From a national CRIS along the road to Green Open Access – and back again: Building infrastructure from CRIStin to Institutional Repositories in Norway

Lars Wenaas[a], Nina Karlstrøm[a], Tore Vatnan[b]

[a] CRIStin, Oslo, Norway

[b] USIT, University of Oslo, Norway

Summary

In Norway the institutional repositories are seen as the most important piece of the Open Access-infrastructure-puzzle. The Norwegian national CRIS is the single point of entry for all Norwegian research, and this can be used to increase the number of Open Access-publications in the institutional repositories. The CRIS is used as a delivering-portal for fulltext, and there is a great potential in identifying the large portion of Norwegian research results that are candidates for self-archiving, but for some reason do not end up in the repositories today. The infrastructure is strengthened through three steps; building an easy route for data from the CRIS to the repositories, identify the portion of candidates for self-archiving, and finally, distribute the content of the repositories to the rest of the world.

1 CRIStin at a glance

CRIStin, (Current Research Information SysTem in Norway), is a national CRIS serving more than 160 institutions within the university/college sector, research institute sector and specialized health care. CRIStin as a national CRIS came into being after merging the CRIS (formerly known as FRIDA), used by four universities, and the CRIS used mainly by the university colleges (formerly known as ForskDok). The research institute sector and the specialized health care reported their publications by separate means.

80% of public funds for R&D in higher education institutions are channelled directly from the Ministry of Education and Research. 15% of these again are divided into two parts: 1) a performance-based part, where funds are distributed among institutions on the basis of benchmarks for publications and competitive research funding, and 2) a strategic component allocating funds to institutions for positions for PhD students and for scientific equipment.

The CRIS is the main reporting tool for research-based funding. All institutions report their academic publications through the CRIS to the Ministry of Education and Research and the Ministry of Health. In addition the CRIS in CRIStin aims to be a useful tool for researchers and research institutions by providing a number of services such as simplifying administrative routines for the researchers, and making widespread reuse of the high quality data registered in the CRIS. CRIStin also has the national responsibility for negotiating consortium agreements for access to electronic

resources for HE libraries and publicly funded research institutions. And last, but not least, CRIStin is the national coordinator for work on Open Access in Norway.

In CRIStin we have examined how to use the CRIS, and the data in the CRIS to promote more and better deposits into the Institutional archives. Since the CRIS is the single point of entry for all research publications in Norway, (Open Access or not), the CRIS can also serve as a delivering portal for fulltext in cases where there are Open Access-permissions. Make note that the CRIS itself does not store any fulltext, the main job for the CRIS is to store metadata of the totality of research done by Norwegian scientists, but in building a better Open Access-infrastructure around the CRIS, we believe that the archives will be populated further with fulltext.

This strategy is divided into three parts;

- Step 1: Make sure it is easy for the researcher to upload a copy of the article, and transfer it to the author's institutional repository.
- Step 2: Identify the (quite large) portion of articles registered in the CRIS, where the publisher allows for depositing in an institutional repository.
- Step 3: Distribute and showcase the content of Norwegian institutional repositories to external services.

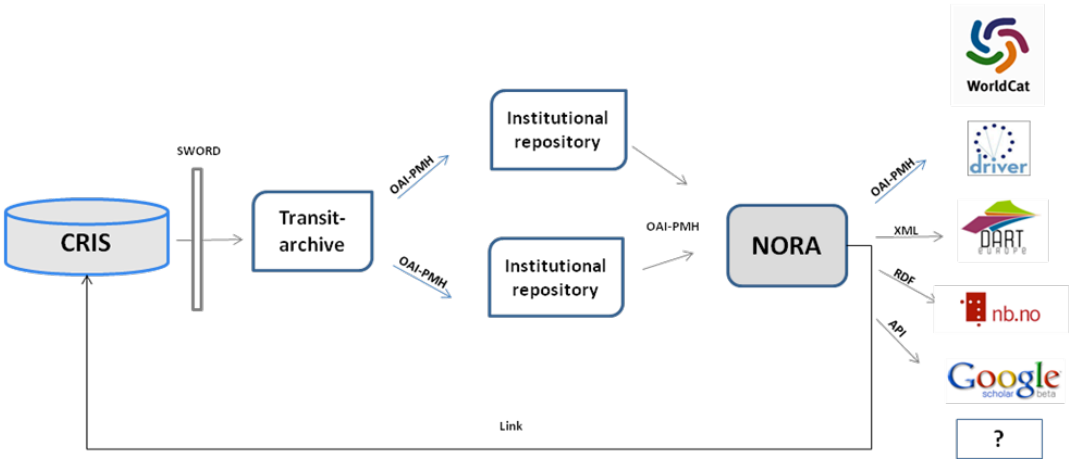

Figure 1: The principles of the infrastructure

All these steps helps to solve challenges Open Access has in Norway today, and further underlines the, in our view, sound connection between a CRIS and Open Access; between a system for metadata and institutional repositories.

2 Building a highway to the archives

All researchers must register all their research publications in the national CRIS, and while doing so they may upload a copy of their article. This copy will be transferred to a dedicated DSpace-installation by the SWORD protocol. This transit archive (as we call it) serves only one purpose; To store the fulltext until the administrators of the institutional archives take their time to collect it and deal with it within their own routines.

The choice of having a DSpace installation has a technical side to it and from a Norwegian point of view a quite obvious one: almost all of the institutional repositories in Norway are based on DSpace. But more important, the technical solutions we use must support the proper standards for transferring data in our domain, namely OAI-PMH[1] and OAI-ORE[2].

The principle for the upload-process is to make things as easy as possible for the researcher, but at the same time gather as much information as possible to make the clearance-job of the repositories administrations as easy as possible. Below is a snapshot of the form used in the CRIS-application.

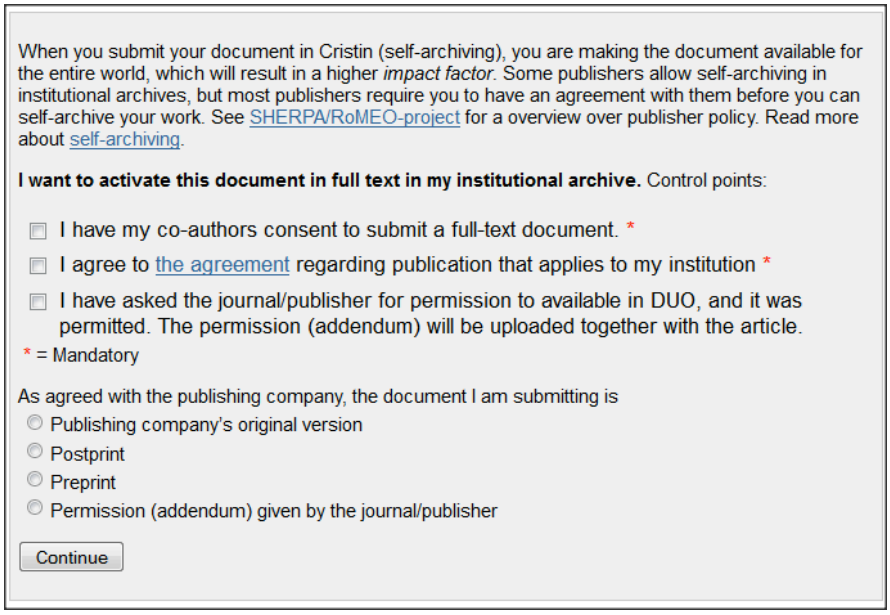

Figure 2: Article submission form

[1] Open Archives Initiative Protocol for Metadata Harvesting

[2] Open Archives Initiative Object Reuse and Exchange

3 Fill the repositories with all possible findings

A Norwegian survey[3] by Sigbjørn Hernes (Lillehammer University College) shows the potential for self-archiving in Norwegian repositories. The survey is based on articles registered in the CRIS and reveals that as much as 40% of the articles in the CRIS could have been self-archived in a repository. This paper will not dwell on the various reasons why the occupancy level in the repositories is at such a low, but it might have something to do with the awareness around self-archiving both within the institutions and the researchers. We need good Open Access-policies on every level to accommodate this. This may take years, in the meantime we are focusing on some practical routines inside the CRIS.

Here a web-based search in the CRIS-application will:

1. Identify the potential candidates for self-archiving, one article at a time. This is done using the Sherpa/Romeo information, a database containing the different publisher's policy on self-archiving.
2. Put together lists of articles with the status "Green Open Access" to be sent to managers of institutional repositories.
3. Make it possible to suppress "false positives". (For example cases where the researcher does not wish to self-archive, and where there is no institutional self-archiving policy.)
4. Generate RSS-feeds to monitor changes (check for new candidates from an extern website).

When this routine is run by an archive administrator, he or she can contact the researchers and ask them to upload a copy of the fulltext through the CRIS. The system will then, as described in step one, pass on the article to the relevant institutional archives.

A note on cases where there are two or more authors: If an article is initially cleared by the Sherpa/Romeo-system, there is still the matter of getting the co-authors allowance to deposit a copy in a repository. When this latter clearance is done, the article will automatically be passed on to the co-authors corresponding institutions as well. The two (or more) institutions will then have to confirm, by themselves or together, whether every legal matter stands as they should be. Although CRIStin will do a check with the Sherpa/Romeo database for a first clearance, further clearances for versions, embargo etc must always be done by the local repository.

To say that this part of the strategy is mostly concerned with the green road[4] is at best a bit misleading. The core service along the green road is to check whether one actually is allowed to deposit an article in a repository. The uncertainty concerning this possibility is non-existing when following the golden road; here we know that everything is Open Access, and thus cleared with respect to all the proper legal issues. Here we have the option of automatically downloading the article and pass in on to the relevant archives. This can be done as long as two demands are met; The CRIS must know the article is published in a proper Open Access-journal and the record must contain the DOI (the direct link to the article).

[3] The survey can be found (in Norwegian only) on the homepage of 'The Norwegian Association of Higher Education Institutions': http://www.uhr.no/documents/Oppdatering_potensiell_OA_2007_.doc

[4] Green Open Access differs from Gold Open Access, the latter is when the article is free to read for all readers and payment for publishing is done by the author. Green Open Access is based on the willingness of the individual publishing house on when (if ever) the article in question can be deposited in a repository.

As a test-case, CRIStin is working with BioMed Central to receive article feeds from BMC with articles written or co-written by Norwegian authors. Once the proper data is in the CRIS, we can pass on the article in fulltext automagically to the archives. This routine is on a prototype-stage at the time of the release of this article.

4 Spread the word

As of January 2011 CRIStin runs the service NORA (Norwegian Research Archives), which harvests Norwegian institutional repositories. The service has been financed as a project for several years, but is now part of the CRIStin portfolio. In its core, NORA is a standalone search-engine with an index based on the content of all Norwegian repositories.

The service has been based on:

- Harvesting and importing research publications from Norwegian repositories (including links to the articles)
- Distributing research publications and master theses to the major search engines
- Showcasing Norwegian Open Access research (and especially the links harvested in the first step)

These tasks are still as valid after CRIStin has taken over the responsibility for the service, but the plans for the future is to fully integrate NORA in the CRIS, and thus strengthen both the CRIS and the searchability of Norwegian archives. We also shift weight over to the distribution-principle; our task is to make Norwegian research visible and available, not necessarily visible and available only within our own services. We believe other services (Google Scholar etc.) are better services for the end user, and definitely more preferred, therefore it makes sense to abandon the idea of being *the* primer entrance to Norwegian research in fulltext. If you find a Norwegian article through Google Scholar, Dart, agents or anywhere else, well, then our job is done.

4.1 Semantic web

Spreading the word is also about CRIS-metadata, which also includes link to the fulltext in the proper repository. We believe enabling a SPARQL-endpoint in CRIStin will further increase the use of both fulltext and metadata and intend to implement semantic web technologies by establishing a SPARQL endpoint that gives access to all relevant metadata related to the articles. By the use of common vocabularies, data from CRIStin can easily be incorporated in external services and information systems. External systems/services will be able to specify what they want from CRIStin, and thereby eliminating problems concerning (proprietary) data exchange formats. As this process is fairly lightweight compared to older ways of importing data, data from CRIStin can be imported/shown when needed, ensuring that data is always updated.

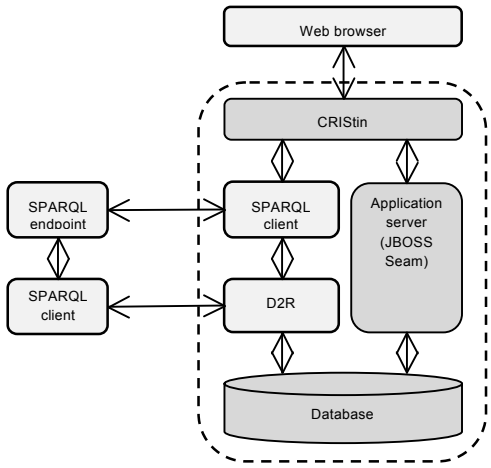

Figure 3: Semantic web-stack

5 Conclusion

We believe there is a great potential in the crossing field between the national CRIS, the Institutional repositories and Open Access. By using the CRIS as a single point of entry for all research there is a potential to increase holdings in the IRs and also identify the large portion of Norwegian research results that are candidates for self-archiving. The process can and should be automated where possible, especially when it comes to gold Open Access-publications. There is nevertheless another important task that the CRIS should perform, this is to distribute and disseminate the research to the end-user and increase visibility of the research results. After all, this is what Open Access is all about.

We have yet to measure the effect of this strategy, it is still uncertain to what extent these efforts actually have the effect we believe them to have. Most of these services have been launched recently or will be shortly, but some components are still on the drawingboard. Even so, we are optimistic on the prospects of filling up the archives with fulltext in the long run. Visit http://www.cristin.no in the near future, to see whether this strategy works or not.

Contact Information

Lars Wenaas – larswen@cristin.no

Nina Karlstrøm – nina.karlstrom@cristin.no
CRIStin
P.O. box 1059
Blindern
0316 Oslo
Norway

Tore Vatnan – tore.vatnan@usit.uio.no
University of Oslo
USIT
P.O. box 1059 Blindern
0316 Oslo
Norway