

# DiVA Publishing System. The Community's Collaborative Development Approach

*Eva Müller, Frida Sandgren, Stefan Andersson, Uwe Klosa, Peter Hansson*

Uppsala University Library  
Box 510, SE-751 20 Uppsala  
{eva.muller | frida.sandgren | stefan.andersson | uwe.klosa | peter.hansson}@ub.uu.se

## Abstract

In this paper we give an introduction to the DiVA publishing system developed at Uppsala University Library, focusing on present functionality and ongoing development activities. We want to share our experiences of a community's collaborative system development approach. Our intention is to contribute to understanding the advantages of a community system development and the necessary organizational framework to support it. The DiVA project provides a good example of how electronic publishing is possible to carry out within a university library organisation and how a strategy of collaborative development between a number of institutions can work in practice.

## 1 Introduction

DiVA—"Digitala Vetenskapliga Arkivet" (Digital Scholarly Archive)<sup>1</sup> is a publishing system and a service developed and maintained within the DiVA project based at Uppsala University, Sweden. DiVA supports workflows for both electronic publishing and printing. The great advantage of DiVA is that it enables publishing in XML. The system treats the electronic copy of the document as the "Digital master" for both the electronic and print versions.

A significant effort has been put into developing practical solutions to support longevity of the electronic documents and to ensure long-term access. Solutions supporting this are integrated into the system. Each document is assigned a persistent identifier based on URN:NBN<sup>2</sup>, is stored along with checksums in the live repository and the archive copy is also stored in the local depository (the DiVA Archive).

The metadata are stored in the DiVA Document Format<sup>3</sup>, a rich, locally developed and XML based schema.<sup>4</sup> The transformation of this schema enables the provision of various metadata services, such as harvesting via OAI-PMH<sup>5</sup>, or the automatic generation of catalogue records for local and national catalogues.

In co-operation with the Royal Library, the national library of Sweden, a workflow for the exchange of preservation packages has been established. This approach assures long-term storage and access to the documents from the Royal Library.<sup>6</sup>

The DiVA publishing system was developed with a focus on how to achieve rational and convenient workflows for both authors and administrative staff working in the publishing process and simultaneously increase efficiency and reduce production costs. The resulting workflows are based on the reuse, in many different contexts, of the structured data originally created by authors.<sup>7</sup>

The system is based on current standards and recommendations and is implemented using Java and XML technologies. From a system architecture point of view, DiVA is built using a component-based design methodology, which enables flexibility and the potential for collaborative system development.

---

<sup>1</sup> For a comprehensive description of the DiVA system and the DiVA project see Müller, E.; Klosa, U.; Andersson, S.; Hansson, P. (2003) and Müller (2004)

<sup>2</sup> RFC 3188 (2001)

<sup>3</sup> DiVA Document Format

<sup>4</sup> Müller, E. Klosa, U. Hansson, P. Andersson, S. Siira, E. (2003)

<sup>5</sup> The Open Archives Initiative Protocol for Metadata Harvesting

<sup>6</sup> Müller, E. Klosa, U. Hansson, P. Andersson, S. (2003)

<sup>7</sup> Müller, E. Andersson, S. Klosa, U. Hansson, P. (2003)

The DiVA system has been in full operation since January 2003 and it is used by a number of Scandinavian universities. These universities formed a consortium in 2002 and today it consists of twelve members.<sup>8</sup> Cooperation in the DiVA project is open to all universities and publicly financed research organizations and the number of consortium members is constantly increasing. The idea behind the consortium is to share products and technical solutions developed within the DiVA project, and to allow the exchange of experiences between people working with individual implementations of the system and electronic publishing in general. Questions of common interest are discussed on a regular basis and a number of agreements supporting interoperability have been achieved within the consortium.

For example, the participating universities have agreed upon a common document format (the DiVA Document Format) and a list of subject categories.<sup>9</sup> These were fundamental pre-requisites for the DiVA portal (<http://www.diva-portal.org>), which is a common interface for local DiVA repositories.

Other issues important for long-term access and preservation are also addressed within the consortium. These issues are the use of persistent identifiers, storage formats and metadata. DiVA gives enormous visibility to the published documents since co-operation with search services such as Scirus and Google Scholar Search have been established.

## 2 Collaborative System Development Approach

How can we accomplish our goals faster and more competently and at same time satisfy our shared needs with the limited resources we have? Can we use the resources more efficiently? Do we have all the resources (money, staff, and equipment) we need within the organization?

These are some of the questions which universities and university libraries are facing when building infrastructures to support the publishing and dissemination of research results.

To build such an infrastructure is a complex task and many factors have to be taken into consideration when choices of how to proceed are being made.

One of the promising approaches is to collaborate in an organized way with others who already have some experience within the field. This approach is becoming more popular and is used in the development of systems within higher education like by Sakai, OSPI and Kuali; projects where open source solutions have been developed and are available for their community.<sup>10</sup>

What is a collaborative system development model and why collaborate? The collaborative model makes it possible to pool resources. There is no doubt about the potential of collaborative development, though not all collaborative projects will succeed. Through our experiences from the DiVA project we have identified a number of conditions, which can ensure success, and sustainability of a collaborative system development. These are: *clear terms of participation*, *an effective project management role* and *good communication among participants*.

Additionally, for better understanding it is necessary to consider the key role of *stakeholders*.

### 2.1 Stakeholders

In our opinion, our success is, to a great extent, due to the fact that the project started as a local enterprise with a focus on the local issues and funded by local sources. Both the university and the university library – the stakeholders – have a shared interest in the success of the tools and workflows developed by the project. Moreover, they gave the project the support which was needed.

From the very beginning the DiVA project was met with positive interest from the local research community and this has helped to bring the library community and researchers from other relevant fields together and the project could benefit greatly from it.<sup>11</sup>

Once this winning model had been introduced at one university, the adoption by other universities was relatively straightforward. It depended, of course, on the shared priorities and goals within the project and self-interests within participating organizations, but also on the fact that the collaboration is extending to make use of a shared maintenance of the system.

---

<sup>8</sup> DiVA consortium members are: Århus University, Denmark; NTNU—Norwegian University of Science and Technology, Norway; Jönköping University, Linköping University, The Royal Institute of Technology in Stockholm, Stockholm University, Södertörn University College, Umeå University, Uppsala University, Örebro University, and Växjö University, Sweden.

<sup>9</sup> Ämneskategorier för vetenskapliga publikationer [Subject categories for scientific publications]

<sup>10</sup> Brooks, Lois (2004)

<sup>11</sup> See section: 2.4 Related Research Projects

## 2.2 Clear Terms of Participation

The participation within the project is open for all universities and publicly financed research institutions and is based on a written agreement between Uppsala University and each participating organization. This agreement states conditions for participation and for regulating costs, staffing and equipment at each institution.

A major principle in the project is that all economic resources, which institutions initially bring to the project, are used for further development of the system and its sub-systems. Additionally, where there is a lack of resources, the requested new development costs will be shared by all the participants. This lack of dependency on external funding guarantees sustainability and the possibility to enhance the system with new functionalities according to demands raised by the community. The decisions on priorities are made after consulting all participants.

This approach results in more rational development planning and at same time makes it possible also to focus on research and experimentations.

## 2.3 Effective Project Management Role and Good Communication

As mentioned earlier, the DiVA project was initiated at Uppsala University in 2000. The work done in Uppsala has been met with positive interest from universities both within Sweden and in other countries. In 2002, a consortium was formed and there are now twelve universities from three countries in this network.

Such a transformation from a single university project to a cooperative network involved a good portion of challenges and innovative thinking.

However, the common vision, shared priorities and goals have helped to develop ways to work collectively and to communicate effectively. Moreover, understanding the mutual benefits of working together allows the project to move on faster. Through working together we can draw on expertise from different institutions to contribute knowledge and skills and work on developing not only technical solutions, but also for example content and institutional policies concerning intellectual property rights. And even more important—others within the network develop skills in these areas.

The efficiency of the project leading and quality of communication is heavily dependent the supporting infrastructure. Examples of such an infrastructure are given below.

### 2.3.1 Channels of Communication

- **Helpdesk**

The question of support often arises when considering the use of open source software. An organization that maintains and develops the open source software further often offers fee-based support packages for the user community. Uppsala University Library, the organization that is maintaining, developing and for some participants hosting the DiVA publishing system, has adopted the same support model as that used by others in the open source community.

The current helpdesk communicates with the user community through several communication channels such as phone, email, a mailing list and a wiki-website (maintained collaboratively by the community). A technical helpdesk is pre-requisite for participants that use the DiVA publishing system at Uppsala University Library. For other participants running the DiVA publishing system locally, with lack of insight in the technical part of the system, a subscription to the existing technical helpdesk service is more cost effective than building up new local expertise.

- **Implementing workshops and training services**

All new participants of the DiVA community are offered an introduction to the publishing system from the user perspective. The aim is to inform newcomers about important features and pitfalls early on to reduce the load of work for the helpdesk.

- **System developers' area**

The great advantage of DiVA is that a component-based system development methodology was used while developing the system. That, in combination with well-defined interfaces, makes it possible to replace components with more advanced versions, or even add new ones. In this way the developers of the DiVA system can continuously add new functionality and make improvements. This model of the further development of DiVA supports a cooperative effort. Moreover, it also offers a great advantage—the components can be developed with a simple functionality, which is sufficient for the dedicated function and as more resources and new demands arise a single component can be replaced by a more advanced one.

To support this collaborative development of the source code an infrastructure for collaborative development has been established. This infrastructure is based on the use of a concurrent versioning system for source files, centralized automatic build function, a mailing list and a DiVA developer wiki website.

The versioning system makes it possible for developers at any of the member organizations to contribute to the project. This is organized by contribution to different modules.

One of the modules which is currently developed collaboratively is a search and index module based on Lucene. This module will replace the commercial database, which is now used for indexing and searching. Though not all of the participating universities have the possibility to contribute with code hopefully we will see more collaboration in this area in the future.

- **User group meetings, reference group meetings, mailing list and news service**

Two major official DiVA-meetings are held each year. At these meetings DiVA participants as well as representatives from organizations interested in the DiVA project meet to update themselves on current activities, listen to reports from the working groups etc. Traditionally, there are also training sessions held for new participants after each DiVA-meeting. Along with the growing number of participants, coordinating information to the group, for which there is an e-mail list as well as a newsletter sent each month, has become increasingly more important.

To assure exchange of experience and ideas between the local research community and the project a reference group was established. This group is a forum for researchers from disciplines related to digital libraries and the project where new ideas, related to the current research agenda, can be gathered. These groups also provide the project with feedback and grant a critical mass and quality filter.

### 2.3.2 Further Development

However, working collaboratively with others requires true commitment and willingness to value perspectives and skills of others. This attitude is necessary when priorities within the project are discussed and new developments are planned.

To achieve smooth decision making clear processes have been established in the project. These include identifying and gathering ideas and issues important for the project as well as managing it.

For gathering, adaptation and processing of ideas a model of working and focus groups has been adopted.

Typically, a working group is established when there is a need for a general agreement on issues pertaining to the development and enhancement of DiVA. As an example, one of the current groups is dealing with policy issues concerning publishing of pre- and post prints in DiVA. Usually, such issues need to be discussed and requirements for changes and future development of the system established before any actual implementation can take place.

There is also a working group, which is engaged in exploring workflows for the publishing of undergraduate theses. The motivation is to carry out a system and usability requirements for the development of a more flexible workflow associated to the publishing of undergraduate theses in DiVA. As an example, surveys on how undergraduate theses are managed before publishing have been conducted at several institutes of higher education.

Finally, one of the groups is focusing on print-on-demand services for end-users and on how to integrate such a service into the DiVA system.

## 2.4 Related Research Projects

As an example of current research projects related to DiVA the project 'Educational Content Markup'<sup>12</sup> has as its overall goal to develop tools and methods for semantic mark-up of educational texts. The project is a collaboration between partners such as the Swedish Royal Library, ourselves at the Electronic Publishing Centre at Uppsala University Library and the research program Digital Literature at Uppsala University.

Within the project, a tool is being developed for authors of scientific documents to use in their creation of semantically marked-up texts. Results from the project will also bring benefits for the DiVA system, as it will provide a basis for building new services for end users of published documents. The end user can be provided with additional context information, connected to the marked semantic elements within the text on the web. This can be accomplished by linking online resources.

---

<sup>12</sup> Educational Content Markup

Such additional information can in turn be used to increase the readability and understanding of the text. It gives a possibility for end users such as undergraduate students or scholars to acquire more detailed information about the content.

### 3 Conclusions

As demonstrated in this paper, the community's collaborative development approach can be a practical strategy used to develop solutions and projects in the area of scholarly communication, especially when it is integrated into the library environment.

Libraries are used to collaborating wherever possible.

The understanding that collaborative model approach is not only sharing, but also accepting and respecting the fact that other perspectives can add value to our own, is an important concept.

Also, the collaboration around the development of an electronic publishing system and services is not so different from other types of collaboration.

It has the necessary pre-requisites—the shared goals and a common vision and since we are not in competition with each other, all parties have something to win.

The possibility to pool resources, skills and equipment helps to develop high quality solutions and at the same time it helps to minimize risks and guarantee sustainability.

### References

- Ämneskategorier för vetenskapliga publikationer [Subject categories for scientific publications], <http://www.ub.uu.se/epub/categories/>, 2005-04-28
- Brooks, L. (2004). Values of Community Source Development, *Syllabus*, September, <http://www.campus-technology.com/article.asp?id=9940>, 2005-04-28
- DiVA Document Format, <http://publications.uu.se/schema/ddf/>, 2005-04-28
- Educational Content Markup, <http://www.skeptron.ilu.uu.se/broadly/dl/mu.htm>, 2005-04-28
- Müller, E. (2004). DiVA – Academic Archive Online. *DigiCULT.Info*, November, 22-25, [http://www.digicult.info/downloads/digicult\\_info\\_9\\_xs.pdf](http://www.digicult.info/downloads/digicult_info_9_xs.pdf)
- Müller, E. Andersson, S. Klosa, U. Hansson, P. (2003). Metadata Workflow Based on Reuse of Original Data. In: *Proceedings of the ETD 2003: Next Steps - Electronic Theses and Dissertations Worldwide*. The Sixth International Symposium On Electronic Theses and Dissertations, the Humboldt-University in Berlin, Germany from 21 - 24 May 2003. <http://edoc.hu-berlin.de/etd2003/andersson-stefan/HTML/index.html>
- Müller, E.; Klosa, U.; Andersson, S.; Hansson, P. (2003). The DiVA project – development of an electronic publishing system. *D-Lib Magazine*, 9, <http://www.dlib.org/dlib/november03/muller/11muller.html>
- Müller, E. Klosa, U. Hansson, P. Andersson, S. (2003). Archiving Workflow between a Local Repository and the National Archive Experiences from the DiVA Project. In: *Proceedings of the Third ECDL Workshop on Web Archives*, in conjunction with the 7th European Conference on Research and Advanced Technologies for Digital Libraries in Trondheim, Norway. August 21st, 2003. <http://bibnum.bnf.fr/ecdl/2003/>
- Müller, E. Klosa, U. Hansson, P. Andersson, S. Siira, E. (2003). Using XML for Long-term Preservation. Experiences from the DiVA Project. In: *Proceedings of the ETD 2003: Next Steps - Electronic Theses and Dissertations Worldwide*. The Sixth International Symposium On Electronic Theses and Dissertations, the Humboldt-University in Berlin, Germany, 21 - 24 May 2003. <http://edoc.hu-berlin.de/etd2003/hansson-peter/HTML>
- The Open Archives Initiative Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>, 2005-04-28
- RFC 3188 (2001). Using National Bibliography Numbers as URNs, <http://www.ietf.org/rfc/rfc3188.txt>, 2005-04-28

