

Repositories recreated – The Finch report versus DiVA in Sweden

Stefan Andersson* and Aina Svensson

Electronic Publishing Centre, Uppsala University Library, Uppsala, Sweden

Abstract. In this paper we give an overview of the development of institutional repositories in Sweden, by addressing the criticism raised in the Finch report. In contrast to how the use and development of institutional repositories in the UK is described we show the ever-increasing importance repositories take in Sweden, with DiVA as our particular example.

Keywords: DiVA, Sweden, electronic publishing, digital publication, institutional repositories, research information, Open Access, the Finch report

1. Introduction

Recently the technological and organizational infrastructures of institutional repositories have been questioned. The British so-called Finch report [1] from last summer argued that further development are needed in order to make them better integrated and interoperable to 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”.

In Sweden we see a very different development of institutional repositories, compared to what is described in the Finch report. The institutional repositories today function as integrated services with many different applications and uses within a University or a research institute. The repositories are built on a thorough technical framework with high quality metadata standards, a prerequisite so that data can be used as a basis for evaluation and for allocating funds on a local basis, as well as to be re-used in other external web services. Moreover, the repositories are initially built for long-term preservation and most successful in disseminating academic publications from Swedish universities, i.e. open access.

In this paper we want to address the criticism in the Finch report and instead demonstrate the ever-increasing importance of institutional repositories in Sweden, with DiVA¹ as our particular example.

2. Background – A brief history of institutional repositories in Sweden

Interest in digital publishing and open access started early at Swedish universities and has primarily been driven by research libraries, with valuable support from the National Library of Sweden. In the mid

*Corresponding author. E-mail: stefan.andersson@ub.uu.se.

¹DiVA Portal: <http://www.diva-portal.org>.

1990s started the first activities for digital publishing, although on a small scale and in the beginning by simple methods. Typically this was done by placing bibliographic information and abstracts on a web site or in a simple database. Some of the earliest initiatives for digital academic publishing came from the universities in Linköping, Lund and Gothenburg together with The Blekinge Institute of Technology (BTH) [3].

When the first databases or software for repositories were created in 2000 it became possible to collect and disseminate digital publications in a completely new way. EPrints is a free and open source software package for building institutional repositories, developed in 2000 at the University of Southampton School of Electronics and Computer Science.² During the same year DiVA was created at the Electronic Publishing Centre at Uppsala University Library and several universities took the opportunity to become members during following years. Two years later, in 2002, DSpace was released at the Massachusetts Institute of Technology (MIT), also an open source repository software package used for creating repositories for scholarly published digital content.³ With several new software and tools adapted for digital publishing, the academic full text publishing took off in Sweden. Universities either chose to participate in the DiVA Consortium or to implement available open source software or developed local solutions.

In addition to the increasing full-text publishing at Swedish universities there also has been a growing need to report on all scientific publications at a single university. These registrations were at the beginning collected in separated databases without the possibility to add a full text. Therefore, many repositories were further developed to become integrated systems, now used both for digital publishing and for statistics and bibliometrics.

3. About DiVA and DiVA Consortium

The development of DiVA was preceded by a well prepared pilot study which facilitated the technological development. From the beginning, the focus was mainly to publish doctoral theses and to define and implement a workflow for posting the theses [2]. The requirement of bibliometrics and evaluation of Swedish research led to a completely new DiVA system, with the extended purpose to cover the entire research output from a University. In 2008 the new version was released, built on open source software and with an extended metadata model to describe different types of academic publications. DiVA is now a combined system used both for full text publications (open access) and for registering and bibliographic references (bibliometrics).

Today, DiVA is a joint collaboration at 34 universities, research institutes and museums. All members contribute to the system development, although the Electronic Publishing Centre at Uppsala University Library is acting as the technical and organizational hub. Publications from the various institutions are searchable in the common DiVA portal as well as in the local DiVA search interface. Although all members use the same system, layout and some functionality can be adapted locally. 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.

²See <http://www.eprints.org/>.

³See <http://www.dspace.org/>.

Table 1

Institutional repositories and different software or publishing system used by universities, research institutes and museums in Sweden

Software/publishing system	Number of users
DiVA	34*
DSpace	4
EPrints	1
Pure	1
Other (locally developed solutions)	3
In total	43

*One of the DiVA members is from Norway (The Norwegian University of Science and Technology).

4. Open access in Sweden – Current situation

All universities and university colleges in Sweden, except a couple of very small and specialized ones, do have an institutional repository. Also some museums and research institutes in Sweden have their own repository. Currently there are a total of 43 institutional repositories in Sweden (Table 1). Since 2009 research publications from the various repositories (DiVA, DSpace etc.) are searchable in the national portal SwePub.⁴

The National Library of Sweden has played an active role in coordinating and promoting open access and the development of institutional repositories in Sweden. The OpenAccess-programme, OpenAccess.se, is a platform for cooperation between research libraries and the major stakeholders in the research sector, such as research funders.⁵

Important achievements for open access in Sweden are the large number of universities (i.e. 17) with an open access policy or recommendation, as well as that 6 of the Swedish research funders have open access mandates. Furthermore, as part of the Swedish research bill for 2013–2016, the Swedish Research Council has been commissioned to develop forms and national guidelines for open access, a work that will begin during 2013.

5. The Finch report versus DiVA

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.

5.1. 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 for 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. In comparison 15% of the files were found by internal links, bookmarks or direct links and 10% from external pages (other web sites except search engines); see Fig. 1.

⁴See <http://swepub.kb.se/>.

⁵See <http://www.kb.se/openaccess/Hjalptexter/English/>.

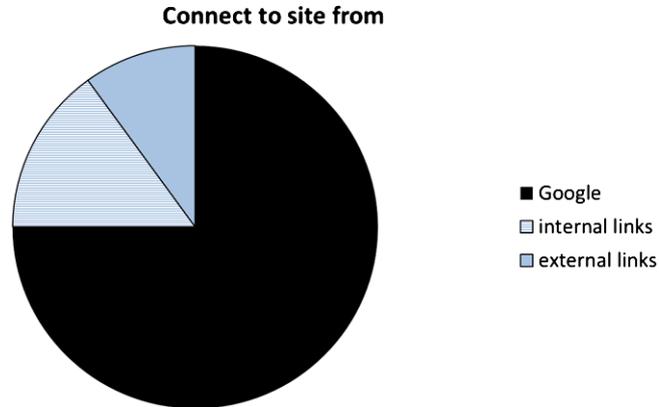


Fig. 1. Origin of connections to full texts in DiVA in 2012 in percent. (Colors are visible in the online version of the article; <http://dx.doi.org/10.3233/ISU-130696>.)

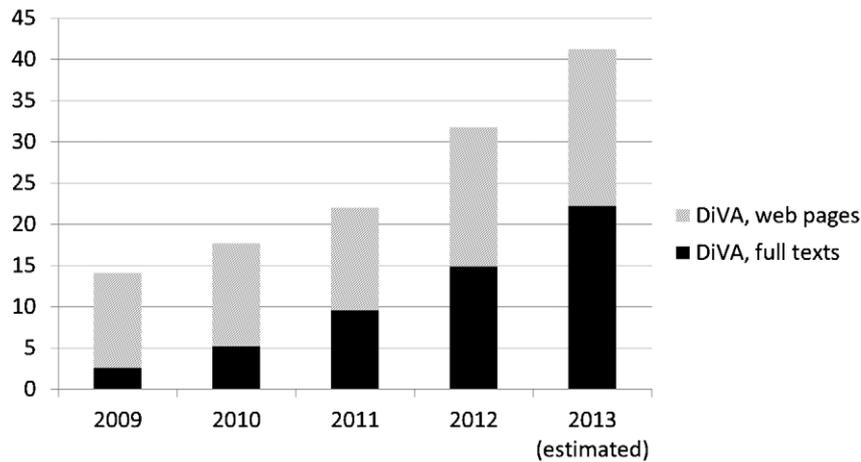


Fig. 2. DiVA: web statistics (in millions per year).

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. About 30% of the files were downloaded from the Swedish.se domain. Large numbers were also downloaded from the United States, China and India and more or less every country in the world was represented as a user.

Judging from the first five months of this year (2013) the usage will continue to rise at the same rate as previous years, estimated to more than 20 million downloaded full texts in total (Fig. 2).

The single institution with the highest rate, Linköping University, counted 1,8 million full-texts. Compared to the UK and according to the Finch report, there were 585,000 downloads from the biggest university repository (UCL)⁶ in 2011. That sum can be matched by medium-sized Swedish university colleges in DiVA like Halmstad or Jönköping; see Table 2.

⁶Cf. University College of London – 585,000, highest in the UK.

Table 2
Downloaded full text files in 2012 per university in DiVA

1	Linköping University	1,485,896
2	KTH – Royal Institute of Technology	1,267,900
3	Uppsala University	1,103,762
4	NTNU – Norwegian University of Science and Technology	995,193
5	Linné University	688,742
6	Umeå University	672,199
7	Stockholm University	642,029
8	Halmstad University	473,423
9	Mälardalen University	433,167
10	Jönköping University	427,798

5.2. *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.

Furthermore, in Sweden, the need for common identifiers for authors has grown increasingly important in order to share publication metadata between different systems, not only within the university itself but also on a national level. ORCID,⁷ which is an open, non-profit, community-driven effort to create and maintain a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers, has been recognized as the main alternative in this case and nationally the Swedish Research Council will actively support the implementation of the system.

5.3. *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

⁷See <http://orcid.org/>.

distributing digital object identifiers in the form of a URN:NBN (National Bibliographic Number) which will ensure working web links.

Additionally, as a consequence of the principle of public access to official documents in Sweden, 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.

5.4. 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.

DiVA can also be used as a low cost e-publishing platform for open access-journals where long term preservation and access can be organized within the system in combination with integrated web pages created by including the information through the APIs. At Uppsala University in particular specialized journals within the humanities have taken advantage of this solution. Some examples are *Futhark – International Journal of Runic Studies*⁸ and *Samlaren*,⁹ a journal for Swedish literature research.

6. 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.

Based on our experience and code from DiVA we are working on a new project, Alvin, focusing on cultural heritage collections. The Alvin platform will handle archives and a wider range of items with different types of metadata such as manuscripts, maps, images and music.

⁸See <http://www.futhark-journal.com/>.

⁹See <http://www.svelitt.se/samlaren/>.

References

- [1] Accessibility, sustainability, excellence: how to expand access to research publications, Report of the Working Group on Expanding Access to Published Research Findings, June 2012.
- [2] S. Andersson, När Digitala vetenskapliga arkivet blev DiVA, in: *I lag med böcker: Festskrift till Ulf Göranson*, Acta Universitatis Upsaliensis, Uppsala, Acta Bibliothecae R. Universitatis Upsaliensis, Vol. 46, 2012, pp. 15–29.
- [3] T. Hedlund and I. Rabow, Open Access in the Nordic countries, State of the Art Report & Workshop Views and Recommendations, Commissioned by The Nordbib Programme, 2007.