

PsyDok: electronic full-text archive for psychological documents

Ulrich Herb

Electronic Publishing, Digital Repositories, Open Access

Saarland University and State Library (SULB), Germany

Abstract

The Saarland University and State Library (Saarlaendische Universitaets- und Landesbibliothek SULB) runs the Special Subject Collection Psychology which is part of an information system for the supra-regional literature supply in Germany. With the increasing opportunities for electronic information the following question emerged for the Special Subject Collection: how can a contribution to the availability and preservation of electronic documents particularly with regards to grey literature be made? Giving an answer SULB established the repository PsyDok in autumn 2002. PsyDok is one of the few subject-specific digital repositories in Germany. It is a registered data provider of the Open Archives Initiative (OAI) and certified by the German Initiative for Networkinformation (DINI), which developed a variety of standards digital repositories have to comply with to achieve this mark of quality. PsyDok offers many interfaces to disciplinary and multidisciplinary retrieval and information systems. The allocation of Uniform Resource Names (URN) guarantees persistent availability and citeability of the documents. Several services and mechanisms guarantee not only a maximum of visibility and a rapid and global scientific exchange and flow of information stored on PsyDok but also its long term

availability and its scientific utilization. PsyDok fits the principles of "Open Access": Many different document types may be published, but PsyDok focuses particularly on diploma theses, dissertations, professorial dissertations and grey literature. Regarding aspects of visibility, information exchange and scientific utilization especially producers of grey literature benefit from PsyDok.

1 Introduction: Scientific information supply in Germany: Special subject collections and virtual libraries

In Germany scientific information supply is coined by a distributed, supra-regional system of special subject collections¹ (in German: Sondersammelgebiet, SSG) which are tied to scientific libraries. There are 121 special subject collections which can be found in more than 40 libraries all over Germany. These special subject collections are funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).² The main function of a SSG is to build a collection of subject related scientific literature which is as complete as possible; this includes domestic and foreign research literature (including electronic media, microfilms and multimedia objects) and especially scientific journals. Within this system Saarland University and State Library (Saarlaendische Universitaets- und Landesbibliothek SULB)³ runs the Special Subject Collection Psychology since 1966. In 1998 the German Research Foundation adopted a memorandum focusing on the future trends of scientific information supply. This memorandum asked for the enhancement of information supply with regard to the means and opportunities of up to date information technology. Accordingly the German Research Foundation started funding so called Virtual Libraries. In 1999 the Virtual

Library of Psychology at Saarland University and State Library went online. Virtual Libraries claimed to be central entry-points to scientific information by bunching subject-specific information on the WWW and making it accessible. These virtual libraries did not try to preserve or publish information but only to link to it. Therefore most virtual libraries are using a subject gateway⁴ and a subject specific meta search engine. They usually did not make any contribution to the availability of scientific information on the www.

2 The availability of psychological information

In autumn 2002 SULB and the Institute for Psychology Information (Zentrum fuer Psychologische Information und Dokumentation ZPID, Germany)⁵ launched a joint project named „Digital Psychology Information (DPI)", which was funded by the German Research Society until spring 2005. The aim of DPI was to improve the availability and accessibility of electronic psychological information. One of the work packages defined within DPI was the implementation of the subject specific repository PsyDok⁶ run by SULB; its home page lives at <http://psydok.sulb.uni-saarland.de>. By establishing a service like PsyDok SULB did not only react to the ephemeral nature of most online information but made also a contribution to the availability and preservation of electronic documents particularly with regards to grey literature.

3 PsyDok: A subject specific scientific repository

3.1 Why PsyDok?

PsyDok is a nationwide and central publication platform for scientific psychological documents. Both publishing and using PsyDok-documents is free of charge, which also means the PsyDok fulfils the demands of Open Access. Unlike scientific documents published on conventional institutional repositories PsyDok-documents reach a much higher, maximized visibility and dissemination. This is based upon its centralized function within the discipline psychology and its manifold integration in both disciplinary and multidisciplinary retrieval systems. This systematic integration offers overwhelming advantages especially compared with the still usual but deficient practice of publishing documents on web servers of academic institutions or faculties. Such information can not be searched (or found) systematically, if they are found at all this happens accidentally by click-by-click-retrieval. In addition publications like these are of a very limited value within scientific discourse: both span of life and availability are short (or are at least unpredictable), permanent access, their immutability and citeability are doubtful. All in all PsyDok-publications do not fall victim to the informational fragmentation and do not vanish in the WWW's noise. By offering a rapid and global flow of scientific information PsyDok applies to the motto "maximizing research impact by maximizing research access".

3.2 What offers PsyDok?

In accordance to the special subject collection principle PsyDok acts as a supraregional, subject specific repository for digital scientific documents. It is the central point for free of charge psychological fulltext publishing. Many different document types may be published and found on PsyDok, but it focuses particularly on diploma and master theses, dissertations, professorial dissertations, preprints, teaching material, psychological scales and grey literature. Monographs, book chapters, journals and articles from journals can also be published as PsyDok is a postprint-server as well. PsyDok gives scientists an infrastructure for self-archiving and facilitates their walk on the green road to Open Access. Most of the documents published on PsyDok can be assigned to the genre grey literature. This category is of great value for the scientists as a study by the German Research Foundation proved⁷ (DFG 2005, p. 22). PsyDok is not a genuine publication platform, all documents published were produced in another context or for another purpose and afterwards deposited on the repository – none of them was originally produced to be published on PsyDok.

3.3 What about technical matters?

PsyDok is based upon the software OPUS,⁸ which was developed at the University of Stuttgart (Germany). Till now 50 German universities are using OPUS to run their digital repositories. It requires UNIX/Linux with a MySQL- or MSOL-Database and an APACHE/PHP-environment. OPUS is also mentioned in the Open Society Institute's guide to institutional repository software (Open Society Institute, 2004). The preferred

file formats to be published on PsyDok are PDF-, PostScript- or HTML-files, the ideal format to date is PDF/A.⁹

3.4 What about time, digital archaeology and 404 errors?

Saarland University and State Library ensures permanent and long-term access, long-term usability and citeability for all documents published. In order to supply a persistent and stable electronic location of its documents, PsyDok offers a persistent identifier in form of an Uniform Resource Name¹⁰ (URN) for each document. Each URN is registered by the German National Library¹¹ (Die Deutsche Bibliothek, DDB). URNs are comparable to the Digital Object Identifier (DOI)¹² or the less known Persistent URL (PURL).¹³ With the allocation of URNs the persistent availability and citeability of the documents is guaranteed. For each document with a persistent identifier (no matter if in the form of an URN, DOI or PURL) permanent access, permanent reference and citeability can be acknowledged.¹⁴ SULB is not going to develop techniques of long-term usability and access, nevertheless these advantages can be ensured. Technologies like these need a lot of research and development which can only be done at a subordinate level, but not by a single university library. In Germany several institutions (as the DDB) and projects (as nestor¹⁵ or kopal¹⁶) are focusing on the aspects of long-term access and availability. Technologies and practices developed and evaluated within these projects will be applied to PsyDok. On the long run long-term usability seems to need the design and implementation of emulation and migration technologies.

3.5 What about quality and standardisations?

As PsyDok is not a genuine publication platform Saarland University and State Library does not have any influence on the genesis of the documents which are published on PsyDok and hence has little chance to apply any kind of quality control.¹⁷ Nevertheless several related principles are taking effect before documents are deposited on PsyDok, so quality control can be derived from the type of document to be published.

- With institutional series or working papers series an editorial review should be assumed
- With dissertations and professorial dissertations a sufficient quality is at hand due to the type of document
- Diploma and master theses will only be published if the referee explicitly states a recommendation
- With preprints the authors revisable reputation can be derived from former publications or the quality of the document can be derived by the later formal publication
- Postprints are controlled by the publishing houses expert staff, mostly by a peer review
- Each psychological faculty or institute is encouraged to use PsyDok's internal peer review module

Within the Open Archives Community (OAI)¹⁸ there are vitally important discussions going on how an extended OAI Protocol for Metadata Harvesting (PMH) could offer a peer review functionality for records on OAI-Servers. This technique is essentially

based on digital social networking models and focuses on preprints, but may be adapted to other document types as well (Rodriguez et al. 2005). This means that in the future digital repositories as PsyDok might offer a genuine quality control. Regarding standardisations, technological and organizational quality control of digital publication services in Germany the German Initiative for Networkinformation (German: Deutsche Initiative für Netzwerkinformation e.V., DINI)¹⁹ has to be mentioned. DINI developed a variety of standards digital repositories have to comply with in order to become a certified DINI-Server. DINI intends to make electronic publishing with digital repositories more serious and reliable; the DINI-Certificate is a mark of quality. The criteria checked by DINI-experts concern organizational matters (as workflow optimization), questions of data integrity and authenticity, guaranteed archival storage periods, detailed description of the service, data export into long-term archival systems, questions of content indexing and question of metadata management and export. PsyDok met all these criteria and is a certified DINI-Server.²⁰ Another and even more important sign of standardization is the compliance with the OAI PMH.

3.6 What about dissemination and visibility?

Discussing the advantages of electronic publishing compared to conventional print publishing two of the main arguments in favour of the e-model are both the accelerated circulation and availability and the maximized visibility of digital publications. Notably relating to visibility PsyDok offers lots of benefits - especially comparing it with other institutional repositories. PsyDok is exceptional proper for the integration into disciplinary and multidisciplinary retrieval systems, since PsyDok is - unlike an

institutional repository - no general dealer but a scientific specialist dealer. Subject specific search engines may index PsyDok as a whole without picking single domain related documents while multidisciplinary catalogues may integrate PsyDok data easily into their browsing scheme because all documents only focus on one subject. PsyDok offers many interfaces to disciplinary and multidisciplinary retrieval and information systems in order to facilitate the integration. First of all PsyDok is a registered OAI-data provider so OAI-based search engines (OAI service providers) can add its Metadata to their index. Other multidisciplinary retrieval systems are the *Karlsruhe Virtual Catalogue* (KVK)²¹, a meta search interface which queries library and book trade catalogues, *OASE*²² (Open Access to Scientific Literature) which allows to search the bibliographical data of the most important document servers in Germany and abroad. PsyDok documents can also be found via the OPUS meta search²³, a search interfaces which offers a simultaneous search in all OPUS-based digital repositories. PsyDok publications can be found via Scirus²⁴ (Elsevier's scientific search engine) and the Google spin-off for scientific documents, Google Scholar²⁵. Usually most scientists pay much attention on disciplinary information systems, so PsyDok is well integrated into retrieval systems focusing strongly on psychology. First to be mentioned is the subject gateway PsychLinker²⁶. PsyDok offers also an interface to ZPID's Psyndex (the most comprehensive abstract database of psychological literature, audiovisual media, and tests from the German-speaking countries).²⁷ All diploma and master theses are additionally listed in the diploma theses database of ZPID. Each PsyDok document is indexed by the robot-based psychology search engine PsychSpider²⁸ of ZPID. Another central access point for psychological information is the OPAC²⁹ of SULB with its

special subject collection, which registers numerous queries for psychological information a day and which of course lists PsyDok documents as well. Even information systems which mainly focus on other subjects than psychology integrate PsyDok data. For instance selected documents can be found via the Collection of Computer Science Bibliographies³⁰ and CiteSeer³¹. Last but not least PsyDok-documents are also listed in WWW search engines where they produce very high rankings. Naturally Saarland University and State Library is eager to expand the number of retrieval systems hinting to PsyDok. PsyDok will be added to the index of the Social Science Search Engine run by the Social Science Information Gateway (SOSIG)³². The same applies to the search engine PsychCrawler³³ of the American Psychological Association (APA)³⁴ and to Scirus Repository Search. The integration into the new APA database for grey literature, PsycEXTRA³⁵ is considered, it will depend on legal matters. Dissemination and visibility of PsyDok-documents are a multiple of what can be reached through conventional library catalogues, www search engines, traditional literature databases or subject mixed repositories. This can be proved by a simple example: During October 2005 the diploma theses “Gesichtsform und Attraktivität” by Markus Becker was viewed online 212 times. The print version of the same diploma theses is held by only one library (and can not be borrowed there), so this document would be nearly invisible if it was not published on PsyDok. This difference is of course very distinct with diploma theses, since these are mostly only held at the library of local psychological institutes. But the usage of electronic vs. print versions is also amazingly different with documents that can be borrowed. For example the report “Professioneller Wandel in der Medizin als Herausforderung für die psychologische

Eignungsdiagnostik” by Hans Uwe Hohner is available at three university libraries in Germany. One of them is the Saarland University and State Library with its Special Subject Collection Psychology, which registers most of the queries, interlibrary loans and requests for document delivery for psychological literature in Germany. The report mentioned is held by SULB since 1998 and was not borrowed one single time since February 2004³⁶; but during October 2005 it was viewed online 166 times. At the end of 2004 the burden on the server became so heavy that PsyDok had to be transferred to another server – another evidence for visibility and dissemination.

4 What’s to come?

Beside the integration into more retrieval systems further enhancements are under development. First to mention are the implementations of a print-on-demand service³⁷ and a Creative Commons³⁸ licence management. Furthermore the APA’s PsycINFO Classification Categories and Codes will be used as a browsing structure. The new multilingual layout developed by SULB will bring another improvement.

Internationalisation will also be enhanced by the implementation of a cross-lingual search option that is based on the system CANAL.

To offer additional information about the quality of a document (which is often derived from its usage) PsyDok already offers anonymized statistics giving evidence of document related downloads without showing any visitor’s IP-address or displaying the reader’s identity. Admittedly these statistics are still quite noisy, but they give prove to the fluctuation and dissemination of digital publications and confirm the benefits of electronic publishing. Discussing the pros and cons of e-publishing there are mainly two

arguments used to describe e-publishing as mediocre compared to conventional print-publishing with traditional publishing houses:

- a lack of citation counts and citation linking
- a lack of reliable, public statistics providing information of a document's usage

Obviously both arguments aim at issues of quality by inherently denying its existence on digital repositories or at least by denying any procedure of quality measurement.

There are already existing models and technical answers to the first argument, especially the already mentioned CiteSeer, Citebase³⁹, OpenCit⁴⁰ and Google Scholar with its citation linking. These technical solutions are based on a distributed network approach: they can not be developed by single operators or administrators of digital repositories. Anonymized download statistics alleviate the second problem. Statistics like these shed light on the significant higher reading rate and the significant higher and faster dissemination of electronic documents compared to print documents - especially if these documents are systematically exposed to scientific and to lay-oriented retrieval systems. Besides: in fact it is proved that higher download rates correlate with higher citations.⁴¹ Of course many technical problems have to be considered and to be solved, until digital repositories will reach a level of standardisation that allows and offers really reliable access statistics per document comparable or equivalent to the Counter Standard.⁴² One of the most promising initiatives in this field is Interoperable Repository Statistics (IRS).⁴³ IRS plans to develop tools, which remedy the deficiencies of incompatible data that is hard to interpret, compare and aggregate (IRS Homepage 2005). Another startling approach is the analysis of usage statistics in order to compare

journal usage and citation impact factor (Bollen 2005).⁴⁴ After all current endeavours tend to mould quality measures by examining correlations between citation counting of electronic documents and usage statistics and thus combining both strategies.

Literature

- Bollen, J., Van de Sompel, H., Smith, J., Luce, J. (2005), Toward alternative metrics of journal impact: a comparison of download and citation data. *Information Processing and Management*, 41(6):1419-1440
- Deutsche Forschungsgemeinschaft (2005), Publikationsstrategien im Wandel? Ergebnisse einer Umfrage zum Publikations- und Rezeptionsverhalten unter besonderer Berücksichtigung von Open Access, Bonn, Germany.
- Open Society Institute (2003), A guide to institutional repository software, 3rd edition. http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf
- Rodriguez M. A., Bollen J., Van de Sompel H. (2005), The Convergence of Digital-Libraries and the Peer-Review Process. A preprint forthcoming from the *Journal of Information Science*. <http://arxiv.org/abs/cs.DL/0504084>
- Interoperable Repository Statistics Homepage, <http://irs.eprints.org/>

Notes

¹ For more information about these special subject collections, see <http://webis.sub.uni-hamburg.de/>

² For more information about DFG, see <http://www.dfg.de>

³ For more information about Saarland University and State Library, see <http://www.sulb.uni-saarland.de>

⁴ In a nutshell subject gateways are human controlled and quality controlled (not robot based) search engines or web catalogues containing links and metadata to scientific information resources in the www. Due to the ephemerality of most scientific websites the maintenance of subject gateways turned out to be quite time-consuming.

⁵ ZPID creates electronic databases on literature, tests, audiovisual media, and Web resources. For more information, see <http://www.zpid.de>

⁶ For more information about PsyDok, see <http://psydok.sulb.uni-saarland.de>

⁷ The study examined both publication and reception preferences of German scientists from all disciplines and it pointed out that especially members of the social science and humanities community rely on grey literature.

⁸ For more information about OPUS, see <http://opus.uni-stuttgart.de>

⁹ For more information on PDF/A, see http://www.aiim.org/pdf_a/

¹⁰ For more information about the URN, see <http://www.persistent-identifier.de/?lang=en>

¹¹ For more information about DDB, see <http://www.ddb.de>

¹² For more information about DOI, see <http://www.doi.org/>

¹³ For more information about PURL, see <http://www.purl.org>

¹⁴ An URN resolving mechanism passes the URN on the actual URL of a specific registered document. An abstract ID – for instance an URN as `urn:nbn:de:bsz:291-psydok-5541` – is assigned to a concrete object (the conference paper “Human ‘sense of space’, moving images and architecture” by Rainer Schönhammer).

¹⁵ For more information about nestor, see <http://www.langzeitarchivierung.de/index.php?newlang=eng>

¹⁶ For more information about kopal, see <http://kopal.langzeitarchivierung.de/>

¹⁷ Therefore SULB reserves the right to decide which documents will be published and which will not be.

¹⁸ For more information about the Open Archives Initiative, see <http://www.openarchives.org>

¹⁹ For more information about DINI, see <http://www.dini.de>

²⁰ Currently there are thirteen servers certified by DINI, a list of them can be found at <http://www.dini.de/dini/zertifikat/zertifiziert.php>. All in all there are about one hundred institutional repositories in Germany: <http://www.dini.de/dini/wisspub/dokuserver.php>

²¹ For more information about KVK, see <http://www.ubka.uni-karlsruhe.de/hylib/en/kvk.html>

²² For more information about OASE, see http://www.ubka.uni-karlsruhe.de/kvk/kvvk/kvbk_en.html

²³ This search interface can be found at http://opus.uni-stuttgart.de/opus/gemeinsame_suche.php

-
- ²⁴ For more information about SCIRUS, see <http://www.scirus.com/srsapp/>
- ²⁵ For more information about Google Scholar, see <http://scholar.google.com>
- ²⁶ PsychLinker is a scientific catalogue to psychological online information which is jointly run by ZPID and SULB, broad categories within PsychLinker are advised by members of the psychological community. For more information, see <http://www.psychlinker.de>
- ²⁷ Psyndex is run by ZPID, so ZPID members select documents to be added Psyndex.
- ²⁸ For more information about PsychSpider, see <http://www.psychspider.de>
- ²⁹ This online catalogue can be found at <http://opac.sulb.uni-saarland.de>
- ³⁰ For more information about the Collection of Computer Science Bibliographies, see <http://liinwww.ira.uka.de/bibliography>
- ³¹ For more information about CiteSeer, see <http://citeseer.ist.psu.edu>
- ³² For more information about SOSIG, see <http://www.sosig.ac.uk>
- ³³ For more information about PsychCrawler, see <http://www.psychcrawler.com>
- ³⁴ For more information about APA, see <http://www.apa.org>
- ³⁵ For more information about PsycEXTRA, see <http://www.apa.org/psyceextra/>
- ³⁶ Statistics of former years are not available
- ³⁷ For more information about the service, see <http://www.proprint-service.de/>
- ³⁸ For more information about Creative Commons, see <http://www.creativecommons.org>
- ³⁹ For more information about Citebase, see <http://citebase.eprints.org/cgi-bin/search>
- ⁴⁰ For more information about OpenCit, see <http://opcit.eprints.org>
- ⁴¹ You may read more about it at the OpenCit web site: <http://opcit.eprints.org/oacitation-biblio.html>
- ⁴² For more information about Project COUNTER, see <http://www.projectcounter.org>
- ⁴³ IRS is funded by JISC since June 2005. For more information, see <http://irs.eprints.org/>
- ⁴⁴ This model uses data mining techniques in various ways: besides offering quality measures, tools like a usage-based recommender system could be offered. More details can be found within the slides of Johan Bollen's lecture "A framework for assessing impact of units of scholarly communication based on OAI-PMH harvesting of usage information" at the OAI4 workshop, <http://oai4.web.cern.ch/OAI4/>