
ProPrint world-wide print-on-demand services for study and research

*Elmar Mittler and
Matthias Schulz*

The authors

Elmar Mittler is Director, Göttingen State and University Library, Göttingen, Germany.

Matthias Schulz is an Informatics Specialist, Computer and Media Services, Humboldt University, Berlin, Germany.

Keywords

Libraries, Digital libraries, Print media, Demand, Germany

Abstract

The libraries of more and more universities and research institutions have local digital repositories, and the amount of material is increasing every day. Users need an integrated retrieval interface that allows aggregated searching across multiple document servers without having to resort to manual processes. ProPrint offers an on-demand print service within Germany for over 2,000 monographs and 1,000 journals. Partners worldwide are now invited to join.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0737-8831.htm

Introduction

ProPrint currently offers an on-demand print service within Germany for over 2,000 monographs and 1,000 journals[1]. It does this by enabling virtual connections between document servers so that information within the metadata of all servers can be searched with a single search interface. This service provides users with unlimited access to distributed electronic documents via the central ProPrint search engine and sets up an efficient print-on-demand workflow for those libraries that publish electronic information.

The idea

The libraries of more and more universities and research institutions have local digital repositories, and the amount of material is increasing every day. Users need an integrated retrieval interface that allows aggregated searching across multiple document servers without having to resort to manual processes. To do this, networking the document servers using standards like the Open Archive Initiative (OAI) metadata harvesting protocol is a prerequisite.

These document servers are not expected to replace printed information. Experience shows that the increased availability of digital information enhances the desire to print. Most people resort to the convenience of desktop printing, but there is also an ongoing demand to have loose pages bound in a book-like format, especially if the material is for long-term use. Such a book may be also unique, for example a compilation of texts on a particular subject taken from multiple sources.

ProPrint is a software solution for this kind of print on demand using OAI-standards and (at least for the next few years) PDF as output format.

Part of the idea behind the use of PDF as a quasi-standardized print format is to enable its use worldwide. Data hosted by one or more remote servers can (in principle) be printed anywhere in the world that has the necessary technical facilities. Instead of shipping a print-on-demand work, users can fetch it locally, whether in Hong Kong, New York or New Delhi.

ProPrint is currently far from realizing this dream, but the Computing and Media Services of Humboldt University in Berlin and the Göttingen State and University Library have taken the first steps into this direction though a grant by the

Received 9 February 2004

Revised 17 March 2004

Accepted 19 March 2004



Library Hi Tech

Volume 22 · Number 2 · 2004 · pp. 227-230

© Emerald Group Publishing Limited · ISSN 0737-8831

DOI 10.1108/07378830410543548

Deutsches Forschungsnetz (DFN) and the Ministry of Education and Research (BMBF). Partners worldwide are now invited to join.

The ProPrint project

Within the last five years most German universities have set up local document servers in order to distribute and archive digitally born scientific literature. Most these universities have also changed or are changing their graduation rules in order to allow the publication of doctoral theses in digital format, rather than on paper. At the same time, university libraries are digitizing historical material in order to make it available via the Internet, and scholarly document servers are hosting more and more electronic journals as a long-term solution to the crisis in scholarly publication. The goal of all these activities is to make information available worldwide in a cost-effective way. The Computer and Media Services of Humboldt University is one of the leading providers of theses in digital format and the State and University Library of Göttingen is one of Germany's biggest Digitization Centers. Together they started the ProPrint project in November 2000 by merging the content of the servers of Humboldt and Göttingen. The first phase was completed in summer 2003.

The ProPrint developers use a "LAMP" system that includes a Linux operating system, an Apache web server, a MySQL database, and PHP for programming. These components were chosen because they were free and freely available for others to use.

Consistent standards for communication, for document formats, and for metadata have been and remain a core concern, and the central ProPrint search engine is an OAI-compliant service provider. The OAI protocol uses Hypertext Transfer Protocol (HTTP), Dublin Core Metadata and eXtensible Markup Language (XML). OAI enables an efficient exchange of metadata between data and service providers using an asynchronous retrieval in which service providers regularly send requests to data providers and then store the metadata in local repositories. Search requests from end users run against these local repositories.

An OAI compliant document server has to deliver Dublin Core as its metadata format. ProPrint has extended the Dublin Core metadata set with specific elements that are defined in a separate (ProPrint) name space. This includes structural elements of the DIEPER metadata set[2] (e.g. chapters, sub-chapters and page formats) and sales metadata. Since the digital

documents of the servers in Berlin and Göttingen had different page formats, the metadata allow for partial structures.

The results of these developments are:

- a service where users can order a single document or compose a new one from parts of others in the ProPrint database; and
- print-on-demand for these documents that can be made available locally.

Technical details

The ProPrint Web service is based on dynamic Web pages. During implementation, the specifications shown below were used.

Metadata

Development of a central uniform metadata[3] set for all document servers that allows extensibility to include other document types and objects (such as maps). Like Dublin Core (DC), a ProPrint name space and a ProPrint application profile have been developed. The metadata are regularly updated between the document servers.

Documents

ProPrint produces a printable document either from single documents or from a patron-selected compilation of multiple documents from more than one ProPrint server (see Figure 1).

Access rights control

Users, administrators and print service providers each have specific profiles and different access rights to parts of ProPrint. Designated administrators control these rights.

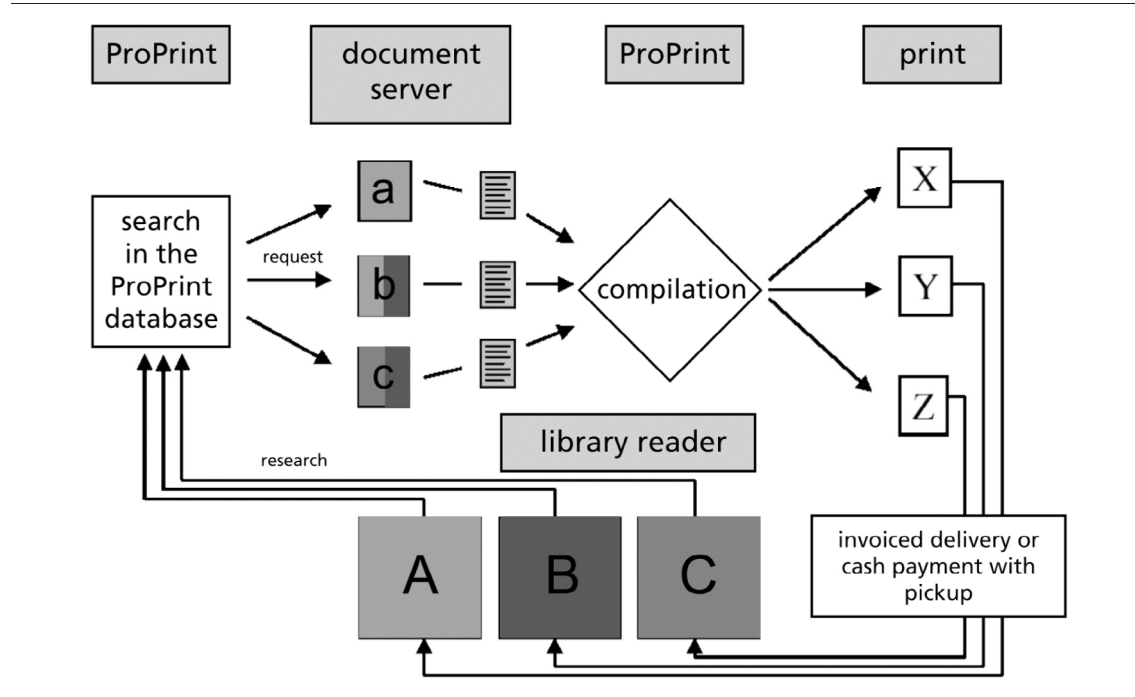
Print

Decentralized professional print shops distribute the documents in printed and bound form to the ProPrint user. ProPrint supports sorting of colour and black-and-white pages, thus allowing the automatic use of separate output devices. Users without access to a local ProPrint outlet copy shop can order their book directly from Göttingen.

Billing and sale

Göttingen prepares the bills for the printing and binding via a workflow that integrates the local print shops and Göttingen's accounting department. Documents are checked for number of pages and colouring. Pricing per page is determined for black and white and for colour pages. ProPrint generates a printable invoice that includes the price plus an invoice number.

Figure 1 ProPrint workflow



System administration

ProPrint can be administered either directly on the server or via a Web interface. The administrative interface includes the following functions:

- text/language (including the help text);
- user administration;
- OAI server;
- calendar/tasks;
- print services (including the integration of professional print shops); and
- invoice numbering.

The use of PDF

ProPrint requires standardized formats and a certain level of technical quality in order to enable the document exchange between servers. ProPrint uses Adobe's PDF format, but the technical quality of documents can vary due to scanning mistakes or upgrades to higher PDF versions. This became clear during a test involving document servers from the Technical University Library in Stuttgart and the University Computing Center at the Technical University of Chemnitz. When representative documents were tested by trying to print them at a professional print shop, recurring failures led to the following recommendations:

- *Embedding of fonts and pre-checking of documents.* If the PDF format is going to be used for archiving, all fonts should be embedded using a distiller or other software tools, even though this expands file size and

increases the load time for Web use.

Documents must be checked thoroughly prior to archiving on a server;

- *No security settings for PDF files.* To enable archiving in PDF format, security settings and user restriction have to be avoided. Converting the PDF file via Ghostscript into a postscript file and back to a PDF file erases the security settings and user restrictions; and
- *Checking for coloured text.* For ease of use and cost effectiveness, documents should be free of redundant colour elements. For example the automatic colouring of URLs in Microsoft Word can cause a document to be considered a "color document" by the printer, and increase the cost. Authors should be warned to avoid automated settings that cause colour-coding.

Electronic publishing in Germany

The German Initiative for Networked Information (DINI) is the association of computer centers, libraries and media centers in Germany, comparable to the Coalition for Networked Information (CNI) in the USA. Its working group on electronic publishing sponsored a study in 2003 that found a lack of standardization in local repositories in German universities and research institutions. The main findings were:

- in most cases, electronic documents are stored on the local server and are accessible via the OPAC;

- the preferred document format is PDF;
- 40 per cent of these local servers have written standards for electronic documents; and
- 23 per cent deliver detailed bibliographic, administrative and technical metadata.

The working group wanted to promote the idea of standardized local repositories. After publishing a booklet about university servers, the working group developed a certification process for OAI compliant document and publication servers, with the main focus on server technology and quality management for electronic documents[4]. This document was published in January 2004. DINI-certified local repositories will be ideal partners of ProPrint in the future. See Figure 2 for a diagram of the ProPrint service in Germany.

The future of ProPrint

Today, over 4,000 different documents are available through ProPrint from the document servers of Humboldt University in Berlin and the State and University Library in Göttingen. These

include monographs, journals, digitized documents from historical archives, dissertations, master's theses, and publications from three national and international conferences. Including these heterogeneous materials, each of which posed a different challenge, has proven Proprint's stability. ProPrint can handle digitally born as well as converted and retro-digitized works. ProPrint was designed intentionally as an open system so that every document server provider can use it.

ProPrint is not restricted to public institutions, and offers useful functionality for commercial content providers. However, it is designed as a co-operative activity of information institutions. ProPrint is striving to build a networked environment for quality publications in local repositories with a distribution system that facilitates customized printing on demand. So far ProPrint combines the document servers of Humboldt and Göttingen. The University Library of Graz will be the next library in the group. New partners from all over the world are welcome. For contact information or the names of project staff, see the Appendix.

Notes

- 1 See <http://edoc.hu-berlin.de/proprint/flyer-en.pdf>
- 2 See <http://gdz.sub.uni-goettingen.de/dieper/>
- 3 See www.edoc.hu-berlin.de/proprint/
- 4 See www.dini.de/documents/Zertifikat.pdf (English version in preparation).

Figure 2 Scheme of the ProPrint service in Germany



Appendix

Website: www.proprint-service.de

E-mail contact: info@proprint-service.de

Project staff

- *Project direction:* Professor Dr Elmar Mittler, Dr Peter Schirmbacher.
- *Management:* Hans J. Becker, Matthias Schulz.
- *Scientific assistants:* Dr Karen Strehlow, Andres Imhof, Susanne Dobratz, Antje Groth.
- *Library staff:* Dr Michael Voss
- *Student workers:* Hans-Werner Hilse, Cliff Richter.
- *Technical assistants:* Jürgen Braun, Markus Enders, Thomas Fischer, Frank Klaproth, Winfried Mühl.