Setting up an institutional archive:
some technical and organisational considerations (Tworzenie archiwum instytucji – kwestie techniczne i organizacyjne)

Jörgen Eriksson, Lund university
2007-09-12

Lund University and Lund university libraries

Lund University was founded in 1666. With eight faculties and a multitude of research centres and specialised institutes, it is today the largest unit for research and higher education in Sweden. [1]

The library organisation is highly decentralised. The libraries at Lund University form a network, Lund University Libraries (LUB), consisting of the common units the Lund University Libraries Head Office (BD) and the University Library (UB), as well as the subject oriented departmental, faculty and centre libraries. The departmental and faculty libraries are financed by the faculties and departments, and fall under their respective board.
The University Libraries Head office is the unit responsible for common electronic services for all libraries like the library OPAC, electronic journals and databases and also electronic publishing and registration of publications [2]. A division at the Head office is responsible for common work with scientific communication, open access issues, publication registration and research performance assessment. [3]

Background: A short history of e-publishing at Lund University.

The only central publishing service before the start of the library-run services was a traditional printing press, Lund university press. This was closed in 1999 and even if there were talk about replacing it with an electronic publishing service nothing came out of that effort at the time. A fairly large number of departments at the university published their own technical report and dissertation series in print. With the growth of the Internet and WWW a lot of this decentralised departmental publishing became electronic.
There was no central publication registration service or any requirements to register research publications.

Doctoral dissertations

The first central electronic publication/self-archiving service was run by the university library and started in 1996. It covered the document type doctoral dissertations and from the start it was a service for announcing the time and place of the defence and register the dissertation while self-archiving was optional. There has been a steady rise in the number of self-archived dissertations over the years, and 2006 we self-archived ca 25% of the doctoral dissertations of that year (ca 100 out of ca 400) which was the best coverage so far. One should also be aware that at the same time an unknown number of doctoral dissertations are self-archived at departmental web-sites and personal home-pages, so the total number of freely available doctoral dissertations is higher then what we can show in the central service. The software was replaced by new in-house developed software in late 2004.

LU:research – our institutional repository

In 2002 the Head office set up an institutional repository that should cover all types of research publications except doctoral dissertations. At the start our focus was on self-archiving only, but we soon discovered that we needed to broaden that and add the possibility to only register publications to generate interest in the service within the university. The software we used was the free EPrints software developed at the University of Southampton, We started by setting up a demonstrator that showed a unified university perspective. We made a survey of what were available already at departmental web sites etc. and added samples from those to have some content to show as examples. We "launched" the demonstrator at a half-day seminar on electronic publishing/scientific communication in May 2002. We held five seminars on self-archiving and oa-issues in general during 2002 and 2003, all which have been fairly well attended (20-40 researchers). We were also part of a range of courses which the Learning and Teaching Development Centre at Lund University [http://www.ced.lu.se/] offers researchers. This kind of meetings with faculty at different levels we continue to do. We also hold regular seminars for department and faculty librarians with the aim to enable them to promote LU:research and to work as first-line support to their researchers when questions on using LU:research arises.

Workflow varies within different parts of the university. The medical faculty uses the library staff to find, register and when possible also self-archive, articles, while the humanities faculty has taken a decision that each researcher should enter their own publication records with the library as a support organisation.

By April 2007 the archive contained ca 14.000 publication records out of which ca 2300 had free full text attached in the archive. This is a fairly well-populated archive by an international comparison but far from our goals [4]. Every year ca 2500 publications with authors affiliated to Lund university are registered in ISI Web of Science and only a fraction of these are self-archived in our repository even if a growing number is registered.

Masters theses and student papers

We started a third central service for the publication of masters theses and other student papers in late 2005. In this service only full-text papers have been deposited and by September 2007 there where ca 6500 full text publications available. The software used is the same as for the doctoral dissertations, with some modifications because of different workflows.

Lessons learned

In this section I will describe some of the lessons we have learned while running these services and after that some of the conclusions we drew.

An early and important lesson was that different actors within the university have different requirements and demands on a repository service. We identified three main perspectives.

1. The University:

The single, unified, entry point is the central University perspective. It was also the University that already from the beginning wanted the adding of bibliographic records to the repository, not just full text. In the longer term the University sees LU:research as a marketing tool, and when the researchers/departments really start to use it, as a tool to help assessing research activities at different departments. Our library director is raising awareness on different university management levels about these issues and that have led to the following results.

The university board adopted an OA policy for Lund University in November 14, 2005.

"In order to maximize the number of open access publications the Board of Lund University strongly recommends that:

- Researchers at Lund University, if possible, publish in journals with open access
- If no equivalent open access journal is available, researchers choose a journal allowing parallel publishing/deposition of the article
- Transfer of copyright be avoided. As a minimum the author's right to parallel publishing must be retained
- Lund University work for the transition of scholarly journals to a publishing model, where articles either are made freely available to the reader directly or through parallel publishing. [5]

In May 2007 the vice-chancellor mandated that all publications by authors affiliated to Lund university should be registered I LU:research . This should be done retrospectively from 2002 and up until present."

2. The faculty/department:

One of the results of the first seminar in 2002 was that we were approached by the information committee of the medical faculty. They wanted to show the output from their research but where not interested in just showing it in the university context. Together with them and their faculty librarian we created an independent user interface to their sub-set of our institutional repository, Lund Virtual Medical Journal. There are mainly bibliographic records but through the support of their library they are also adding full text to the bibliographic records. This experience has been repeated when we have talked to other leaders at different organizational levels within the university. Their interest for the service grows with their possibilities to put their own brand on their publications.

3. The Individual researcher:

The researcher wants as little, or no job at all in adding his publications and want to be able to show them on his personal homepage, reuse them in various formats for new applications, and as references in publications and when reporting projects. More and flexible ways to enter publications then through a web-form where you enter record by record is needed and the same goes for the output of a search, which should be possible to extract in a number of different formats.

Action plan
The OA policy for Lund University adopted by the university board in November 14, 2005 led to the creation of an action plan (available in Swedish only) which can be summarized as follows.

- Co-operation between three university offices with a common action plan
- The Information Office
    – Dissemination of information and marketing of the university
- Legal Office

- Copyright issues, the university-researchers, researchers-publishers
- Head Office
    - Publishing and registration services

All of these offices at the university have a common interest in cooperating on scientific communication issues. The Information office and the Head office both have a common goal to disseminate information about research done at Lund university and cooperation with the Legal office is essential since copyright issues is such a large and difficult part of the self-archiving work, and also in regulating copyright issues between the university and the researchers.

Considerations for a new software

The action plan together with the requests and comments we had received regarding our services over the years where the basis for a functional requirements specification for an upgrade of our services. The most important general requirements are listed here.

- 1 technical platform for all of our services
    - Running three different software systems for three similar services is inefficient when it comes to technical support and maintenance and not very user-friendly when it comes to users who need to use more then one of the systems. Using the same system for both registration and self-archiving is also important. A record should only have to be entered once.
- Possible to change and adapt easily
    - Our experience from running our services is that to maintain credibility it is important to be able to respond quickly and positive to the requirements from the different players within the university. In today's changing world of scientific communication it is especially important to be able to meet new needs. In this there is also an ambition to move as many tasks as possible from programming level to librarian configuration. It should be simple to add new document types and their metadata templates or to do a batch upload of records from another database to take two examples that have been repeated requests from our users.
- Simplified and extended possibilities to integrate in and export records to other environments and to national and international search services.
    - Our users want to be able to brand their own publications and integrate them in their own web environment. The university wants the publications to be disseminated as widely as possible to different search services.
- Flexible options for importing records from other sources to the system
    - Import from different reference management systems and database searches in the major index- and reference services should be done in a standardised way.
- Unique identification of authors
    - This is important for two reasons. One is the reuse of records for research assessment purposes. The other is to make it possible to make a sub-set of an authors publication records in a simple way, for reuse on his personal homepage, in project applications etc.
- The possibility to express relations between objects

- It becomes more and more common that a publication is linked to supplementary material or is part of a context and we wanted the system to be able to express a number of such relations.
- Co-ordination with research assessment, project databases and other activities
  - Our goal is that a publication record should be created once in one place and then it should be possible for the record to be reused for any other purposes within and outside the university.

After we had finalised our specification we started to look into existing software [7] and how they lived up to our requirements. This was done in early 2006. All of the software solutions we looked at have attractive features but none of them covered all of our requirements. The most important thing that we felt was lacking was the possibility to change and adapt quickly to new requirements. This is the main reason why we decided to go for an in-house development from scratch. And we judged that we had the in-house competence to succeed with such a development. The first stage of development is almost finished as of writing this in September and we plan to release the system in late September or early October.

Existing software solutions

A good starting point when looking for suitable software is the Budapest Open Access Initiative (BOAI). Guide to Institutional Repository Software v 3.0. [8]
It covers nine institutional repository software's in depth. It is a bit old by now and you will need to study the software home pages to catch the developments since 2004 but it is also a great help to guide you what to look for in a system.
An interesting solution might be to let another institution host your repository. This can leave you free to concentrate on the content by out-sourcing the technical maintenance part.
Some examples where this option is offered is the Eprints software at University of Southampton [9] or DSpace, which is offered through the BioMed Central publishing house [10]. We plan to offer a similar solution some-time next year.

To quote from the BOAI site, the most important thing is not the software itself. It's the overall strategy, which should bear on what functionalities and needs the software should fulfil. "An institution's system needs will be driven by its content policies and by the procedures required to implement those policies. A well designed and carefully planned repository program can function well with any of the systems discussed here-but none of these systems can help a poorly designed or inadequately planned repository succeed. "
This quote is summing up our experiences very well and what we are trying to achieve with our action plan and system upgrade is a "user benefit" situation that will help to overcome the researchers' reluctance to add the new task of self-archiving to all of the others they have to do.

[1] More facts about Lund University are found at http://www.lu.se/o.o.i.s/450
[2] More about Lund University Libraries http://www.lub.lu.se/en/about-lub/organisation.html
[3] The division Scientific communication http://www.lub.lu.se/en/about-lub/organisation/lund-university-libraries-head-office/scientific-communication.html
and it's services   http://www.lub.lu.se/en/publish/publish.html
[4] Find other repositories in OpenDoar http://www.opendoar.org/index.html
[5] http://www.eprints.org/openaccess/policysignup/fullinfo.php?inst=Lund%20University

[6] http://lvmj.medfak.lu.se/

[7] EPrints (http://www.eprints.org/software/), PURE (http://www.atira.dk/en/pure/) and FEDORA (http://www.fedora-commons.org/) where the ones that went to the final round. In short EPrints is an out-of-the-box free IR software, PURE a commercial solution and FEDORA a free software that is not so out-of-the box as EPrints but instead offer more open development possibilities.

[8] http://www.soros.org/openaccess/software/

[9] http://www.eprints.org/services/sales/

[10] http://www.openrepository.com/