Extended Abstract: Repositories Recreated – Working Towards Improved Interoperability and Integration by a Co-operative Approach in Sweden

Stefan ANDERSSON\textsuperscript{a,1} and Aina SVENSSON\textsuperscript{a}

\textsuperscript{a}Electronic Publishing Centre, Uppsala University Library

Introduction

Recently the technological and organizational infrastructures of institutional repositories have been questioned. For example the British so-called Finch report\textsuperscript{2} from last summer argued that further development, as well as higher standards of accessibility of repositories, are needed in order to make them better integrated and interoperable to ultimately bring greater use by both authors and readers. Not only the technical frameworks and presumably low usage levels are criticized but also the lack of “clear policies on such matters as the content they will accept, the uses to which it may be put, and the role that they will play in preservation”. The report concludes that: “In practice patterns of deposit are patchy”.

As in the UK, today, all universities and university colleges in Sweden, except a couple of very small and specialized ones, do have an institutional repository. A majority (around 80%) are working together on a co-operative basis within the DiVA Publishing System\textsuperscript{1} with the Electronic Publishing Centre at Uppsala University Library acting as the technical and organizational hub. Because the system is jointly funded, and the members contribute according to their size, it has been possible even for smaller institutions with limited resources to run a repository with exactly the same functionalities as the biggest universities.

In this presentation we want to demonstrate the ever-increasing importance of institutional repositories in Sweden. Starting more than a decade ago the DiVA Consortium has, for some time, been addressing the problems now raised by the Finch report in a number of areas:

\footnote{1 Corresponding Author.}
\footnote{3 See: \url{http://www.diva-portal.org/}}
1. Search engines, visibility and usage (downloads)

Our firm experience is that indexing in major search engines like Google is most crucial for the visibility and consequently the number of downloads from the repository. Therefore large efforts have been directed towards facilitating the crawling of search engines, for example by adding HTML metadata or using sitemaps, to make the individual pages or files easier to find. Since about 75% of the full-text documents are found directly from Google improvements in DiVA in this regard have led to substantially increased levels of usage. Interestingly, looking at the referring links, inclusion in search portals or library catalogues composed of harvested metadata records, on the other hand, show no or very little significance for the number of hits in the repository. Web statistics of full-texts from DiVA for last year show that the collection of 115,000 files was downloaded as many as 15 million times in total. The single institution with the highest rate, Linköping University, counted 1.8 million full-texts. Compared to the UK where, according to the Finch report, there were 585,000 downloads from the biggest university repository (UCL) in 2011 that sum can be matched by medium-sized Swedish university colleges in DiVA like Halmstad or Jönköping.

2. Bibliometrics and current research information

In the last few years the practice of bibliometrics in evaluation of Swedish research has strongly increased. As a result DiVA has since 2008 acted as both a bibliographical catalogue of the entire research output from a university and as a repository, with the two sides completely integrated in the system. Whereas the usage of the research results, i.e. the publications themselves, is without borders (Swedish research publications are predominantly written in English and downloads occur from more or less every country in the world) bibliometrical studies are most often restrained to local or, at the most, national conditions. The handling of metadata for statistical calculation is based on reliable and persistent identifiers for authors, departments, subjects, and publication types, and the hierarchical relations between all of these. Because of this DiVA can, as opposed to widespread repository software like DSpace or Eprints, treat complex metadata relations as well as a number of controlled identifiers. Within the DiVA framework the aim is to harmonize the description format and to create common policies for registration procedures as what material to include in the database or not. A common practice will eventually lead to the possibility to compare the different universities with each other from a mutual point of view. Furthermore, the university libraries play an important role in reviewing the metadata records. Workflows for examining and rectifying the records have been set up in the system which, in the end, will result in a much higher quality of content. The use of persistent identifiers also makes it possible for DiVA to interact by APIs with university web sites and provide a showcase for their research where lists of publications can be automatically created for researchers’ or departments’ homepages.
3. Infrastructure and preservation

Right from the start DiVA has been directed towards long-term preservation and availability of the publications. Since 2008 we are using the open-source digital object repository system Fedora to manage the digital collections. From early on a connection was set up to the national library’s system for distributing digital object identifiers in the form of a URN:NBN (National Bibliographic Number) which will ensure working web links. Additionally, in Sweden, as a consequence of the principle of public access to official documents, university archives must keep theses and student papers available to the general public. Since these documents increasingly are published primarily in electronic formats a number of universities have set up routines where the digital files and their metadata are finally harvested from DiVA also to the local university archive for long-term preservation. Using a common repository system as DiVA will, obviously, also in itself create a higher level of interoperability on a national level.

4. Content and publishing

Contrary to what is stated in the Finch report the Swedish repositories, or universities, may well act as publishers themselves. From the beginning the focus on repository publishing in Sweden has lied on grey and unique material, like theses or working papers, rather than parallel versions of original articles. The doctoral theses published in the system are also considered to be the “original” version of the document. The process of producing both the printed and electronic versions is largely integrated and DiVA can also be used for dynamically creating links to bookselling in web-shops for the printed items. As most Swedish universities impose some kind of obligations to publish the doctoral theses online the level of completeness is very high.

5. DiVA and future development

It is clear that open institutional repositories are well established in Swedish academia and widely used for many different purposes like open access publishing, bibliometrics and long term preservation. Data from DiVA is reused in many different ways and for various purposes. DiVA has an obvious and well known role as an integrated system within the University infrastructure.

Next steps within the DiVA Consortium will be to further collaborate to improve (or follow up) metadata standards and quality and to further customize DiVA to the users evolving needs (e.g. administrators at the University management, librarians or researchers). We have already started discussions about sharing metadata and full-texts in the DiVA publishing system between members, with the aim to facilitate for researchers and administrators (and increase quality). Another thread is to look into the upcoming demands to make research data freely available and connect it to the publications.