

Serving Innovation in Scholarly Communication with the Open Platform "Digital Peer Publishing"

Wolfram Horstmann, Peter Reimer, Jochen Schirrwagen

'hbz' – Publishing Systems Group
Juelicher Strasse 6, 50674 Cologne, Germany
email: horstmann@hbz-nrw.de

Abstract

The internet causes a continuous emergence of novel forms of scholarly communication and collaboration. Electronic publishing provides a means for representing eventual outcomes of these processes, i.e. all types of content such as papers and advanced forms of media. Electronic journals are often chosen as an adequate publishing format because they simultaneously deliver content in a well-known manner but, at the same time, allow extending traditional publishing with innovative features. The initiative Digital Peer Publishing (DiPP) provides technological, organizational and legal frameworks and tools that help to incubate and proliferate such innovative publishing projects. The hosting platform reflects principles of a Service Oriented Architecture. It combines, via Web Services, already established components such as an OAI repository (Fedora) and a Web Content Management System (Plone) with customized workflows for document processing, conversion and distribution. As an open platform it is capable of integrating external tools and services or acts itself as a service provider. It is therefore disposed for supplementing research infrastructures with electronic publishing.

Keywords: scholarly communication; information management; workflow system; text technology; repository; web service; OAI; DiPP

1. Introduction

The use of web based tools for review, production and distribution of scholarly information move genuine publishing processes away from publishers towards the scholar. Automatic Email-alerts of reviewing deadlines, advanced document templates and conversion mechanisms as well as end user compliant Web Content Management Systems and technical distribution standards such as OAI [1] disperse the traditional editorial office to scholars' desktops. Even though individual human expertise will never be dispensable in a qualified publishing process, the benefits of seamless communications and workflows between scholars might supersede the benefits of the external editorial back-office. In any case, reengineering and structural change of publishing through electronic tools are irreversible processes.

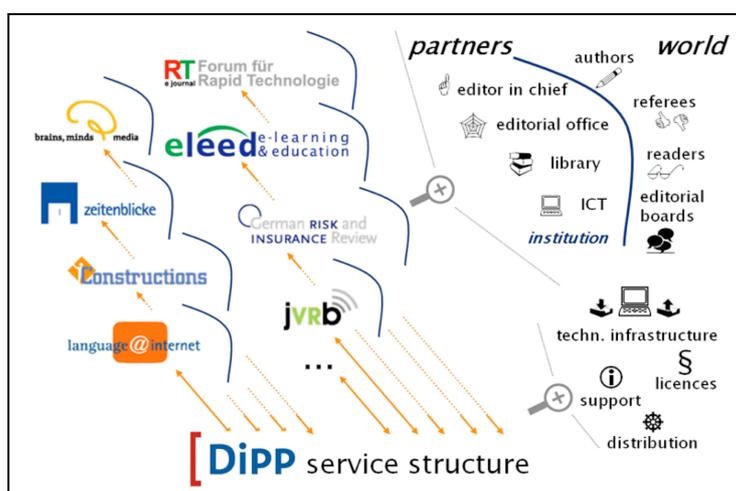


Figure 1: Organizational approach in the initiative Digital Peer Publishing

Some scholars criticize this trend for the extra work imposed, some scholars appreciate the chance of better controlling how scholarly information is processed and eventually published. Some scholars even introduce new forms of publishing themselves, independently. The latter cases build the main user group of open publishing solutions because independent initiatives are scarcely in the position to pay software licenses or services of for-profit vendors and providers. Open publishing solutions therefore offer open-source software such as Open Journal Systems (OJS) [2] or non-for-profit hosting concepts such as DiPP [3].

2. Methodology: The DiPP Approach

The German state North-Rhine-Westphalia (NRW) has launched the initiative “Digital Peer Publishing” to provide – accessible through the browser – infrastructure and service networks for scholars and their peers when building and operating independent publishing projects. The organizational approach (see Fig. 1) assumes that the editorial team is a virtual organization that is rooted at the institution of the editor-in-chief, often with the support of the local academic library. The platform is developed, maintained and supported by the public academic library centre *hbz* in Cologne. Managers of publishing projects (such as eJournals) can apply the DiPP platform in two different ways: In the primary application scenario, DiPP hosts the project. Alternatively, managers may set up a DiPP platform, themselves (see 2.4). Unlike most installed software packages (e.g. OJS), DiPP also supports text-technology for uploading prepared documents (RTF for standard text processors such as Open Office / MS-Word™ or TeX as well as XML) to be automatically converted to XML-Docbook in order to generate automatically presentation formats such as HTML and PDF (see also 2.3).

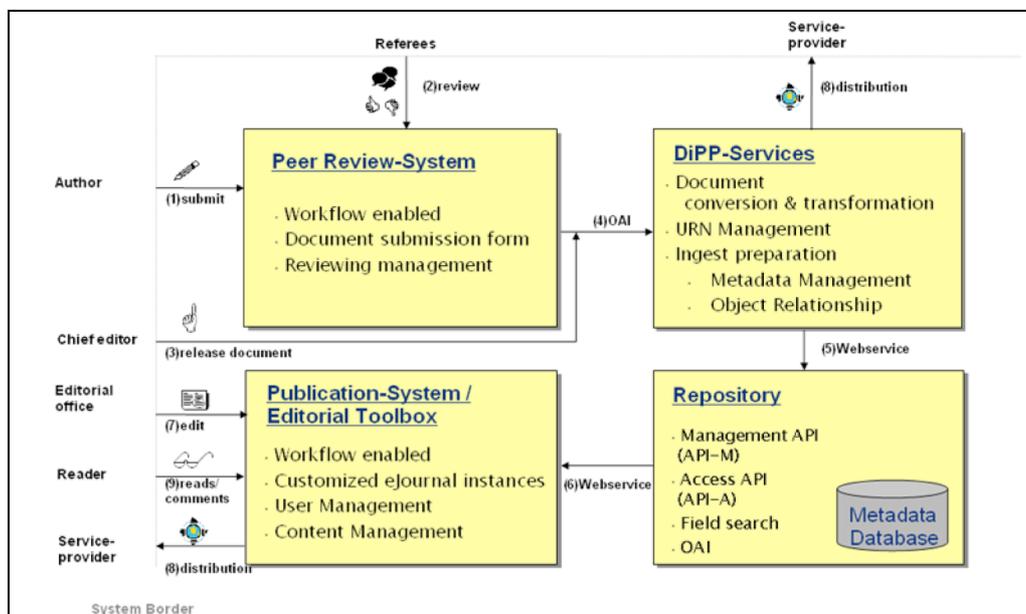


Figure 2: Overview of workflows supported in the DiPP platform

2.1 Services

The DiPP-platform offers a wide range of services (see Fig. 2). Ultimate goal is to provide managers with a means to operate profound online publishing projects simply by using a browser, email and the text-processor their community is used to – without the necessity to develop, install and maintain software and operate servers. The services include hosting and archiving, a comprehensive production system, distribution services, licensing and knowledge exchange (see also Table 1).

Category	Service	Specs	URL
Hosting and archiving	attribution of a "persistent identifier" to primary publications - URN (W3C)	persistent identifier, URN	http://www.persistent-identifier.de/ http://www.w3.org/Addressing/
	powerful and secure hardware		
	solid repository	Fedora	http://www.fedora.info/
	open archival format	XML, Docbook	http://www.w3.org/XML/ http://www.docbook.org/
	preparation for long-term access and archiving	Nestor / Kopal	http://www.langzeitarchivierung.de/ http://kopal.langzeitarchivierung.de/
Production System	easy editing of WWW-Pages (WYSIWYG)	Plone	http://plone.org/
	Individual homepage layout	cf. DiPP-eJournals	http://www.dipp.nrw.de/journals
	automatic article conversion with style sheets	RTF/LaTeX ->XML <- HTML/PDF	http://www.latex-project.org/ http://www.w3.org/MarkUp/ http://partners.adobe.com/public/
	advanced news and messaging functions	email Lists / RSS-Feeds	http://www.w3.org/RDF/
Licence-Management	Digital Peer Publishing Licence – ,construction kit‘ for variable affordances, committed particularly to scholarly communication (author retains rights)	DPPL	http://www.dipp.nrw.de/lizenzen
	automated licence management - integration in metadata und articles		
Distribution	Indexing services for search engines and databases, bibliografic exact indexing and retrieval, optimized full text search through HTML-presentation of articles	ISSN, DOAJ, OAISTER, SCHOLAR-GOOGLE	http://www.issn.org/ http://www.doaj.org/ http://oaister.umdl.umich.edu/ http://scholar.google.com/
	open data provider, automatic notification for data and articles	OAI-PMH / eMail-Alert	http://www.openarchives.org/
	customizable forms for metadata, also advanced, qualified and disciplinary	DC, DDC, PACS etc.	http://dublincore.org/ http://www.oclc.org/dewey/about/ http://publish.aps.org/PACS/
	tools for detailed usage statistics and article based usage analysis	AWstats and , custom analytics	http://www.awstats.org/
Workflow and Review System	traditional peer-reviewing: choice of workflows (blind, double-blind etc.)	GAPworks	http://www.gapworks.de/
	post-processing workflow (imprimatur scheduling etc), if reviewing system is used, accepted articles are automatically transferred to the production system (via OAI)	CMFopenflow	http://plone.org/
Knowledge Exchange	steady exchange between DiPP team and managers/editors	Workshops, Mailings, WIKI	

Table 1: Overview of the services offered [can be adapted and compiled to the needs of the specific project]

2.2. Technical Approach

Rather than “reinventing the wheel”, the technical approach in DiPP is to network and customize distinguished open source solutions and apply internationally accepted standards (see also Table 1). The approach reflects principles of a Service Oriented Architecture [4] in that it applies the logic of self-contained functionalities via Web-Services that ‘orchestrate’ a technically heterogeneous environment with platform independent standards (here: WSDL, OAI etc., see Fig. 3) rather than providing a monolithic, proprietary system. Beside the massive re-use of existing resources, the advantages lie in the open character of the architecture: components can be exchanged or omitted more easily and new components can be integrated seamlessly, without changing the complete architecture. The main components are the Fedora repository [5], the Plone Web Content Management System, the reviewing system GAPworks and the DiPP Services that complement the components with specific functions for scholarly communication.

Fedora is used as the central instance that manages and delivers digital content. The main reason for choosing Fedora was the flexible digital object model and the focus on tidy technical implementation of interfaces (e.g. Web-Services). The alternative D-Space [6] provided more pre-defined functions for installation and GUIs. But because the architecture in DiPP foresaw solely the role of a pure data management solution (without end user interface), Fedora fitted best. End user interfaces (GUIs) are provided via Plone. The reason for preferring Plone against alternatives such as TYPO3.com was the possibility to incorporate ZOPE as an application server and extend functionalities with Python-based web applications running as a separate process. Finally, GAPworks was chosen because it was the only system providing genuine peer-review functions as a self-contained tool.

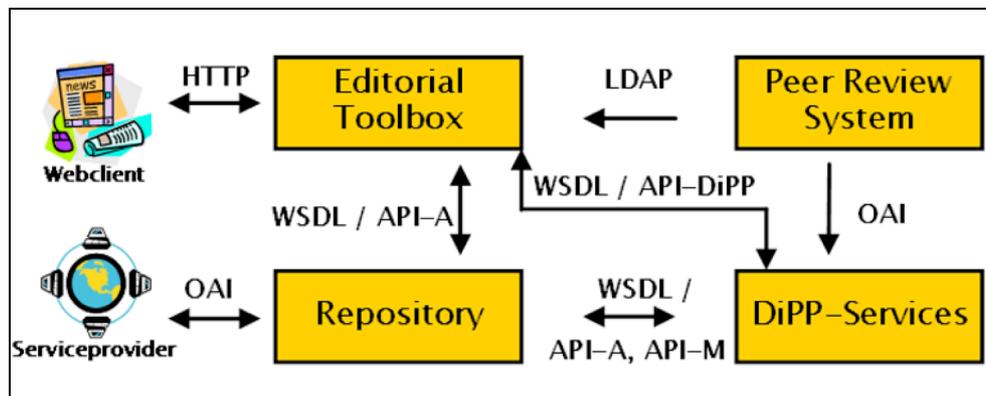


Figure 3: Schematic of the architecture highlighting the interfaces

2.3. Document Processing and Conversion

Automated document processing and conversion (see Fig. 4) is a distinctive feature of the DiPP platform that is not offered by most other platforms such as OJS. The motivation for this approach is that authors usually produce their documents in standard text processors such as MS Word and Open Office or in typesetting languages such as TeX. In order to reduce workload in editorial offices, the preparatory work of authors should be optimally usable by editorial officers to perform post-processing on the basis of the submitted document in a standard text processor. Therefore, DiPP recommends customized stylesheets for authors. The specific choice might vary within the scholarly community: technical disciplines may prefer TeX based systems while humanities may prefer MS Word. In order to provide an open standard for the latter case, DiPP offers RTF; the former requires TeX-Styles.

As a result, editorial offices already receive submissions in the intended format. The following reviewing process varies from case to case: some offices prefer reviewing with email attachments, some use the web based process in GAPworks. After acceptance, submissions usually have to be formally revised, since authors will not comply 100% with the styles. However, once trained, editorial officers can perform formatting fast and comfortably in a text processor. The final version is uploaded via web-forms, where also metadata and hierarchy definitions (issue, volume etc.) are included, and automatically converted to Docbook-XML using a server side Upcast™ installation. Docbook-XML is the central archival and conversion format for a publication that receives an URN and is transferred to the Fedora repository. XHTML-pages are then automatically generated via CSS/XSLT. If desired, a PDF can also be generated using dbletx (Some officers produce their PDF directly on a RTF or TeX basis). The final version can then be scheduled and released for publication.

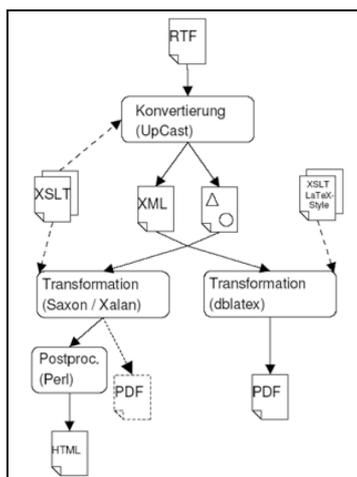


Figure 4: Schematic of the document conversion process [For the sake of brevity, only the pathway based on RTF is shown; The TeX based pathway is not shown here]

2.4. How to Use the DiPP-Platform

Managers of publishing projects (such as eJournals) can apply the DiPP platform in two different ways: DiPP hosts the project or managers set-up a platform, themselves. If DiPP hosts the project, managers receive an individual instance with individual URL. No own server or installation is needed to operate the project. Powerful hardware and secure data handling and storage are provided by the hbz. A test-account can be requested and will be available instantaneously. The managers can then decide how to proceed. The degree to which managers self-configure and self-manage their instance depends on the technical skills managers can bring up:

- In a plain hosting solution, managers self-configure most features.
- At the basic service-level, managers provide exact specifications for workflow frameworks that will be implemented by the DiPP-Team. All content related, 'editorial' work is organized or performed by the managers.
- At extended service-levels, managers simply provide content which is integrated by the DiPP-Team.

Experiences with this approach show that highest quality and fastest progress is achieved when managers co-operate intensively with the DiPP-Team in the initial phase of setting up the instance (2-10 net working days). Then, in the phase of regular operation, a plain hosting solution is sufficient given that basic technical skills are within the manager's team. Hosting and working days have to be cost-covering for partners outside the state but can be waived in co-operative development projects. The whole organization is non-for-profit.

As an alternative to the hosting approach, managers can set up the platform themselves. Since almost any component in the system is open-source, no fees have to be paid for DiPP. Since every partner preferred the hosting approach so far, no experiences can be reported. However, a technical documentation helps to set-up an own implementation.

3 Interim Results

The first phase of DiPP (until the end 2005) focused on the building of the infrastructure and resulted in ten journal start-ups and pick-ups. Several other new publishing projects or migrations are in preparation. Already the first completed year brought more than 110.000 article views (overall more than 200.000 visits, 120.000 visitors, 700.000 pages, 6 Mio. Hits) – showing an increasing tendency. Numbers of article views correspond well to the numbers for per-year usage of hundreds of thousands of subscribed electronic articles that are provided at a complete university [7]. This suggests to attribute this strong usage to the 'Open Access' approach, the principle of providing scholarly articles freely accessible in the internet to anyone from anywhere (see 4.1).

3.1. Statistical Analysis

Data sample included usage events from ten domains: www.dipp.nrw.de/afrika, www.brains-minds-media.org, www.constructions-online.de, eleed.campussource.de, www.rtejournal.de, www.risk-insurance.de,

www.jvrb.org, www.languageatinternet.de, www.socwork.net, www.zeitenblicke.de. Data were collected with AWSTATS: Automatic hits and requests (Robots, Crawlers, Spiders etc.) were excluded. Overall number of published articles was 183 since December 2004. 102 were published over the DiPP platform with URN in the OAI repository. These articles were analyzed for article based statistics (others were published before completed migration). The analysis was performed on by-months base. The overall number of analyzed months is reduced from theoretically 120 to 108 due to running migrations (in case of article based statistics to 99). Article views are initial usage events of HTML source files only. Usage events within articles (PDFs, links, supplementary materials etc.) were not counted. PDFs could not be analyzed representatively, since not all analyzed eJournals offer PDFs. Means and variance are shown as insets (see Fig. 5, top).

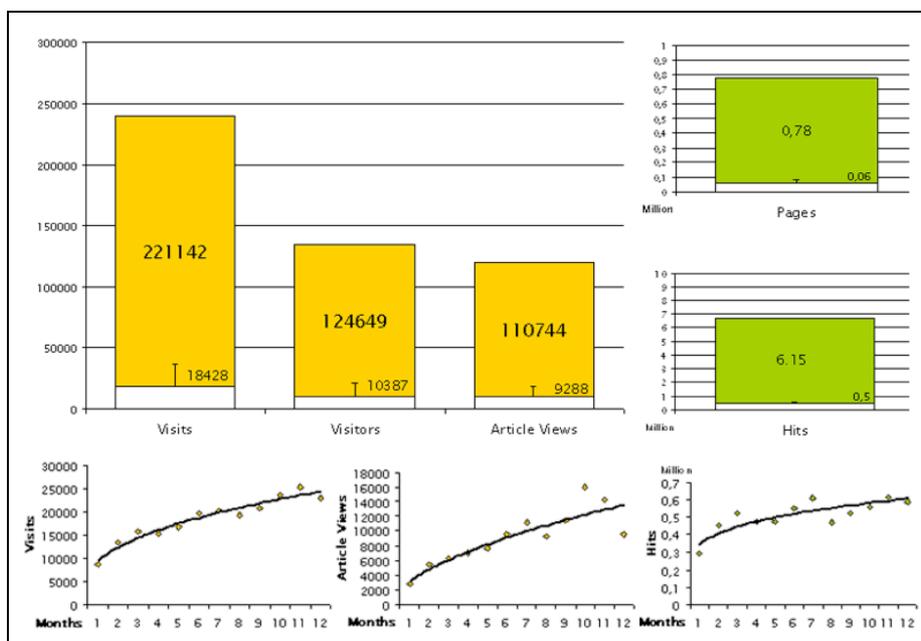


Figure 1: Usage statistics in 2005 [Top: Sums and averages. Bottom: Time courses]

Values of www.socwork.net were collected with WEBALIZER (partly computed on by-day basis), for article based statistics computed from visitors (one article view per visitor, conservative assumption with respect to the statistical average ('mode') computed for other journals). Fluctuations in the time courses (see Fig. 5, bottom) are due to coincidences of publication dates of eJournal issues. Solid lines in the time courses were computed with a model, a regression analysis that showed best fits for power functions: Visits: $y = 9414 x^{0.38}$, $R^2 = 0,95$. Article views: $y = 3221 x^{0.58}$, $R^2 = 0,88$. Hits: $y = 347336 x^{0.2}$, $R^2 = 0,71$.

4 Discussion: Tackling Innovation

Two basic innovation vertices were successfully achieved after 18 months: (1) Accomplishing incubation and proliferation of publishing projects in new disciplinary profiles and (2) providing cost and labour efficient e-only publication infrastructure (for operating publishing projects in an open access mode).

However, concerning innovation, this is only the beginning: The traditional publishing system fails to incorporate many significant aspects of contemporary scholarly work. A simple and well known example is the informal usage of early versions (e.g. preprints in physics on arXiv.org or working papers in economics on repec.org) that is to a large part isolated from the actual publishing process. Today's progress in electronic publishing makes possible a seamless presentation of the whole publication workflow. Some of the new eJournals in DiPP will explore this chance by applying an open review. The DiPP platform will be extended appropriately. Other approaches even incorporate the post processing history by continuous revisions of publications (e.g. livingreviews.org).

Another, even more severe weakness of the traditional publishing system concerns the failure to adequately represent genuine and essential parts of the contemporary research process, namely the *replication of results*. In many disciplines traditional text-image publications provide no means anymore of falsifying or verifying what the researchers have actually done. In many disciplines, research happens beyond language and static images in

applying massively dynamic and interactive media. Good scientific practice mandates that published results have to be replicable. Thus, electronic publishing faces the challenge to incorporate these in developing adequate representation formats and processes for quality assurance. Primary data (see e.g. National Institute of Health, 2003) are only the start for a more complete representation of the research underlying a publication – tools for processing primary results, visualizations and simulations are still to be integrated. Many of the eJournals in DiPP therefore explore the development of new formats: brains-minds-media.org, for example, even prefers submissions providing supplementary material over straightforward articles. The DiPP platform already offers features supporting such novel formats. But the management of compound-media and the corresponding reviewing, licensing as well as exploitation will demand a deeper integration of upcoming standards such as MPEG-21 [9].

4.1 Open Access

In the realm of scholarly electronic publishing, Open Access [10,11] is rather an intrinsic property of adequate representations for scholarly communication processes than a question of faith and politics. But Open Access (OA) introduces innovations in electronic publishing because it highlights the overcoming of barriers such as subscriptions, rights management and business models (see also 4.2), thereby opening a corridor for novel approaches to enter traditional publishing. However, contemporary discussions on OA stick very much to the current status of traditional publishing because they either propose an alternative distribution channel for conventional publications by depositing author's copies in institutional repositories ('green road', e.g. eprints.org) or flip the business model from subscription fees to author fees ('golden road', e.g. plos.org). None of these approaches take into account the chance to provide scholarly communication with novel forms of electronic publishing. The polarized discussion between those two poles even prevents constructive discussion on how to further exploit the internet as a fast, efficient, versatile and powerful communication channel for scholars [12]. For these reasons, the DiPP initiative stresses the positive effects of OA: (a) the improved chances for impact of publications as compared to restricted access approaches and (b) the chance to provide seamless communication workflows from production and quality assurance over distribution and usage to re-use of published material that overcomes the (in the long run: artificial) divide between repositories and eJournals. For (a), DiPP will extend its approaches on article based analysis, once valid, web-based metrics [13] for a more detailed and immediate impact assessment on the article level developed [14]. Metrics such as the Journal Impact Factor [15] are excellent guidelines, but much too coarse with respect to temporal resolution and the level of content aggregation, since articles are averaged over journals over several years. For (b), DiPP will pursue both extending repositories with review functions and strengthen the archival character of eJournals. Open reviews, for example, inevitably lead to a closer integration of repositories with genuine journal structures, since pre-publication archives must be included. DiPP will also continuously extend publishing workflows in direction of what has been termed 'eScience' [16], i.e. the intense collaboration and communication within and between scholarly communities on the basis of advanced internet infrastructure or "Grid Computing" [17].

4.2 Business Models

Independent publishing projects provide the main driving force for structurally integrating innovations in scientific communication. Open publishing solutions and infrastructures such as DiPP can provide a biotope for incubating and proliferating innovations. It is an inherent property of innovative projects not to be a major player on the conventional market. Therefore, the DiPP initiative does not act as a publisher and not as a competitor to publishers. Instead DiPP acts as a fosterer of thematic and methodological innovations in publishing, which is expressly interested to support the transfer of projects and methods to the conventional system. However, developing business models [18] is not within the primary scope of DiPP – uptake is left to the conventional system.

5 Conclusion

Independent scholarly publishing projects are a major driving force for structurally integrating innovations in scholarly communication. Open publishing solutions and infrastructures can provide a biotope for incubating and proliferating such innovation bearing projects. The initiative Digital Peer Publishing (DiPP) offers an open platform with technical, organizational and legal tools and frameworks that serves as a host for eJournals and further publishing projects. DiPP expressly encourages exploring innovative approaches such as entering new disciplinary scopes, publishing new media and providing novel workflows. In DiPP, Open Access is understood as a means to fully exploit the potential of the internet as the, at present, most efficient information channel for scholarly communication. Further developments in scholarly communication such as those envisioned with the buzzword 'eScience' are about to be integrated. But single initiatives will not be enough to fulfil the needs of

scholarly communication. DiPP therefore stresses the character of an open platform – not only as an open archive with open standards and open sources but also with an open architecture ready to include new external tools and frameworks and to provide services for other platforms.

References

- [1] Van de SOMPEL, H.; LAGOZE, C. The Santa Fe Convention of the Open Archives Initiative. *D-Lib Magazine*, 2002, vol. 6, no. 2. URL <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>, 1.3.2006.
- [2] WILLINSKY, J. (in press) Open Journal Systems: An example of Open Source Software for journal management and publishing. *Library Hi-Tech*, 2006. URL http://research2.csci.educ.ubc.ca/eprints/archive/00000047/01/Library_Hi_Tech_DRAFT.pdf, 1.3.2006.
- [3] HORSTMANN, W.; REIMER, P.; SCHIRRWAGEN, J. Multi-level eJournal support structures in the initiative "Digital Peer Publishing NRW". *Joint Workshop on Electronic Publishing Organised by Delos*. Lund, Sweden. 2005. URL http://www.dipp.nrw.de/publikationen/horstmann_lund_full.pdf/download, 1.3.2006.
- [4] ERL, T. Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services. Upper Saddle River, NJ, USA : Prentice Hall PTR, 2004.
- [5] STAPLES, T.; WAYLAND, R.; PAYETTE, S. The Fedora Project: An Open-source Digital Object Repository Management System. *D-Lib Magazine*, vol. 9, no. 4. URL <http://webdoc.gwdg.de/edoc/aw/dlib/dlib/april03/staples/04staples.html>, 1.3.2006.
- [6] SMITH, M. DSpace: An Institutional Repository from the MIT Libraries and Hewlett Packard Laboratories. *Lecture Notes on Computer Science*, Springer-Verlag, 2002, vol. 2458, p. 543-549. URL https://dspace.mit.edu/bitstream/1721.1/26706/1/Smith_2002_DSpace.pdf, 1.3.2006.
- [7] MONTGOMERY, C. H.; KING, D. W. Comparing Library and User Related Costs of Print and Electronic Journal Collections. *D-LIB Magazine*, 2002, vol. 10, no. 8. URL <http://www.dlib.org/dlib/october02/montgomery/10montgomery.html>, 1.3.2006.
- [8] National Institute of Health. NIH Data Sharing Policy. NIH GUIDE 2003. URL http://grants.nih.gov/grants/policy/data_sharing/, 1.3.2006.
- [9] MPEG-21 Overview v.5. ISO/IEC JTC1/SC29/WG11/N5231. Edit. by J. Bormans, K.Hill. Shanghai. 2002. URL <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>, 1.3.2006.
- [10] Open Society Institute. Budapest Open Access Initiative. 2002. URL <http://www.soros.org/openaccess/read.shtml>, 1.3.2006.
- [11] Max-Planck-Society. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. 2003. URL <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>, 1.3.2006.
- [12] BRODY, T.; HARNAD, S. Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals. *D-Lib Magazine*, 2004, vol. 10, no. 6. URL <http://www.dlib.org/dlib/june04/harnad/06harnad.html>
- [13] BJÖRNEBORN, L.; INGWERSEN, P. Toward a Basic Framework for Webometrics. *Journal of the American Society for Information Science and Technology*, 2004, vol.55, no. 14, p. 1216–1227. URL <http://www.db.dk/lb/>, 1.3.2006.
- [14] GARFIELD, E. The Impact Factor. *Current Contents*, 1994, vol. 25, p. 3-7. URL <http://scientific.thomson.com/knowtrend/essays/journalcitationreports/impactfactor/>, 1.3.2006.
- [15] HARNAD, S. Fast-Forward on the Green Road to Open Access: The Case Against Mixing Up Green and Gold. *Ariadne*, 2005, issue 42. URL <http://www.ariadne.ac.uk/issue42/harnad/>, 1.3.2006.
- [16] CROW, R.; GOLDSTEIN, H. Guide to Business Planning for Launching a New Open Access Journal. 2nd ed. Open society Institute. 2003. URL <http://www.soros.org/openaccess/oajguides/>, 1.3.2006.
- [17] de ROURE, D.; JENNINGS, N. R.; SHADBOLT, N. R. *The Semantic Grid: A Future e-Science Infrastructure*. In Grid Computing. Edit. by F. Berman, G. Fox, T. Hey : John Wiley & Sons, Ltd, 2003, p. 437-470. ISBN 0470853190
- [18] Grid Computing. Edit by F. Berman, G. Fox, T. Hey : John Wiley & Sons, Ltd, 2003.