appear, and large preprint archives such as the *arXiv (*created 1991 as "hep-th"), *NCSTRL* (launched at Cornell in 1995), and *CogPrints* (launched 1997 by Harnad), became the main sources of scientific information in many fields. Many large research institutions also began to deposit various reports, dissertations, and other types of non-journal literature and data electronically. [19]

Naturally, as various initiatives came under way to integrate research information in large *Current Research Information Systems,* as described above, it made sense to add links to the

---

16 Palca (1991)
17 E.g. Clayton (1983)
18 Gardner (1990a)
19 Compare the discussion in Crow (2002). Crow distinguished between four kinds of repositories publisher-centric, discipline specific, scholar specific and institutional digital repositories. Crow and other authors such as Harnad (2003), Pinfield & al (2002) emphasized the increased institutional visibility as an important motivation for institutional repositories. This is because they compare institutional repositories to the other kinds of digital repositories. But a CRIS with direct links to the published documents would clearly provide the same kind or even greater visibility.

13

full text of research publications. At the same time, the lack of control over the publishers' web sites frequently made this a very work-intensive and sometimes impossible task. The publisher could move the article to another URL without notice, or use complex schemes with login, post variables and cookies to make direct linking impossible. Also, the access restrictions normally made harvesting and URL-checking impossible, and various users of the research information system could have different access rights to the linked content. User X could have access via provider A, user Y access via provider B, and user Z no access at all.

Another difficulty was the distributed nature of full-text archiving. The maintainers of a CRIS would have to determine the location and bibliographic format for every full-text publication to be included. This problem will be further discussed in the next subsection.

The desire to create efficient research information systems was thus an important impetus for the creation of dedicated and standardized publication repositories. A CRIS could be seen as dysfunctional if the actual research results were difficult to access by the CRIS users.

Based on the desire to create functional CRIS systems, and some of the improvements in electronic publishing mentioned above, it is possible to list different functional specifications that pose slightly different demands on the underlying computer systems. Some examples of such functional specifications and the corresponding technical requirements are listed in the table below:

1. Stable location	2. Uniform access	3. Easy harvesting & indexing	4. Content-based queries	5. Many different types of data	6. Open access	7. Quick and easy publishing
Permanent url, separate links to each version, standard document identifiers. Stable document formats.	IP-independent and path-independent access. Standardized document formats. Browser independent access.	Standardized simple protocols, e.g. OAI-PMH etc. Standardized metadata and content Fast high capacity system.	Standardized open content formats. Structured document data. Efficient retrieval mechanisms. Fast system.	Basic structure of the interface and underlying database independent of data types. Flexible metadata standards. Facilities to aid the browsing and finding of non-textual objects.	Easy access to metadata and especially peer review status. Simple interface for self-archiving. Support for different work flows.	Easy interface for authors. No red tape. Tools to assist peer review and quality control. Support for work flows, version management, access rights etc.

Table 2: Some examples of functional requirements for full-text repositories,

Table 2 is only intended to illustrate the most basic requirements of research repositories and is clearly not complete in any sense. Various additional parameters will be discussed in more detail below. Suffice it here to observe that requirements five to seven demonstrate the difficulty of using centralized repositories, and that the first three requirements may easily be met by widely distributed systems. Beginning with requirement five, it is very difficult to design a system that is equally suitable for all forms of data. Thus research institutions with different research profiles tend to favour different systems that are suited to their types of data. Moving on to requirement six, many large publishers allow self-archiving of published papers or monographs, but do not allow secondary publishing. As for the seventh requirement, it is usually easier to publish quickly in a local repository where the work flow is better adopted to existing regulations and organisational parameters.

There are of course also some advantages with using central repositories. An argument in favour of centralised solutions has recently been put forward by Romary and Armbruster.[20] From the overview presented below, it should be clear that both institutional and central repositories play an important role, and also that the difference between the two may not be as great as one may first believe.

3.2.3 Open Access and Institutional Repositories

As we saw above, many of the CRIS systems were integrated with institutional or national repositories. There were also many research institutions with separate full-text repositories. The electronic publishing of theses and dissertations had begun in the late 1980s[21] Yet, so far there had been very little national or international coordination, and many repositories were only used for internal reports and dissertations. A number of initiatives came under way during the 1990s to improve this situation.

Between 1994 and 1998, the *National Science Foundation* in the US funded the *Digital Library Initiative* (later called DLI-1). Among the projects funded were *Stanford Integrated Digital Library Project (SIDLP)*, which, among other things, was instrumental in the development of the *Google* search engine. There were also other similar initiatives in other countries, albeit with smaller budgets. In the UK, for example, *JISC* (Joint Information Systems Committee)

---

20 Romary & Armbruster (2009)
21 Some notable early work in this area was carried out at Virginia Tech, where Yuri Rubinski created a *DTD* (Document Type Definition) to produce documents in *SGML* (Standard Generalized Markup Language).

funded a number of projects under its *eLib* programme. Most of these projects focused on fairly fundamental issues, but they also contributed directly to the creation of digital libraries, such as e.g. Schatz's *Interspace* at the University of Illinois, and the University of Michigan digital library which cooperated with *JSTOR*.

A JISC-funded project that should be mentioned in this context was *NetEc*, an international effort, which started in 1993, and aimed at improving research communication in Economics. Rather than creating a single subject-based archive, or some digital library at a HE-institution, *NetEc* created *RePec*, which was a widely distributed database that besides preprints also contained journal articles, software components, together with author and institutional data. A number of different interfaces in different locations and with different functionality were built on top of the *RePec* database. Any institution could easily create its own archive by simply depositing the documents into local ftp or http directories. All they had to do was to provide text files with document descriptions in the *ReDIF* format. These institutional repositories could thus very easily become integrated (using the Guildford protocol) into the global *RePec* archive, with no need for any advanced software solutions at the institutional level.

A similar system had been developed in parallel through the *CS-TR* (Computer Science Technical Reports) project. This was a joint research project between *CNRI*, Carnegie Mellon, Cornell, MIT, Stanford and Berkeley and it was funded by *ARPA*. The primary aim of the project was to provide uniform access and search facilities for technical reports in various departmental archives. These reports were normally stored in FTP directories, and the only descriptions were provided through text files in the same directory named "index", "readme" or something similar. Different institutions used different conventions which made it difficult to develop indexing and search services on top of such repositories. Different bibliographic formats were used for metadata, the reports and other files were in different formats such as e.g. Postscript, TIFF or plain text, and the dates of the files in the FTP directories may not correspond to the actual creation date of the documents.

James Davis at Xerox and Carl Lagoze at Cornell University developed the *Dienst* system.[22] It used the HTTP protocol to retrieve documents in various formats from departmental repositories. All the documents had to be given a unique persistent *docid*, which included a unique code for the publishing institution. Every document was also required to have bibliographical information attached. Bibliographic metadata were stored in a simple format

---

22 Davis & Lagoze (1994)

previously used in email messages (RFC 1357) . *Dienst* allowed users to search the abstracts and bibliographic fields of all documents at all participating institutions through a single web interface. *Dienst* was, with some additional improvements, used as the basis for the *NCSTRL* repository at Cornell University.

In 1997 the *Association of Research Libraries* created *SPARC*, (Scholarly Publishing & Academic Resources Coalition). SPARC was to find ways of combating the spiralling costs for scientific literature, and promote the use of new technology for improving scientific communication. SPARC came to play a very important role in the development of institutional repositories in coordination with research libraries.

In 1999, the Nobel Price winner Harold Varmus, then director of the *NIH* (National Institutes of Health), proposed the creation of "E-biomed". *E-biomed* was to be an electronic publishing site in the biomedical field. The site should contain both peer-reviewed papers and working papers. Working papers should be accepted without quality control and only rejected if they contained "extraneous or outrageous material". The main benefits of the new system would be open access to all reports and the ability to create personalized journals, more rapid dissemination of results, reduced costs, access to more and enriched data. (It would no longer be necessary to publish some data in an external archive, images could be provided with higher resolution, related documents could be linked, etc Varmus also mentioned other possibilities, previously discussed by Harnad and others, such as bulletin boards, open peer review, and multiple revisions. *E-biomed* was launched that same year as *Pubmed Central*. (The name was intended to convey the close relationship to *Pubmed*, the version of the *Medline* database which was made freely accessible via the internet in 1997.)

The same year work also came underway to integrate the main existing preprint archives through means similar to those used in *Dienst*. This lead to the creation of *OAI*, the Open Archives Initiative.[23] OAI created a standard protocol that allowed repository data to be uniformly harvested from institutional repositories.

The protocol *OAI-PMH* allowed bibliographic metadata in any format (as long as the format could be specified in an XML schema), but it required all repositories to support the *DC* (Dublin Core) standard, first proposed in 1995 at a workshop in Dublin, Ohio, sponsored by *OCLC* and *NCSA*. *DC* was constructed to be simple, flexible, and easily integrated with other bibliographic standards. Thus, for example, whereas the RFC 1357 mentioned above, had a

---

23 For details see Van de Sompel & Lagoze (2000)

repeating field labelled "CR-CATEGORY", which required special computer science report categories, *DC* had a field called "Subject" which could be specified according to a number of different schemes (LCSH, MeSH, Sears, AAT, INSPEC, ERIC and DDC), as well as free-form ("Scheme=Other"). Another important feature was the "Relation" field, which allowed the specification of different relationships to other documents (e.g. "supersedes", "contained in", "cites" etc).

OAI-PMH did not specify any standard for unique persistent references to digital objects. The objects were referred to using a normal *URI,* and unique identifiers should be put in the metadata records in accordance with whatever bibliographic format being used.

The protocol allowed documents to be retrieved using the identifiers from bibliographic records. It did not contain any methods for depositing or updating documents. It was simply intended to allow easy harvesting of repository data.

In 2002 *SPARC* published a very influential "position paper" supporting institutional repositories (*IR*).[24] Institutional repositories were seen as the basis of a new more efficient "scholarly publishing paradigm", and as an important way for research institutions to present a window to the outside world. The essential features of an *IR* was that it was institutionally defined, had scholarly content, had sufficient permanence, and was open and interoperable.[25] An IR would " aim to preserve the entire intellectual output of the institution ". *SPARC* expected library consortia to coordinate and assist the development of institutional repositories.

The director of *SPARC* had also in February 2002, together with Stevan Harnad and others, signed the *Budapest Open Access Initiative.* The Initiative recommended two complimentary strategies to achieve open access. One was *self-archiving* in IRs made interoperable through the OAI-PMH protocol,, and the other was publishing in open access journals. Budapest was followed in 2003 by the *Bethesda* and *Berlin* declarations which also promoted open access publishing through self-archiving and were signed by many leading research organizations all over the world.

The *European Commission* published in 2006 a report commissioned by its *Directorate General for Research*.[26] The report recommended that all publicly-funded research should be deposited in open access repositories, and further that a policy *mandating* open access archiving

---

24 Crow (2002)
25 Ibid p 16
26 Dewatripont & al (2006)

should be established for all research funded by the European Commission.[27]

The same year the *EC* also launched the *DRIVER* (Digital Repository Infrastructure Vision for European Research) project. *DRIVER* intended to make any form of scientific content, including scientific/technical reports, research articles, experimental or observational data, freely accessible via the internet. It aimed to build on existing institutional repositories and networks in Europe and integrate them into one large- scale virtual content resource . The project was supported by ten major universities and research organizations. The initial *DRIVER* project was followed in 2008 by *DRIVER II,* which continues to promote an interconnected repository infrastructure in Europe.

In 2007 a JISC-funded project called *SWORD (*Simple Web-service Offering Repository Deposit) was launched. *SWORD* aims to develop a simple deposit protocol, a complement to *OAI-PMH* to facilitate and standardize depositing across different repository systems.

4 The Current Situation in the Nordic Countries

As we have seen in the brief historical overview above, the development of Current Research Information Systems and Institutional Repositories in the Nordic countries may be seen as part of a global development, and also as part of various European initiatives to integrate research information for the use of various groups besides the producers of research themselves.

In this chapter, the current situation in the Nordic countries will be discussed. Most of the data comes from our surveys, but data have also been used from other sources, notably the *ROAR* and *OpenDOAR* databases and various *DRIVER* reports.

The first section will discuss the task of research information from the perspective of legal requirements and the goals of the *Berlin Declaration*. The second section will deal with financial constraints. The third section discusses software systems and functional requirements. Finally the fourth section will cover content-related issues.

4.1   Legal and regulatory framework

---

27 Ibid p 87

This section will, for obvious reasons, not deal with all the laws, regulations, and agreements relating to scholarly communication in the Nordic countries. The focus will be on the issues the respondents of our surveys found to be most important.

4.1.1 University Laws and Research Information

The basic legal framework for research information is quite similar in the Nordic countries. There are government mandates for all universities and similar institutions to make their research results as accessible as possible.

In Denmark, clause 1:2(3) of the *University Act* states that universities shall exchange "knowledge and competences with society and encourage its employees to take part in the public debate."

According to the Finnish *University Act*, the universities must similarly promote the "impact on society" of research results and artistic activities.

Iceland's *Higher Education Institution Act* mandates the "dissemination of knowledge and skills to students as well as to society in general."

In Norway the law speaks about informing "public administration, cultural life and business and industry."

In Sweden, the *Higher Education Act*, (1:2.2) stipulates that all higher education institutions must inform society about their activities and endeavour to make research results accessible. Recently (*SFS 2009:45*), an additional emphasis was placed on the responsibility of HE-institutions to facilitate the exploitation of research findings. In the preparatory legislative documents, however, the issue of usability is never explicitly tied to the infrastructure of research information. The focus is upon issues such as the protection of immaterial rights and access to risk capital.

4.1.2 Copyright and Publisher Licences

There are, however, also some legal complications regarding research information, and these pertain mainly to institutional repositories. One issue concerns publisher licences.

Many publishers provide authors with standard copyright assignment forms, which require them to transfer their copyright. Because of such agreements, the right to deposit a publication

in an institutional archive becomes dependent upon the publisher's permission. More recently there has been many attempts to replace these copyright assignment forms with *licences to publish*. In such a licence the author transfers certain specified rights to the publisher. All other rights remain with the author, e.g. the right to deposit in an open archive.

Major scientific publishers have begun to provide authors with standard licences to publish that do not require authors to assign their copyright to the publisher.[28] Yet this is a very complex issue, and it is extremely difficult for individual researchers to know what type of licences are compatible with funder or institutional policies. In fact, there are several studies showing that many researchers are not even aware of the current copyright status of their papers.[29] This is undoubtedly one of the areas where additional higher-level (national or Nordic) coordination would be the most immediately beneficial. One example of work that is currently being done is the transfer of the *SURF/JISC* model licence into the Nordic countries. This is carried out as part of a Nordbib project.

4.1.3   Mandating Open Access

Many major universities and research organizations in all Nordic countries except Iceland have signed the *Berlin Declaration*, which requires signatories to "encourage" making all research results available open access. Some major international funders have made OA a requirement for their support. Thus there is considerable interest in examining legal issues in relation to OA mandates.

These issues are somewhat different for universities and funders. An example of a university mandate is provided by *Helsinki University*.[30] Unless researchers deposit their articles and book chapters, the department may lose funding corresponding to the missing number of publications. Marjut Salokannel, a legal expert who chairs a working group tasked with the implementation of the OA mandate, predicted that "[t]he most complicated part with regard to the implementation of the self-archiving mandate will, at least in the beginning, probably be the relations with the publishers.". The challenge is to make sure all researchers sign the right kind of licences, and that these licences are accepted by all publishers.

---

28 For examples see Nature: http://www.nature.com/authors/editorial_policies/license.html and Science:
      http://www.sciencemag.org/feature/contribinfo/prep/license.pdf
29 See e.g. Swann (2005)
30 Salokannel (2008)

An example of a funder mandate is provided by the Norwegian Research Council. Their policy requires researchers to deposit in OA repositories but allows an exception in cases where the publishers do not allow self-archiving. This is a much softer stance than that of the European Research Council, whose policy (from 2007) clearly states that "all peer-reviewed publications from ERC-funded research projects be deposited […] and subsequently made Open Access within 6 months of publication." A recent report commissioned by the Swedish Research Council did not find any legal obstacles preventing an OA funder mandate. [31] It should be noted in this context ,that according to a recent *DRIVER* inventory, 50% of the respondents saw the lack of policies requiring OA depositing as one of the most important obstacles preventing the development of digital repositories.[32]

The table below gives an overview of the present situation of OA mandates and policy recommendations based on the survey responses.

	Denmark	**Finland**	**Iceland**	**Norway**	**Sweden**
Mandatory OA	None	University of Helsinki	None	A few inst + Res. Council	Res. Council
Recommended OA	Some	Some	None	Around half of the institutions	All major institutions

Table 3: Open access mandates and recommendations

4.1.4 Copyright and Other Obstacles to Scholarly Communication

Assistance needed

Figure 1: Need for assistance for IRs
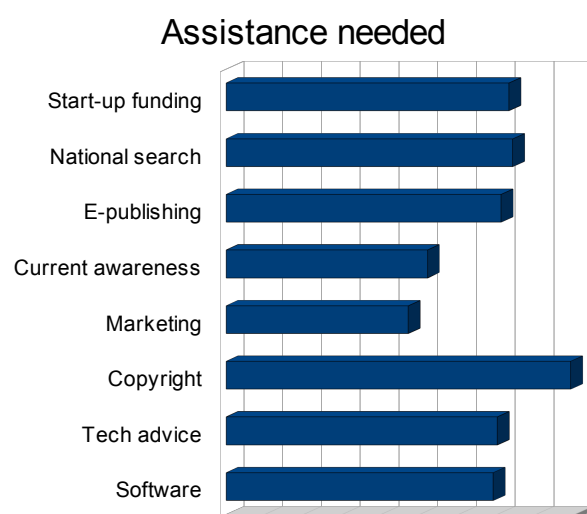
In contrast to the *OA* situation, according to our survey, major HE-institutions in all Nordic countries except Iceland mandate research registration in a *CRIS*. This strongly suggests that the legal difficulties relating to copyright agreements is the principal reason why so few institutions in the Nordic countries mandate OA .

This conclusion is further supported by the answers in our survey to questions relating to

31 Library Head Office, Lund University (2008) In October 2009, the Research Council adopted a new policy of Mandatory OA. http://www.webcitation.org/5kRf4Ywbg
32 Graaf & Eijndhoven (2008)

copyright.

The library directors were asked to rate the relative benefit of coordination and assistance in eight different areas relating the operation of repositories and research information systems. The possible choices were "not important", "important" and "very important".

These answers have been assigned the values 0, 1, and 2 and added together in order to derive Figure 1 which thus shows the relative importance of assistance in these areas. *Copyright* was clearly seen as the area were national or higher-level coordination would be the most useful.

The respondents were also asked what they saw as the greatest challenge or problem regarding electronic research publishing. This was a question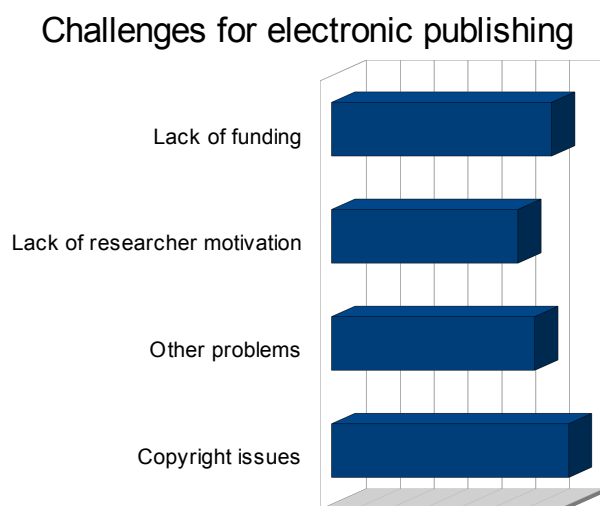 with free answers, but most of the responses fell into three categories: missing funding/resources/staff, difficulties motivating researchers, and issues relating to copyright. Figure 2 is intended to show the relative importance of these problems. It shows how copyright issues are seen as a major obstacle not only to filling the institutional repositories but to electronic publishing in general.

Figure 2: Challenges for electronic publishing.

4.1.5   Some Concluding Remarks

As we have seen, there is considerable uncertainty among both library staff and researchers concerning copyright and licensing issues.. Various publishers have different policies in regard to preprints and postprints, and these policies may change from day to day. There is a database called SHERPA/RoMEO which attempts to keep track of these changes. Sometimes, however, the policies from publishers are not entire clear, and it is difficult to know the exact limits of author rights. There may be many complicating details regarding such issues as embargo intervals, last depositable version, distribution and use of the archived papers, etc. At times authors risk unknowingly signing conflicting agreements with funders and publishers or

institutions.

4.2   The Financial Situation

From Figures 1 and 2, above, it is clear that Nordic library directors view the lack of funding as one of the main obstacles to developing the electronic infrastructure of scholarly communication. This section will discuss the economic aspects of the research information infrastructure in some detail.

4.2.1   Funding Issues and the New Publishing Models

It was mentioned above, on page 18, how, when *SPARC* was created in 1997, one of its principal aims was to combat the spiralling costs of academic research information. In their announcement, the *Association of Research Libraries* noted that during the previous decade, "despite canceling hundreds of thousands of dollars worth of serials, research libraries are spending 124% more on serials to acquire 7% fewer titles."[33]..

In 1998 Brendan J. Wyly, from a research library at *Cornell University,* noted that the "costs of acquiring scholarly information are relatively substantial in research libraries, and even in terms of entire universities' budgets".[34] He discussed the large profit margins of major publishers, and observed that in the area of scholarly publishing, "the major impediment to competition is that all of the direct incentives for authors lead them to publish in well known channels to reach the largest number of readers without regard to the cost of those channels to the readers.". Wyly concludes by noting that the current system of publication was simply too expensive and cannot handle the volume of published research findings. .Technological innovation in publishing could not only solve such problems, but would also serve the main aim of academic research, since research institutions could afford to make all their work visible and accessible.

In the 2002 *SPARC* report on institutional repositories, Raym Crow listed four essential components of scholarly communication: *registration* – establishing intellectual priority, *certification* – certifying the quality or validity of the findings, *awareness* – making others

---

33 ARL (1997)
34 Wyly (1998)

aware of the results, and *archiving* – preserving the cultural heritage.

Crow observed that the melding together of these components in the traditional paper-based publishing model tended to disguise both the fact that the academic community has stood for most of the labour and costs involved, and what the actual cost of each component really is. Thus, when technological innovation has reduced the cost of a communication component, the publishers have been able to keep the same price level and increase the profit margin. Crow argued that institutional repositories, by disaggregating the traditional publishing model, would lead to much lower overall costs and more efficient cooperation, not only between libraries, but especially between libraries and researchers..

Crow noted that most of the costs associating with creating and maintaining institutional repositories were related to content management and policy issues.[35] Among the expensive tasks specifically mentioned were the development of policies and agreements for access and submission, the crafting of educational material, and metadata considerations. The cost of such tasks could clearly be greatly diminished by increasing coordination and cooperation between different institutions.

More recently John Houghton and others have examined the economic implications of alternative publication models in some detail.[36] They concluded that "[o]pen access publishing and self-archiving (with overlay services) appear to be more cost- effective systems for scholarly publishing, with cost savings available throughout the scholarly communication process". [37]   "[T]here are gains to be realised from moving towards open access publishing models and, despite the lag between the costs and the realisation of benefits, the transition may be affordable within existing system- wide budgetary allocations".[38]

In his report to the *Knowledge Exchange*, published in June 2009, Houghton has examined the costs and benefits of open access publishing in the UK, Netherlands, and Denmark.[39] His estimates of the costs for various publishing models have been reproduced here as Figure 3. He calculates that switching to "Gold" publishing, where the authors publish in channels paid for through author fees, would bring net savings of around EUR 70 M per year in Denmark. "Green" publishing, where no subscriptions are cancelled, but authors self-archive in

---

35 Crow (2002)
36 Houghton & al (2009)
37 Ibid p XXII
38 Ibid
39 Houghton (2009)

institutional repositories could save around EUR 30 M per year in Denmark.

Houghton observes that the relative savings from Green publishing are much greater for Denmark than for the UK or Netherlands. This is due to the fact that the Danish repository system has lowered overhead costs per published paper.

4.2.2   Resources Required to Operate Research Information Systems

As Crow noted, most of the costs of maintaining research information systems normally do not pertain to the maintenance of hardware and software systems (which will be discussed below), , but to the work relating to the content. Therefore a good approximation of the actual resources required may be gathered by looking at how much work is necessary at the library level. What is missing from that picture is, besides the technical costs, mainly the work carried out by researchers when they enter data. On the other hand, in a good system, even with very productive researchers, this work should not take many hours per year. Alma Swan and Sheridan Brown found, in their large survey from 2005, that most authors needed only a few minutes to self-archive an article, and many systems have become more user-friendly since then.[40]
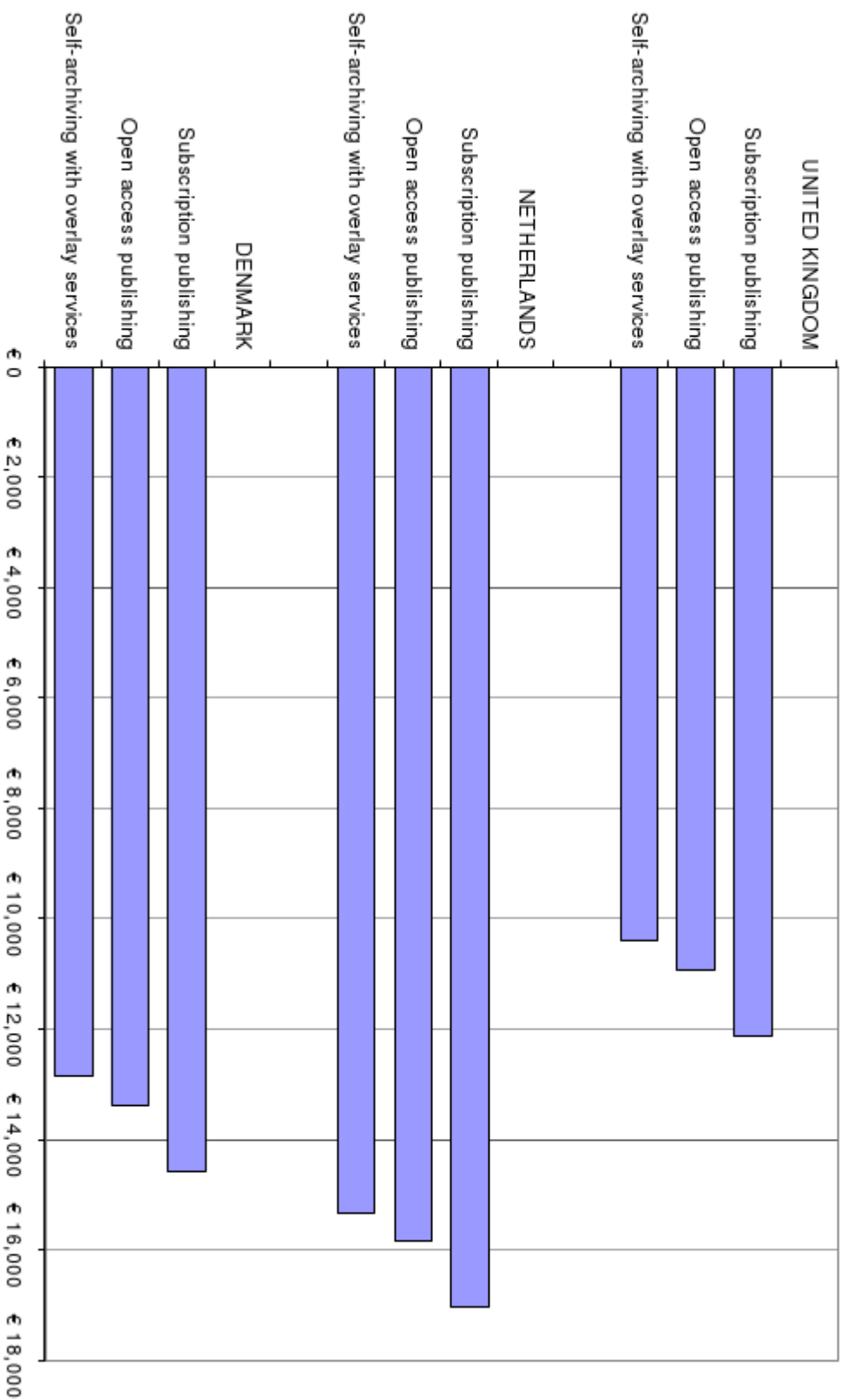
---

40 Swan & Brown (2005)

Figure 3: This figure has been copied from Houghton (2009)

Among the institutions included in our surveys, some operate an integrated CRIS, whereas others (the majority) have separate IR and CRIS. The figures below show the staff required to operate these systems.
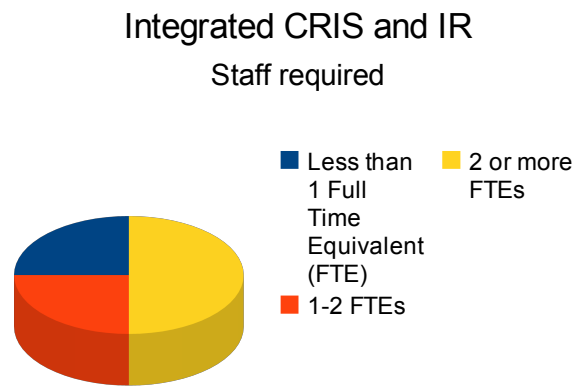
Integrated CRIS and IR
Staff required

Figure 4: FTEs required to operate an integrated CRIS and IR.

Separate IR
Staff required
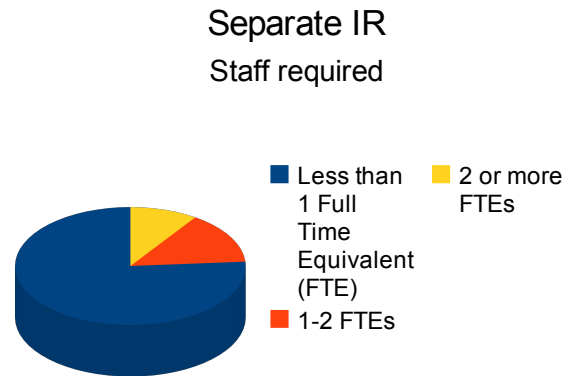
Figure 5: FTEs required to operate a separate IR.

Separate CRIS
Staff required
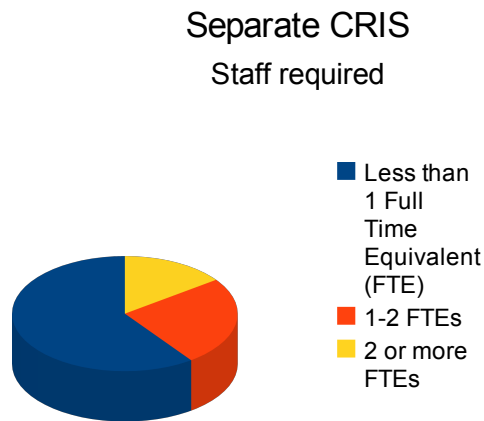
Figure 6: FTEs required to operate a separate CRIS

As would be expected, more staff is needed to operate an integrated system, than separate, but for the institutions that operate dual systems, it is clear that (in general) the two systems together require more staff than a single integrated one.

Obviously, the resources needed to operate a CRIS or IR, will depend on the size of the university, and on the number of items entered into the system every year. It will also dependent on what phase the system is in, since the initial launch of a new system or major changes clearly

will require more work than continuing maintenance. Yet these relationships are far from simple, and some major universities operate their institutional repositories with less than one FTE. The University of Tampere, for example, had around 1100 researchers and produced 1480 peer-reviewed publications in 2007. They operate a self-developed IR (in fact there are three different systems[41]). One person manages all tasks relating to their IR, including promotion, support, and technical issues, on less than full time.

4.2.3   Budget Issues

The respondents who operated a dedicated CRIS or IR were asked how it was funded. Most institutions let the costs be absorbed into the library budget as routine costs. This figure was slightly higher for CRIS, perhaps, as we will see below, since most of the CRISs had been in operation for a very long time. Around 20% used a specific item in the library budget for CRIS and IR. Dedicated funding from the central administration was equally common. Grant's from external sources were extremely rare. See Figures 7 and 8 below.
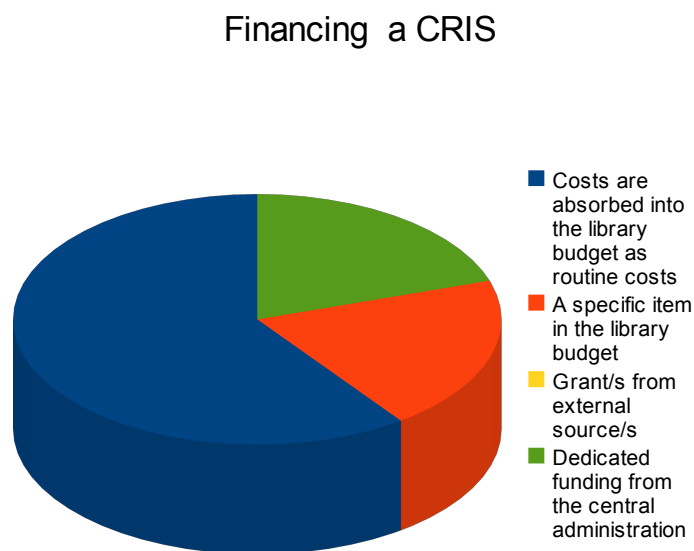
Financing a CRIS

Figure 7: Sources of funding for operating a CRIS.

---

41 http://acta.uta.fi/ http://tampub.uta.fi/ http://tuktelmat.uta.fi/
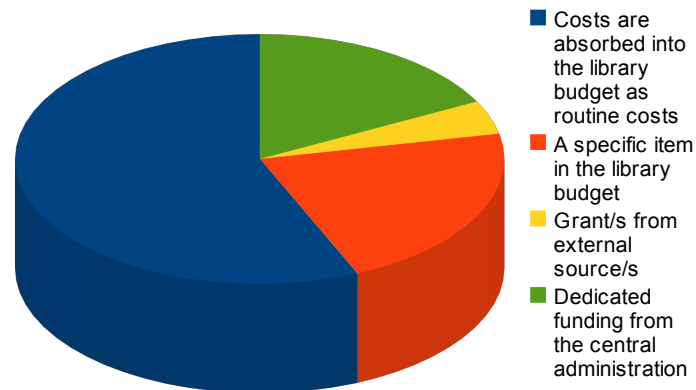
Financing Institutional Repositories

Figure 8: Sources of funding for operating an IR

Houghton and others have demonstrated how new models of publication, where institutional repositories play a crucial role, have considerable impact on the economy of scholarly communication. It is difficult to determine from our survey to what extent this has had any effect on the funding streams. With Green (parallel) publishing, there are initially only costs for maintaining the institutional repositories. Thus it seems reasonable to expect that some of the funding used for research should be used to improve the infrastructure of scientific communication. If Houghton's calculations are correct, this would be investments with a very high rate of return.

Alma Swan has listed five different "business models" for repositories.[42] The repositories at Nordic HE-institutions could theoretically use any of her first three models: the "institutional", the "publicly funded", and the "community model". The survey results seem to show that in actuality, the institutional model is the one being used.

One issue, which should be mentioned here, even if it was not included in out questionnaire (we had too many questions already), concerns *publication fees*. As more and more researchers can be expected to switch to Gold publishing, some funds will move from subscription costs to

---

42 Sawn (2008)

publication costs. At the same time, the benefits of OA publishing are generally considered to be much larger than the savings from cancelled subscriptions.

Lund University may perhaps be a useful example in this regard. The University was early to adopt an OA policy and was the first European member of BioMed Central. In the 2009 LU has allocated $163 000 earmarked for publication fees for OA (Gold) publishing.[43] This fund will be managed by the Lund University Libraries, Head Office. This way the Head Office may sign general agreements with major OA publishers, thus relieving authors of any associated worries. The fund will not be used to pay for hybrid journals, since that would, in a sense, mean paying for the same content twice.

There are of course many other possible financial models, but this example illustrates some of the main questions for consideration.

A recent report by *Universities UK* and the *Research Information Network* recommends that all HE institutions should "establish dedicated budgets to which researchers can apply for funds to meet the costs of publication fees".[44]

In a previous Nordbib report concerning open access in the Nordic countries, Ingegerd Rabow and Turid Hedlund discussed some of these issues in more detail. [45] They concluded that "[t]he employer role of the universities in an author pay model creates a need to find means to finance publishing of the research output of the university employees. An important question for future discussions is therefore finding possible solutions either in the form of funding included in the research grants or specifically established funds in the universities to finance publishing.

4.3   Software Systems and Functional Requirements

The fundamentals of the technologies and systems relating to scholarly publication systems have already been discussed above. In this section, we will therefore only briefly describe the most important features of some commonly used software packages, and give an overview of the systems currently being used at Nordic HE-institutions.

---

43 Eriksson & Bjørnshauge (2009)
44 Universities UK etc (2009) p 19
45 Hedlund & Rabow (2007)

4.3.1   Research Information Software

Since software systems for institutional repositories, such as DSpace and FEDORA, already have facilities for entering metadata, and it is fairly easy to add additional related information, such as information about organisations, researchers or funders, it may well be asked why there is a need for dedicated CRIS systems. As we saw above, it is generally more work-intensive to maintain two separate systems than one integrated.

There are two main reasons for this situation. One is that many popular packages for institutional repositories are still difficult to use for a CRIS. The only major exception is the *PURE* system, developed by Atira A/S in Aalborg, Denmark. *PURE* is popular, especially in the Nordic countries, both as an integrated system, and as a separate CRIS. As will see in the next subsection below, it is very common to use *PURE* for CRIS functionality and DSpace for the repository functions. One of the major weaknesses of software like DSpace is the lack of support for metadata profiles. There is support for basic *Dublin Core,* as part of the *OAI-PMH* functionality, but this is normally not enough. There are also often other problems with using repository software by itself for a CRIS, such as e.g. the lack of user-friendly input forms and work-flows, difficulties of handling project data from external sources, and difficulties of dealing with complex organisational structures.

The other main reason is the work load involved in migrating data, work-flows, and interfaces to a different system. It may often be more efficient to simply develop a connector to the repository system. Such work is usually fairly easy to do in-house when using an open-source system, and for proprietary software, large customers may be able to convince the vendor to provide a connector.

Table 4 below contains an overview of commonly used software systems for institutional repositories. It aims to show the great variety of solutions (these are only a small sample of all systems used), and how different systems have different strengths and weaknesses. The table is not entirely fair, since some features mentioned for one software package may also exist for another, without being mentioned.

33

System	D O A R	R O A R	Profile	Features	Import	Search	Meta data	O S	Platform
CDS Invenio	11	14	Developed at CERN, strong in Switzerland	Flexible metadata. Uses MARC21 internally. Support for user groups	The bibconvert utility lets administrators import metadata using xslt or bfx format specifications	Keyword, boolean, full-text, field-based	MARC XML, MARC21, DC XML	Yes	Python
CONTENTdm	16	N/A	Created by OCLC, which also provides hosting	Contains a Windows-based, digital collection builder , OCR extension available, hosting at OCLC available, indexed by WorldCat. Various tools to create reports directly from repository data	Imports directly from OCLC databases. metadata can be imported from tab-delimited files and matched to field names with a graphical interface	Keyword, form-based, boolean, allows proximity operators, full-text, field-based	DC, XML, Z39.50	No	Runs on most Unix and Windows servers
Digital Commons, bePress	63	82	Set of tools and hosting solution from Berkeley Electronic Press, cooperation with Proquest	Provides hosting at bepress, contains peer review module	Many media types, metadata from pubmed ID, conversion to PDF	Keyword, boolean, full-text, field-based	MODS, DC, Some support for SWORD	No	mod_perl, Apache Unix
Digitool	11	4	Provided by ExLibris, emphasis on user interface, digital collections	Built for digital collections. Facilitates integration with e-learning and institutional portals	DC XML, MARC XML, CSV tables	Keyword, boolean, full-text, field-based, searches may be limited by document types. Allows search inside XML files	MARC, DC, METS, EAD	No	Oracle and Unix
DiVA	14	16	Developed at Uppsala University Library. Used in Scandinavia. The new version is integrated with OPUS for advanced CRIS functionality	Provides templates in MS Word, and TeX for authors to create document structure	Conversion from MS Word and XML. Full unicode and Math support in input forms.	Keyword, boolean, full-text, field-based	Contains 99 elements in internal XML and may generate MARC, DC, METS, MODS etc	No	Oracle on Windows server for database backend, Tomcat, Apache. Unix for application

System	D O A R	R O A R	Profile	Features	Import	Search	Meta data	O S	Platform
DSpace, Open Repository	431	429	Developed at MIT by MIT Libraries and HP to use for the MIT repository	Supports many different types of digital content. Contains many features for data protection and preservation.	DC XML, Supports SWORD	Keyword, boolean, field-based	DC.	Yes	Java, Apache, Unix
Eprints	251	323	Developed at the University of Southampton under the leadership of Stevan Harnad.	Has many facilities to aid registration of data. Easy for users to export to reference managers. Automatically tracks journal policies on self-archiving.	Imports directly from PubMed and CrossRef via PMIDs and DOIs. XSLT import plugins	Keyword, boolean, full-text, field-based using web form	MARC, METS, DC	Yes	Perl, MySQL, Apache, Unix
ETD-db	22	25	Developed at Virginia Tech to handle theses and dissertations	Simple system. Easy to install and customize.	Conversion functions not included	Keyword, boolean, field-based	Allows harvesting in DC, ETD-ms and MARC. Contains script to export USMARC records	Yes	Perl, MySQL, Unix
Fez, Fedora	16	19	Originally developed as a research project at Cornell University. Strong RDF support. Widely used by libraries, museums, and similar organisations. Fedora by itself requires considerable technical resources to install and maintain.	Handles many types of objects and metadata. Can handle very large databases. Good data security, with the ability to rebuild a damage database. Easy to create customized interfaces	Can import via Submission Information Packages. Batch import requires METS or FOXML documents	Keyword, Field-based, boolean. Has its own search engine Gsearch developed in Denmark and funded by DEFF.	DC, METS, MARC. May be easily adapted to most metadata standards	Yes	Fez is written in PHP, Fedora Java, Apache Tomcat

35

System	D O A R	R O A R	Profile	Features	Import	Search	Meta data	O S	Platfor m
Greenstone	19	5	Developed in New Zealand at two universities in coop. with UNESCO and Human Info NGO in Belgium		For textual documents PDF, PostScript, Word, RTF, HTML, Plain text, Latex, ZIP archives, Excel, PPT, Email (various formats), source code. Image formats including GIF, JIF, JPEG, TIFF, Audio incl mp3, Ogg Vorbis audio, MPEG, MIDI, etc. Imports from METS and directly from DSpace collections, also from Z39.50 serve, metadata from CVS file etc	Keyword, boolean, supports stemming	MARC, DC XML SWORD support in next version	Yes	Java, Apache Tomcat
HAL	11	10	Developed at the *CNRS* in France. Researcher oriented.	Provides hosting. Automatically updates ArXiv for supported disciplines. Version control, author verification, links to published versions	Imports XML documents according to the HAL schema.	Keyword, boolean, full-text, field-based	OAI_DC	No	PHP, MySQL, Apache
PURE	2	N/A	Developed by Atira A/S in Aalborg. Used by many Nordic HE-institutions	Includes full CRIS functionality. Modular. Interfaces to DSpace, FEDORA etc	Users may import from PubMed, ArXiv and reference managers.	Keyword, boolean, full-text, field-based Uses Apache Lucene.	DC. May import Marc through PXA XML. PURE uses a metadata model that is an extension of CERIF.	In part	Java, Jakarta, Apache Axis. Standard SQL server

Table 4: An overview of commonly used repository software. The DOAR column contains the number of repositories in OpenDOAR using the system. ROAR contains the same figure from the ROAR database. The OS column lists whether the software is open-source or not.

Since the most popular software packages for institutional repositories are all open-source, it is usually fairly easy to create connectors to existing research information systems.

4.3.2   The Current Nordic Infrastructure

As mentioned above, almost all Nordic HE-institutions with integrated CRIS and IR use some combination of *PURE* and *DSpace* or *PURE* by itself. In Sweden most institutions use the *DiVA* system. The second most common solution is to use self-developed software. For separate IR systems, the situation is very similar. Most institutions use *DSpace*, and among the remainder, self-developed systems are the most frequent. Universities who reported using a separate CRIS system mainly use *Forskdok, Opus, Sordino, Trip* and *Frida.* See Figure 9 below for an overview.

Forskdok, used in Norway, was derived from the BIBSYS system at the University of Bergen. (Compare above page 9.) *OPUS* is a publication database developed at the University of Uppsala. It is used by many Swedish HE-institutions. It is completely compliant with the SVEP recommendations for metadata descriptions.[46] *Sordino* has been developed by Timo Tahvanainen at *Sordino Information Systems Ltd*, and is used by some Finnish universities. *TRIP* is not exactly a dedicated CRIS system, but is a database system developed in Sweden. It was one of the first such systems to make full text searching a simple task. (The acronym stands for "Text, Retrieval, Input and Presentation".) The system has been used for CRISs and similar systems in several Nordic countries. *Frida* is a system for research documentation developed at the University of Oslo, and is used at some Norwegian HE-institutions. (It was built on top of an Oracle database.)

---

46 The SVEP recommendations may be found in Andersson & al (2005)

Separate CRIS systems

- ForskDok
- Frida
- Opus
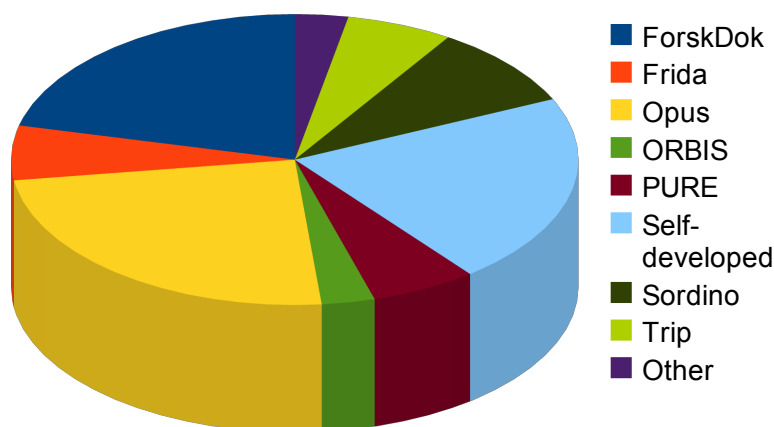- ORBIS
- PURE
- Self-developed
- Sordino
- Trip
- Other

Figure 9: Relative proportion of CRIS systems from the main survey . (This does not include the integrated CRIS and IR where PURE and DSpace was the most common solution.)

The respondents were asked whether the CRIS systems follow national recommendations for metadata standards. Among the integrated systems, less than one percent were reported as deviating from national standards. Among separate CRIS systems, that figure was considerably higher: close to 20%.

Finally there were a number of questions relating to the functionality of the CRIS. The percentage of systems supporting a particular function are listed in Table 5. The figures should not be taken too literally, since they may not have been answered by anyone directly involved with the CRIS, but they should be useful as a general indication of functionality available. (Since a certain functionality had to be selected actively, the figures are probably a bit to low.) Please note that the questions about connecting the CRIS to an IR did not apply to any integrated system.

The answers are not exhaustive in terms of possibilities. As an example, according to the responses, 93% of the CRIS systems allow "researchers and administrators" or "only administrators" to submit information. If we assume that this figure is about 2% too low, that

would leave around 5% where another category submits information. We don't know what this category is. It could be that only project managers may submit informations (i.e. not all researchers in the projects) or that only administrator and some other category (e.g. librarians) may submit information.

Functionality	Percentage
Researchers and administration can submit information to the CRIS	83
Only administration can submit information to the CRIS	10
Information submitted to the CRIS is validated by a librarian	65
Information submitted to the CRIS is validated by a department secretary	25
CRIS can be used to submit publications metadata	95
CRIS can be used to submit publications full-text	55
Harvested data can be used in publication submission	28
CRIS and IR have interface enabling data transmission from CRIS to IR	10
CRIS and IR have interface enabling data transmission from IR to CRIS	5
CRIS can be used by the researcher to create CV	70
CRIS is open online for general public	83
CRIS is used for university's external reporting (e.g. annual reporting for the Ministry/Research Council)	80
CRIS is used for university's internal needs (e.g. strategic planning)	83

Table 5: Functionality of the CRIS as reported in the surveys. The true values are probably a few percent higher.

Institutional Repositories
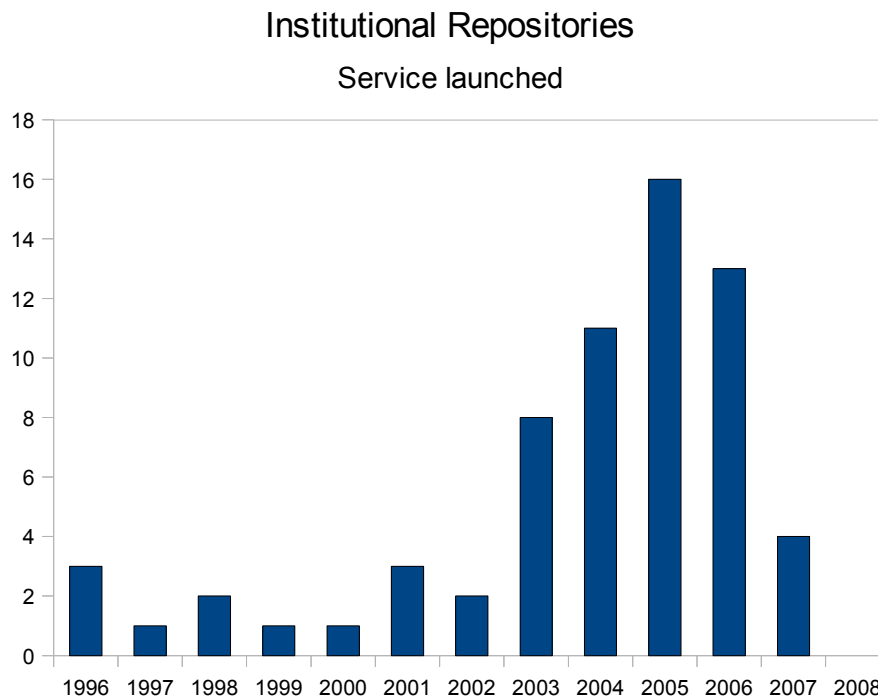### Service launched

Figure 10: This figure shows which years current institutional repositories at Nordic HE-institutions were first launched. We did not get this information from all respondents. 2008 has been excluded, since the survey was sent out during that year.

The figure above shows how most of the institutional repositories were set up in the present decade. Almost all of them support *OAI-PMH*. The only exceptions are two older self-developed systems. The CRIS systems are generally older than IRs, and most of them were launched the previous millennium. Some have been in continuous operation for two decades or more.

The survey did not contain any specific follow-up questions regarding the choice of systems for CRIS and IR. Such information may not be readily available in cases when the main decision-making process took place many years ago, and in any case, the different time frames with continuously evolving software systems, would make such answers very difficult to evaluate. It would, however, be interesting to compare the priorities and functional requirements of different HE-institutions.

4.3.3   Some Concluding Remarks

The choice, development and modifications of systems for CRIS or IR may be a quite complex question, and the decision-making process, when choosing a new system may take a long time and involve many people.

As an example, Lund University fairly recently decided to switch to a new repository system. After consulting with a number of users with different roles, as well as with technical experts, a preliminary functional specification was drafted. This initial specification was sent out for additional feedback, and this process was repeated again with a smaller group. The final document listed a number of functional requirements, and to this was added a number of additional recommendations. Clearly the results of such a process, including specifications and feature lists, and test results could be useful for other HE-institutions with similar needs.

Repositories, especially when integrated with CRIS-data, may be used for e-learning, group collaboration, research visibility, bibliometric evaluations, research maps, easy generation of publication lists and many other tasks. This is definitely an area where many existing systems could be used more efficiently.

4.4   Content and Usage

The previous sections have dealt with legal, financial, and technological aspects of the scholarly communication system. Despite the importance of all these aspect, the crucial part of any communication system is obviously the actual communication. To operate a CRIS or IR without data would be rather like operating a railway without any trains. This section will describe the contents of CRIS and IRs in the Nordic countries.

The survey contained some questions about what kind of data institutions register in their Current Research Information Systems,. As was mentioned above, when discussing CRIS functionality, the answers to these types of questions may have a larger margin of error than the others in our survey. Also, for similar reasons, the reported frequencies will probably be slightly lower than the actual.

The reported data in CRIS are listed in Table 6 below. It is very difficult to understand exactly how the questions have been interpreted by the respondents. It seems very unlikely, for

41

example, that a CRIS does not include the names of the researchers in a research project. Perhaps the respondents meant that not all researchers were included. The table should therefore not be taken too literally but only used as an indication of the relative frequency of various kinds of data in Nordic CRIS systems.

It may be noted that some of these types of data correspond to fields that were proposed as mandatory by EuroCRIS in 1998. According to the C*GP* ("Code of Good Practice for Current Research Information Systems"), every CRIS should include references to the funding organisation or body, the project title and the "most specific" research unit.[47]

.

---

47 EuroCRIS (1998)

Data Type	Percentage
Researcher name	95
Researcher role (e.g. principal investigator, graduate student, etc)	51
Researcher contact information	54
Researcher affiliations	87
Researcher areas of expertise	36
Researcher publications	92
Researcher other activities/output (apart from publications)	44
Funding project name	44
Funding project description	38
Funding project amount of funding	21
Funding project time-frame	36
Organisations name	95
Organisations description	28

Table 6: Types of data in a CRIS as reported in the surveys. The true values are probably a few percent higher.

The survey also contained a number of questions about the content of institutional repositories. It was interesting to find out what proportions of major document types that had been deposited. Another question was if substantially more documents had been deposited in 2007 than in 2006. As can be seen from the figures below, few repositories had more than 50% of all publications of any document type, although there was a substantial improvement from 2006 to 2007.

The charts do *not* show the total proportions of each *document type* in the repositories. Instead, the figures show what proportion of all *repositories,* that have predefined proportions of documents. Figure 12, for example, shows the proportion of all repositories that don't have any PhD theses (0%, the largest blue slice), have 1 to 10% of all published PhD theses (the orange slice), 11-20% of PhD theses (the yellow slice), and so on. The pie chart shows that only about 1/3 of the repositories had more than half of that institutions' PhD theses deposited in 2006. In 2007, the situation had improved markedly with above 50% of the repositories having more than 1/5 of the PhD theses in their respective repositories.
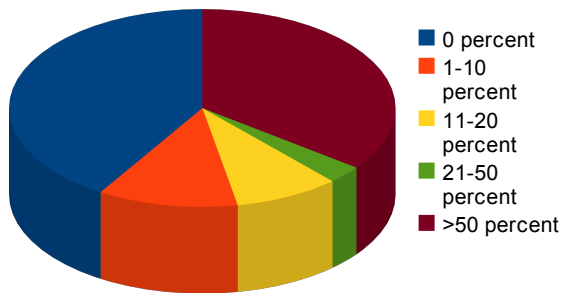
PhD Theses 2006

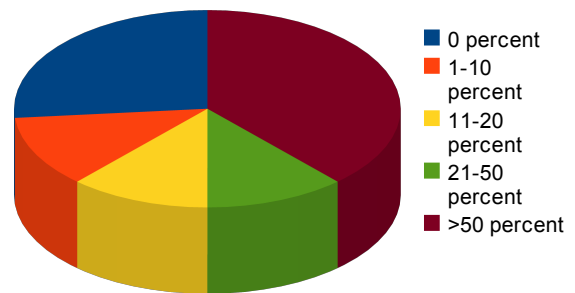
Figure 12: PhD theses in IRs 2006

PhD Theses 2007

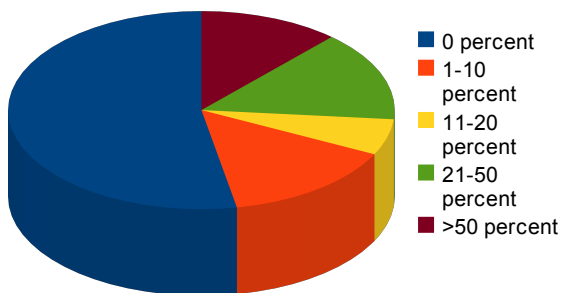*Figure 13: PhD theses in IRs 2007*

Exam papers 2006

Figure 14: Exam papers in IRs 2006
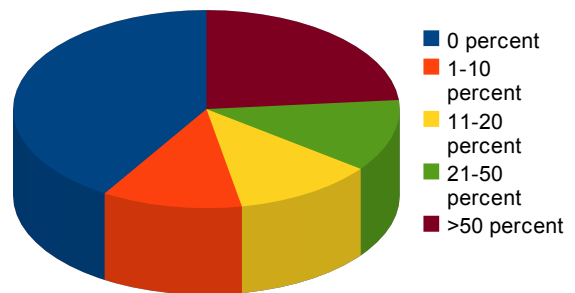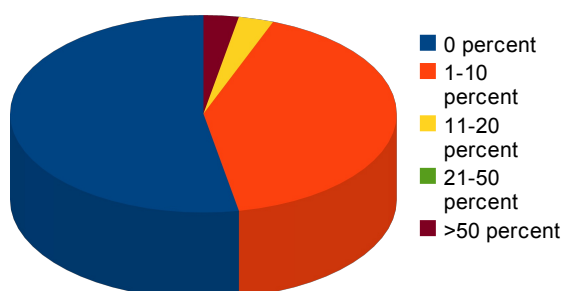
Exam papers 2007

Figure 15: Exam papers in IRs 2006

Articles 2006
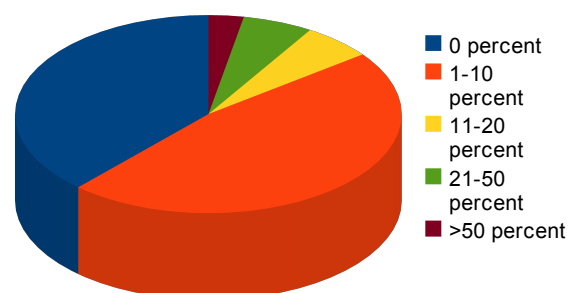

Figure 16: Articles in IRs 2006

Articles 2007

Figure 17: Articles in IRs 2006

Reports 2006

Reports 2007

0 percent
1-10 percent
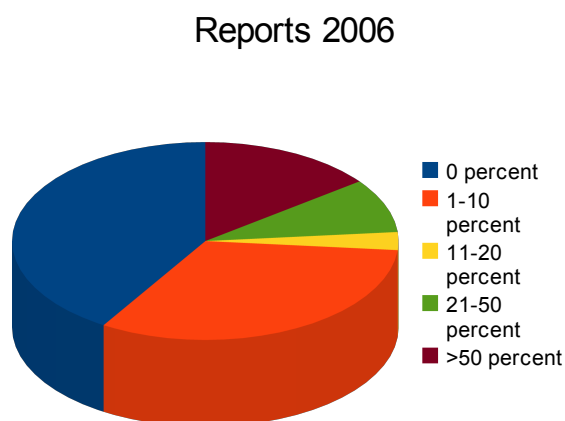11-20 percent
21-50 percent
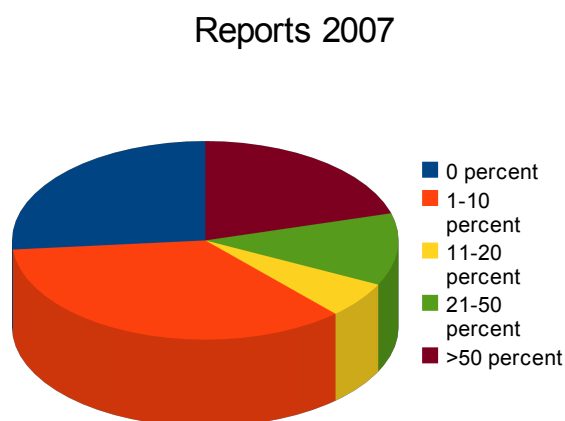>50 percent

Figure 19: Reports in IRs 2006   *Figure 18: Reports in IRs 2007*

Although there are marked difference between 2006 and 2007, there are still, for most publication types only a fraction of all repositories that have more than 10% of 2007 publications. Only 15% of the repositories had more than 10% of the articles from 2007.

This leads to a number of questions about how to populate the repositories with more full-text documents.   Vanessa Proudman has recently tried to identify the most important factors that lead to a good coverage. [48] Based on her investigations she produced four recommendations:

1) Communicate with the research community based on a thorough knowledge of their particular work processes and motivations.

2) Make sure that the repository and its services are tailored to the needs and problems of the researchers.

3) Create document collections that reflect the academic output of the institution's disciplines.

4) Try to identify other locations where researchers have deposited papers, and harvest from those location, either using some standard protocol or by special agreement with the maintainers of the external source.

Many respondents have through their comments and responses indicated difficulties motivating researchers to self-archive. A survey by Alma Swan has shown that the vast majority of all

48 Proudman (2008)

researchers would readily comply with a funder or institutional mandate to self-archive in repositories. Yes many are reluctant to do so without any immediate incentives. One important factor which prevents researchers from depositing, apart from copyright or licensing issues, is the perceived difficulty or work load. In another survey Swan and Sheridan Brown asked researchers who did self-archive about their motivation for doing so. They found that most of them were motivated by their support of open access, and by the desire to make their research more visible.[49]

As was mentioned above, it is of course also very important to make the process of depositing as quick and effortless as possible for the researchers. Useful data may be harvested and sent from and to other sources, in order to make sure the researchers never have to enter the same data twice into different systems. AJAX or similar technologies may be used to facilitate input and to ensure data quality.

---

49 Swan & Brown (2005) p 50

5 Summary and Discussion

The final overall response rate for the whole survey was 71%. This figure is a bit misleading, however, since the sizes of the HE-institutions vary a great deal, and most of the missing institutions are quite small with little or no research output. Some respondents also failed to reply to all questions that should be answered, but this was fairly uncommon (between 1 and 8 respondents), and again usually for smaller institutions, for which some questions may have been difficult to answer correctly.

The somewhat low response rate does not seem to be directly related to any technical aspects of the survey design, since the response rates for some countries was much higher. It seems that the amount of previous awareness of the project, and its goals, among library directors, was of paramount importance for getting completed surveys.

Because of the low response rate in some areas together with the skewed distribution of missing responses, and since we noted that some questions had been interpreted differently by different respondents, we have abstained from attempting any kind of statistical analysis of the answers. Such an analysis would require us to hypothesize about the underlying distribution without sufficient basis in facts.

The survey has identified the main strengths and weaknesses of the Nordic CRIS and IR systems. In general, the Nordic HE-institutions are at the forefront of the international developments. Yet there are many areas where improvements are of great importance. The population of full-text in IRs was far from sufficient in many institutions. Hopefully the situation has since improved, but there is still much to do in this area. Many HE-institutions operate CRIS systems that are not entirely up-to-date, and they often do not contain enough data to be truly useful.

Intellectual property rights, funding models, and technical coordination have been highlighted by respondents as key subjects for high level assistance. It is clearly not cost-efficient if individual libraries spend time and staff resources developing their own repository systems. A multitude of differing systems also makes integration and efficient usage of the repository services much more difficult.

47

There are a number of areas that would clearly benefit from national and/or Nordic coordination. Much useful work is currently being carried out within the European DRIVER II-project, but there are also many areas where a specific Nordic cooperation would be advantageous. The conditions are very similar, and, although there may be exceptions, most major HE-institutions in the Nordic countries are in the front-line of the international development. They have already well-established networks of stakeholders and appropriate central organisation for handling this cooperation, e.g. Nordbib.

The same factors that create special conditions for Nordic scholarly publishing may also provide a special role for Nordic scholarly repositories.[50]

The historical overview in the first chapter showed how slowly the scholarly communication system has evolved. It also showed how research libraries have played a crucial role throughout this development. We hope that, with proper cooperation, coordination, and funding, the Nordic HE libraries will continue to improve the research infrastructure, and solve the difficulties revealed in the present report.

.

---

50 On this issue see e.g. Hedlund & Rabow (2009)

6 References

ARL. 1997. *ARL Newsletter* (195). Available at: http://old.arl.org/newsltr/195/osc-spar.html.

Beyer, Janice M., and Harrison M. Trice. 1982. "The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings." *Administrative Science Quarterly* 27(4):591-622.

Bourne, C. P., and T. B. Hahn. 2003. *A history of online information services, 1963-1976*. MIT Press.

Clayton, Robert. 1983. "Technology Lecture: Developments in Scientific Information Systems." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 388(1794):21-48.

Closs, José S., and Francine M. Cheater. 1994. "Utilization of nursing research: culture, interest and support." *Journal of Advanced Nursing* 19(4):762-773.

Crow, Raym. 2002. *The case for institutional repositories: a SPARC position paper*.

Danelid, Hans et al. 2008. *Kartläggning av högskolans elektroniska publicering och rekommendationer för nationell samverkan*. Stockholm, Sweden: The Association of Swedish Higher Education .

Davis, James R., and Carl Lagoze. 1994. "A protocol and server for a distributed digital technical reportlibrary."

Dewatripont, M. et al. 2006. "Study on the economic and technical evolution of the scientific publication markets in Europe.[Sl]: European Commission. 2006." *europa. eu. int/comm/research/science-society/pdf/scientific-publication-study_en. pdf*.

Eriksson, Jörgen, and Lars Bjørnshauge. 2009. "Lund University Supports Publishing in OA Journals." *ScieCom Info* 1(1).

Gardner, William. 1990. "The electronic archive." *Scientific publishing for the 1990s. Psychological Science* 1(6):333-341.

Ginsparg, P. 2008. "The global-village pioneers." *Physics World*:22-26.

van der Graaf, M., and K. van Eijndhoven. 2008. *The European Repository Landscape: Inventory Study Into He Present Type and Level of OAI-compliant Digital Repository Activities in the EU*. Amsterdam University Press.

Hansen, Inge Berg, Morten Bennum, and Peter Reumert Korsby. 1983. "Planning and development of a Danish I & D policy." *Journal of Information Science* 7(2):59-65.

Harnad, Stevan. 2003. "Open access to peer-reviewed research through author/institution self-archiving: maximizing research impact by maximizing online access." In *Digital Libraries: Policy, Planning, and Practice*, eds. Derek Law and Judith Andrews. , p. 63-98.

Harnad, Stevan. 1990. "Scholarly skywriting and the prepublication continuum of scientific inquiry." *Psychological Science* 1:342-343.

Hedlund, T., and I. Rabow. 2009. "Scholarly publishing and open access in the Nordic countries." *Learned Publishing* 22(3):177-186.

Hedlund, Turid, and Ingegerd Rabow. 2007. *Open Access in the Nordic Countries State of the Art Report* Copenhagen: Nordbib.

Herbert Van de Sompel, and Lagoze, Carl. 2000. "The Santa Fe Convention of the Open Archives Initiative." *D-Lib Magazine* 6(2).

Hiltz, S. R., and M. Turoff. 1978. *The network nation: Human Communication via Computer*. Reading, MA: Addison-Wesley.

Houghton, J. et al. 2009. *Economic implications of alternative scholarly publishing models*. JISC. Available at: http://www.jisc.ac.uk/media/documents/publications/rpteconomicoapublishing.pdf.

Houghton, John. 2009. *Open Access – What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark*. Knowledge Exchange. Available at: http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOA_What_are_the_economic_benefits_-_a_comparison_of_UK-NL-DK__FINAL_logos.pdf.

Kreysa, Frank J. 1972. "SSIE-An Information Center Which Stores Foresight." *Journal of Chemical Documentation* 12(1):19-22.

Landry, R., N. Amara, and M. Lamari. 2001. "Utilization of social science research knowledge in Canada." *Research Policy* 30(2):333-349.

Palca, Joseph. 1991. "New Journal Will Publish Without Paper." *Science* 253(5027):1480.

Pinfield, S. 2002. "Creating institutional e-print repositories." *Serials: The Journal for the Serials Community* 15(3):261-264.

Proudman, Vanessa. 2008. "The population of repositories." In *A DRIVER'S Guide to European Repositories*, Amsterdam: Amsterdam University Press, p. 49 - 101.

Salokannel, Marjut. 2008. "University of Helsinki opens its research vaults: a few words on open access and the new research environment in Finland." *ScieCom Info* 4(2).

Senders, John. 1976. "The scientific journal of the future." *The American Sociologist* 11:160-164.

Swan, Alma. 2008. "The business of digital repositories." In *A DRIVER's Guide to European Repositories (Amsterdam, 2007)*, , p. 15-47.

Swan, Alma, and Sheridan Brown. 2005. *Open access self-archiving: An author study*. Cornwall, UK: Key Perspectives Ltd.

UNESCO. 1971. *Study Report on the feasibility of a World Science Information System*. Paris: UNESCO.

Universities UK and the Research Information Network. 2009. *Paying for open access publication charges*. Research Information Network.

Wyly, B. J. 1998. "Competition in scholarly publishing? What publisher profits reveal." *Association of Research Libraries Newsletter* 200.

Wysocki, Adam. 1976. "The UNISIST Programme: A Framework for International Information Exchange." In *Proceedings*, Paris: Unesco, p. 401-405.

Zimmerman, E. 2002. "Cris-cross: Current research information systems at a crossroads." *CRIS2002: Gaining Insight from Research Information, Kassel, Germany*.

7 Appendix – A list of all questions sent to the respondents.

The following questions were sent to all respondents:

Number of research active staff (full-time equivalents, teachers included)

Does your institution have statistics for research output (number of publications) for the year 2007?

Number of peer-reviewed articles 2007

Number of non-peer-reviewed articles 2007

Articles in newspapers or magazines 2007

Books 2007

Book chapters 2007

PhD theses 2007

Licenciate theses 2007

Papers in conference proceedings 2007

Reports 2007

Preprints and working papers 2007

Student theses/exam papers 2007

Other publications 2007

Does your institution have some kind of Institutional Repository for full text documents (IR) or Current Research Information System for bibliographic publication data (CRIS).

Does your institution have an integrated system comprising both IR (full text documents) and CRIS (bibliographic publication data)?

Does your institution have a separate Institutional Repository (for full text documents)?

Does your institution have a separate CRIS (registration of publication data for institutional output)

The following questions were sent to institutions without any IR or CRIS:

Why have you decided not to offer an institutional repository service? Please rate the following alternatives from 1 to 5, where 5 is the most accurate or important.

[Our institution does not publish any research]

[The question has not been raised]

[It has been discussed but resources/competence has not been available]

If you lack competence or resources, please specify in which area(s). More than one alternative may be selected.

[Funding]

[Staff to set up and operate the service]

[Technical expertise]

[Other (please describe in the comments field below)]

Comments

Why have you decided not to offer a Current Research Information System (CRIS)? Please rate from 1 to 5, where 5 is the most important

[Our institution does not publish any research]

[The question has not been raised]

[It has been discussed but we lack resources/competence]

[Funding]

[Staff to set up and operate the service]

[Technical expertise]

[Other (Please describe in the comment field below)]

Comments

Consider the task of setting up an Institutional Repository and/or a CRIS. In which areas would a national coordination/service be beneficial? Please rate the importance below:

[Shared software/technical solution for all local systems?]

[Shared resources for technical advice ?]

[A national centre for advice and guidelines regarding copyright management]

[Joint production of materials for marketing and teaching]

[A national centre for current awareness and information dissemination]

[Support for e-publishing of journals and other periodicals]

[A national search service for the research publications from your country]

[Start-up funding]

Comments

The following questions were sent to institutions with an integrated CRIS & IR:

What type of software (or program platform) do you use for your integrated service?

[DiVA]

[DSpace]

[GNU Eprints]

[ORBIS]

[PURE]

[Self developed]

[Other (name)]

The year the service was launched

Does the system/software support OAI-PMH?

What is your integrated system/service called?

URL to the service

What unit in your organisation is responsible for the service? (more than one alternative may be selected)

[The Administration]

[The Library]

[The Information Department]

[Other]

Comment on the organisation

Staff resources for operating the service (all tasks from marketing to technical support)

[Less than 1 Full Time Equivalent (FTE)]

[1 – 2 FTEs]

[More than 2 FTEs]

How do you finance the operation of your IR (more than one alternative may be selected)

[Costs are absorbed into the library budget as routine costs]

[A specific item in the library budget]

[Grant/s from external source/s]

[Dedicated funding from the central administration]

[Other]

What types of publications are deposited ?

[Peer reviewed articles]

[Non-peer reviewed articles]

[Newspaper or magazine articles]

[Books]

[Chapter in book]

[PhD theses]

[Licentiate theses]

[Conference proceedings]

[Reports]

[Preprints and Working papers]

[Student theses/exam papers]

[Other]

Comments on publication types

Do you follow any national recommendations/decisions regarding metadata standards, (e.g. publication types, author names, titles etc,) ?

Comment

A9.3 Do you apply quality control (e.g. manual or automatical enhancement of metadata) ?

Comment

Do you follow any national recommendations/decisions regarding subject classification?

If YES, what system do you use?

Comment

Has your institution taken any decision either recommending or mandating Open Access?

If the answer is YES please write the URL to the decision or email the document to nordbib.survey@gmail.com

Comment

Questions for institutions with a separate IR:

What software (or program platform) do you use for your IR?

What software (or program platform) do you use for your IR? [Other]

The year the service was launched?

Does the system/software support OAI-PMH?

What do you call your IR?

Please write the URLto the IR. .

What unit in your organisation is responsible for the service? (more than one alternative may be selected)

[The Administration]

[The Library]

[The Information Department]

[Other]

Comment on the organisation

Staff needed for operating the service (all tasks from marketing to technical management and support)

[Less than 1 Full Time Equivalent (FTE)]

[1 – 2 FTEs]

[More than 2 FTEs]

Staff needed for operating the service (all tasks from marketing to technical management and support) - Comment

How do you finance the operation of your IR (more than one alternative may be selected)

[Costs are absorbed into the library budget as routine costs]

[A specific item in the library budget]

[Grant/s from external source/s]

[Dedicated funding from the central administration]

[Other]

Comments on financing

What types of publications are deposited in your IR?

[Peer reviewed articles]

[Non-peer reviewed articles]

[Newspaper or magazine articles]

[Books]

[Chapter in book]

[PhD theses]

[Licentiate theses]

[Conference proceedings]

[Reports]

[Preprints and Working papers]

[Student theses/exam papers]

[Other]

Comments on publication types

Do you follow any national recommendations or decisions regarding metadata standards (e.g. publication types, author names, titles etc,)

Do you apply quality control(e.g. manual or automatical enhancement of metadata) ?

Comments on national recommendations and quality control

Do you follow any national recommendations or decisions regarding subject classification

If YES, what system do you use?

Comment on subject classification

Has your institution taken any decision either recommending or mandating Open Access?

Has your institution taken any decision either recommending or mandating Open Access? - Comment

Questions for institutions with a separate CRIS:

What software do you use for your CRIS ?

[OPUS]

[ORBIS]

[PURE]

[Self-developed]

[Other]

Comment on software

The year the service was launched

Does the system support OAI-PMH?

Comment

What is your CRIS called?

Please write the URL to the service.

What unit in your organisation is responsible for the service?

      [The Administration]

      [The Library]

      [The Information Department]

      [Other]

Comments on the organisation

Staff needed for operating the service (all tasks from marketing to technical management and support)

      [Less than 1 Full Time Equivalent (FTE)]

      [1 – 2 FTEs]

      [More than 2 FTEs]

Staff needed for operating the service (all tasks from marketing to technical management and support) - Comment

How do you finance the operation of your CRIS (more than one alternative may be selected)

      [Costs are absorbed into the budget of the organisation responsible for the service as routine costs]

      [A specific item in the budget of the organisation responsible for the service]

      [Grant/s from external source/s]

      [Dedicated funding from the central administration]

      [Other]

Comments on financing

What type of information is registered in the CRIS?

      Researcher information: [Name]

      Researcher information: [Role (e.g. Principal Investigator, graduate student)]

      Researcher information: [Contact information]

59

Researcher information: [Affiliations]

Researcher information: [Expertise]

Researcher information: [Publications]

Researcher information: [Other scientific activities/output]

Funding projects information: [Name]

Funding projects information: [Description]

Funding projects information: [Amount of funding]

Funding projects information: [Time-frame of project]

Organisations information: [Name]

Organisations information: [Description]

What type of functionality does the CRIS support?

Submission and validation: [Researchers and administration can submit information to the CRIS]

Submission and validation: [Only administration can submit information to the CRIS]

Submission and validation: [Information submitted to the CRIS is validated by a librarian]

Submission and validation: [Information submitted to the CRIS is validated by a department secretary]

Publications: [CRIS can be used to submit publications meta-data]

Publications: [CRIS can be used to submit publications full-text]

Publications: [Harvested data can be used in publication submission]

Cris/IR interface: [CRIS and IR have interface enabling data transmission from CRIS to IR]

Cris/IR interface: [CRIS and IR have interface enabling data transmission from IR to CRIS]

Exploiting collected information: [CRIS can be used by the researcher to create CV]

Exploiting collected information: [CRIS is open online for general public]

Exploiting collected information: [CRIS is used for university's external reporting (e.g. annual reporting for the Ministry/Research Council)]

Exploiting collected information: [CRIS is used for university's internal needs (e.g. strategic planning)]

General comments on CRIS

Do you follow any national recommendations or decisions regarding CRIS standards (e.g. EuroCRIS, national level) ?

Do you apply quality control (e.g. manual or automatical enhancement of data)?

Comments on national recommendations and quality control

Do you follow any national recommendations or decisions regarding subject classification?

If YES, what system do you use?

Comments on subect classification

Has your institution taken any decision either recommending or mandating registration of research output?

Has your institution taken any decision either recommending or mandating registration of research output? - Comment

Questions for all institutions with an IR and/or CRIS:

What percentage of the following document types were deposited in full text during 2006 and

2007?

[PhD theses 2006]

[PhD theses 2007]

[Exam papers 2006]

[Exam papers 2007]

[Articles 2006]

[Articles 2007]

[Reports 2006]

[Reports 2007]

[PhD theses 2006]

[PhD theses 2007]

[Exam papers 2006]

[Exam papers 2007]

[Articles 2006]

[Articles 2007]

[Reports 2006]

[Reports 2007]

Do you plan to develop your services for electronic research publishing in the future?

If the answer is YES, please describe your plans

What do you see as the greatest challenges/problems regarding electronic research publishing?

Does your institution publish any OA-journals?

If the answer is YES, please write the titles

Has your institution adopted a policy recommending/encouraging publishing in OA-journals?

If the answer is YES, please describe or write the URL to the policy or email the policy to nordbib.survey@gmail.com.

Please consider the task of operating a Repository for the research publications of your

institution. In which areas would a national coordination/service be beneficial? Please rate the importance below

[Shared software/technical solution for the local systems?]

[Shared resources for technical advice?]

[National advice and guidelines regarding copyright management?]

[Joint production of material for marketing and teaching ?]

[National centre for current awareness and information dissemination]

[Support for e-publishing of journals and other periodicals?]

[A central search service for national research publications?]

[Start-up funding?]

Comments regarding national coordination/support