ISSN 0167-5265
ISUSDX 28(2) 77–168 (2008)
Volume 28, Number 2, 2008

INFORMATION SERVICES & USE

SPECIAL ISSUE:
APE 2008 Academic Publishing in Europe, Quality and Publishing

IOS PRESS

INFORMATION SERVICES & USE
Volume 28, Number 2, 2008

SPECIAL ISSUE: APE 2008 Academic Publishing in Europe, Quality and Publishing

CONTENTS

Information Services & Use 28 (2008) 77–80
DOI 10.3233/ISU-2008-0587
IOS Press

Foreword

APE 2008 Proceedings

The 3rd APE Conference "Academic Publishing in Europe", held 22–23 January 2008 in the Berlin–Brandenburg Academy of Sciences, concentrated on 'Quality & Publishing'. APE 2008 was again organised unter the Auspices of the EU Research Directorate-General, and supported and sponsored by a large number of associations, societies and publishing companies. The large number of invited speakers and sponsored guests contributed significantly to the high level of presentations, panels and discussions. The sponsors are listed in the last pages of these Proceedings.

APE 2008 was overbooked with a long waiting list which shows the growing interest in this annual conference. For the first time we had satellite meetings for university presses and information portals. The Pre-Conference Day with the Education & Training Course for Younger Academic Publishers on 21 January 2008 in the NH Hotel had more participants than in 2007 and was again excellently prepared and moderated by Anthony Watkinson.

On behalf of the Organisers and the Programme Committee we would like to thank the Chairmen, the Speakers and last but not least the participants for an excellent conference. APE conferences aim at a better understanding of scholarly communication through value-added publishing and the increasingly important role of relevant information in science and society. They provide an inventory and show the way ahead in a collaborative effort involving all stakeholders in the information society.

Academic publishing is facing technological, financial and political challenges. These might have major consequences for scholarly communication. The public/private interface might change in the coming decade. There is lot of uncertainty and little coherent knowledge about all aspects of the value chain. There is as yet no complete understanding about the economics of publishing. We have to be aware that cultures and habits that deal with scientific information differ substantially between the scientific disciplines and that they are different on all stages of work with information (in creating, access, working with content and publication).

APE conferences provide an international and independent platform, where academic, professional and educational publishers, university presses, librarians, scientists, authors and editors, teachers, learned and professional societies and associations, funding agencies, politicians and policy makers, subscription agencies and booksellers, recruiting agencies as well as technology providers can present visions and views, exchange experiences, and discuss opportunities and challenges. One participant wrote to us, that this was a real "Open Mind" conference, referring to some fierce discussions about the economics and sustainability of Open Access.

All APE 2008 presentations can be found at the APE 2009 website under 'APE Literature'. Here you can also find the 'APE 2008 Short Report'. We are grateful to the rapporteurs Dr. Svenja Hagenhoff and Dr. Chris Armbruster, who managed to capture the synthesis of many different presentations.

Speakers at APE 2008 had no obligation to write papers. For this reason it has taken several months to collect a fine selection of articles, which are presented in this Special Issue of the journal *Information Services & Use*, published by IOS Press in Amsterdam. With the help of many notes and audio recordings, we were able to 'reproduce' the full text of some presentations.

The 4th APE Conference has been announced for 20–21 January 2009 in the Berlin–Brandenburg Academy of Sciences. The focus will be on "Impact & Publishing", which is all about awareness, usage, branding, impact, added value, the role of different search engines and a better understanding of academic publishing.

The website is open and first registrations are coming in. For actual information please visit: www.ape2009.eu.

Looking forward to seeing you in Berlin in January 2009.

Arnoud de Kemp
Editor

Final Programme APE 2008 Full Conference APE 2008 – Academic Publishing in Europe

"Quality & Publishing"

Tuesday, 22 January, 2008

Welcome and Opening: Einar Fredriksson, IOS Press, Amsterdam

Greetings: Karl-Peter Winters, German Association of Publishers and Booksellers, Frankfurt

Opening Keynote: **"Innovation in Scholarly Communication. Visions and Projects from High-Energy Physics"**, Rolf-Dieter Heuer, DESY, Hamburg, newly elected Director-General, CERN, Geneva

Keynote 2: **"Open Access – A never ending Story"**, Arne K. Richter, European Geosciences Union

Keynote 3: **"STM Publishing: the known Knowns, the known Unknowns and all Points in between"**, Michael A. Mabe, International Association of STM Publishers

Session: Peer Review and Quality
Chairman: Mayur Amin, Elsevier, Oxford

Experiments with alternative Peer Review Approaches – What have we learnt? With contributions from Ulrich Pöschl, Max Planck Institute for Chemistry, Mainz – Catriona MacCallum, PLoS – Linda J. Miller, Nature, New York

"Peer Review in Scholarly Journals: Perspective of the Scholarly Community – Results from an international Study (PRC)", Mark Ware, Mark Ware Consulting

Session: All about Money
Chair: Gertraud Griepke, Springer, Heidelberg

"What means rich in publishing? Competition between for-profit and not-for-profit publishers from an economist's point of view", Alexis Walckiers, ECARIS, Brussels

"What is the Value Chain of Scientific Publishing?", Thomas Connertz and Guido F. Hermann, Thieme, Stuttgart

"Scientific Publishing from a Funding Body Perspective: Views from the European Commision", Deirdre Furlong, European Commission, Brussels

Wednesday, 23 January, 2008

Panel Discussion: What matters? The Future Role of Libraries in Science and Society: Swallowed by OA Repositories, turned into University Presses or kept as Book Museums?
Chairman: Michael Seadle, Humboldt University, Berlin
Participants: Ann Okerson, Yale University Library – Nol Verhagen, Amsterdam University Library – Wolfram Neubauer, ETH Libraries, Zürich

Panel: Visible or Invisible? What Tools do Academic needs?
Chairman: Ehrhardt F. Heinold, Heinold + Spiller, Hamburg

"A Comparison of Information Portals for Academics", Tamara Pianos, German National Library of Economics, Kiel

Panel: Thijs Willems, Scopus – Vanessa Proudman, NEEO/Economists Online – Wilma Mossink, Knowledge Exchange – Tamara Pianos, Vascoda. With Comments from Holger Kersten, University of Magdeburg

Session: Innovation & Enabling Technologies
Chairman: Eefke Smit, STM, Standards & Technology, Amsterdam

"Comparing the Scientific Impact of Conference and Journal Publications in Computer Science", Erhard Rahm, University of Leipzig

"Semantic Markup in Scientific Publishing", Peter Murray-Rust, University of Cambridge

"The Impact of ACAP on STM Publishers", Jens Bammel, International Publishers Association, Geneva

"Electronic Expression of Licensing Terms", Fiona Bennett, Oxford University Press

Round Table: University Presses and Books in the HSS (Humanities and Social Sciences)
Chairman: Eelco Ferwerda, Amsterdam University Press
Participants: Birgit Schmidt, University of Göttingen – Marianne Alenius, Museum Tusculanum Press – Patrizia Cotoneschi, Firenze University Press – Gérard Wormser, Presse Universitaire de Lyon

Session: New Content, New Models, New Roles
Chairman: Dirk Lens, Swets Executive Board, Lisse

"New Methods to Access Scientific Content", Stefan Geißler, TEMIS, Heidelberg

"New Scientific Communication Models in the MPG: The Case of Data Journals", Laurent Romary, Max Planck Digital Library, Berlin

"Libreka: A Collaborative Publishers Approach to present Full Text Books on the Internet", Ronald Schild, MVB, German Association of Publishers and Booksellers, Frankfurt

"DAISY – An Opportunity to improve Access to Information for All", Thomas Kahlisch, German Central Library for the Blind, Leipzig

Closing Keynote: "**Scholarship in the Digital Age. Are we prepared at all? Considerations and Recommendations"**, Rudi Schmiede, Darmstadt University of Technology

Closing Panel: **"Information in Science and Society"**

Chairman: Arnoud de Kemp, AKEP, German Association of Publishers and Booksellers.
Participants: Barbara Casalini, Casalini Libri, Fiesole – Gary Coker, MetaPress, Birmingham (USA) – Annette Holtkamp, DESY, Hamburg – Elisabeth A.L. Mol, Springer Science + Business Media, Dordrecht – Rudi Schmiede, Darmstadt University of Technology – Herman P. Spruyt, International Publishers Association, Geneva

Pre-Conference Day: **"The Purpose of Publishing" Education and Training Course for Younger Academic Publishers (YAP)**
Chair: Anthony Watkinson, University College London

Monday, 21 January 2008

Setting the Scene: Paul Evans, Elsevier, Amsterdam

Tutorial 1: **Giving Authors what they want (Authors Experience)**, Paul Peters, Hindawi, Cairo

Tutorial 2: **Providing Readers with the Google Experience (User Experience)**, Matthias Wahls, Brill, Leyden and Lorna Berrett, Wiley–Blackwell, Oxford

Tutorial 3: **Outsourcing Production and Other Functions (Saving Money)**, Gary Coker, MetaPress, Birmingham (USA) and Gertraud Griepke, Springer, Heidelberg

Workshop I: How to get organized?

Workshop II: How to get the right people with the right skills?

Summary and Presentation of 'Certificate of Attendance'

Information Services & Use 28 (2008) 81–82
DOI 10.3233/ISU-2008-0572
IOS Press

APE 2008 Academic Publishing in Europe, Quality and Publishing

January 22–23, 2008, Berlin

(Reported by Dr. Svenja Hagenhoff and Dr. Cris Armbruster)

Dr. Karl-Peter Winters
German Association of Publishers and Booksellers (Börsenverein), Frankfurt am Main, Germany

Ladies and gentlemen, dear colleagues. I have the pleasure again today of welcoming you here. The Academic Publishing in Europe Conferences of the last two years have shown just how fascinating and necessary interchange can be between European professionals from libraries, publishing companies, academic institutions and international organisations – irrespective, of course, of the different points of view taken by individual conference delegates when it comes to the issue of digital publishing. As you surely know, since January 1st this year, a new law has been in place in Germany for the regulation of copyright in the information society. As a publisher of academic specialist literature and chairman of the publishers' committee within the Börsenverein des Deutschen Buchhandels, the Association of Booksellers and Publishers in Germany, I have, of course, been closely following the details of discussions surrounding the implementation of the law. For us as publishers the law has brought certain changes because of which, in part together with the libraries, we have been involved in considerable political struggle. One positive aspect or practical relevance for publishing companies is for example the new ruling on the possibility for acquisition of rights for unknown types of use. For the future this gives us some safeguards for our planning. We publishers continue, of course, to monitor the ways in which user behaviours changed by the internet and the measures we need to take so that we are ready for the digital future. It is no longer a matter of being involved in electronic publishing but of how products are to be offered and of their quality. And that is why the motto of this conference "Quality and Publishing" is so appropriate in the current debate. In his talk, Ronald Schild will introduce the Libreka platform which enables German publishing companies to make their text available for searches worldwide on the internet. That is a contribution of our association to quality. Last year we dealt above all with the general conditions for electronic publishing in all the organisational and legal detail. We have also negotiated on this with the libraries – with good results. I am particularly pleased that a solution has been found in the dispute with the organization Subito – Dokumente aus Bibliotheken in Germany – over document distribution and delivery. At the end of last year the Börsenverein, the International Scientific Publishers Association STM and Subito reached an agreement on an outline arrangement for the licensing of electronic document distribution within Germany, Austria and Switzerland. On the basis of this outline agreement separate right agreements are to be concluded in the coming weeks and months between Subito and the individual scientific publishing companies. This is, above all, important for the distribution of documents in Europe and in the international arena. Issues raised in cooperation discussions with the German National Library – Deutsche Nationalbibliothek – included the practical implementation of

the legal deposit of internet publications and the definition and handling out of print titles. Here too, we are in the process of arriving at a joint agreement in line with the proposals already made by the International Publishers' Association and the Federation of European Publishers. Furthermore, together with the German Collecting Society VG Wort, we have agreed on arrangements that make it easier for publishing companies to provide their books in a special audio format for the visually impaired and the blind. The talk later on by Thomas Kahlisch will give an insight into the issue in the European context. As you can see, the European directives and requirements preoccupy us. And we try to implement them in Germany in a sensible way. Now, I am eager to hear the reports of practical experience in Europe and look forward to a lively dialogue with European colleagues. I find it especially laudable that a workshop was held again yesterday for the upcoming generation in publishing because I consider it particularly important that it is not just the industry's professionals who think about the future but that young employees are also given an idea of the radical changes linked to electronic publishing. Finally, I would like to thank the organisers of this conference – and above all, Arnoud de Kemp – who have made it possible for us to have two very meaningful days here and it looks to me sometimes as we have an out-of-the-camp-festival. I hope that you gain many new insights and wish us all a very enjoyable conference. Thank you.

Information Services & Use 28 (2008) 83–96
DOI 10.3233/ISU-2008-0570
IOS Press

Innovation in scholarly communication: Vision and projects from High-Energy Physics [1]

Rolf-Dieter Heuer [a], Annette Holtkamp [a] and Salvatore Mele [b]
[a] *Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, D-22607 Hamburg, Germany*
[b] *CERN, European Organization for Nuclear Research, CH1211, Genève 23, Switzerland*

Abstract. Having always been at the forefront of information management and open access, High-Energy Physics (HEP) proves to be an ideal test-bed for innovations in scholarly communication including new information and communication technologies. Three selected topics of scholarly communication in High-Energy Physics are presented here: a new open access business model, SCOAP^3, a world-wide sponsoring consortium for peer-reviewed HEP literature; the design, development and deployment of an e-infrastructure for information management; and the emerging debate on long-term preservation, re-use and (open) access to HEP data.

Keywords: Open access, data preservation, subject repository, high-energy physics

1. Preamble

Research in High-Energy Physics (HEP), also called Particle Physics, is motivated by the goal of attaining a fundamental description of the laws of physics, such as explaining the origin of mass, and understanding the dark matter in the universe. Although fundamentally driven by the quest for knowledge, the ensuing research is performed at the edge of what is feasible technologically and hence drives the development of technology in many areas. The knowledge gained from studying the microcosm of particle collisions at the highest energies ever attained provides also insight into the early universe and its development since its creation. To further this understanding, experimental particle physicists build the largest instruments ever to reach energy densities close to the Big Bang, teaming up in collaborations of up to several thousands of scientists. At the same time, theoretical particle physicists, who represent the other half of the community, build hypotheses and theories to accommodate and predict experimental findings.

HEP experimental research takes place in international accelerator research centres in Europe, such as the European Organization for Nuclear Research (CERN) in Geneva or the Deutsches Elektronen-Synchrotron (DESY) in Hamburg; in the United States mainly at the Stanford Linear Accelerator Center (SLAC) in California and the Fermi National Accelerator Laboratory (Fermilab) in Illinois; and in Japan at the High Energy Accelerator Research Organization (KEK) in Tsukuba. Canada, China, and Italy host other laboratories. HEP theoretical research takes place in hundreds of universities and institutes

[1] Based on a keynote talk by Rolf-Dieter Heuer at APE2008 [1].

worldwide, which also host experimental teams building parts of the large detectors used at the large accelerator laboratories and analyzing the data collected with them.

With the start-up of CERN's Large Hadron Collider (LHC) in 2008 and preparations for the International Linear Collider (ILC) in full swing, we expect revolutionary results explaining the origin of matter, unravelling the nature of dark matter and providing glimpses of extra spatial dimensions or grand unification of forces. Any of these insights would dramatically change our view of the world.

The LHC will collide protons 40 million times a second and reproduce the conditions at the origin of the universe. These collisions will be observed by large detectors, up to the size of a five storey building, crammed with electronic sensors: think a 100 MegaPixel digital camera taking 40 million pictures a second.

This contribution will not describe the exciting intellectual, scientific and technological endeavour of particle physics. Rather it will attempt to describe some solutions that, on the long wave of its track record in international collaboration, HEP has proposed and implemented for its infrastructure in scholarly communication. Although this is a very discipline-specific approach to the present evolution of scholarly communication, it might be of interest for an audience learned in academic publishing.

2. Introduction

Progress in information and communication technologies is driving evolving needs and profound changes in scholarly communication. Scientists, and not only HEP scholars, have come to expect:

- easy and unrestricted access to comprehensive scientific information in their field and cognate disciplines;
- state-of-the art information tools to optimize their research workflow, with powerful discovery tools and limited noise;
- quality assurance, at the intersection of three centuries of tradition in peer-review, but with a twist of 21st century communication for immediacy of feedback and dissemination.

At the same time, these desires have to be balanced against budget efficiency and optimization of resources for research. HEP has been proposing solutions to these needs since decades, as described in Section 3, while HEP *ante-litteram* open access tradition, which dates back half a century, is discussed in Section 4.

With the intention of informing, and possibly inspiring, the ongoing debates in the wider arena of innovation in scholarly communication, and its intersection with academic publishing in Europe and beyond, this contribution discusses the vision of HEP along three axes of innovation: a new open access business model (Section 5); the design, development and deployment of an e-infrastructure for information management, a next-generation repository (Section 6); the emerging debate on long-term preservation, re-use and (open) access to HEP data (Section 7).

3. Scholarly communication in HEP

To set the scene, it is useful to quote five numbers and a concept. The five numbers, which parameterize scholarly communication in HEP, are:

- 20,000, a lower limit to the number of active HEP scholars;

- 10,000, an upper limit to the yearly number of HEP articles;
- 80%, the fraction of these articles produced by theoretical physicists working alone or in small teams in hundreds of universities worldwide;
- 20%, the fraction of these articles authored by large collaborations of experimental physicists, centered around half a dozen international laboratories [2];
- 50:50, the ratio of active experimental and theoretical HEP scholars.

The concept is the one of *preprint culture,* which is discussed in the following.

The *preprint culture* in HEP pioneered the free distribution of scientific results. For decades, theoretical physicists and scientific collaborations, eager to disseminate their results in a way faster than the distribution of scholarly publications, printed and mailed hundreds, even thousands, of copies of their manuscripts at the same time as submitting to peer-reviewed journals [3]. While assuring the broadest possible dissemination of scientific results, albeit privileging scientists working in affluent institutions, this *ante-litteram* form of "author-pays" or rather "institute-pays" open access implied non-negligible financial liabilities for research centres: as an example, in 1990s DESY used to spend about 1 million DM (€500,000, not corrected for inflation) a year for the production and mailing of hard-copies of these preprints. CERN used to spend about twice as much.

From the very beginning of the preprint culture, HEP libraries classified preprints received from all over the world and regularly distributed worldwide information about the latest accessions. This was a pioneering instance of a literature database later evolving in email alerts, and the only way for HEP scholars to keep track of the fast-paced advancement of the field. As an example, the DESY library started in 1963 to publish the biweekly "HEP Index". It contained bibliographic information about the newest HEP preprints, journal articles and conference proceedings, and in addition it grouped these items according to standardized keywords.

In this scene, three revolutions mark the advances in scholarly communication in HEP, with repercussions in the contemporary innovations affecting other disciplines.

3.1. 1974, information technology meets (HEP) libraries

The SPIRES database, the first grey-literature electronic catalogue, saw the light at SLAC [4,5]. Shortly thereafter the SLAC and DESY libraries joined forces to cover the complete HEP literature including preprints, reports, journal articles, theses, conference talks and books. In 1985, the database contained already more than 140,000 records. It now contains metadata for about 760,000 HEP articles, including links to full-text, standardized keywords, publication notes. It offers additional tools like citation analysis and is interlinked with other databases containing information on conferences, experiments, authors and institutions.

3.2. 1991, the first repository

arXiv, the archetypal repository, was conceived in 1991 by Ginsparg [6], then at the Los Alamos National Laboratory in New Mexico, and is now hosted at Cornell University in New York. It evolved the four-decade old preprint culture into an electronic system, offering all scholars a level playing-field from which to access and disseminate information. Today arXiv has grown outside the field of HEP, becoming the reference repository for many diverse disciplines beyond physics, from mathematics to some areas of biology. It contains about 450,000 full-text preprints, receiving about 5,000 new articles each month.

3.3. 1991, the web is woven

The early history of the web and its invention at CERN in 1991 by Sir Tim Berners-Lee are today household stories [7]. What is less known is that the first web server outside Europe was installed at SLAC to provide access to the SPIRES database, which had then the honour to be the first database on the web [8]. In Summer 1992, SPIRES linked to the arXiv for full-texts, starting a close partnership, and bringing preprints on the web, accessible through a detailed indexing including reference to the ensuing published versions.

Even now, in the era of electronic journals, the preprint culture lives on in HEP thanks to the speed and ease of access. Journals have to a large extent lost their century-old role as vehicles of scholarly communication. However, at the same time, they continue to play a crucial part in the HEP community. Evaluation of research institutes and researchers, especially young ones at the beginning of their career, is largely based on publications in prestigious peer-reviewed journals. The main role of journals in HEP is perceived as the one of "keeper-of-the-records", by guaranteeing a high-quality peer-review process. The HEP community needs high-quality journals as its "interface with officialdom".

4. HEP and open access

Thanks to decades of preprint culture, today, in open-access speak, HEP could be defined as an almost entirely "green" discipline, where authors self-archive their research results. Posting an article on arXiv even before submitting it to a journal is common practice. Even revised versions incorporating the changes due to the peer-review process are routinely uploaded. It is interesting to remark that this comes without mandates and without debates: very few HEP scientists would not take advantage of the formidable opportunities offered by the discipline repository of the field, and the linked discovery and citation-analysis tools.

The synergy between HEP and open access[2] extends beyond preprints. In 1997, HEP launched one of the first peer-reviewed open access journals: the *Journal of High Energy Physics* (JHEP), courtesy of the International School of Advanced Studies (SISSA) in Trieste. It was followed in 1998 by *Physical Review Special Topics Accelerators and Beams*, published by the American Physical Society, and the *New Journal of Physics*, published by the Institute of Physics Publishing, which carries HEP content in a broader spectrum of content covering many branches of physics.

After preprints, arXiv and the web, open access journals appear to be the next logical step in the natural evolution of HEP scholarly communication. It is remarkable that in this field, this call for open access journals is not only originating from librarians frustrated by spiralling subscription costs and shrinking budget, but is a solid pillar of the scientific community. At the beginning of 2007, the four LHC Collaborations ATLAS, CMS, ALICE and LHCb, counting over 5,000 scientists, declared:

> We, [...] strongly encourage the usage of electronic publishing methods for [our] publications and support the principles of open access Publishing, which includes granting free access of our publications to all. Furthermore, we encourage all [our] members to publish papers in easily accessible journals, following the principles of the open access paradigm [9].

[2] There are many definitions and flavours of open access. For the sake of clarity, in this contribution we will assume something along the lines of "grant anybody anywhere and anytime access to the (peer-reviewed) results of (publicly-funded) research"... in HEP, of course.

The Helmholtz Alliance 'Physics at the Terascale', a German network comprising theorists, experimentalists, computer and accelerator scientists from 17 universities, two Helmholtz institutes and one Max Planck institute, issued a similar statement in January 2008:

> The Strategic Helmholtz Alliance 'Physics at the Terascale' fully supports the goal [...] of free and unrestricted electronic access to peer-reviewed journal literature in particle physics. [It] will benefit scientists, authors, funding agencies and publishers alike. Unrestricted access to published scientific results is essential for wide dissemination and efficient usage of scientific knowledge. [It invites its members to] raise awareness on open-access publishing in their communities and [...] to publish in open-access journals [10].

In order to meet these goals the SCOAP^3 initiative, subject of the next section, was started.

5. The SCOAP^3 initiative

SCOAP^3, the Sponsoring Consortium for Open Access Publishing in Particle Physics, is an initiative that aims to convert to open access the HEP peer-reviewed literature in a way transparent to authors [11]. Its business model originates from a debate involving the scientific community, libraries and publishers [12]. The essence of this model is the formation of a consortium to sponsor HEP publications and make them open access by redirecting funds that are currently used for subscriptions to HEP journals. Today, libraries (or the funding bodies behind them) buy journal subscriptions to support the peer-review service and to allow their users to read articles, even though these mostly access their information by reading preprints on arXiv. The SCOAP^3 vision for tomorrow is that funding bodies and libraries worldwide federate in a consortium that will pay centrally for the peer-review service and that journal articles will be free to read for everyone. This evolution of the current "author-pays" models for open access attempts to make the transition to open access transparent for authors, by removing any barriers.

SCOAP^3 will sponsor articles through a tendering procedure with publishers of high-quality journals. It has therefore the potential to contain the overall cost of journal publishing by linking price with quality and injecting competition into the market.

In practice, the open access transition will be facilitated by the fact that the large majority of HEP articles are published in just six peer-reviewed journals from four publishers [2], as presented in Fig. 1.

Five of those six journals carry a majority of HEP content. These are: *Physical Review* D(published by the American Physical Society), *Physics Letters* B and *Nuclear Physics* B (Elsevier), *Journal of High Energy Physics* (SISSA/IOP) and the *European Physical Journal C* (Springer). The aim of the SCOAP^3 model is to assist publishers to convert these "core" HEP journals entirely to open access and it is expected that the vast majority of the SCOAP^3 budget will be spent to achieve this target. The sixth journal, *Physical Review Letters* (American Physical Society), is a "broadband" journal that carries only a small fraction (10%) of HEP content; it is the aim of SCOAP^3 to sponsor the conversion to open access of this journal fraction. The same approach can be extended to another "broadband" journal popular with HEP instrumentation articles: *Nuclear Instruments and Methods in Physics Research A* (Elsevier) with about 25% HEP content. Of course, the SCOAP^3 model is open to any other, present or future, high-quality journals carrying HEP content. This will ensure a dynamic market with healthy competition and a broader choice.

The price of an electronic journal is mainly driven by the costs of running the peer-review system and editorial processing. Most publishers quote a price in the range of €1,000–2,000 per published article.

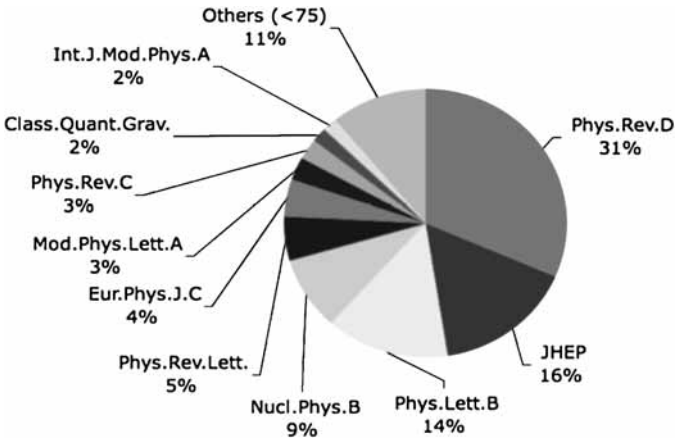

Fig. 1. Journals favoured by HEP scientists in 2006. Data from the SPIRES database. Journals that attracted less than 75 HEP articles are grouped in the slice named "Others".

On this basis, given that the total number of HEP publications in high-quality journals is between 5,000 and 10,000, according to how one defines HEP and its overlap with cognate disciplines, the annual SCOAP^3 budget for the transition of HEP publishing to open access would amount to a maximum of 10 Million Euros per year.

The costs of SCOAP^3 will be distributed among all countries according to a fair-share model based on the distribution of HEP articles per country, as shown in Fig. 2. In practice, this is an evolution of the "author-pays" concept: countries will be asked to contribute to SCOAP^3, whose ultimate targets are open access and peer-review, according to their use of the latter, a.k.a. their scientific productivity. To cover publications from scientists from countries that cannot be reasonably expected to make a contribution to the consortium at this time, an allowance of not more than 10% of the SCOAP^3 budget is foreseen.

The SCOAP^3 initiative fits in the European-wide debate on the access to results of scientific research. The Council of the European Union, in the conclusions of its 2832nd Competitiveness Council, recognized "the strategic importance for Europe's scientific development of current initiatives to develop sustainable models for open access [...]", underlining "the importance of effective collaboration between different actors, including funding agencies, researchers, research institutions and scientific publishers, in relation to access [...to] scientific publications". These principles are precisely the pillars of the SCOAP^3 model. Finally, it "invite[d] Member States to enhance the co-ordination between [...] large research institutions and funding bodies on access [...] policies and practices" [14].

It appears at first glance to be a formidable enterprise to organize a worldwide consortium of research institutes, libraries and funding bodies that cooperates with publishers in converting the most important HEP journals to OA. At the same time, HEP is used to international collaborations on a much bigger scale. As an example, the ATLAS experiment, one of the four detectors at the LHC, has been built over more than a decade by about 50 funding agencies on a total budget of €400 M (excluding person-power), placing about 1000 industrial contracts. In comparison, the SCOAP^3 initiative has about the same number of partners, but a yearly budget of only €10 M, and will handle less than a dozen contracts with publishers. Therefore, the aim is to operate SCOAP^3 along the blueprint of large HEP collaborations, to profit from their experience.

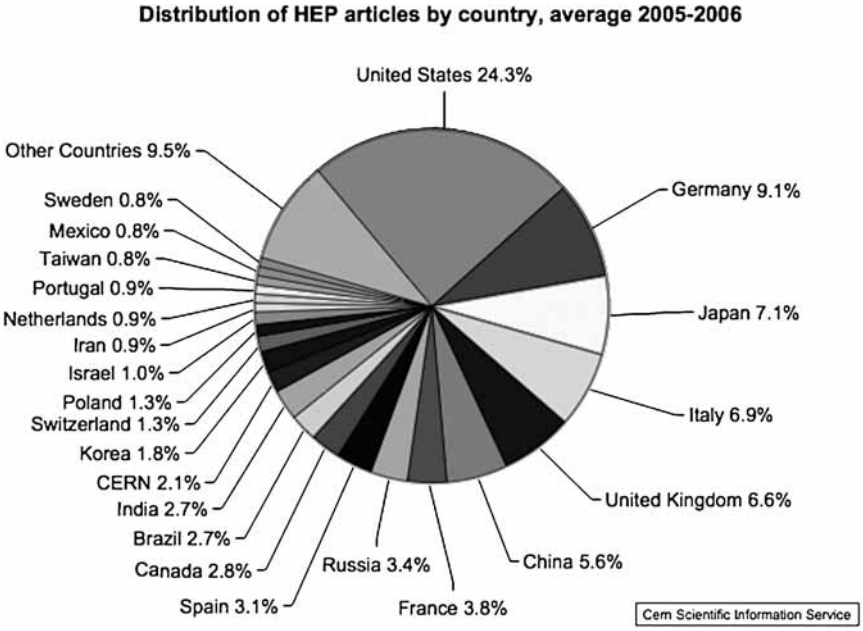

Fig. 2. Contributions by country to the HEP scientific literature published in the largest journals in the field. Co-authorship is taken into account on a pro-rata basis, assigning fractions of each article to the countries in which the authors are affiliated. This study is based on over 11,000 articles published in the years 2005 and 2006. Countries with individual contributions less than 0.8% are aggregated in the "Other countries" category [13].

SCOAP[3] is now collecting Expressions of Interest to join the consortium. Once it will have reached a critical mass, and thus demonstrated its legitimacy and credibility, it will issue a call for tender to publishers, aimed at assessing the exact cost of the operation, and then move quickly forward with the formal establishment of the consortium and its governance, then negotiating and placing contracts with publishers.

To date, most European countries have endorsed the project and major library consortia in the United States are in the process of completing the American share: SCOAP[3] has already received pledges for about a third of its budget envelope,[3] with another third having the potential to be pledged in the short-term future, as presented in Fig. 3.

6. Towards a future HEP Information System

For many years now almost all journal literature has been electronically available, the entire web is readily searchable, and commercial online databases provide metadata about all scientific literature. In addition, online services are changing more and more rapidly as new tools are developed and new ways of interacting with users evolve. In light of this fast-changing world, it is important to assess the usage by HEP researchers of the information resources that the community has pioneered in the last decades, as described in Section 3. Such an assessment serves two purposes: within the field, it informs on the need for HEP-specific community-based resources and their real role in the present

[3]The evolution of the SCOAP[3] fundraising and membership can be followed at [15] and [16].

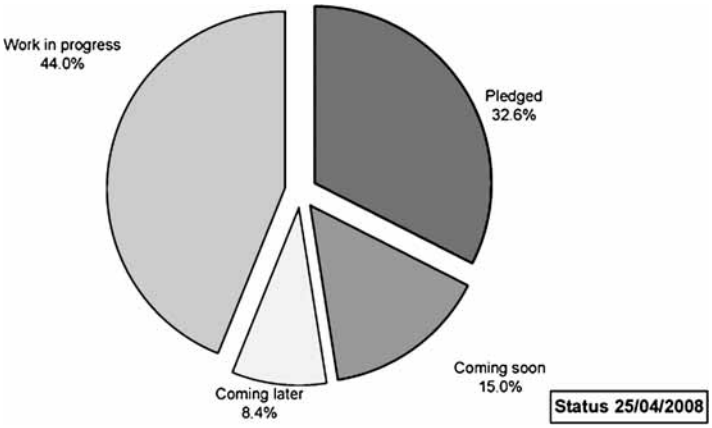

Fig. 3. Status of the SCOAP^3 fund-raising at the time of writing. About a third of the funds have already been pledged, 15% are expected to be pledged in the coming weeks, while discussions and negotiations are in progress for another 44%.

internet landscape, inspiring their future evolution; globally, it provides an in-depth case study of the impact of discipline-based information resources, as opposed to institution-based information resources or cross-cutting (commercial) information platforms. This information is particularly relevant in light of recent worldwide moves towards self-archiving of research results at the institutional or disciplinary level and the need to effectively incorporate these resources in the research workflow.

This assessment was performed in mid 2007 through a user survey [17] that was filled by about 10% of the practitioners in the field: an overwhelming response whose results are discussed in the following. The main question of the survey concerned the most-used information systems in the field: for 91% of the participants these are services maintained by the community. The most popular systems are the SPIRES database with 48.2% and arXiv with 39.7%. Google scores 9% though within the group of scientists with less than two career years this fraction rises to 22%. The role of commercial databases is negligible with 0.1%. The results are depicted in Fig. 4. It should be noted that the use of Google benefits strongly from the fact that community-based systems have made their content available for harvesting. At the same time, Google also acts, as in many other fields, as a broader alternative to publisher portals, given that indexing of many publisher websites has taken place in recent years.

Similar findings are observed for the most used systems to look for articles for which either the authors or the reference are known, as well as theses. Commercial systems never exceed a few percent share.

In addition to inquiring about the most heavily used systems for different tasks, the survey aimed to assess the importance of various aspects of information resources. Respondents were asked to tag the importance of 12 features of an information system on a five-step scale, ranging from “not important” to “very important”. The results are presented in Fig. 5. Access to full-text stood out clearly as the most valued feature, following close behind are depth of coverage, quality of content and search accuracy.

The survey explicitly inquired about the level of change that HEP scholars would expect, and require, from their information resources in the next five years: 75% expected “some” to “a lot of ” change and 90% of the users tagged three features as the most important areas of change: the linked presentation of all instances of a result, centralization, and access to data in figures and tables.

The survey also collected thousands of free-text answers, inquiring about features of current systems and their most-desired evolution. Some of the most inspiring free-text answers were along the following lines:

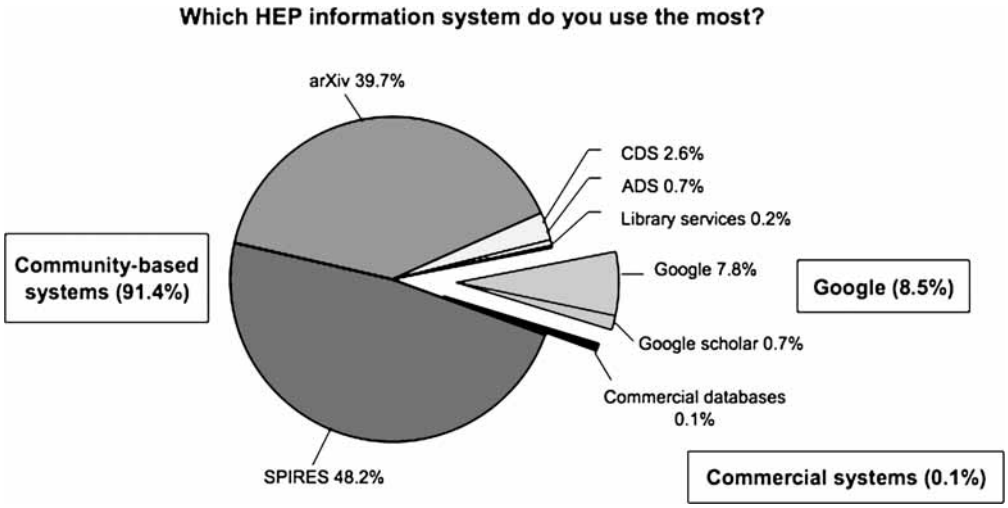

Fig. 4. Information resources favoured by HEP scientists. Community based systems dominate the landscape, even though among younger scholars there is an onset of Google. The usage of commercial systems (SCOPUS, INSPEC, the WebOfScience and similar products) is negligible.

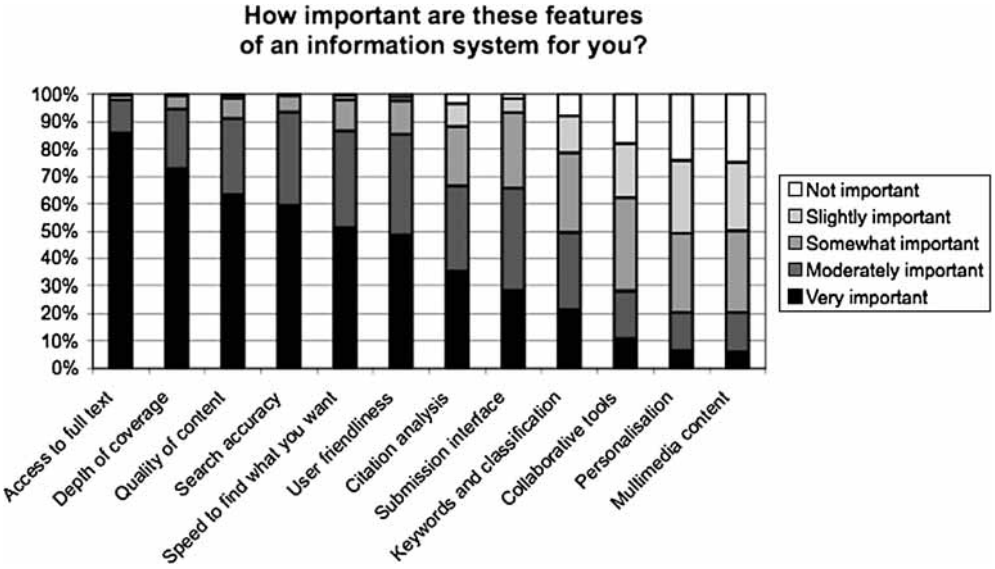

Fig. 5. Features of an information system most relevant for HEP scientists.

- Desire for seamless open access to older articles, prior to the onset of arXiv in 1990s.
- Improved full-text search and access to research notes of large experimental collaborations. These are a crucial grey-literature channel where large amounts of information and details about the results of large experiments transit.
- Indexing of conference talks and long-term archiving of the corresponding slides, beyond the lifetime of conference websites. Interlinking these slides with the corresponding conference proceed-

ings, in preprint form with reference to published volumes, and possibly other instances describing the results.
- Use of the HEP information resources as fora for the publication of ancillary material, crucial in the research workflow, and in particular:
  - numerical data corresponding to tables;
  - numerical data corresponding to figures;
  - correlation matrices and additional information beyond these presented in tables to allow an effective re-use of scientific results;
  - fragments of computer code accompanying complex equations in articles, to improve the research workflow and reduce the possibility of errors;
  - primary research data in the form of higher-level objects.
- "Smarter" search tools, giving access to articles related to articles of interest.
- Establishment of some new sort of open peer-review, overlaid on arXiv.

The survey also tried to assess the potential for the implementation of Web2.0 features to capture user-tagged content. Respondents were asked how much time they would spend on a tagging system to give a service to the community: 63% would spend between five minutes a day and an hour a week. There is an immense potential for user-generated, or rather user-tagged and user-curated, content in the field of information provision in HEP, as in many other fields of web-based communication.

These results inform the future evolution of information management in HEP and, as these researchers are traditionally "early adopters" of innovation in scholarly communication, can inspire developments of disciplinary repositories serving other communities.

The results of this survey and strategic discussions between four leading HEP laboratories (CERN, DESY, Fermilab and SLAC), in synergy with other partners (notably arXiv) and in a continuous dialogue with major publishers in the field, led to a roadmap towards a future HEP information system, consisting of the following steps:

1. Build a complete HEP information platform;
2. Enable text- and data-mining applications;
3. Demonstrate and deploy Web2.0 applications;
4. Preserve research data and ensure their re-usability.

Work on step 1, the inception of the next generation of HEP information system, is in progress, blending the current SPIRES database with a modern platform, the Invenio[4] open-source digital-library software. This new information system, under the working name of Inspire, is being developed by a collaboration of four leading HEP laboratories: CERN, DESY, Fermilab and SLAC. It will integrate the content of present repositories and databases to host the entire body of metadata and the full-text of all open access publications, past and future, including conference material, and will embody the one-stop shop HEP researchers are waiting for, encompassing all content of arXiv as well as decades of previous articles. In addition, it will offer advanced tools for citation analysis, for example the "cited with" option, which often allows serendipitous discovery of related articles. The steps 2 and 3 have been charted, and will be further refined during 2008, leading to the ultimate creation of a next-generation repository for HEP. Interestingly, the technical solutions to be deployed are independent of the content, and therefore

[4]Invenio was developed at CERN, where it now powers a 1-million records digital library ranging from articles to preprints, from multimedia to the institutional archives of the Organization, http://cdsware.cern.ch/invenio/index.html.

could be readily exported to other disciplines. Indeed, the Invenio software is now used in a dozen other repositories carrying content in other sciences and the humanities.

The next step on the roadmap for a future HEP information system will be to enable text- and data-mining applications that

- detect relations between documents carrying similar information;
- create datasets to exercise new hybrid metrics to measure the impact of articles, authors and groups;
- extract numerical information from figures and tables within published articles.

The mid-term future will see the development of Web2.0 applications that

- engage readers/authors in subject tagging, altering automatically-assigned classifications;
- allow community-based aggregation of related objects (articles, preprints, conferences, lectures);
- enable the possibility to review and comment on articles, adding links to additional documents or other digital objects.

It is interesting to note that the last features are already available in many services "overlaid" on arXiv, as a proto-form of alternative peer-review, but their acceptance is limited, due to the reduced usage of these sites when compared with the main access points to the literature. An inspiring experiment will be the deployment of these Web2.0 features in the production systems that the vast majority of HEP users adopts for their daily access to the literature: will this naturally lead to these additional means of communications entering the mainstream of the research workflow?

A long-term target will be, as a natural evolution of repositories, to link to data, simulations and computer programs behind each record, as discussed in Section 7.

7. The next frontier: Research data

As discussed in the previous sections, HEP is a community of pioneers in scholarly communication. It would come as a surprise to an external observer, that few steps have been taken in opening up its data for re-use, a subject which is becoming more and more mainstream in the debate on the evolution of scholarly communication towards an "open society". An immediate explanation of this phenomenon is the sheer size and the complexity of HEP primary data, as well as the limited scope for their immediate interdisciplinary fruition, when compared, for instance, to data from astronomy or Earth observation. While the first is epitomized by the mental picture of one year of LHC data which, were it written on CDs, would result in a stack 20 km tall, the latter gives space to more reflections. In recent years, powerful synergies arose between HEP and astrophysics, both studying our universe from different points of view, and simultaneous analyses of some data set will become an issue in the very near future. Beyond this opportunity, and based on first principles, HEP data should be preserved for their future re-analysis, being generated by a global enormous financial and human investment: they are a scientific legacy.

In HEP, the cases of a re-analysis of older data are only few but important. An example in which some is the re-analysis of data collected in 1980s at DESY together with data collected in the late 1990s at CERN, in the light of improved theories [18]. Without going into technical details, this work allowed to improve the understanding of the force which binds quarks in the nuclei. These studies were only possible through the strong desire and motivation of a few colleagues who took care to preserve and repeatedly copy the old JADE data onto up-to-date data carriers, to preserve the detector specific

software such that it still runs on modern computer systems, and to realise the value and significance of old data to actual questions of modern physics.[5]

Beyond anecdotes, there is a growing awareness of the need to preserve HEP data for their re-use and for (open) access [19,20]. There are five distinct continua for which data should be preserved or for which access would enable the production of more science, corresponding to very different user profiles and time scales after the data collection:

1. The same researchers who took the data, after the closure of the facility, with a time scale which can be a year or a decade;
2. Researchers working at similar experiments, with a time scale between days and years;
3. Researchers at future experiments, with a time scale of several decades;
4. Theoretical physicists who may want to re-interpret the data, with a time scale of months or some years;
5. Theoretical physicists who may want to test future theories, with a time scale of decades.

These cases are all different. In (1), the knowledge on the content of the data is in principle available, but software and hardware problems might be the limiting factor. In (2) and (4) a synergy between the user and the producer of data could lead to an immediate re-use of data, without the need to devise preservation strategy. However, (3) and (5), which can imply crucial consequences for the advancement of the field, make it evident that knowledge, in addition to technical solutions, has to be preserved together with the data. Enabling all these continua to re-use HEP data implies an indissoluble link between preservation strategies and access strategies. Indeed, preservation, in HEP, is not entirely a technical or an archival issue: during the long life-time of experiments, sometimes two decades, computing centres routinely copy old tapes onto new facilities and software migration can and does occur. This is made possible by a curation of the data by the producer themselves, which implies that HEP data from facilities recently stopped or about to be discontinued remains vaguely readable, although only understandable to the scientists who produced them in the first place. With the disbanding of the experimental collaborations that collected the data, though, software migrations stop and, worse, the insider knowledge needed for a re-analysis at a later stage is scattered or lost.

In a sentence, the issue with the preservation, re-use and (open) access to HEP data is the complexity of the data themselves.

A prerequisite for the re-use of HEP data is to embed in the data the knowledge about the data themselves, creating higher-level objects which can be understood later in time and by scientists not involved in the creation of the data. This "parallel" data format would have to emerge in addition to the ones used internally by the experiments. The benefits of such model are obvious. But there are formidable obstacles to overcome, technical and sociological. From a technical point of view, these "parallel" high-level data should be defined, possibly standardized, and created at the same time as the standard data reconstruction [19,20]. Even if this process would only cost a fraction of the human and financial capital invested in HEP experiments, a small fraction of a large number (thousands of person-years) still translates in a major project. This leads to the sociological barriers: this further investment would be in competition with research, and there is little to no academic incentive to put data preservation higher on the agenda than the data perusal and the preparation for further experiments. Scientists would need enor-

[5] To continue the story they bought a juke-box to store CDs of OPAL data after the completion of this later experiment.

mous academic incentives or additional funds. A debate[6] should take place concerning data ownership, (open) access, credit, accountability, reproducibility of results, depth of peer-reviewing. While each of these issues would require a full-length article to be laid out, it is clear that decades of a traditional and monolithic way of doing research need rethinking. Due to the complexities of these issues, HEP may be considered as a worst-case scenario in the topic of data preservation, re-use and (open) access, but a scenario that has the potential to inspire other fields of science, as in the other endeavours of HEP in the field of scholarly communication.

8. Conclusions

With 50 years of preprints and 17 years of repositories, not to mention the invention of the web, HEP has spearheaded (open) access to scientific information and is now in a period of change at two frontiers: the cross road of open access and peer-reviewed literature and the inception of a next-generation repository which has to adapt the current technological advances to the research workflow of HEP scientists.

In the spirit of their collaborative tradition, HEP scientists are now proposing to pool together resources from libraries and HEP institutes worldwide to sponsor the transition to open access of the entire literature of the field, through the SCOAP^3 initiative (Sponsoring Consortium for Open Access Publishing in Particle Physics). This open access publishing model is gathering growing international consensus, being non-disruptive to authors and, to a substantial degree, to publishers and societies. It has the potential to fundamentally alter the role of libraries in the publishing process and re-think the role of high-quality journals in the open access era. SCOAP^3 should demonstrate its potential in the coming months.

At the same time a new e-infrastructure of HEP Scientific Communication is being set up. A complete information platform is being built, enabling text- and data-mining as well as Web2.0 applications. This new platform is entirely user-pulled, building on a tradition of authors strongly supporting and advocating the use of repositories and meeting the expectations of HEP physicists in light of the new opportunities offered by technological development to scholarly communication. The new e-infrastructure might provide inspiration to many other communities which are currently exploring ways to improve the dissemination, discovery and organization of research results, primarily focusing on author self-archiving.

A new challenge for the future, which has however to be tackled as soon as possible, is the preservation, re-use and (open) access to HEP data. This is both a technical issue and a sociological one. On one side sits the sheer amount and relative complexity of HEP data, on the other a culture that surprisingly did not foster efforts to disseminate and exchange HEP data. Awareness on this issue is rising, and the next years, together with new, landmark, HEP data from the Large Hadron Collider at CERN, might bring new ways to preserve and re-use its data, another example for other scientific disciplines.

We are at the onset of a new era of innovation in scholarly communication in HEP and beyond.

References

[1] R.-D. Heuer, in: *International Conference on "Academic Publishing in Europe"*, Berlin, January 22nd–23rd, 2008; http://www.ape2007.eu/presentations/2.Heuer.pdf.

[6]The FP7 PARSE.Insight project (Insight in Permanent Access to the Records of SciencE) has among its objectives to understand the implications, not only technical, for HEP to start a process of preserving its data. PARSE.Insight will deliver its report in 2010. http://parse.digitalpreservation.eu/.

[2] S. Mele et al., *Journal of High Energy Physics* **12** (2006), S01; arXiv:cs.DL/0611130.
[3] L. Goldschmidt-Clermont, *Communication Patterns in High-Energy Physics*, 1965; http://eprints.rclis.org/archive/00000445/02/communication_patterns.pdf.
[4] L. Addis, 2002, http://www.slac.stanford.edu/spires/papers/history.html (last visited May 1st, 2008).
[5] P.A. Kreitz and T.C. Brooks, *Sci. Tech. Libraries* **24** (2003), 153–172; arXiv:physics/0309027.
[6] P. Ginsparg, *Computers in Physics* **8** (1994), 390–396.
[7] T. Berners-Lee, *Weaving the Web*, HarperCollins, San Francisco, 1999.
[8] P. Kunz et al., http://www.slac.stanford.edu/history/earlyweb/history.shtml (last visited May 1st, 2008).
[9] S. Bianco et al., Report of the SCOAP^3 Working Party, http://www.scoap3.org/files/Scoap3WPReport.pdf.
[10] http://www.terascale.de/news/scoap3_initiative (last visited May 1st, 2008).
[11] S. Bianco et al., Report of the SCOAP^3 Working Party; http://www.scoap3.org/files/Scoap3WPReport.pdf, http://scoap3.org.
[12] R. Voss et al., Report of the Task Force on Open Access Publishing in Particle Physics, http://www.scoap3.org/files/cer-002632247.pdf.
[13] J. Krause et al., http://www.scoap3.org/files/cer-002691702.pdf.
[14] http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97236.pdf.
[15] http://www.scoap3.org/fundraising.html.
[16] http://www.scoap3.org/whoisscoap3.html (last visited May 1st, 2008).
[17] A. Gentil-Beccot et al., arXiv:cs.DL/0804.2701.
[18] The JADE and OPAL collaborations, *Eur. Phys. J. C* **17** (2000), 19–51; hep-ex/0001055.
[19] S. Mele, Preservation, re-use and (open) access to HEP data, in: *Tools & Trends in Digital Preservation*, The Hagues, 31 October 2007.
[20] J. Engelen, in: *Conference of the Alliance for Permanent Access*, Brussels, 15 November 2007, http://www.alliancepermanentaccess.eu/power/Engelen_Alliance_151107.ppt.

Information Services & Use 28 (2008) 97–103
DOI 10.3233/ISU-2008-0558
IOS Press

Open access – A never ending story? *

Arne K. Richter **
European Geosciences Union (EGU), Germany

1. Introduction

Like all good stories, also the story of "open access" is occasionally rather confusing, and in which direction the story will develop and where and how it will end will depend on the point of view and the position of the person actually telling the story. Thereby, one has to remember that the story of "open access" is a trilogy rather than a monograph, since at least three sides with their own, independent ideas are involved: the subscribers (readers and researchers), the authors and the publishers.

The present report is therefore based on our own experience as scientists serving as readers, researchers, authors, editors, referees and officers for the European Geosciences Union (egu.eu) and its open access publications, with our open access publisher Copernicus Publications (copernicus.org), and with the software house for open access publishing Copernicus Systems + Technology (copernicussystems.net).

After a short retrospect on the history of the open access mission, we will discuss the advantages as well as the stumbling blocks of open access for researchers, authors and publishers and possible business models for open access publishing.

2. Retrospect on the open access mission

Only a few years ago the academic community was facing the following, rather hopeless situation:

- A steady increase in the subscription rates of academic work by about 250% from 1983–2003 or of about 13% per year (journalprices.com) on the one hand, and, at the same time, a substantial decrease in the financial support of academic work in general on the other hand.
- A substantial surplus for commercial publishers from academic publications on the one hand, and a substantial transfer of "value adding" work from the publishers back to the academic community serving as authors, editors and referees free of charge anyhow.
- An explosive expansion of the internet and its accessibility worldwide and of academic work available on the internet, even though only toll-free.

Based on this situation, the academic community reacted in different ways:

*Lecture presented at the International Conference on Academic Publishing in Europe 2008, Berlin, Germany, 22–23 January, 2008, www.ape2008.eu. Published under the creative commons attribution licence.

**Address for correspondence: Dr. Arne K. Richter, Executive Secretary, European Geosciences Union, Buchenweg 30, D-37191 Katlenburg-Lindau, Germany. Tel.: +49 5552 7109; Mobile: +49 170 9078220; E-mail: arne.richter@egu.eu.

- To require "Free Access to all Academic Work on the Internet" like the free access to all the other work presented on the web.
- Unrefereed, self-publishing of own work in the old fashion pre-print style (arXiv.org).
- Self-archiving of copies of published work (sherpa.ac.uk/romeo) or its archiving in public repositories and digital libraries (opendoar.org) for free public use.
- Cancellation of subscriptions and/or passover of entire editorial boards from expensive to less expensive journals and/or notices of termination from the position of editor and/or referee by individual scientists.

The request for "free access" was later on phrased more precisely as "open access" in the Bethesda Statement (2002) and in the Berlin Declaration (2003) in the following way:

- Barrier-free access to all academic work on the internet.
- License to copy, distribute, transmit and display the work publicly and to make and to distribute any derivative works to proper attribution of authorship (creative commons attribution license).
- Deposit the work and any supplemental material in at least one online repository for open access and long term archiving.

A survey by the Deutsche Forschungsgemeinschaft (DFG) on the other hand showed that the scientific community prefers to publish its work in proper, thematic scientific journals of high reputation and international, worldwide distribution, which guarantee a fair but rigorous peer review process of all academic work prior to publication, and which are included in the Citation Index with reasonable Impact Factors.

This survey clearly disapproves the preprint and self-archiving philosophies. Moreover, the distribution of author-generated copies of published work will be questioned with regard to its originality, and in the long term it will undermine the availability of genuin publications. And, finally, cancellations of subscriptions will most of all hurt the scientists themselves as readers and researchers but also as authors (I cannot read but my work is not read either).

Thus, the early messages of the open access mission are:

1. From readers and researchers: free access to and free downloads for derivative works of academic work of high quality.
2. From authors: publication in refereed open access journals of high reputation and international, worldwide distribution.
3. And from scientists in general: both services, in principle, at no cost, since scientists that cannot pay expensive subscriptions can also not pay expensive page/article charges.

The open access mission must therefore find a way which is a win–win situation for all three sides: for the scientists as readers and authors as well as for the open access publishers.

3. Advantages of open access for subscribers and publishers

Here "Open Access" benefits, first of all, from the explosive development and availability of a barrier-free internet worldwide, tending to become the only platform for any digital information and any digital work.

In addition, all electronic media being able to store digital work for purposes of information and/or for performing derivative works are steadily decreasing in price but increasing in its capacities: besides

the PCs, Laptops and Palmtops, we will shortly have the TG stick, TG flash memories and fusel cells for powerful mini-workstations, as well as digital paper in user-friendly sizes. Moreover, more and more software, which is needed to perform professional original and/or derivative work, can be downloaded from the internet free of charge.

In parallel, the academic institutions and organizations have established a worldwide network of open access data bases, repositories, digital libraries and archives for open access academic work. This also includes data bases and archives for well defined topics, such as e.g. ADS, CAS, CSA, GeoRef etc., as well as the first long-term digital archive PORTICO of the US Library of Congress. Moreover, Alert Services, RSS feeds and DOI are active tools to inform the scientists online about the latest work of interest or its location on the internet.

Several organizations are working on the problem to improve the internet search-engines, such as Google or Google Scholar, to become more personalized and sophisticated with regard to the actual search as well as with regard to the personalized listings of the results (Get-What-You-Really-Want Search Engines). Another feature describing the quality of an open access publication could be the comparison of its content with the content of similar open access publications with regard to plagiarism and IPR (Intellectual Property Rights) violations to classify also the originality of a publication. Certainly, only academic work which is fully open accessible will profit from such advantages.

ISI – Thompson will also include quality, open access and purely digital journals in the Citation Index; and it has already been shown that the higher the percentage of open access publications in a given journal is the higher its Impact Factor.

A further important advantage for researchers is that open access work is published under the creative common attribution license, i.e. that any open access work may be copied, distributed, transmitted and displayed publicly and may be used to make and to distribute any derivative works to proper attribution of authorship without any annoying questions regarding copyright regulations any longer.

To help readers and researchers in studying open access literature online, certain open access publishers have started to change the format of their journals from portrait, double column (print) to landscape, on-screen publications with off-screen readable fonts (e.g. www.atmos-chem-phys-discuss.net).

Finally, the scientific community is more and more requesting that also the essential supplemental material of an original publication becomes publicly available, including reports, movies, numerical codes, experimental set-ups etc. Here only the internet with links to (self-)archived materials can serve as an answer.

In summary we may conclude:

- Open access – internet publishing provides the best overall services for the academic community for studies and derivative works, and
- Open access work is available free of charge worldwide.

And the message back to the publishers:

- Only quality, open access work will be cited and/or used for derivative works in future, i.e. non-open access work will not be read and therefore slowly forgotten within the scientific community.
- Publishers should use the numerous advantages of the free internet more rigorously, since typical tasks such as publishing, distribution and marketing of academic work as well as its storage and its problems with subscriptions, back issues, annual page budgets, investments in new publications etc. will not exist in their old fashion manner any longer in the open access world of publishing and will therefore represent a good chunk of expenses to be saved.

4. Stumbling blocks for open access

Open access as the barrier- and subscription-free way of publishing academic work may be advantageous for readers and researchers; however, it is less advantageous for authors and publishers.

In the open access world the sources of income for publishers reduce to just page/article charges to be paid by the author. Thus, in the open access world the philosophy of payment is inverted from "subscription-paid and author-free" to "subscription-free and author-pays". In turn, these charges are therefore, in general, rather high. According to the conversion and the hybrid models offered by academic publishers, the article charges amount to about EUR 1,500–2,100 per article, depended on the size of the publisher and whether only the subscription fees or the overall turnover is converted into article charges.

In the past the subscription fees have been paid by the libraries or directly by the organizations. Page or article charges, however, have to be paid by the authors or their departments or institutes. Very often these departments and institutes have no explicit account for "page or article charges".

"Active" departments and research institutes would pay presently less for their subscriptions than for their article charges, if the price per article would amount to the value given above.

Even the signatories of the Berlin Declaration continue to pay subscription fees or a similar amount for their own authors to publish open access and are therefore still bound to the overall increase in the payment for published academic work. And the amount provided by some organizations to authors that want to publish open access is only a fraction of the article charges mentioned above.

And according to our experience in the geosciences, the old model of "subscription and author free" seemed to work more "ethical" than the open access model, since we were always able to distribute our own "pre-prints as accepted for publication" or even our off-prints to our friends and their departments and institutes, whereas for most of our colleagues a payment of article charges of the amount mentioned above will be impossible.

Finally, all major organizations in Europe will continue to support the classical, commercial publishers in the one or the other way. During the transition phase from toll-access to open access the amount of money for open access will remain relatively small.

Thus, serious open access publishers serving an international, worldwide community must investigate the details of applying the internet and the modern, digital technologies in their publication strategies for reducing the article charges for authors and yet earn an income sufficient for completing their tasks and responsibilities in the way asked for by the academic community.

5. Open access online publishing for authors and publishers

Due to the steep cuts in their budgets, scientists have learned to use their computers, the associated software and the internet to compile and to edit the results of their research ready-for-publication (called camera-ready some years ago), and to submit or to upload them online by themselves. This includes not only simple text but also equations, tables, plots, figures, photographs, movies, sound tracks etc. In this way, they are also able to interact with server-aided macros and templates to produce their articles online in the respective journal style ready for publication on the internet. Moreover, in the open access world readers and researchers will loose their interest in elaborated and expensive formats and will prefer simple, author-generated pdf or HTML files, wikis or blogs. Thus, the actual production of articles will rest mainly in the hands of the authors and of intelligent servers, and the staff at the production offices of

publishers will just serve as a kind of "supervisor and quality control" rather than performing typesetting, editing, lay-outing, formatting etc. as in previous times. Purely author-generated, high-quality online publications are already included in our discussion journals (e.g. www.atmos-chem-phys-discuss.net).

Once an article has "passed", the server generates automatically customized XML files of the article and sends them to the respective data bases, archives, repositories and libraries worldwide, and via the alert services and RSS feeds also to the individual customers. In this way the article is published immediately worldwide and becomes at once part of all search engines and catalogues practically free of charge for the publisher.

In this way, an author receives prompt publication, a gigantic large readership, highest impact factors, and, last but not least, the copyright remains with the author. And for the publisher the cumbersome printing and distribution and tracing of printed issues has passed, while print-on-demand allows to distribute printed material at any time. Here a substantial part of past expenses of publishers can again be saved.

Remains the more complicated and time consuming and therefore more expensive part in the publication procedure, the procedure of the actual peer review of scientific work prior to publication. Here we are applying two different models:

1. The online yet classical review based on 2–3 independent referee reports – anonymous or eponymous.
2. The "open online peer review plus public discussion" procedure.

Since in both cases the entire review process rests in the hands of the editors and referees, it entirely can be serviced and monitored by intelligent servers, reducing the actual work of the editorial support office of the publisher more or less to a monitoring support of the work of the authors, editors and referees. Thereby, the second procedure has shown to have the following additional advantages:

- The manuscripts submitted for publication are more carefully compiled and of higher standard, so that editors and authors have less work to do for the review.
- Referees' and public comments are open access and published alongside the original article. Thus, value is added to the work of the referees and editors and the reports are generally of greater value to the author.
- The actual work of the editorial support office is roughly cut in half.
- Manuscripts are more intensively check against plagiarism and IPR violations.

At any time, however, either the author, the editor or the publisher can decide which services have to be added by the publishing staff for a successful publication of the manuscript. This also holds for copy editing, which is performed online by professional staff and which has to be paid by the author as an extra service.

From our model of publication we may therefore conclude:

- The entire process of submission, registration, peer review, production, publication and distribution of academic work can be handled, in principle, by the authors, editors and referees in cooperation with intelligent servers together with the publisher providing just the overall management, monitoring and quality control, the costs of which are merged into one fixed flat-rate per article.
- In addition, the author pays for the services asked for from the editorial support and the production offices of the publisher, page by page.

This service charge model is already applied by the EGU to its open access journals; however, not yet fully automated, since the articles in all main journals are still edited and formated by the staff in the production offices in the old fashion portrait, double column, print-on-paper style.

6. Business models for open access

Due to an increasing pressure on the part of the science community towards open access, academic publishers will be urged to become, sooner or later, open access publishers. In the transition phase, which is already going on, different business models are discussed in parallel:

1. The Hybrid Model: in a journal, which is not open access in general, an author may choose the option that his/her article is open access in accordance to the creative commons attribution license. For this service the author has to pay an extra article charge (Springer's Open Choice Model).
2. The Conversion Model: a smaller, well-defined community organized through a dozen or so institutes/organizations and using only a small number of journals for research and publication, arrange with the corresponding publishers that by payment of the regular subscription fees the journals become open access (the CERN et al. Model).
3. The Article Charge Model: articles are publish in accordance to the classical procedure but are open access and the charge per article is minimized but fixed for the journal in question (BiomedCentral, PloS, IOP – New Journal of Physics).
4. The Service Charge Model: the less service an author requires from the editorial support and the production and distribution office of the publisher, the less the price per page (European Geosciences Union).

According to the first model, an institute subscribing to journals but asking its authors to publish open choice, pays actually twice. The conversion model, on the other hand, puts the same fiscal burden on the formerly subscribing institute, even if other institutes will have free access to their work. Models 3 is a rigid model leading still to article charges of the amount of about EUR 1,500, which is too high for the majority of the worldwide geosciences community. Model 4 is an automated, adoptive and flexible model, and in view of the very fast development of the software on the internet and for the web, this model will lead eventually to very low flat rates. Our dream is even that the makers of OpenOffice, ADOBE etc. develop an Open Access Publishing Tool to further reduce the page and finally the article charges to values, which will then become affordable for all authors worldwide. The mission statement of the EGU for publications is: dedicated to the pursuit of excellence and free and universal accessibility and affordability of scientific publications in all areas of geosciences and planetary and solar system sciences for the benefit of the scientists worldwide.

7. Reputation

The essential work for a publication is performed by the scientists acting as author, editor, referee, subscriber, reader and researcher. The decision about the quality, acceptance and value for further promotion purposes, and thus of the reputation of a journal also rests in the hands of the scientific community, including some of the parts of reputation, such as excellent editors, strong and competent referees, a steady flow of excellent, up-to-date and timely work, diversity etc.

Thus, learned societies and organizations with their devoted communities and their tradition to listen to their grass-roots members and to follow and to incorporate their feedback in their work will be able to build up publications of highest standards, reputation, and excellency under the flag of open access, and by applying the automated, service charge model even at low investment and maximum services for their members and their worldwide communities.

Some years ago, the EGU has started a new open access journal with an open peer review process and public discussions right in the middle of a group of traditional, high ranked atmospheric and environmental journals. After only 4 years this journal was the number 1 publication according to its Impact Factor. A later journal started with an Impact Factor of 2.13 right from the beginning!

8. Summary

Open access is a mission for modern, online publishing of academic work with great potential and advantages for all sides involved: the researchers (subscribers), the authors and the publishers.

Based on the continuous increase in the subscription fees by about 13% per year and the resulting decrease in the availability of quality publications in the academic libraries, the general opinion was voiced that at least the work on the internet should be open accessible, i.e. barrier-free accessible and freely usable for any derivative work to proper attribution of authorship. Certainly, such a mission guarantees maximum worldwide distribution and impact of academic work and a number of simplifications and savings with regard to the tasks and responsibilities of publishers.

On the other side, however, it causes new, even more severe problems which have to be solved: in a strictly open access world the sources of income for publishers would reduce to page or article charges to be paid by the authors. Converting the income from subscription fees or even the overall income to article charges, one would presently get fees of the order of EUR 1,500–2,500 or even more per article. For most of our scientists worldwide it will be absolutely impossible to pay such a high price. Even more, in the previous subscription model in which publishing was free for authors, the "richer" colleagues could distribute copies, preprints and off-prints of their work to their "poorer" colleagues. In the new open access world reading would be free but publishing practically impossible for most of our colleagues. Thus, as long as the number of genuin publications in journals of high reputation is one of the key measure for the quality of a scientist, open access so far would only favour the rich but not necessarily the best scientists. Even though the open access publishers are taking a 10% margin of free-of-charge pages in their annual budget into account, this would not be sufficient to serve the worldwide community in many areas in the geosciences.

Thus, if we want to put the open access mission on sustainable grounds, we have to continue to make rigorously use of the internet and the new web 2.0, the freely available software and intelligent servers and to develop the "Automatic Online Publication and Distribution System" to further reduce the tasks and responsibilities of publishers with regard to submission, peer review, production, distribution and marketing of academic work to an absolute minimum, just providing monitoring and quality control of the processes involved and the overall management of a journal. In this way, we envisage that article fees can be reduced to affordable flat-rates covered, may-be, by interested sponsors.

In this way, "Open access" is indeed a story that, at least for the moment, has not yet come to its end: to the battle call "free to read for everyone" we have to add "and free to publish for everyone" if we want to serve our entire community in a fair way!

Information Services & Use 28 (2008) 105–107
DOI 10.3233/ISU-2008-0567
IOS Press

Interactive open access publishing and collaborative peer review for improved scientific communication and quality assurance

Ulrich Pöschl [a] and Thomas Koop [b]
[a] *Max Planck Institute for Chemistry, D-55128 Mainz, Germany*
[b] *Bielefeld University, D-33615 Bielefeld, Germany*

Abstract. The traditional ways of scientific publishing and peer review do not live up to the needs of efficient communication and quality assurance in today's highly diverse and rapidly developing world of science. Therefore, new opportunities for enhanced scientific quality assurance are among the most important advantages and perspectives of open access to scientific publications. Open access gives referees more information to work with; it enables interactive and transparent forms of review and discussion open to all interested members of the scientific community and the public; and it facilitates the development and implementation of new metrics for the impact and quality of scientific publications. The effects and advantages of open access can be efficiently and flexibly combined with the strengths of traditional scientific publishing and peer review. Among the successful initiatives pursuing this approach are the interactive open access journal *Atmospheric Chemistry and Physics* (ACP, www.atmos-chem-phys.net) and a growing number of sister journals published by the European Geosciences Union (EGU, www.egu.eu). They are practicing a two-stage publication process with public peer review and interactive discussion, which has been designed to resolve the dilemma between rapid scientific exchange and thorough quality assurance.

1. Introduction

Recent high profile cases of scientific fraud have fuelled the discussion of scientific quality control. A problem of similar, if not greater, importance is the large proportion of carelessly prepared scientific papers that dilute rather than enhance scientific knowledge. Both problems are indicative of shortcomings in the traditional peer review system, yet many scientists and publishers believe that peer review remains the best available approach for scientific quality assurance. Nevertheless, requests for improvements are commonplace. Among the suggestions are the public exchange of referee comments and author replies [1,2] and public discussion of manuscripts before formal publication [3,4]. We think and demonstrate that these concepts and their advantages can be efficiently combined with the strengths of traditional peer review.

Quality assurance of scientific publications usually proceeds through two pathways: a pre-publication short term assessment by designated referees during the peer review process, and a post-publication long term assessment by the scientific community through comments, citations, review articles and monographs. Both can be combined in a collaborative peer review process where members of the scientific community can participate in the assessment of scientific manuscripts through interactive comments in addition to designated referees' reports.

2. Interactive open access journal concept

This approach is pursued by the open access journal *Atmospheric Chemistry and Physics* (ACP, www.atmos-chem-phys.net) and a growing number of sister journals published by the scientific service provider Copernicus (www.copernicus.org) on behalf of the European Geosciences Union (EGU, www.egu.eu). These journals have a two-stage publication process with public peer review and interactive discussion [5–8]. In the first stage, manuscripts that pass a rapid pre-screening (access review) are immediately published as "discussion papers" on the journal's website. They are then subject to interactive public discussion for a period of eight weeks, during which the comments of designated referees, additional comments by other interested members of the scientific community, and the authors' replies are also published alongside the discussion paper. While referees can choose to sign their comments or remain anonymous, comments by other scientists must be signed. In the second stage, manuscript revision and peer review are completed in the same way as in traditional journals (with further rounds of non-public revisions and referee review where required) and, if accepted, final papers are published in the main journal. To provide a lasting record of review and to secure the authors' publication precedence, every discussion paper and interactive comment remains permanently archived and individually citable.

3. *Atmospheric Chemistry and Physics*

ACP was established in 2001 and now publishes about 500 papers per year. On average, one in four papers receives a comment from the scientific community in addition to the comments from designated referees (in traditional journals this happens for about one in 100 papers). And there are typically 0.5 pages of comments and replies per page of original discussion paper. These cover the full spectrum of opinions – from harsh criticism to open applause – and provide a wealth of additional information and evaluation that is available to everyone.

ACP statistics confirm that collaborative peer review facilitates and enhances quality assurance. The journal has a relatively low overall rejection rate of less than 20%, but only four years after its launch ACP had already reached – and has since then maintained – the highest ISI journal impact factor in the field of "Meteorology and Atmospheric Sciences" (48 journals) and one of the highest in "Environmental Sciences" (144 journals) and "Geosciences, Multidisciplinary" (131 journals) [9]. These numbers support the anticipation that public peer review and interactive discussion deter authors from submitting low quality manuscripts and, thus, relieve editors and referees from spending too much time on deficient submissions.

This is particularly important, because refereeing capacities are the most limited resource in the publication process. While peer review depends crucially on the availability and performance of referees, it has traditionally offered little reward for those providing careful and constructive reviews. In public review, however, referees' arguments are publicly heard and, if comments are openly signed, referees can also claim authorship for their contribution. Over six years at ACP, we have found that about two-thirds of our referees prefer to remain anonymous. There are, however, interesting differences between subdisciplines: $\sim$50% of modellers sign their referee comments, while only $\sim$30% of the experimentalists do so. It appears that modellers more often provide suggestions and ideas for which they like to claim authorship as a reward [5–8].

4. Conclusions and outlook

We think that collaborative peer review with a two-stage publication process and interactive public discussion effectively resolves the dilemma between rapid scientific exchange and thorough quality assurance. It has proven to foster scientific discussion, deter submission of sub-standard manuscripts, save refereeing capacities, and enhance information density in final papers. Moreover, it can be flexibly integrated into existing journals as well as large scale publishing systems and repositories (such as arXive.org) – simply by adding an interactive discussion forum.

In the geosciences, seven sister journals are already successfully practicing the interactive open access journal concept of ACP (*Biogeosciences*, *Climate of the Past*, *e-Earth*, *Geoscientific Model Development*, *Hydrology and Earth System Sciences*, *Ocean Science*, *The Cryosphere*, www.egu.eu), and the launch of further sister journals is foreseen. Moreover, the concept has been adopted by the recently launched journal *Economics*, which is also aimed at top quality papers involves some of that discipline's most prominent institutions and scientists (www.economics-ejournal.org). Modified concepts of public peer review and interactive discussion are pursued by the open access publications *PLoS One* (www.plosone.org) and *Biology Direct* (www.biology-direct.com) in the life sciences.

Overall, open access enables not only the maintenance but substantial improvement of scientific quality assurance, and it provides the basis for efficient usage and augmentation of scientific knowledge in a global information commons [10]. Moreover, public review, discussion, and documentation of the scientific discourse can serve as an example for rational and transparent procedures of settling complex questions, problems, and disputes, i.e. as a model for the further development of the structures, mechanisms, and processes of communication and decision making in society and politics [11,12].

References

[1] D.K. Lahiri, Discourse among referees and editors would help, *Nature* **439** (2006), 784; doi:10.1038/439784d.
[2] D. Kennedy, Nick Cozzarelli, *Science* **312** (2006), 159; doi: 10.1126/science.1128103.
[3] E.G. Van Meir, Opening the chamber of peer-review secrets, *Nature* **429** (2004), 803; doi: 10.1038/429803c.
[4] L.B. de Carvalho, Pressure also leads to worthless publications, *Nature* **439** (2006), 784; doi:10.1038/439784a.
[5] U. Pöschl, Interactive journal concept for improved scientific publishing and quality assurance, *Learned Publishing* **17** (2004), 105; www.ingentaconnect.com/content/alpsp/lp/2004/00000017/00000002/art00005.
[6] T. Koop and U. Pöschl, An open, two-stage peer review journal, *Nature Web Debate on Peer Review* (2006); doi:10.1038/nature04988, www.nature.com/nature/peerreview/debate/nature04988.html.
[7] www.atmospheric-chemistry-and-physics.net.
[8] www.atmospheric-chemistry-and-physics.net/general_information/public_relations_and_background_information.html.
[9] www.atmospheric-chemistry-and-physics.net/news_acp_isi_impact_factor.pdf.
[10] P.A. David and P.F. Uhlir, *Creating the Information Commons for e-Science*, Paris, 2005; www.codataweb.org/UNESCOmtg/Book-of-Abstracts-InfoCommons2005-final.pdf.
[11] U. Pöschl, Open access & collaborative peer review, DFG-iFQ Working Paper No. 1, 2006, p. 43; http://www.forschungsinfo.de/Publikationen/Download/working_paper_1_2006.pdf.
[12] E. Bodenschatz and U. Pöschl, Qualitätssicherung bei open access, in: *Open Access Handbuch*, B. Malina, ed., UNESCO Kommission Deutschland, Bonn, 2007; www.unesco.de/fileadmin/medien/Dokumente/Kommunikation/Handbuch_Open_Access.pdf.

Information Services & Use 28 (2008) 109–112
DOI 10.3233/ISU-2008-0568
IOS Press

Peer review in scholarly journals: Perspective of the scholarly community – Results from an international study

Mark Ware
Mark Ware Consulting, UK
E-mail: mark@markwareconsulting.com

This summary is extracted from the report of the same title published by the Publishing Research Council (PRC) and reproduced here by kind permission of the PRC. The full report and a shorter edited version can be found on the PRC website at http://www.publishingresearch.net/PeerReview.htm.

Peer review is seen as an essential component of scholarly communication, the mechanism that facilitates the publication of primary research in academic journals. Although sometimes thought of as an essential part of the journal, it is only since the second world war that peer review has been institutionalised in the form we know it today. More recently it has come under criticism on a number of fronts: it has been said that it is unreliable, unfair and fails to validate or authenticate; that it is unstandardised and idiosyncratic; that its secrecy leads to irresponsibility on the part of reviewers; that it stifles innovation; that it causes delay in publication; and so on. Perhaps the strongest criticism is that there is a lack of evidence that peer review actually works, and a lack of evidence to indicate whether the documented failings are rare exceptions or the tip of an iceberg.

The survey reported here does not attempt directly to address the question of whether or not peer review works, but instead looks in detail at the experiences and perceptions of a large group of mostly senior authors, reviewers and editors (there is of course considerable overlap between these groups). Respondents were spread by region and by field of research broadly in line with the universe of authors publishing in the journals in the Thomson Scientific database, which covers the leading peer reviewed journals. The survey presents its findings in two broad areas: attitudes to peer review and current practices in peer review.

Attitudes to peer review

1. **Peer review is widely supported.** The overwhelming majority (93%) disagree that peer review is unnecessary. The large majority (85%) agreed that peer review greatly helps scientific communication and most (83%) believe that without peer review there would be no control.
2. **Peer review improves the quality of the published paper.** Researchers overwhelmingly (90%) said the main area of effectiveness of peer review was in improving the quality of the published paper. In their own experience as authors, 89% said that peer review had improved their last published paper, both in terms of the language or presentation but also in terms of correcting scientific errors.

3. **There is a desire for improvement.** While the majority (64%) of academics declared themselves satisfied with the current system of peer review used by journals (and just 12% dissatisfied), they were divided on whether the current system is the best that can be achieved, with 36% disagreeing and 32% agreeing. There was a very similar division on whether peer review needs a complete overhaul. There was evidence that peer review is too slow (38% were dissatisfied with peer review times) and that reviewers are overloaded (see #13 below).
4. **Double-blind review was preferred.** Changes to peer review in recent years (such as the growth of double-blind review, and the introduction of open and post-publication review) have attempted to improve the system. Asked which of the four peer review types was their most preferred option, there was a preference for double-blind review, with 56% selecting this, followed by 25% for single-blind, 13% for open and 5% for post-publication review. Open peer review was an active discouragement for many reviewers, with 47% saying that disclosing their name to the author would make them less likely to review.
5. **Double-blind review was seen as the most effective.** Of the four types of peer review discussed, double-blind review had the most respondents (71%) who perceived it to be effective, followed (in declining order) by single-blind (52%), post-publication (37%) and open peer review (26%). Respondents did not have personal experience of all types of review and tended to rate more highly the systems they had experienced. It is notable, though, that although 37% of respondents said that post-publication review was effective, only 8% had had experience of it as authors.
6. **Double-blind review faces some fundamental objections.** Double-blind review was primarily supported because of its perceived objectivity and fairness. Many respondents, including some of those supporting double-blind review, did however point out that there were great difficulties in operating it in practice because it was frequently too easy to identify authors from their references, type of work or other internal clues.
7. **Post-publication review was seen as a useful supplement to formal peer review.** In terms of recent developments facilitated by technology advances, some 37% thought that post-publication review was effective but only 5% preferred it over other approaches. It is clear that this was because researchers tended to see it as a useful supplement to formal peer review rather than a replacement for it (53% agreed compared to 23% disagreeing). Interestingly, they saw this usefulness despite a clear view that it tends to encourage instant reactions and discourage thoughtful review.
8. **No support for replacing peer review with metrics.** There was strong opposition to replacing peer review with post-publication ratings or usage or citation statistics to identify good papers, with only 5–7% of respondents supporting these approaches.
9. **Mixed support for review of authors' data.** A majority of reviewers (63%) and editors (68%) say that it is desirable in principle to review authors' data. Perhaps surprisingly, a majority of reviewers (albeit a small one, 51%) said that they would be prepared to review authors' data themselves, compared to only 19% who disagreed. This was despite 40% of reviewers (and 45% of editors) saying that it was unrealistic to expect peer reviewers to review authors' data. Given that many reviewers also reported being overloaded, we wonder, however, whether they would still be as willing when it actually came to examine the data.
10. **Limited support for payment for reviewers.** Respondents were divided on whether reviewers should be paid, with 35% in favour and 40% against payment. A majority, however, supported the proposition that payment would make the cost of publishing too expensive (52% for, 18% against) and the large majority of reviewers (91%) said that they reviewed to play their part as a member of the academic community.

Current practices in peer review

11. **Single-blind review was the most commonly experienced.** The average respondent had published 60 papers in their career to date, suggesting they were fairly experienced and productive researchers, and 8 papers in the last 24 months. As authors, respondents' experience of peer review was mainly of single-blind reviewing (84% said they had experienced this kind of review), followed at some distance by double-blind reviewing (44%). Less than a quarter (22%) reported experience of open peer review, while experience of post-publication review was limited to 8% of respondents.
12. **Longer review times was a cause of dissatisfaction.** Authors said the peer review of their last published paper took an average of 80 days. They were evenly balanced on whether or not this was satisfactory. There was a clear correlation between the reported time taken for peer review and the author's satisfaction: 67% were satisfied provided the time was under 30 days, but this dropped to 10% for 3–6 months, and to 9% for longer than 6 months.
13. **The most productive reviewers were overloaded.** Some 90% of authors were also reviewers, acting regularly for about 3.5 journals and a further 4.2 journals occasionally. They reported reviewing an average of 8 papers in the last 12 months, compared to the maximum of 9 that they said they were prepared to review. Active reviewers, defined as those doing 6 or more reviews in the last 12 months, completed an average of 14 reviews per year, nearly twice the overall figure. This means that although Active reviewers make up 44% of all reviewers, they are responsible for 79% of all reviews. So when this group reports it is over-loaded – doing 14 reviews per year compared to their preferred maximum of 13 – there is clearly a problem.
14. **About 20% of invitations to review are declined.** As well as completing 8 reviews per year, the average reviewer declined about 2 invitations to review, mainly because of a lack of time. Active reviewers, despite doing more reviews, if anything decline slightly fewer invitations proportionately.
15. **The average review takes 5 hours and is completed in 3–4 weeks.** Reviewers say that they took about 24 days (elapsed time) to complete their last review, with 85% reporting that they took 30 days or less. They spent a median 5 hours (mean 9 hours) per review.
16. **Altruistic reasons for reviewing were preferred over self-interested ones.** Substantially the most popular given was "playing your part as a member of the academic community". Self-interested reasons such as "to enhance your reputation or further your career" or "to increase the chance of being offered a role in the journal's editorial team" were much less frequently advanced.
17. **The average acceptance rate was 50%.** Editors reported that the average acceptance rate for their journals was about 50%, which is consistent with other studies. About 20% are rejected prior to review (either because of poor quality (13%) or being out of scope (8%)) and another 30% are rejected following review. Of the 50% accepted, 40% are accepted subject to revision. Acceptance rates were lower in humanities and social sciences, and higher in physical sciences/engineering journals.
18. **Use of online submissions systems.** Three quarters of editors (76%) reported that their journal used an online manuscript submission and tracking system. Their use was most common in life sciences (85%) and least common in humanities and social sciences (51%).
19. **Access to journals literature.** Some 69% of respondents described their access to the journals literature as good or excellent, with 7% describing it as poor or very poor. This probably represents an improvement in overall access compared to the CIBER 2004 survey [1], which reported 61% with

good/excellent and 10% poor/very poor (though a different geographical distribution of responses makes direct comparison difficult).

The survey thus paints a picture of academics committed to peer review, with the vast majority believing that it helps scientific communication and in particular that it improves the quality of published papers. They are willing to play their part in carrying out review, though it is worrying that the most productive reviewers appear to be overloaded. Many of them are in fact willing to go further than at present and take on responsibility for reviewing authors' data.

Within this picture of overall satisfaction there are, however, some sizeable pockets of discontent. This discontent does not always translate into support for alternative methods of peer review; in fact some of those most positive about the benefits of peer review were also the most supportive of post-publication review. Overall, there was substantial minority support for post-publication review as a supplement to formal peer review, but much less support for open review as an alternative to blinded review.

Reference

[1] I. Rowlands, D. Nicholas and P. Huntingdon, *Scholarly Communication in the Digital Age: What Do Authors Want*?, Centre for Information Behaviour (CIBER), 2004.

Information Services & Use 28 (2008) 113–119
DOI 10.3233/ISU-2008-0559
IOS Press

What means rich in publishing? Competition between for-profit and not-for-profit publishers from an economist's point of view

Alexis Walckiers
ECARES – Université Libre de Bruxelles, CP 114, Av. F.D. Roosevelt, 50, B-1050 Bruxelles, Belgium
E-mail: awalckiers@gmail.com

Abstract. In our report for DG-Research of the European Commission entitled "Study on the economic and technical evolution of the scientific publication markets in Europe", published almost two years ago, we showed that, on average, journals published by for-profit publishers are three times more expensive than journals published by not-for-profit publishers. Not-for-profit publishers are, in general, not subsidized, and our results are thus consistent with impediments to effective competition on the market for scientific publications. Relying on these observations I address two questions sequentially. First, I review the potential, natural and strategic barriers to expansion and entry. Second, scientific associations have large market shares in some scientific disciplines. I discuss the objectives of not-for-profit publishers and study whether they help innovate and drive prices down. Finally, I conclude that there is a need to promote pro-competitive, i.e. much more flexible, pricing strategies, scrutinize mergers, and refund VAT to research institutions.

Keywords: Scientific publications, price, competition, not-for-profit

1. Introduction

In our report for DG-Research of the European Commission,[1] we showed that, on average, journals (with similar characteristics) published by for-profit publishers are three times more expensive than journals published by not-for-profit publishers. Another finding of our report is that not-for-profit publishers have large market shares in some scientific disciplines and are, in general, not subsidized.

These findings shed light on the specificity of the market for scientific publications, because if this was observed in most other sectors of the economy, consumers would purchase the cheap goods (with similar characteristics), and the producers of the expensive good would have to lower their prices or would be driven out of the market through competitive pressure. This is not what happens on the market for scientific publications, where for-profit publishers have successfully expanded their portfolio of journals and enjoy large profit margins; they are in this sense *rich in scientific publishing.*

Our results are thus consistent with impediments to effective competition on the market for scientific publications. This market, where libraries purchase scientific journals on behalf of academics (who need to access a broad range of journals), is characterized by a number of barriers to entry and/or expansion.

[1] "Study on the economic and technical evolution of the scientific publication markets in Europe", published by DG-Research of the European Commission in 2006.

- There are economies of scale (first copy costs) and scope in the production of scientific journals;
- Scientific journals are platforms on which authors and readers meet and, as a consequence, switching from one journal to the other requires both sides to coordinate;
- Researchers are happy to give their output for free to publishers because they are rewarded through scientific reputation, which lowers price sensitivity, especially for 'unavoidable' journals;
- Pricing practices favour large publishers at the expense of small ones (and entrants).

However, to some extent, large scientific associations can overcome these barriers, as their publications and their membership are valuable assets; they are also potentially *rich in scientific publishing.* They can reach economies of scale and can contact authors and readers, as well as referees and editors.

Most scientific associations have been quite conservative in launching new journals. This article discusses the objectives of scientific associations and to what extent they are compatible with launching new journals and competing to help decrease prices.

The remainder of this article is structured as follows. Section 2 presents in more detail our empirical findings on the pricing of scientific publications. Section 3 reviews the potential, natural and strategic barriers to entry. Section 4 discusses the objectives of not-for-profit publishers and whether they help innovate and drive prices down. Section 5 concludes that there is a need to promote pro-competitive, i.e. much more flexible, pricing strategies, scrutinize mergers, and refund VAT to research institutions.

2. Key empirical finding: Large price differences between for-profit and not-for-profit publishers

Section 2 presents in more detail our empirical findings on the pricing of scientific publications. They were first published in the *Study on the economic and technical evolution of the scientific publication markets in Europe* and further developed and published in the *Journal of the European Economic Association*.

The database we use comprises more than 2600 journals, of 22 large domains (as defined by the Journal Citations Report, JCR from ISI) of sciences and social sciences.[2] The prices and publisher information are provided by SWETS, and the 2003 edition of the JCR is used for the number of citations. Using the data on publishers, we classified the journals into three categories: (a) FP journals published by for-profit publishers, (b) NFP journals managed by not-for-profit publishers (e.g. scientific societies, University presses), and (c) NFPP journals published and distributed by FP firms on account of scientific societies. To do this we looked for the publishers on the World Wide Web.

The main results are:[3]

- There are large differences of prices across fields.
- FP journals are, on average, 3 times more expensive than NFP journals for given field, age and citation count. So, if the price of an NFP journal is normalized to €100, on average an FP journal would be sold for €300.
- Similarly, NFPP journals are, on average, 2 times more expensive than NFP journals for given field, age and citation count.

[2]The 22 domains studied are: 1. Biochemistry and Molecular Biology; 2. Cell biology; 3. Chemistry, multidisciplinary; 4. Chemistry, physical; 5. Clinical neurology; 6. Economics; 7. Education and educational research; 8. Engineering, chemical; 9. Engineering, electrical and electronic; 10. Law; 11. Materials science, multidisciplinary; 12. Mathematics applied; 13. Mathematics; 14. Neuroscience; 15. Pharmacy; 16. Physics, applied; 17. Physics, multidisciplinary; 18. Plant science; 19. Psychology, clinical; 20. Psychology, multidisciplinary; 21. Sociology and 22. Surgery.

[3]See Dewatripont et al. [3] for technical details.

- NFP journals are more cited.
- Prices of journals increase with the number of citations, i.e. journals with higher number of citations (usually more widely circulated journals) are more expensive.
- FP journals are more expensive per citation than NFP journals.
- Journals are more expensive in more concentrated domains, i.e. prices are higher in domains in which the largest publishers have larger market shares.

These results are consistent with impediments to effective competition. Indeed, on more competitive markets, on which entry and expansion are likely, large price differences (up to 200% more expensive) for a given quality, will lead consumers to switch to cheaper goods expanding market shares of new entrants or existing less expensive producers. Market shares of expensive goods will decrease accordingly, and competition will drive their prices down, or force them to exit the market.

Consequently, long-lasting price differences are possible because the market for scientific publications is unusual: it is intermediated, subsidized and characterized by natural and strategic barriers to entry and expansion. Hence, some FP publishers are (and are likely to remain) '*rich in scientific publishing*' since they enjoy market power and charge high prices.

3. An unusual market, where large price differences can be sustained

The large price differences identified in the previous section can be sustained, because the market for scientific publications is unusual. Indeed, many direct and indirect market participants are publicly subsidized, there is some intermediation in the sense that readers usually do not pay directly what they consume, and entry and expansion on the market is impeded by natural and strategic barriers.

3.1. Public subsidies in an intermediated market

Large price differences and the lack of switching behaviour on the side of consumers can be partly explained by intermediation and public subsidies. Indeed, research output as well as referees and other market participants are to a large extent subsidized by public funds. Researchers are happy to give their output for free to publishers, because they are rewarded otherwise, e.g. through promotions, outside offers and prestige. Moreover, scientific reputation reduces price sensitivity, because good journals are 'unavoidable', both for authors and readers.

Besides, researchers do not pay for scientific publications themselves, libraries do. This intermediation further reduces price sensitivity on the side of consumers, as the ones who assess the value of a journal as input for their work, usually have a very poor view of their cost.

Other impediments to switching by consumers to take advantage of the large price differences are usually referred to as barriers to entry. Economists generally distinguish natural and strategic barriers to entry or expansion. The former are linked to characteristics of the products, i.e. are imposed to market participants, whereas the latter follow from choices and actions of market participants.

3.2. Potential natural barriers

There are many ways in which powerful incumbent publishers are favoured: both in terms of publication costs and in terms of coordination of the sides of the market. An important part of these comparative advantages are not related to an anti-competitive behaviour of the actors, but to the nature of scientific publishing.

First, the production of scientific journals involves a good deal of economies of scale and scope. Producing the first copy of an issue is much more expensive than producing the next ones. There are also economies of scope in the production of scientific journals, as some costs can be shared by journals across different fields (e.g. costs related to distribution of journals, contacts with customers and with suppliers, as well as the development of electronic platforms for journals). The presence of high first copy costs and economies of scope lead publishers with a high reputation and a large installed readership to enjoy lower average costs per copy.

Second, the nature of scientific publications makes it difficult for customers to switch from a journal to another (e.g., a cheaper or newly launched journal). As argued by a large body of economic literature[4] an important feature of scientific publications is that they are platforms on which (at least) two sides of the market meet: authors on the one side and readers on the other (one could add editors, referees, and librarians). Bergstrom [1] describes a situation in which the scientific community is not satisfied with an existing journal (for instance because prices are considered to be too high) and wants to switch to a newly launched journal. It requires coordinating authors, editors, referees, readers and librarians at the same time, which may prove difficult. The good reputation of a journal for a subset of these actors (editors and referees for instance) may be sufficient to prevent the others from switching.

Hence another dimension of being '*rich in scientific publishing*' is the installed base of authors and readers. In this sense, scientific associations are rich. Moreover, as discussed by Edlin and Rubinfeld [4], if they want to launch new journals, scientific associations may overcome coordination problems, since they can contact authors, readers, referees and editors and may encourage them to lobby librarians.

3.3. Potential strategic barriers

Whether large publishers use strategies to deter entry or reduce expansion of rivals on the market (or even to exclude them from the market) through means other than competition on the merits is an open question. Many papers (among which, Edlin and Rubinfeld [4]) have addressed this question and relying on their arguments, mergers have, recently, been denied by competition authorities. I next present briefly some issues pushed forward by the tenants of an action by competition authorities.

The main competitive concerns relate to a pricing strategy that is usually referred to as the 'Big Deal', or its more recent versions. Under this type of deal, buyers have the opportunity to access a substantial part of the electronic collection of a publisher for a price that relates to the budget spent on printed journals. Most versions of this deal presuppose that libraries commit to spending the amount they spent in the past on the publisher's collection, sometimes increased by several percents above inflation. Librarians complain about cancellations being unattractive and have the impression to be locked-in when they negotiate these deals.

Under this type of deal, electronic journals are bundled, favouring publishers with large portfolios, who can leverage the success of some key journals in other disciplines (see e.g. Jeon and Menicucci [5]). This further reinforces economies of scope.

Similarly, when fixed costs are high, producers need to recoup them. When possible, a form of price discrimination enables the seller to, simultaneously, charge a high price to some users (those who benefit the most from the journal) and a lower price to others (those who's benefit is lower). This helps recouping fixed costs. In the 'Big Deal', publishers use past spending of libraries as well as on-line usage by academics through downloads, to charge different prices for the same access. The ability to do so depends on whether the publisher can collect and use this information. Big incumbent players are better

[4]See e.g. Bergstrom [1], Edlin and Rubinfeld [4], McCabe and Snyder [7], Jeon and Rochet [6] or Dewatripont et al. [2].

equipped to do so than a potential entrant, since entrants have less information about usage and past spending than established publishers have.

Finally, large long term contracts tie-up significant parts of the libraries' budgets, leaving them less able to react quickly to new offerings in the market, or to price differences, which again favours incumbent publishers at the expense of entrants.

Using strategies such as bundling and price discrimination may also lead to be '*rich in scientific publishing*', but it requires large portfolios and information on past spending and downloads.

4. Scientific associations have launched few journals, why?

Scientific associations can often overcome natural and strategic barriers to entry due to their contacts with authors, readers, referees, and indirectly with libraries as well as their widely diffused journals that attract most citations. This makes them potentially '*rich in scientific publishing*' and places them in favourable position to expand.

Hence, some decision makers emphasise the role scientific associations could play in launching new journals to accompany the growth of scientific production, discipline price setting and help funding bodies to get better value for money. However, this requires:

- Consumers to switch for cheaper goods, reducing market shares of more expansive publishers and drive their prices down.
- That launching new journals to follow the growth of scientific production and drive prices of competitors down fits in the objectives of scientific associations.

With few notable exceptions, scientific associations have been more conservative than for-profit publishers in launching new journals. Indeed, the statistics presented in Dewatripont et al. [2] show that, out of 22 representative domains:

- The 'Chemistry, physical' domain is the only exception where for-profit journals are, on average, older than not-for-profit journals;
- The 'Materials science, multidisciplinary' and 'Physics, multidisciplinary' domains are also exceptions to the extent that for-profit journals are, on average, as old as not-for-profit journals;
- But in domains such as 'Clinical neurology', 'Law', 'Psychology, multidisciplinary' and 'Surgery' not-for-profit journals are, on average, twice as old as for-profit journals.

With journals published by for-profit publishers being on average one third more recent than journals published by not-for-profit publishers, the field of economics fits the general framework. In economics, scientific associations hold widely cited journals but few have launched new (subfield) journals in the last decades, whereas for-profit publishers have launched many.

It has been underlined in this article that scientific associations could discipline price setting and help scientists (as well as funding bodies) to get better value for money by launching new journals, but not all associations view this as being in line with their main objectives. In the third chapter of my PhD dissertation [8], I study the objectives of scientific associations in their relation to the creation of new journals and show that some associations do not have the best incentives to launch new journals and accompany the growth of scientific production.

If scientific associations try to maximise the circulation of their core journal, they may fear that their historical journal will suffer from the competition of the newly launched journal, when the practice of bundling journals is prohibited. Hence they may not be keen to launch unbundled journals. However, the

same associations favour the introduction of bundled journals because this helps increasing the circulation of their existing journals. When they do so, Walckiers [8] finds that the journals launched tend to replicate FP journals, which may be good to the extent that it puts pressure on price setting by FP publishers. On the other hand, however, it does not help to develop new subfields and diversify the supply of scientific journals.

5. Policy implications

We concluded our "Study on the economic and technical evolution of the scientific publication markets in Europe", for the DG-Research of the European Commission [2] by three groups of recommendations. Let me briefly summarise them:

- **Access issues:** Guarantee public access to publicly-funded research results through *open repositories* shortly after publication; allow for experimentation with new business models, including *open-access journals*.
- **Barriers to entry and concentration:** Promote pro-competitive, i.e. much more flexible, pricing strategies, scrutinize mergers, refund VAT to research institutions.
- **Subjects to be further investigated are:** Copyright provisions and economic sustainability of alternative business models.

These recommendations remain valid, as the main regulations surrounding scientific publishing have not been significantly modified since then. The open-access-through-repositories recommendation is the most popular one. A petition encouraging the adoption of the first recommendation *as a matter of urgency* has attracted more than 20,000 signatures, within a few weeks.[5]

However, its implementation may come along with some difficulties. First, the cost of open repositories, both to organize and preserve current and future access is not well established. Moreover, it involves additional work for academics to the extent that they have to upload their papers on one or several open repositories. This may reduce their support to open repositories, if they perceive them as an additional administrative burden. Second, open access should help reduce reliance on scientific publications in the medium run. As a consequence, it is expected that price sensitivity of libraries will increase, but the desired effects on pricing of scientific publications may not accrue in a timely manner. Third, as argued by ALPSP "if policies are adopted which risk damaging journals or even putting them out of business, it is inevitable that small publishers who operate on modest margins will be damaged first; many if not all of these are non-profit publishers".[6] These publishers help providing a wide range of journals at relatively low costs.

This article has shown that, with some notable exceptions, scientific associations have been timid in launching new journals to accompany the growth of scientific production. I have argued that this may, partly, be the consequence of their objectives that may conflict with launching journals. However, I had previously noted that there were a range of natural and strategic barriers to entry that do not help scientific associations to expand their range of publications. As a consequence, the recommendations of Dewatripont et al. [2] on barriers to entry and concentration are most important, in particular, those that relate to mergers and flexible pricing strategies.

[5] Petition for guaranteed public access to publicly-funded research results, available at: http://www.ec-petition.eu/.

[6] ALPSP's response to the "Study on the economic and technical evolution of the scientific publication markets in Europe": http://www.alpsp.org/ForceDownload.asp?id=121.

References

[1] T.C. Bergstrom, Free labor for costly journals?, *Journal of Economic Perspectives* **15**(4) (2001), 183–198.
[2] M. Dewatripont, V. Ginsburgh, P. Legros, A. Walckiers, J.-P. Devroey, M. Dujardin, F. Vandooren, P. Dubois, J. Foncel, M. Ivaldi and M.-D. Heusse, Study on the economic and technical evolution of the scientific publication markets in Europe, European Commission, Brussels, 2006.
[3] M. Dewatripont, V. Ginsburgh, P. Legros and A. Walckiers, Pricing of scientific journal and market power, *Journal of the European Economic Association, Papers and Proceedings* **5**(2/3) (2007), 400–410.
[4] A.S. Edlin and D.L. Rubinfeld, Exclusion or efficient pricing? The "Big Deal" pricing of academic journals, *ABA: Antitrust LJ* **72**(1) (2004), 128–159.
[5] D.-S. Jeon and D. Menicucci, Bundling electronic journals and competition among publishers, *Journal of the European Economic Association* **4** (2006), 1038–1083.
[6] D.-S. Jeon and J.-C. Rochet, The pricing of academic journals: A two-sided market approach, IDEI working paper, 2007.
[7] M.J. McCabe and C.M. Snyder, Open-access and academic journal quality, *American Economic Review, Papers and Proceedings* **95**(2) (2005), 453–458.
[8] A. Walckiers, Three essays on the economics of science and higher education, PhD dissertation, Université Libre de Bruxelles, 2008.

Information Services & Use 28 (2008) 121–122
DOI 10.3233/ISU-2008-0560
IOS Press

About the future of libraries

Wolfram Neubauer
ETH Libraries, Zürich, Switzerland

Figure 1 shows a simplified picture of the information infrastructure of my library. It is by no means something special: I suppose any university library, or at least many of them can show similar pictures. What is special is that 80% of it did not exist 10 years ago. What the library usually is identified with is this little part in the lower left side of the chart: the catalogue, representing mainly the printed collections of the library. Do not think these are outdated, now or soon. But indeed, the usage of the printed collection is changing, if not decreasing so much. And however printed books are still published as manifold as ever before, it is clear that in science the role of print is fading.

So what did libraries do? They developed digital libraries in which they make available vast amounts of primary and secondary information – an average university library nowadays offers its patrons hundreds of databases and thousands of electronic journals, e-books being on their way up. Library users now have easier access to more information than ever before.

Libraries have been developing, and still are developing tools to use that information in teaching and research. In Amsterdam we built a connection between the contents of the library and the virtual learning environment, to help students getting easy access to the contents they need.

Today we are building a tool for e-readers, to help teachers to facilitate their students with electronic information, even if that information is not yet electronically available. Maybe you think that all content is coming available in Open Access and that you do not need a library to get it. Well, Open Access still has a long way to go, but even in an OA paradise libraries will continue to organise, to connect, to disclose information, and to facilitate and to accommodate information users. And, for the time being, to integrate electronic and not electronic sources in one information environment.

Moreover and maybe even more important, libraries have become centres of production of digital information themselves. Look at the right side of the figure: you find there a long list of products in which the library is involved as a producer. If there are so many Institutional Repositories nowadays, and if they slowly begin to be successful, it is because of the libraries that have built them and that organise the workflows to deposit research outputs of their universities in their institutional research archives. Sometimes we do this with a University Press as partner.

In Amsterdam we developed a production pattern for e-dissertations together with Amsterdam University Press that can deliver print versions of any dissertation on a Printing on Demand basis anywhere in the world. Together with AUP we produce dictionaries, journals, e-books. As a library we make available image databases from our special collections. As a nation wide consortium we developed DARE, the Dutch Academic research REpositories. Today we work on so-called collaboratories to support research (SHARE) and we are developing production tools for so-called enriched or enhanced publications. At many universities the library is made responsible for the registration of research outputs of the university, because they know better than anyone else how to handle metadata.

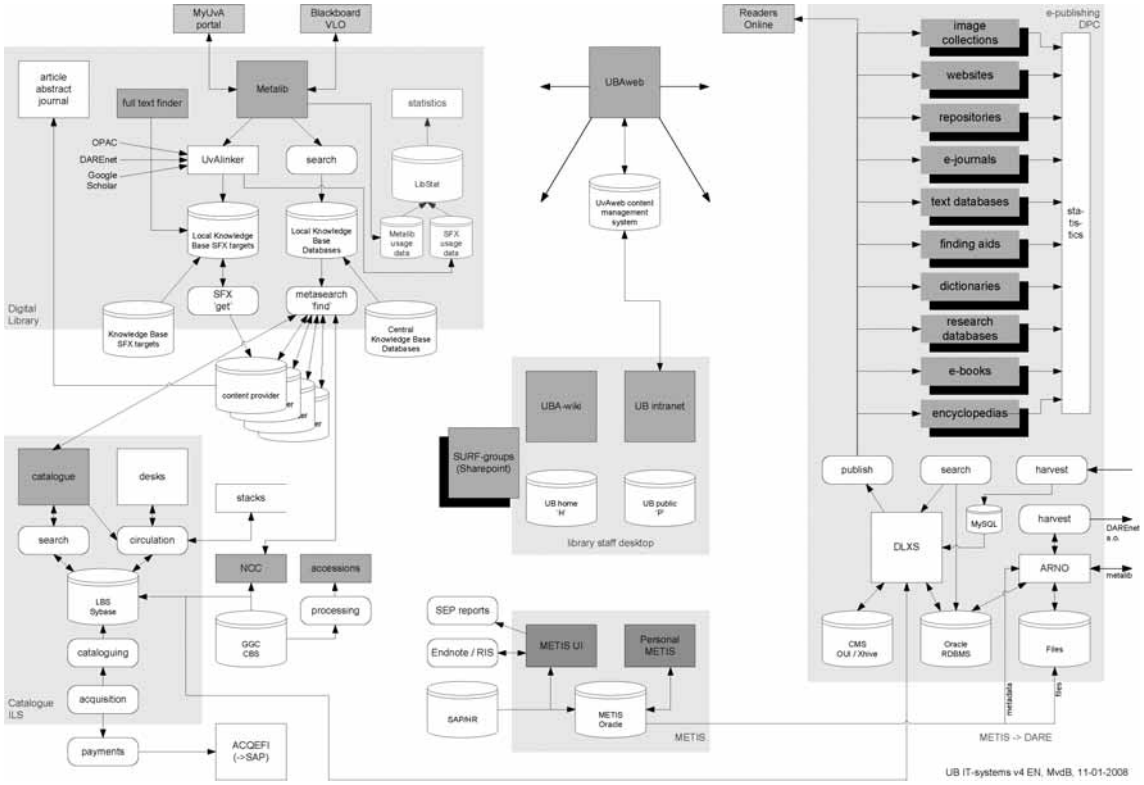

Fig. 1. Information infrastructure university library, University of Amsterdam.

So far from becoming obsolete, libraries are becoming ever more important partners for research and teaching. We are not going to be swallowed by Open Access: we are swallowing Open Access ourselves!!

Information Services & Use 28 (2008) 123–125
DOI 10.3233/ISU-2008-0564
IOS Press

A comparison of academic information portals

Tamara Pianos *
German National Library of Economics, Kiel, Germany

1. Introduction: Information portals in Germany

In Germany, there are a great number of portals. The academic information portal vascoda covers about 40 individual subject portals in many different fields. A study was conducted in 2007 to find out if and how these portals serve the needs of their target groups. In a first step, researchers were asked about their specific needs and wishes.[1] In addition, the contents and search functions of several portals were evaluated. The major findings of this study and previous studies will be presented in this paper.

Vascoda is an interdisciplinary academic information portal. It started its services in 2003, and it was first based on meta-search-technology. Since the relaunch in November 2007, new search engine technology (FAST) searches in over 100 million datasets from over 100 databases of different providers. The search engine focuses exclusively on academic and scholarly contents. The German Research Foundation (DFG) initialized the creation of Virtual Libraries (ViFa) in 1998. Since then more than 30 Virtual Libraries were created at the special collections libraries and other institutions. The Virtual Libraries and other subject portals funded by the Federal Ministry of Education and Research can be accessed through vascoda.

In 2008 there are about 40 subject portals for many different subjects from Anglo-American Culture to Veterinary Medicine. The subject portals offer access to all kinds of information from specific databases to library catalogues and from internet resources to electronic journals. Most of the virtual libraries still offer a meta-search of different databases, they include different document supply services and a couple of other services.

Typical Contents of Subject Portals would be databases, publications and different services. Depending on the subjects the databases will be factual databases, library catalogues, bibliographical databases, full-text databases, journal databases, reference databases and subject directories. Different sorts of publications such as digitized content, full-text repositories, digital encyclopaedias, handbooks, journals and so on can be found. Mailinglists, event calendars and weblogs are also included in many subject portals. Many portals offer different services like: helpdesks, online reference, lists of new publications, newsletter, RSS-Feeds and more.

*Address for correspondence: Dr. Tamara Pianos, Organizational and Human Resources Development vascoda-Projects, ZBW – German National Library of Economics, Leibniz Information Centre for Economics, Düsternbrooker Weg 120, 24105 Kiel, Germany. Tel.: +49 431 8814-365; Fax: +49 431 8814-520; E-mail: t.pianos@zbw.eu; URL: http://www.zbw.eu.

[1] A small number of librarians and other information specialists was also asked about their preferences in information portals. The focus of this paper will be on the findings from the interviews of academics.

Apart from the subject portals organized in vascoda, there are also a couple of other subject portals in Germany created by learned societies or other special groups. Many other countries also offer a number of gateways. In many cases there are international co-operations on a subject level.[2] In 2007, a study was conducted by Heinold, Spiller & Partners[3] to find out if and how the subject portals meet the needs of their target groups.

2. What do academics want?

The answers to this question sound very simple sometimes even banal, yet the realization is often quite difficult. A number of studies have already tried to find out different aspects of what scientists and academics want.[4]

First and foremost: Keep it simple and straightforward. Academics want a straight forward and self-explanatory navigation, preferably one search field plus an optional expert search. If there are too many options on the homepage, the users might be confused and leave right away. Academics want a comprehensive search covering all relevant information, a complete set of metadata plus abstract. They want complex and comprehensive search options, for example they want to be able to choose individual databases. They want one result set – not several sets sorted by database. This particular wish was mentioned very often, yet it is quite difficult to fulfil as long as many portals still work with meta-searches instead of search-engine technology. The same is true for sorting and drill down options for result sets or for full-text-search-options. The problem with these requirements is not only the availability of the right technological solution but also a question of rights and licences.

Many academics want the information "now or never and for free – or at least easy and not too expensive". Open access seems the best choice, but if the desired article etc. cannot be accessed directly

[2]The other portals are not part of this presentation. An overview over different portals in different countries by Eleanor Frierson can be found here: http://www.science.gov/communications/library/icstiJune72006NationalSciencePortals.ppt.

[3]Website of the firm: http://www.hspartner.de/jsp292/contentPath=hauptmenue/englisch.

[4]Studie der Firma Heinold, Spiller & Partner zum Them "Virtuelle Fachbibliotheken im System der überregionalen Literatur- und Informationsversorgung: Studie zu Angebot und Nutzung der Virtuellen Fachbibliotheken", November 2007, not published. Other Papers and Studies in German: Weiterentwicklung der überregionalen Literaturversorgung – Memorandum 1998, http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/memo.pdf. Nutzung elektronischer Fachinformation, -publikation und -kommunikation in der Hochschulausbildung, Barrieren und Potenziale in der innovativen Mediennutzung im Lernalltag der Hochschulen Stefi Sozialforschungsstelle Dortmund, BMBF 2001, Kurzbericht und Endbericht unter: http://www.stefi.de/. Zukunft der wissenschaftlichen und technischen Information (Strategiekonzept), Arthur D. Little (A.D.L.), BMBF 2002, http://www.bmbf.de/pub/zukunft_der_wti_in_deutschland.pdf. Nutzungsanalyse des Systems der überregionalen Literatur- und Informationsversorgung: Teil I: Informationsverhalten und Informationsbedarf der Wissenschaft, Teil II: Zur Nutzung der SSG-Bibliotheken, Poll, Universitäts- und Landesbibliothek Münster, infas 2003, http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/ssg_bericht_teil_1.pdf; http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/ssg_bericht_teil_2.pdf. Das DFG-System der überregionalen Sammelschwerpunkte im Wandel; Anhang "Weiterentwicklung des SSG-Plans", DFG 2004, http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/strategiepapier_ueberreg_lit_versorgung.pdf; http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/strategiepapier_ueberreg_lit_versorgung_tab.pdf. Strategische Erfolgsfaktoren von wissenschaftlichen Portalen: Content-StudieZB MED, Mummert 2004, http://www.dl-forum.de/dateien/Endbericht_Content-Studie_DL-Forum.pdf. Evaluation von vascoda.de aus BenutzersichtErgebnisse der Nutzerbefragung 2005Nutzer-befragung, IwFB ULB Münster 2005, http://www.dl-forum.de/dateien/Evaluation_vascoda_Ergebnisse_Befragung_2005.pdf. Evaluation von vascoda.de aus BenutzersichtErgebnisse der Fokusgruppenbefragung 2005Fokus-gruppenIwFB+ULB, Münster 2006, http://www.dl-forum.de/dateien/vascoda_Ergebnisse_2005_Fokusgruppen.pdf. Richtlinien zur überregionalen Literaturversorgung der Sondersammelgebiete und der Virtuellen Fachbibliotheken, DFG 2006, http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/richtlinien_lit_versorgung_ssg_0607.pdf. Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015, DFG 2006, http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/positionspapier.pdf.

and for free, access through document delivery must be fast, easy and reasonably priced. Most people are not willing to pay for metadata, but in many databases one has to pay for the display of the metadata. Hardly surprisingly, many academics hesitate when it comes to registrations. They want no registration at all – or only if one registration will suffice for many different services. Some quotes nicely illustrate the problem many people share today: "The problem is that I already have four pages full of passwords and logins in my diary. I think at least twice before I register for any other service ..." or "I would not want to register for individual portals but if I could register once and could use all services in Germany – like I can use my library card for library loans in Germany – that would be useful".

3. Why is it difficult to build the perfect portal?

Comprehensiveness is desired but not easy to define and also difficult to realize. Probably every academic has a different view of what the perfect portal might look like depending on personal preferences and focus of research. A search including a vast amount of different databases cannot be tailor-made for an individual researcher. If complex search-options are needed there is a need for basic or better advanced search skills. Comprehensiveness is wanted by most people but when asked to define the aspects it is hard to grasp. Mostly important authors and journals are mentioned as indicators of comprehensiveness.

There are technical and licence restrictions that make it hard to realize the perfect portal. Licensed content cannot easily be covered. Licence situations vary from federal state to federal state, on-campus vs. off-campus etc. thus making it very difficult to understand why certain content can be found in one place but not in another – even if one uses the same portal. National licences and authentication tools will hopefully bring improvements in the near future.

4. Examples of subject portals

The subject portals covered by vascoda present themselves in many different ways.[5] Many subject portals started out as portals that support navigation to different services. Up to now, a great number of subject portals still has many links on the homepage. MedPilot (www.medpilot.de) is the portal that has radically reduced the homepage to the basics. It has one search field and hardly anything else. In this case, this seems to be a very successful strategy.

5. Success factors

For a subject portal to be successful most of the following factors need to be fulfilled: one search covers most of the relevant content. The search-field needs a prominent position on the homepage. A common desire that is not often easily fulfilled is: metadata (and if at all possible abstracts) can be viewed free of charge. The content should be easily available – either through open access or through the integration of document supply.

It seems also very important to enhance information competence through tutorials, guided tours etc. Not every subject portal can be used as easily as Google, yet many searchers are very much used to the way Google works. If subject portals offer more complex services, there is a need for training and instruction.

[5] All the individual subject portals are listed under www.vascoda.de.

Information Services & Use 28 (2008) 127–128
DOI 10.3233/ISU-2008-0562
IOS Press

Comparing the scientific impact of conference and journal publications in computer science

Erhard Rahm
University of Leipzig, Leipzig, Germany
E-mail: rahm@informatik.uni-leipzig.de

The impact of scientific publications is often estimated by the number of citations they receive, i.e. how frequently they are referenced by other publications. Since publications have associated authors, originating institutions and publication venues (e.g. journals, conference proceedings) citations have also been used to compare their scientific impact. For instance, one commonly considered indicator of the quality of a journal is its *impact factor* [1]. The impact factors are published yearly by Thomson ISI in the Journal Citation Report (JCR) by counting the citations from articles of thousands of journals.

However, research results in computer science are often published in high-quality conferences which are not covered by the JCR citation databases [2]. Other commercial citation data sources such as Elsevier Scopus also focus on journals and contain comparatively few conference publications. Hence these data sources cover only a fraction of quality scientific publications in computer science. Furthermore, they miss many citations even for journal articles since all references to them are not captured which originate from conference papers or other papers not included in the publication database.

Several recent system developments capture citation numbers for both journal and conference publications especially in computer science, e.g. Citeseer, the ACM Digital Library, Microsoft Libra (Libra) and Google Scholar (GS). For example, Libra holds more than 900,000 computer science publications and more than 3.5 million citations to them as of December 2007. As shown in Table 1, the majority of papers appeared in conferences and workshops, not in journals. Furthermore, the total number of citations is higher for conferences and workshops than for journals. While there are many workshops and conferences with comparatively little scientific impact the top-cited conferences are highly significant and need to be considered for a meaningful citation analysis in computer science. For example, in the Libra dataset the average number of citations per paper is similar for the 100 most cited conference series than for the 100 most cited journals. These 200 venues account for 78% of all citations.

In [4] we used cleaned citation data from GS for an in-depth citation analysis for database research, a subfield of computer science research. We analyzed all publications over a period of 10 years (1994–2003) which appeared in top database conferences and top database journals. It turned out that the two top conferences (ACM Sigmod, VLDB) not only publish many more papers than the top journals (ACM TODS, VLDB Journal) but that they receive many more citations in total and per paper. The original study used GS data from August 2005. We recently confirmed the findings with GS citation data from December 2007.

Table 1
Journal vs. conference papers and citations in computer science

	#Venues	#Papers (all)	#Cited (all)	#Papers (top 100 venues)	#Cited (top 100 venues)	#Citations per paper (top 100 venues)
Journals	471	321,000	1,655,000	190,000	1,434,000	7.5
Conference/workshop series	2,297	585,000	1,752,000	167,000	1,216,000	7.3

Source: MS Libra.

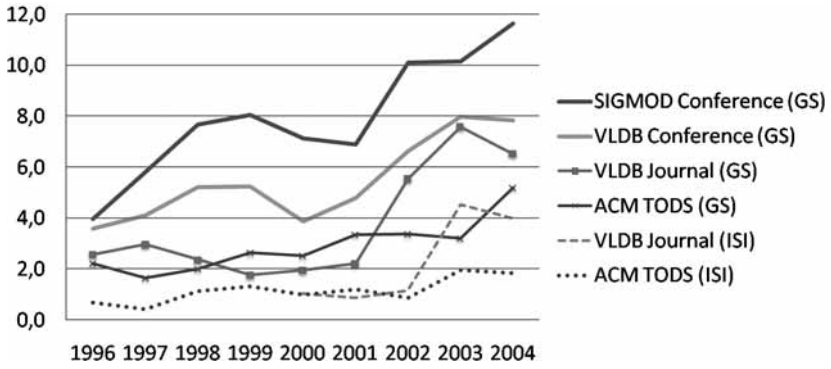

Fig. 1. Impact factors for conferences vs. journals based on data from Google Scholar (GS) and ISI Web of Science.

We also determined impact factors for the considered venues using cleaned GS citation data. The impact factor of a venue (journal, conference series) for year X, IF(X), is defined as the average number of citations in year X publications for a article published in the considered venue in the two preceding years X-1 and X-2. Figure 1 shows the impact factors from 1996 to 2004 for the two conferences and two journals based on GS data from 2007. For comparison, we also show the impact factors for the two journals as recorded in the ISI Web of Science JCR (dashed lines). The results show that for all years the impact factors for the two conferences are significantly higher than for the journals. Furthermore, the GS-based impact factors are much higher than the "official" ones from ISI Web of Science. This confirms that the latter data source misses many citations even for journal articles.

A drawback of data sources such as MS Libra and especially GS is that they incur a significant amount of postprocessing for data cleaning, e.g., to match citations, removing duplicate entries etc. These tasks were supported by several new data integration tools and aligning the data with reference bibliographies such as DBLP [3,5].

References

[1] M. Amin and M. Mabe, Impact factors: Use and abuse, *Perspectives in Publishing* **1** (2000), 1–6.
[2] H.F. Moed and M.S. Visser, Developing bibliometric indicators of research performance in computer science: An exploratory study, Technical Report, CWTS, University of Leiden, 2007.
[3] E. Rahm, A. Thor et al., iFuice – information fusion utilizing instance correspondences and peer mappings, in: *Proc. 8th WebDB*, 2005, pp. 7–12.
[4] E. Rahm and A. Thor, Citation analysis of database publications, *ACM Sigmod Record* **34**(4) (2005), 48–53.
[5] A. Thor and E. Rahm, MOMA – A mapping-based object matching system, in: *Proc. CIDR (Conf. on Innovative Database Research)*, 2007, pp. 247–258.

Information Services & Use 28 (2008) 129–132
DOI 10.3233/ISU-2008-0575
IOS Press

APE 2008 Academic Publishing in Europe, Quality and Publishing

January 22–23, 2008, Berlin

Jens Bammel
International Publishers Association

Good afternoon ladies and gentlemen. I assume I am the last thing that stands between all of you and lunch which is never a happy place to be. So I will try to be short and brief and talk to you about the Automated Content Access Protocol – what it is and what it isn't, because there are a lot of misunderstandings. Before I start I would like to have a quick raise of hands: how many of you have ever heard before today of ACAP or the Automated Content Access Protocol? Is there anybody who has never heard of it before? Okay. That is extremely useful. So I will have to give some basic information as well as more detailed information. I will be talking today about what it is actually, ACAP. More important is also, what it isn't because there are a lot of misunderstandings. Why we are creating this new kind of protocol and put it into the context of all the various standards, movements that we have from ONIX-PL, which we just heard about to International DLI Foundation etc. I will also talk about the next steps and then finally I will ask the publishers here to do some homework and I will give you some homework what I think what you should be doing after you have come home from this conference. Who you should speak to and what you should tell them. So let me start with what is ACAP? It is actually quite simple. It is an open business model neutral standard tool for the expression of permissions and other terms in machine-readable form. I cannot say these words slow enough because they all have a specific meaning and they also say something about what it isn't. When I speak to other people and ask them: "What have you heard about ACAP", they think: "Well, this is all about newspaper publishers, these old dinosaurs and their last attempt to lock away content and to prevent search engines from actually ruining their outdated 19th century business model. That is, of course, not true. First of all, and let me talk about these misconceptions, it is not just about search engine communication. I will be talking about the first use case. The most publicised use case that we have, which is, in fact, improving the communication between websites and search engines. But that is just a first use case. It is a first starting point. It is the place where we started because that is where we have go interest. That is where we got attention and funding to get this project moving. The second use case which is currently under way actually is about communicating between the British Library and UK websites where the British Library has the right to collect data for legal deposit but the law allows it only a very very restricted use of this harvested content and ACAP will make it much easier for British Library to get much broader and more generous use out of the website owners. So that is the second one that is already being worked on and that will be completed hopefully in the next few months. There are further use cases which I will be talking about as well. It is not just about newspapers nor is it just about publishers. At the moment the consortium which governs and runs the ACAP protocol consists of the International Publishers Association, that is me.

It consists of the European Publishers Council which are basically big European media companies and the World Association of Newspapers. But that is just because we had to start somewhere. The membership includes the World Blind Union; it includes the Motion Pictures Association. There are many organisations that are also participating plus which is about photographers. And, most important, it is not about locking away content. Yes, again, like ONIX-PL we are expressing terms. But at the moment ACAP cannot even claim that what we are expressing are simple licensing terms. They are claims. And the claims can refer to anything. I think that one of the important things, and I feel very passionately about that, is, we should stop focusing either on the one hand on copyright and licensing terms and on the other hand on useful information that people need to act on in their daily lives. I actually believe that these two areas are very very closely linked and, just to give you one example, last week I was speaking with colleagues from the International Federation of Library Associations. And we discovered how much joint interest we have in accurate author information. And also information is on the one hand about identifying who has written what, making sure you have complete records. It is also about copyright and how long copyrights subsist because only if you know when somebody was born and, even more importantly when he died, do you know whether there is something in copyright or not. So legal terms and practical terms are actually often quite closely intermingled. It is also not a technical protection measure. Again: nothing. It is all claims and what you can do with it. This is quite different. So, what are we doing with it? Or, why are we doing with it. Ultimately, we have heard a lot about the semantic web. I heard the word "web 3.0" for the first time today I must admit. I will now use the word "web 4.0". Just so that I can claim I was the first one to use it. Let's see if we can hear "web 5.0" this afternoon. But, in the future we have to communicate much better through machines about all kinds of things about content. When we hear that the music world has just disbanded DRM or the DRM dogma at least in some area of its market that does not mean that they do not need information about content. Quite the opposite and that includes information about authorship, about rights ownership, about access terms or access claims. We need these communications whether we are librarians or lawyers and anybody else – researchers. The reason why we are getting involved is because we cannot, if we know that communication in an expression of rights of descriptive terms is so important, we cannot afford that it is owned by one person. We all have paid much too much already to Bill Gates to allow another person to create a proprietary rights expression language and then all of us having to pay that person some kind of fee. We also cannot all afford to create new tools again and again and again as new people come up with new ideas, new little applications which again you find useful. And we have to be aware that we don't get put into a situation where something is developed in another industry or another area and it simply doesn't fit to our needs. So it is important that this is developed and it has to be developed at a global scale because standards . . . , there is no such thing in national standards really anymore. And the final reason why we are doing it is actually because the publishing industry knows a lot about how it should be done and at this point I would like to pay tribute to two people – two kinds of people. One is the actual people who are working on standards development in publishing and, like Brian Green here, you will have noticed over those who know him for some time that his hair may have greyed over time and I am sure that it is not just the passing of time it is also really the kind of effort that has been put into developing standards in this world. And the second group I have to pay tribute to are the librarians again. Simply, because standards development has all been about standards up and down the distribution and value chain of publishing and the librarians bring a unique perspective and that is let's not just fix a short-term problem, let's actually create something with a vision that will help us for decades, if not for centuries. So this long-term perspective has made it possible that models have emerged in the publishing world which are actually applicable and useful around and for other content industries as well. This

brings me to ONIX and ACAP and ONIX-PL. ACAP is in a way taking a step back and says: we want to develop use cases. We want to develop a language in many different cases, in many use cases. But let's not make these use cases insular. Let's make sure there is a shared dictionary. Let's make sure if somebody is called an agent for example, everybody understands exactly what the word agent means. Or when the word "display" is used or something like that, everybody understands what the word actually means. So using the same kind of words, agreeing what they mean, having a compatible data model is incredibly important and will save all of you a lot of money. And that is what ACAP is trying to do. ONIX-PL is about publishers and libraries but there are so many other relationships which require similar information and where interoperability is a key issue. So from a publishing perspective ACAP enables those people who are now using ONIX-PL and other ONIX based systems to be assured that the investment they are doing there will actually be useful at a later time when other applications come up, when other use cases come up. They do not have to re-invest and re-invent everything. Where is ACAP now? We have talked about the search engines very briefly. In the past 12 months we have worked hardest and most successfully on the use case for search engines. So we have created a protocol which you can put into a specific place on your website called robots.txt which every search engine goes to first to find out more about what it can and cannot do on that website. And we have created a language which allows publishers to express in much more detail exactly what they believe and what they think these search engines should do. We all thing of only three search engines. But we should not forget there are hundreds of them out there. And behind each crawler there is somebody who is really interested in your content and wants to use it and display it and show it to somebody else. So there is a real business opportunity there. The current structure is going to be extended for another six months. We are just working in a way on a string to get the use case for the British Library finished, to get the use case for the search engines actually fully finished that includes tools for example for authenticated search engines making sure that you know, when somebody is knocking on your door, that he is actually the one you want to let in. But there are new use cases which we are about to embark on. One will certainly involve syndication. Another one will involve compliance. Yes, the newspaper publishers are very keen on finding all the places where their content is being displayed and finding out whether these people are licensed or not. But non-publishing and public interest projects are also raising their head. I am particularly delighted of the beginning dialogue with creative comments. And it was interesting to note how much we actually have in terms of our whole approach. How much we have in common with creative comments. There are few technical things and one or two political issues, where we disagree. But overall the idea of expressing it and making it simple for people to know what they can and cannot do. And if they cannot do something straight away, to make it easy for them to find out where and how they can get the rights they need, is a very useful tool and everybody wants to work on that. When we get the new stakeholders on board we will have to change our governments. IPA, EPC, World Association of Newspapers, we all agree that we are just one pawn among hundreds of pawns and that we have to get all of them on board and that we will take a role and have to give away the hold on the governments that we currently by shear necessity have. What about search engine support? This is the one question we always get while we are doing all these things. What happens to the top three search engines. Well, it is a fact; none of the three are formally participating in ACAP. It is also a fact that all are aware of the technical work. We have sent all our stuff to them. We had people speak about how they operate so we know that what we have done is technically feasible. We are also quite aware that in all the major search engines there were clearly signs of heavy internal debate. And individuals do can see that, while it is comfortable for them at the moment to simply be able to do what they please more or less, that, in the future, there is a value in more sophisticated communication tools. And the publishers who are participating in the pilot,

namely Elsevier and Wiley, and some of the newspaper publishers have actually found that it would be a very useful tool and it would be a good thing to be able to control your content while making the more valuable stuff behind your firewalls more visible to the public through search engines. The important thing about search engines is of course, at this point, that they have no real incentive to participate. They would have to make a big investment. They would have to change the way they behave in the way their crawlers behave. And why would they want to do to that. There are only two reasons why they would do it. Either because there is a stick or because there is a carrot. Either because they have to or because they will have some kind of advantage. And at the moment that is really what we would like to work on with you. So what is the question? This is all very interesting but why am I telling you this? Why should you be interested in this issue? What does this mean actually for you? And here is where I come to my homework for you. ACAP will become the flexible universal communications tool for all content users if you support it now. Critical mass is the important thing. And what does support now really means? What it means is that where ACAP or ACAP-compatible applications are available, start using them. It does not mean you have to change your policy. To be honest, most publishers I speak to don't even know that they have a website policy or that they could have a website policy. Tell your administrators about it. Actually, implementing ACAP in the so-called robots.txt file takes a medium-skilled website administrator about five minutes. It actually has no impact on the way search engines behave at the moment. But by adding the ACAP translated information that you have already on your website in the robots.txt file, by adding that in an ACAP format, you are actually making the point. And at some point you might want to change your policy. And that takes me to the next point. Actually, talk to your business strategists and ask them about website policy and start thinking about it. As I was saying, there are hundreds of search engines and crawlers coming. Some of them are from search engines, some of quite different people. Why are they coming to your site? Why are they crawling these things? How can you intensify that relationship? How can you make more of it? There is a lot of work that can be done there. And the short tail, or the long tail, sorry, starts after three. Because everybody says you know, we have deals with the big three search engines. But that is quite a long tail if it starts with the fourth crawler that knocks onto your door, your website door. So, the real question that you have to take home and ask yourself is: "Where do you need better content communications?". The publishers, the librarians, the blind, the people in creative comments. Where do we all need to speak better and let's develop use cases around that which actually fit the needs of that particular community. Thank you and "Guten Appetit".

Information Services & Use 28 (2008) 133–139
DOI 10.3233/ISU-2008-0569
IOS Press

APE 2008,
Round Table "University Presses and Books in the HSS"

Göttingen University Press: Publishing services in an Open Access environment

Margo Bargheer * and Birgit Schmidt **
Göttingen State and University Library, 37070 Göttingen, Germany
E-mails: {bargheer, bschmidt}@sub.uni-goettingen.de

1. Background

Over the past decade, the landscape of scholarly publishing in Germany (and other parts of continental Europe) has seen a new wave[1] of not-for-profit university presses coming into the market: Kassel University Press in 1997, Hamburg University Press in 2002, followed by the Universitätsverlag Karlsruhe, and Universitätsverlag Göttingen in 2003, Potsdam 2004, Ilmenau 2006 and several others. And still now new presses are in the course of formation such as Bamberg or Heidelberg. A common trait of these presses is their close relationship to academic institutions, especially libraries, and a strong commitment to Open Access publishing.

It was in the mid-1990s, when academic libraries embraced the new distribution modes for electronic information while facing ever-tightening economic challenges such as rising journal subscription prices. These circumstances lead research institutions to think on alternative value chains for scientific information, namely production, publication and dissemination. Certain fields such as physics easily adopted

*Margo Bargheer is a trained media scientist and head of the electronic publishing department of the Göttingen State and University Library including the "Universitätsverlag Göttingen". She coordinates several national Open Access projects and teaches on the subject of electronic publishing.

**Dr. Birgit Schmidt, MA LIS, has a background as mathematician and a training in library and information sciences. She has worked on long term archiving of digital objects, economic issues of internet-based publishing and currently works for EU-funded project DRIVER (Digital Repository infrastructure Vision for European Research) and a national Open Access project, both aiming to intensify the networking of digital repositories.

[1]The first German university presses emerged from activities of privileged printers in the 17th and 18th century usually granted with exclusive rights to work for the local university. Many of these former university presses still exist as scholarly and independent commercial publishers f.e. Vandenhoeck & Ruprecht, Herder or Ch. Beck. The "second wave" falls into the late mid-20th century and the expansion of the German academic system. The third wave came up with the emergence of electronic publishing and the internet.

digital repositories, and nowadays these repositories such as the arXiv[2] serve as central hubs in the dissemination of scholarly information. It is not long until we will see each research institution in Europe running a digital repository to collect and distribute the institution's intellectual output, either on an institutional or topic-oriented base.[3] These repositories typically hold preliminary scholarly publications such as reports, white papers, preprints or working papers as well as already elsewhere published materials like post prints of articles and books.

The new generation of university presses aims at filling the gap between professional publisher activities, doubtlessly adding value to publications and digital repositories mainly providing infrastructure. These presses therefore take up the challenging double role of a publisher economically pursuing its interest for scientific content and at the same time acting in the name of scientists and universities willing to disseminate their publications as freely as possible. Therefore, it is often the university press offering independent consulting on publishing options for scholars, taking into account intellectual property rights questions, financial constraints or discipline-specific challenges.

In December 2005, with the support of the university press, the University of Göttingen released a resolution[4] encouraging its scholars to use the full range of new opportunities in Open Access publishing: to distribute parallel versions of their publications via the institutional or thematic repositories and to make use of quality-controlled Open Access publishing channels such as journals or presses. Offering professional digital publishing in an Open Access mode, therefore, is a crucial driving force for the Göttingen University Press (Universitätsverlag Göttingen, GUP). This is combined with high-quality print-on-demand books to enable broad awareness and easy access as well as meeting traditional publishing and reading behaviour.

2. Description of Göttingen services

2.1. History

Since its founding in summer 2003, Göttingen University Press is run as a service of the Göttingen State and University Library. It was started as an additional service for the library-operated institutional repository of the university which started as early as 1995 to disseminate electronic theses online. The press is part of the Electronic Publishing department covering the fields of electronic theses online, the institutional repository for scholar's parallel publications, Open Access advocacy and in general offers a publication consulting service on behalf of the university. The institutional repository and its integration into the library catalogue serves as the electronic outlet of the press' publications, accompanied by its website for marketing and distribution.

2.2. Profile

The press offers its innovative publishing services primarily to scholars of the Georg-August-Universität Göttingen. Its publishing program therefore reflects the diversity of the university with

[2] http://arxiv.org.

[3] For an orientation in the landscape of digital repositories see the publications of the EU-funded project DRIVER (Digital Repositories Infrastructure Vision for European Research), http://www.driver-support.eu/en/studies.html. In fact, with regard to the upcoming repository networks like DRIVER it does no longer matter that much, where the actual deposit takes place: well-defined virtual collections can showcase specific disciplines and research topics.

[4] http://www.sub.uni-goettingen.de/ebene_2/pub/resolution.pdf.

13 faculties and several scientific centers covering sciences, life sciences, humanities, arts and social sciences. The press' publications originate from all scientific disciplines (except engineering) with a slight majority in Humanities and Social Sciences. The broad publishing scope requires different needs and challenges for the respective communities to be taken into account, be it discipline-specific layout styles, reading patterns or distribution channels. This diversity also influences the necessary quality control. If the press could only draw on their own personnel, quality control would have been either neglected or confined to those disciplines brought into the press by the staff's own scientific background. To guarantee a high-quality publication profile the Georg-August-Universität has send high-ranking delegates (usually active professors) from each faculty into the editorial board of the university press. The editorial board acts as a steering committee, additionally its members undertake discipline-specific peer review and recommend publications and serial or journal concepts to the board, which decides on the majority principle. However, an objection from a board's member is not simply overruled but thoroughly analysed and leads to editorial recommendations and advices for authors and editors.

Not every publication needs to pass this peer review process which indeed poses an extra workload on each member of the editorial board. The press publishes within two categories: the concept "Universitätsdrucke" is based on the assumption that the Georg-August-Universität Göttingen is a high-ranking scientific institution and therefore reaches a fair amount of quality, as a basis to be published without further qualitative review of the content. However the staff of the press thoroughly controls for formal quality such as typesetting, make-up, layout and image quality. Cover and preliminary matter are designed by the press' staff taking into account aesthetic and economic aspects as well as bibliographic necessities. The category "Universitätsverlag" is exclusively reserved for publications of high quality. The formal quality control lies within the press, the review process either rests with serial editors or is exerted by the faculties representatives. Cover design for "Universitätsverlag" publications receives extra attention by the press' staff. All other publication aspects remain similar, in particular author fees and retail prices of both categories are strictly based on production costs.

The press publishes monographs, anthologies, proceedings, scientific exhibition catalogues as well as serials and journals. Dissertation theses constitute less than a third of the publishing output. Although each publication project is developed individually on the author's or editor's needs, their budget and the publishing expectations of their target group, we observe a strong trend for the "hybrid" publication model that combines Open Access publishing with high-quality print-on-demand and initial print runs between 50–100 copies. The "long tail" of book sales is served by re-orders paid through print sales. The combination of a free online version and its reverberation in the "print" world (high-quality appearance, ISBN, distribution via book retail) brings together the best of both worlds and meets the present needs of scholars relying on books as a publication and reading asset. Some disciplines or research "generations" seem to pick up the opportunities offered by Open Access publishing more easily than others. Young academics at the dawn of their career are in general eager to have their results out and citable and are willing to finance the upfront costs of the publication without worrying about book sales. Established scholars are socialised with traditional offset prints and its large and therefore costly print runs. They are on the one hand used to raise funds for publishing but at the same time expect to finance the publishing project with book sales. There is a slight tendency that members of the "book sciences" such as history or archaeology are convinced of cannibalistic effects of the free online version on book sales. Sometimes these economical worries are underlined by a certain reluctance to expose the scientific results to a broad and therefore anonymous audience. Members of disciplines as mathematics or theoretical physics are more familiar with the preprint culture and communicate preliminary results as part of the scientific process. Therefore, scientists from these fields tend to presume positive marketing effects of the "full

range teaser", the freely available book on the internet. Our data collection is not deep enough yet to robustly support one of these hypotheses. However, there is some evidence that a free online version combined with carefully designed books at moderate retail prices does result in negative effects on book sales.[5]

The legal sciences program of Göttingen University Press deserves special attention. Germany has a long tradition of independent, commercial scholarly publishers. Many of them only reluctantly pick up electronic, let alone Open Access publishing. So far only a handful of German-based publishers have enrolled themselves on the SHERPA/RoMEO database that displays publisher copyright and self-archiving policies. None of the German law-specialised publisher discloses that he allows self-archiving in any form. Law scientists are used to pay for the publishing of their dissertations but usually do not need to worry about the costs of any following publication. However, the Göttingen law faculty has already established three series with the press that are all available Open Access and nevertheless achieve reasonable book sales.

Authors or editors are getting more and more convinced that the immediate world wide and long-term availability of the electronic version promotes their scientific careers and will even outlast the print version. Nowadays, the majority of all GUP publications are available Open Access. In the run of four years the share of full Open Access books has grown from 50% to more than 80% by convincing former reluctant authors and editors to publish their books now Open Access as well and by setting up a stricter publishing policy.[6] To make this policy more visible, Göttingen University Press managed to be listed as a "green publisher" in the SHERPA/RoMEO database.[7]

2.3. Author's rights

As the press is run by the university it works under the policy that it only asks for those rights needed for its purposes and leaves authors and editors as many rights as possible, on the contrary to the buy-out principles of commercial publishers. For the electronic publishing in the internet the press only requires a non-exclusive license which leaves authors and editors the right to publish elsewhere. In most cases the GUP publications are licensed with a Creative Commons Licence 2.0. Users are allowed to download, distribute and print in small numbers given that they mention the author and distribute the work unchanged. The license does not cover commercial use of the free online version and allows authors and editors to use the publisher's version wherever they need it. For the sake of the unambiguous existence of the printed book and as a protection of the press' (and therefore the university's) investment, the press requires an exclusive license for the print version but handles reprint requests as generously as possible.

2.4. Business model

The press' business model is based on a cost-recovery approach. Costs include prepress, production, distribution and marketing as well as a part of the net personnel costs. General overhead costs are ac-

[5]It has to be taken into account that the respective target groups for university presses of the "third wave" are rather small and confined to specialised scientific communities. We do not expect the described positive effect to happen for textbooks with a general approach and a large audience. The upfront investment for putting the material together and updating it on a regular base so far seems to require large amounts of sales at maximum prices combined with restrictive copyright arrangements.

[6]For more detail see the "Leitlinien des Publizierens", http://www.univerlag.uni-goettingen.de/pub/policy_uv_goettingen.pdf.

[7]http://www.sherpa.ac.uk/romeo.php?type=publisher&search=G%F6ttingen.

counted for but not calculated into retail prices and author's and editor's publication fees as the press almost mainly serves members of the university. To minimise capital lockup authors and editors contribute to the production process: with the assistance of the press they produce a PDF file ready for publication while the press takes care of cover design, preliminary matter, quality control (where applicable including peer review) and organises production, distribution and marketing. Authors and editors in exchange pay a share of the printing and the press' processing cost. This share is supplemented by an institutional contribution. Göttingen University and its library support the press by infrastructure and staff for organising the review and production process, as well as the dissemination and archiving of publications. To achieve a modest price setting the retail price of books is calculated on the basis of short print runs for re-orders on the basis of external print-on-demand facilities.

3. Larger context

3.1. Integration into the library

As the last wave of university press foundings has largely been initiated by the respective research libraries these presses draw on library-specific dissemination facilities. Catalogue entries of the Göttingen University Press titles for instance can be combined with cover image and order facilities. All titles receive special care and handling during the indexing and integration process into the library's collection. Our subject specialists provide their contact lists or advise the press on how to win subject specialist's attention. At the same time the press places extra copies at the disposal of the dissertation and exchange department and seeks to provide additional copies for titles with textbook potential. For many of the institutions running a university press with a strong focus on Open Access the publishing activities are usually one facet among others in their portfolio of services. In Göttingen operating a professional publishing service led to a far more differentiated and user-oriented concept for the university's repository. It is the press and the superordinate electronic publishing department that handles ISBN, ISSN or DOI issuance for the university. Moreover, the press offers independent advice on legal and economic aspects of publishing. Consultancy does not necessarily lead to a contract with the press but rather in a publishing solution best for the author's needs. The experience gathered in the day-to-day work as a professional publisher serving different scientific communities steers the Open Access activities for the university at large. They include information and advocacy actions for the university members and the development of larger e-publishing networks to enhance the visibility of Göttingen' research output in the electronic publishing world.

3.2. Integration into information and repository networks

These networks in general expand on two levels. They facilitate the development of common strategic visions and the sharing of technical infrastructures. Awareness building and a greater level of common understanding lies at the base of an advanced and standardised international electronic publishing infrastructure. The information platform open-access.net supports scientists and research institutions to implement Open Access in practise and offers information on publishing strategies, costs and legal aspects. Harmonised metadata and a common level of technical quality on the other hand enhances technical performance and allows the integration of separated data-sets into virtual collections. The EU-funded project DRIVER (Digital Repository Infrastructure Vision for European Research)[8] develops

[8]http://www.driver-community.eu.

the infrastructure for a pan-European repository network. A main organisational aim of DRIVER is the confederation building of all relevant stakeholders in the area of digital repositories. On a national level this is complemented by the development of an "Open Access Network of DINI-certified Open Access Repositories".[9] The "Deutsche Initiative für Netzwerkinformation" DINI (German Initiative for Networked Information) developed a certification process to improve the quality of publication services on the basis of Open Access repositories by referring to international standards and quality criteria. The certification process also provides the basis for the integration of these repositories into the DRIVER network. The above described networks up to now focus on already published content in order to provide the scientific communities with parallel Open Access versions of copyrighted and fee-based materials. Primary publishing however raises different questions and requires its own networking activities even if the same infrastructure for displaying the material comes into use.

3.3. Working group of university presses

To strengthen the press' position in the German-speaking publishing landscape several of the small-scale university presses have allied in the working group of university presses. It consists of 10 German, two Austrian and one North-Italian presses.[10] These presses already offer Open Access publishing options to their authors – with different uptakes of the model depending on the respective faculty policies and actual publication behaviour. Members of the working group cooperate in practical issues such as information exchange on production, dissemination or the legal and economic framework. They meet on an annual base and run a common exhibition at the Frankfurt Book Fair and the "Deutsche Bibliothekartag" (Congress of German Library Associations). Cooperation might also mean practical common projects. The Göttingen and Bozen university presses have just recently agreed to jointly publish an anthology that takes South Tyrol as a geological, geographical and cultural region into focus.

3.4. OAPEN

With the project "Open Access Publishing in European Networks" (OAPEN),[11] presented at the round table "University Presses and Books in the Humanities and Social Sciences in a digital future" at the APE 2008, Göttingen University Press seeks to intensify its collaboration with presses, scholars and service providers in the development of Open Access publishing models for high-quality content. OAPEN offers new opportunities of sharing visions and infrastructure with other European presses.[12] All involved presses are open-minded with regard to Open Access and aim to analyse publication and business models for Open Access e-books. For example, there are some practical experiences in OA book publishing, but no systematic analysis of print sales or usage statistics.

On a technical level the project OAPEN aims to develop a publication platform and a digital library on the basis of a distributed network of digital repositories. OAPEN's vision is to use a subset of the DRIVER information space – a set of aggregated and cleaned metadata about publications in European digital repositories – as a basis for its publication and end user services. In the building of its publication

[9] http://www.dini.de/oa-netzwerk.

[10] http://www.ag-univerlage.de.

[11] http://www.oapen.com.

[12] For a recent assessment and vision of the role of universities and their presses in scholarly publishing see also the "Ithaka Report": Laura Brown, Rebecca Griffiths and Matthew Rascoff, University Publishing in a Digital Age, Ithaka, July 2007; http://www.ithaka.org/strategic-services/university-publishing.

and data model OAPEN seeks to cooperate with other initiatives like the European Digital Library project EDLnet.

4. Conclusion

If universities in the German-speaking[13] countries run and let alone launch their own presses in a highly competitive market for scientific publishing they might seem to trifle with public money. Unlike the Anglo-American university presses most of them operate as small-scale enterprises and rely on direct or indirect subsidies of their mother institution. Staff often is recruited in-house and therefore has to acquire professionalism in a learning-by-doing process. To establish scientific journals that can compete with products of the large commercial publishers seems to be out of reach for most of these presses. Therefore, their main outputs are books and often those ones that could not be realised with profit-oriented publishers as they do not promise a significant return of investment.

At the same time having a dark figure of how many work hours of scientists and their staff, how much third-party funding moves into the book production via the commercial publishing system might prove risky as well in the long run. This goes especially for the situation that universities see their research results disappear behind the walls of subscription fees, high retail prices and a strict copyright regime.

The described university presses have settled at the intersection of professional publishing, ruled by economic interests and an institution-owned infrastructure set up to serve scientist's needs. Their dedication to high-quality Open Access publishing at science friendly conditions is an outcome of this double role. A recent study on the role of university presses in scholarly communications[14] points out that university presses should on the one hand be managed professionally in the economic sense and on the other hand play an important role as part of the service portfolio offered by any university. The author states that it is the university press that should be involved when creating new research programs in order to develop matching publishing strategies. The "e-humanities" in Germany for instance, a program developed in the grid computing context, will require new publishing concepts. Those concepts need to be developed in "think tanks" for the time being without economic pressure but nevertheless based on solid publishing experience. We see this role perfectly taken up by the new university presses. Their small size and flexibility allows them to oscillate between a scholarly publisher and a university owned service thus bringing together the best of both environments. This also includes cooperation with presses that would otherwise be considered as competitors.

New concepts of cooperation like OAPEN will provide practical strategies for innovative services and economic expansion as well as arrangements with commercial service providers. As the core group of OAPEN consists of not-for-profit publishers their multi-facet cultural mission combined with careful consideration of economic aspects will assure a balanced view on Open Access in the book sciences and beyond. In this sense not-for-profit presses have a strong potential to improve the state of scholarly publishing while offering alternative community adapted venues of publishing.

[13] This probably applies to other continental European countries as well although there is little empirical data to support these hypotheses. OAPEN seeks to integrate other European partners during the course of the project, especially from the newer members of the European Union.

[14] J.J. Esposito, The wisdom of Oz: The role of the university press in scholarly communications, *Journal of Electronic Publishing* **10**(1) (2007), http://hdl.handle.net/2027/spo.3336451.0010.103.

Information Services & Use 28 (2008) 141–146
DOI 10.3233/ISU-2008-0563
IOS Press

New methods to access scientific content

Stefan Geißler
TEMIS Deutschland GmbH, Kurfürstenanlage 3, D-69115 Heidelberg, Germany
E-mail: stefan.geissler@temis.com

Abstract. The availability of systems capable to perform accurate and large-scale analysis and indexing of scientific content has a profound effect on the scientific publishing sector. Many scenarios about the extension of existing products or the definition of new ones that would have otherwise resulted in significant or even prohibitive costs, now become feasible.

The present paper introduces several case studies of the application of methods of Text Mining in the scientific publishing sector and outlines the motivation and additional benefits that arise from such approaches.

Keywords: Text Mining, information extraction, automatic indexing

1. Introduction

The world of scientific publishing is currently experiencing a multitude of changes that is likely to reshape the way information is handled more profoundly than what we have seen for many decades. Technological progress, changing user expectations and requirements, growing volumes of data and an evolving legal landscape are currently challenging long standing business models and market positions and opening up options for smart new offerings.

Among these areas where significant changes are going on, there's one that addresses the question what the most convenient unit of information in scientific communication should be. Traditionally this has always been the document: most of the infrastructure as well as the habits of most actors in the field are centred around the notion of documents: it's the document that is published, stored, sold, searched and downloaded.

It can be argued however than in quite a number of cases the document as such does not represent the ideal granularity: a researcher looking for clues about potential targets in drug development or about ways to improve the efficiency of an algorithm in computer science may accept the fact that in order to arrive at the required answer, she may have to search, download, print and read documents, but ultimately in many of these scenarios the document is just the traditional form of delivery, the container for pieces of information that our researcher will consider a solution.

As evident as this observation may be, it has to be noted that much of today's information infrastructure does not draw the consequences following from it. A search engine will in most cases stop after delivery a list of potential hits (documents) and leave it to the reader to digest these; a publisher will charge per document.

The label Text Mining has been used for a number of years as a generic term to refer to approaches that address these issues. An active research area for many years, one might have the impression as if a large-scale adoption of these approaches in the mainstream is still some time down the road. In this paper, however, we will list a number of successful case studies from the scientific publishing sector where Text Mining is employed in production scenarios today.

2. A software framework for Text Mining

Before listing the case studies mentioned above, we outline in this chapter the framework that was employed in all of them. It is evident that processing textual data in the context of the publishing sector is a very challenging task. Typically, both the heterogeneity of the topics as well as the volumes of data that need to be considered go beyond what one finds in other sectors.

In can be observed that algorithmic approaches to the management of unstructured data have recognized these challenges. While the fundamental methods in NLP such as part-of-speech tagging, information extraction, statistical language learning and others have been successfully applied under lab conditions for many years [3], there are now serious suggestions available that bring the level of standardization and scalability that is needed for large scale applications in production scenarios. A very important topic in this respect is the UIMA framework [1], initially started by IBM and now handed over to the apache project.

UIMA (for "Unstructured Information Management Architecture") in particular addresses two issues that are key in industrial applications:

- A UIMA-compliant software allows for scalability by virtue of the underlying UIMA backbone that takes care of dispatching tasks to available computing resources. A setup tested and verified with limited resources can therefore be extended to much heavier tasks by just connecting additional CPUs to the backbone.
- UIMA-compliant components, especially so-called "Annotators", respect a well-defined interface such that interoperability is facilitated whereas in the past components coming from different institutions often implemented their own interfaces and respected proprietary conventions, making the task to combine them in a concrete project tedious and complicated.

Besides adopting UIMA as the underlying framework, the approaches employed in the cases studied in the remainder of this paper also maintain a strict separation between the code and the content inside text analysis components: An architecture called "Skill Cartridges™" allows for a defined place for domain- and scenario-specific resources within the same generic analysis software, making deployment of a uniform software core and using a heterogeneous set of analysis components with it manageable in complex setups.

The software framework that exhibits these properties and that has been employed in the contexts described below is the TEMIS platform Luxid® or parts thereof.

In what follows we go through four case studies in which the setup outlined above has been employed in efforts to process large amounts of scientific text.

3. Case studies

3.1. Thomson Scientific

The Thomson Corporation is one of the largest content providers in the world. In the course of reorganisations, mergers and acquisitions, the organisation often is confronted with the task to index large amounts of textual data in order to allow for a uniform treatment of data coming from different backgrounds or in order to apply updated indexing schemata. Since the effort to perform tasks like that on a larger scale often makes processing of extensive sets of data complex and expensive, Thomson Scientific (TS) decided to employ an automatic solution based on the TEMIS software Luxid®.

In a first application of the software several million documents from the Biosis Database were processed in 2006, annotating a range of entity types from geographical locations to chemical substances and organism names. The modular setup of the processing chain, using the Skill Cartridge™ architecture mentioned above, allowed to freely combine these different resources independently and efficiently, some of which corresponded to existing components that were ready to be used at the start of the projects while others were built from TS resources during the deployment.

When applying the complete processing chain on the document corpus, processing proceeded at a rate of ∼500 ms/document in a setup the comprised among other things an organism taxonomy of more than 2 mio entities. The approach puts TS in a position to reindex legacy data in larger quantities within a few days, thus greatly facilitating updates and relaunches of related products. A thorough description of the project can be found at [2].

3.2. Elsevier

While in the case study listed above the setup has so far been restricted to entities, it is easy to see that it can also be of high interest to automatically process information about the *relations* between entities as mentioned in textual data. An example for this is an application of information extraction approaches to chemical data at Elsevier that contributed to the launch of the Patent chemistry Database (Fig. 1).

Patent literature is today maybe the most important source of up-to-date information in chemistry – in many cases, information such as chemical reactions will not be found elsewhere in the literature. Systematically using this source, however, has to cope with the complex nature of patent documents and the high volumes of data generated here. At Elsevier, therefore, one decided to consider auto-

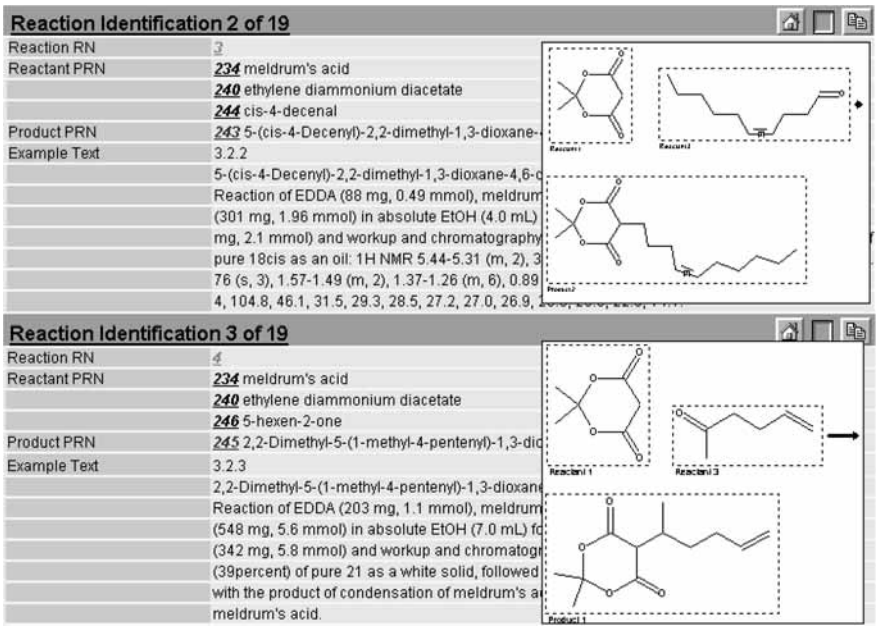

Fig. 1. Information on chemical reactions described in patent literature and processed with automatic Information Extraction approaches is used to fill records in the Patent Chemistry Database from Elsevier.

matic methods in order to harvest such data when planning to launch the Patent Chemistry Database.

Using the Skill Cartridge™ architecture inside a complex processing chain called "Reading Machine", information about substances taking part in chemical reactions as well as side conditions was detected automatically and transformed into a structure format that allows for the creation of a respective database.

3.3. Springer Science + Business Media

In the "Semantic Linking" project at Springer Science + Business Media in Heidelberg, the TEMIS technology is employed to transform online journal content into a densely interlinked network of scientific knowledge. From the viewpoint of the reader of a scientific publication it is often perceived as helpful if certain technical terms and expressions are linked with further information, for instance the respective definition of the term in encyclopaedic content or "Major Reference Works". Evidently seen the volumes of documents involved here, this process needs to be a fully automatic one.

Whereas this approach enriches a journal publication with hyperlinks for specific terms, another kind of added value for the reader is the information that on the topic of a given document there are other documents around that can be assumed to also be of high relevance to the reader because of their semantic closeness. A reader accessing a document on the pharmacological aspects of a neurological disease in a pharmacology journal may not be aware that the same syndrome is also discussed in a journal focussing on psychological issues. Being able to consistently link these documents across the usual boundaries of scientific subjects, journals and institutional publication patterns represents interesting information for the reader and at the same time allows the publisher to attract attention to existing content that would otherwise have remained unnoticed.

Finally, it can be observed that existing vocabularies in many scientific fields would quickly become outdated due to the evolution of scientific progress if it weren't for a constant effort to keep them up to date by continuously adding new terms as they become part of the community's vocabulary. This, again, calls for an approach that automates as far as possible the detection of potentially new technical terms.

The above-listed efforts, currently in the process of deployment, illustrate cases where the appropriate set of automatic natural language analysis approaches can allow for new services around the publishing of electronic content.

3.4. LexisNexis France

One last example for the kind of use of the TEMIS technologies in the publishing industry comes from the KEPLER project at LexisNexis France where documents from the legal domain are structured and analysed. Other than in the various subject fields above it can be noted that the legal domain is a more or less well-defined one, populated by rather clearly defined procedures, actors, competencies and terms. As such the legal domain lends itself to the use of an underlying ontology where this knowledge is formalized: in jurisprudence a *law* is typically applied to a *case* by a *court* headed by a *judge* who arrives at a *sentence* etc. etc., and where all these concepts have a rather static position with respect to an ontology.[1]

LexisNexis therefore decided to specify such an ontology and then to populate it with instances derived from legal documents coming in from courts, legislation and other sources. The resulting structured

[1]The ontology part of the project here was addressed by an application of the French software company Mondeca.

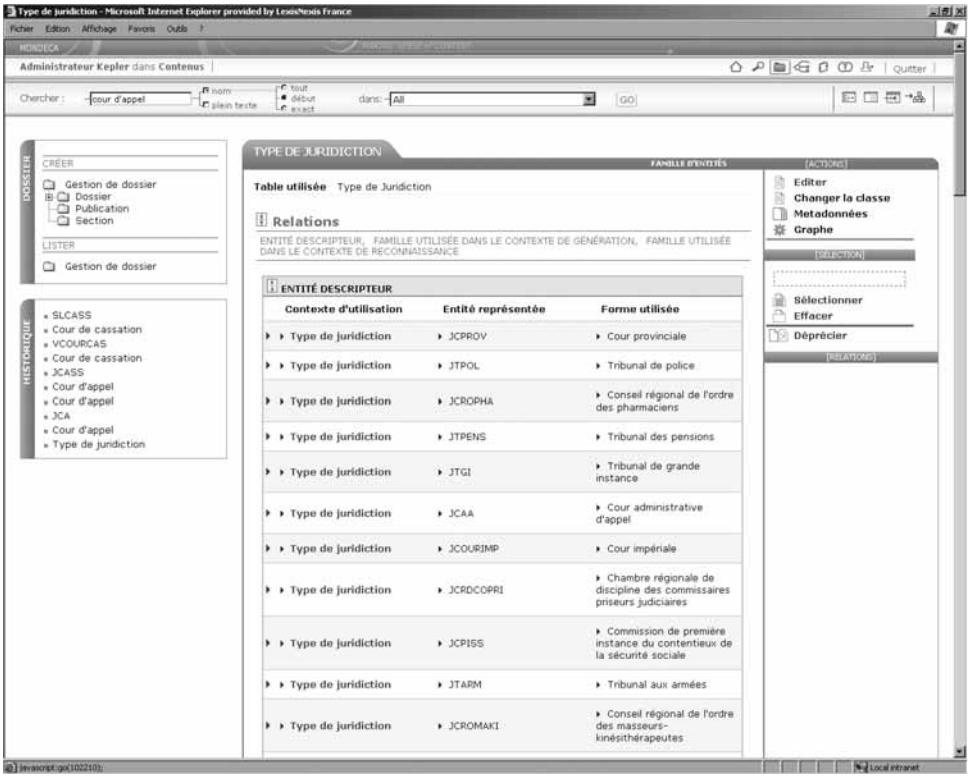

Fig. 2. The application deployed in the KEPLER project give access to the information extracted from legal documents and mapped onto a legal ontology modelling entities such as courts, laws, cases, etc. and which offers rich options for searching, browsing, compiling and publishing complex content.

representation (Fig. 2) facilitates the management of the workflow from the raw information to specific products. If for instance one product is an up-to-date overview on the legislation in a certain field then detecting the fact that in some relevant input document it is declared that a certain act A modifies an act B has direct consequence on the content of the final product.

4. Conclusion

It is not by accident that publishing companies are among the most active users of Text Mining approaches today: while textual content plays a vital role in the business processes of most industry sectors, the situation is different in publishing in that the texts are by definition the very product on which the business is based which gives any approach which directly affects and improves quality, efficiency, comprehensiveness and cost-benefit-ratios of the respective workflows and the resulting products a particular relevance.

We have outlined the way in which Text Mining has consequently been adopted to be used in production in four very diverse scenarios at four large publishing companies. This diversity of scenarios may serve to emphasise the assessment that the automatic processing of natural language content has a great potential in these and neighbouring fields that has just started to be explored.

References

[1] D. Ferrucci and D. Lally, UIMA: An architectural approach to unstructured information processing in the corporate research environment, *Natural Language Engineering* **10** (2004), 327–348.
[2] T. Miller, Rescuing lost data, in: *Text Mining in the Life Sciences Symposium 2006*; http://scai.fraunhofer.de/fileadmin/download/vortraege/tms_06/Thomson_Scientific_Miller.pdf.
[3] D.R. Swanson, Historical note: Information retrieval and the future of an illusion, *Journal of the American Society for Information Science* **39** (1988), 92–98.

Information Services & Use 28 (2008) 147–149
DOI 10.3233/ISU-2008-0576
IOS Press

APE 2008 Academic Publishing in Europe, Quality and Publishing

January 22–23, 2008, Berlin

Ronald Schild
MVB, German Publishers and Booksellers Association, Frankfurt am Main, Germany

Those of you who were here for last year's conference might remember that I was here too and chaired one of the sessions. Last year, I mentioned that we had just started a new project called "Volltextsuche-Online" which stands for "full-text search online". Today, I am here to present this project again. I am quite proud to be able to tell you that it is now up and running, following the launch of the website at the last Frankfurt Book Fair. Furthermore, we have changed the name of the project, because full-text search, or searches in the full text, is just one small functionality in the whole set of features we want to offer. The project's new name is libreka!. We also wanted to make the point that we are not a "YACS" – "Yet Another Californian Start-up". On the contrary, MVB is based at the heart of the German publishing industry and is wholly owned by the Börsenverein, the German Publishers and Booksellers Association.

Now, what do we want to achieve with libreka!? To put it very simply: we want to put every German-language book in print online. We want to be the leading internet platform for the German book trade or, figuratively speaking, we want to be the Swiss army knife of *e*-content. First of all, we have a marketing objective. We want to increase the reach of publishers and booksellers. We want to open up new business models and sales channels for booksellers and publishers and want to protect intellectual property. We want to give publishers control over how much and what they publish on the internet.

Now, what do I mean when I say: "We want to increase the reach of the publisher or bookseller"? First of all, by putting every book in print online with libreka, we want to give end-users access via the internet to culture, to knowledge, to education. At the same time, we offer every bookseller a service similar to the "Search Inside" offered by a large online retailer. Furthermore, we want to integrate the content on libreka! in the main search engines Google, Yahoo and Microsoft. If you look for a book that is on libreka! today, you will probably find it on Google as well. But what we want is to drive initiatives like "Google Book Search" or "Microsoft Life Search" by adding our content and therefore making it more easily available. At the same time, libreka! gives booksellers and publishers the opportunity to offer their customers a search inside the book-function and a chance to read examples of the books on libreka!.

The integration in the search engines basically works in two ways. One way is very easy: the search engine indexes our html-pages. This is what they are already doing today. The second model is more complicated: the data we get from publishers is in pdf-format, though it is not a perfect format to use. We want to move to xml-format, but this data is rather difficult to get. That is why we are stuck with pdf for the moment. What we do is to extract the text from the pdf and build our own text index – this is purely text, no images and no meta-information about the book. Then we give access to that text

index to the search engines that enrich their own search index with it. They generate search results and display a search list where you see basic information on the books, such as the title. But in order to view and to access the book, you have to go back to libreka!, you read and view the book on our platform. This gives the publisher the biggest possible control over the content. But we still have to suit different publisher's needs. There might be publishers who prefer open access models and want to give the content away freely. Other publishers favour more restrictive models. We are not in the position of telling the publishers what model to use. But we want to be open to any business model that a publisher might want to use. At the moment, we are negotiating with the big search engines and we have arrived at a certain understanding that they will at some time integrate our data.

As I have already mentioned, to offer full-text search is not our main objective. What is more important is the relevance of the search results. Already now, with over 42,000 books online, we get at least several hundred search results for more general search terms. That is obviously too much. We know from usability studies that people who use a search engine usually stop looking after the first three listings or at least after the first or second page of the result list. Search results on page 10 are not really important. With this in mind, the real challenge is to show the most relevant search results on the first page of our result list. We do this by combining the search of the full texts of the books on libreka! with their metadata. We have access to the metadata of the German-language books in print that are available via the "Verzeichnis Lieferbarer Bücher (VLB)", which is published by MVB. It lists more than 1.2 million books. This gives us the opportunity to add to the search results the author's name, the publisher, the publishing year and more information on the title. Furthermore, we have a research project running at the moment with the Ludwig-Maximilians-Universität (LMU), a Munich university, to use linguistic methods to increase the relevance of search results.

What I have wanted to point out so far: libreka! means more marketing for publishers and booksellers. They can increase the reach of their target audience and give more information to their customers. But, of course, they want to do business as well. What we are planning to introduce this year are various *e*-commerce models. The easiest model is direct consumer sale of *e*-books. You can sell *e*-books that are downloadable to PCs or reading devices. This model is the most discussed at the moment. Personally, I think that online access to information will become more important and might be one main driving force when talking of *e*-content. And that applies as much to individuals as it does to institutions – libraries, research centres and so on. We can think about different subscription models and pay per use models. Here, again, the publisher decides, what business model he wants to implement for his *e*-content.

And, of course, we are open to booksellers too. Our objective is very clearly to offer every bookseller, regardless of size or online affinity, the ability to sell *e*-content. We don't want a monopoly or very few dominant players to control the *e*-content-market. That will be neither good for publishers, nor for cultural diversity nor for access to information.

And, last point, what we want is to protect intellectual property. Our mission is to provide a distribution platform; we are not a proprietary system like many other book search engines. libreka! is a platform where the publisher can list his products, put up his content and we give it away in accordance with the publisher's needs and wishes. They keep full control over their content – on a per page basis: the publisher decides how many pages of each of his books in libreka! can be viewed: for one book he may grant access to the general public only for pages 1–10. For the next book, everyone is allowed to see everything. And the books can be distributed across all sales channels that cooperate with the publisher. Of course, we offer DRM systems or DRM functionalities, but I am not going into detail about that here.

I would now like to show very briefly how libreka! works. Search for any given term and you get a list with your search results. For every book in this list, you see the cover, the title and a short description.

When you click on one of the books you get access to it. The virtual book is displayed the same way as the printed book. If one of the pages is blank in the printed book, it's displayed the same way in libreka!. More importantly, in a second step, you can look for your search term "inside the book". libreka! then gives you the pages of the book, where your search term appears. The pages the publishers wants to exclude from being seen by the general public are not displayed in libreka!. I can do many more things, like browse various book categories, such as fiction, non-fiction or children's books, but I am not going to show you that now, it would take the whole afternoon.

As I have already said, we are on the way to implementing *e*-commerce functionalities, they are going to follow pretty soon – as well as a privileged access for booksellers. The objective is to rationalise the interaction between publisher and bookseller. At the moment, publishers spend lots of money on transporting sample books to the booksellers, like reading samples. With libreka! they do not need this anymore.

At the moment more than 850 German-language publishers are partners of libreka! We ran a special promotion, advertising free scans for books where no data is available. This brought us 60,000 books to scan – which we are doing now. Furthermore, we were promised 30,000 books in digital format. At the end of this presentation the last slide shows some of the participating publishers. Some of the biggest publishers in Germany are on board with libreka!.

Thank you very much for listening.

Information Services & Use 28 (2008) 151–158
DOI 10.3233/ISU-2008-0565
IOS Press

DAISY: An opportunity to improve access to information for all

APE 2008 (Academic Publishing in Europe)

Thomas Kahlisch *
Deutsche Zentralbücherei für Blinde zu Leipzig, Gustav-Adolf-Straße 7, 04105 Leipzig, Germany
Tel.: +49 341/7113 150; E-mail: thomas.kahlisch@dzb.de
URL: www.dzb.de; www.kahlisch.de

Abstract. Since 1996 public library services for the visually impaired round the world work closely with the ITC and the publishing industry in the DAISY consortium (www.daisy.org) towards international standards and technologies which are needed to improve access to information for the elderly and people with different reading disabilities.

The consortium's vision is "that all information, at time of release to the general population, be available in an accessible, highly functional, feature rich format and at no greater cost, to persons with print disabilities".

DAISY (Digital Accessible Information System) is an open standard for multimedia publishing, the version 3 has already become a national standard in the United States DAISY/NISO 2005 Z3986. The DAISY XML applications DTBook (to structure content in a Digital Talking Book) and NCX (for navigating multimedia contents) are used in OPS (Open Publishing Structure) 2.0 V. 1.0, published in summer 2007 by the IDPF (International digital Publishing Forum, *www.idpf.org*).

Today there are already more then 200,000 DAISY books available to people with reading disabilities all over the world. DAISY content is created from analogue tapes, read by narrators or converted from text files directly into the DTBook format. DAISY books can be played on special hardware players, different software applications and on accessible portable devices like PDAs or mobile phones.

In my presentation I will:

- describe the current projects of the DAISY Consortium,
- demonstrate examples of DAISY content,
- explain the benefits the publishing industry could gain from taking part in developing accessible information systems.

Overview

- Motivation
- MEDIBUS
- DAISY Consortium
- Technology and standards
- Projects
- Content and applications

*Dr. Thomas Kahlish is a Director of German Central Library for the Blind (DZB).

Fig. 1. DZB.

– Playback
– Example
– Cross-publishing

1. Motivation

DZB (Deutsche Zentralbücherei für Blinde zu Leipzig, see Fig. 1) is the oldest public library service for blind people in Germany. DZB was founded in 1894. Today, DZB is moving from being a special library service to a centre of accessible media design. The following timeline shows milestones of developments in accessible technology for blind and visually impaired people, see Fig. 2.

By using common technology like structured document processing and other cross publishing technologies it is possible to establish cooperative projects to increase access for all people with and without disabilities.

2. MEDIBUS

MEDIBUS is coordinating the production of accessible books for all German speaking countries (Germany, Austria and Switzerland).

This map shows all MEDIBUS sites in these countries, see Fig. 3.

Braille and talking books are produced by the member organizations of MEDIBUS. Currently, there are 42,000 titles in Braille, 50,000 talking books on compact cassettes and 20,000 DAISY books are available to the users. The books are delivered free of charge by the post. Patrons can use a catalogue to find new titles: *www.medibus.info*.

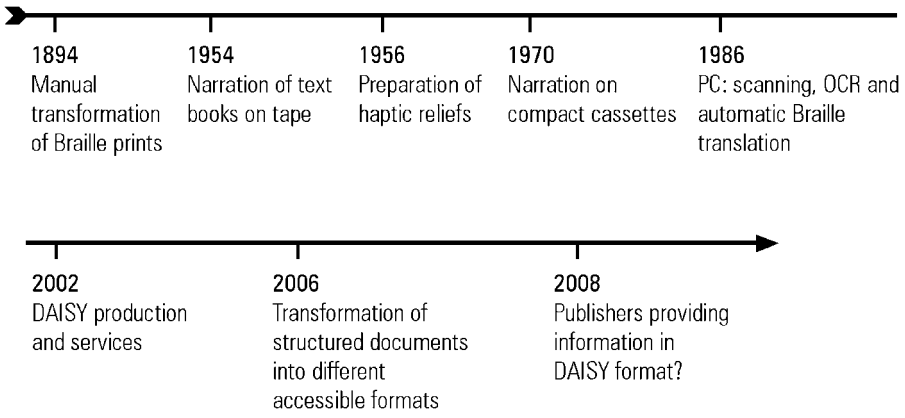

Fig. 2. Timeline.

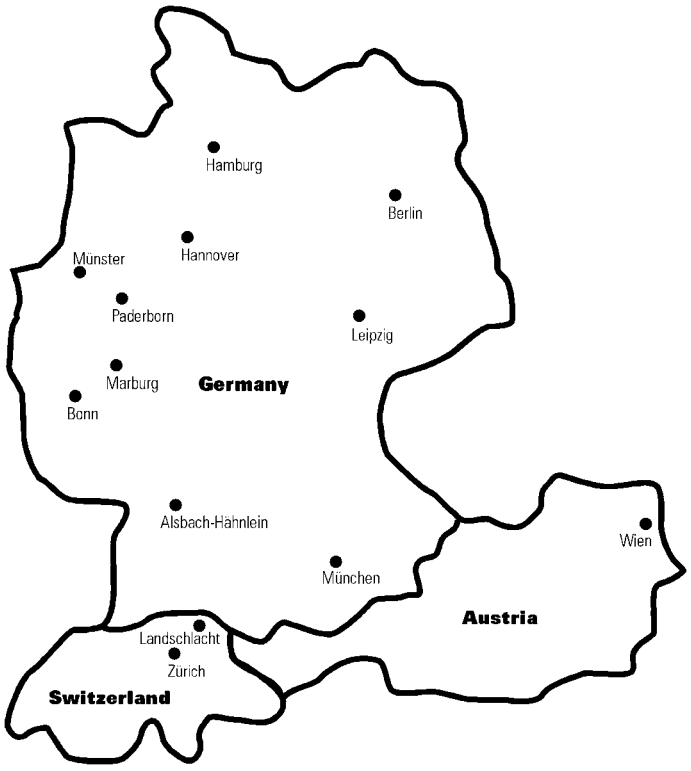

Fig. 3. MEDIBUS sites.

3. DAISY Consortium

DAISY (Digital Accessible Information System) is the name of a standard and the related technology that has been developed by the DAISY Consortium. The Consortium was founded in May

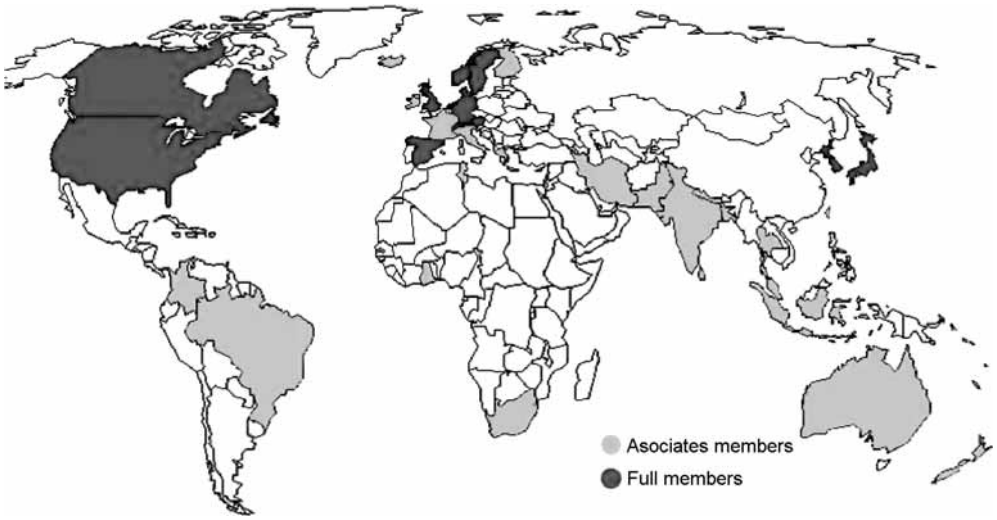

Fig. 4. DAISY member organizations.

1996 by talking book libraries to lead the worldwide transition from analogue to Digital Talking Books.

Now nearly 70 non-profit organizations make up the Consortium and actively promote the DAISY Standard. Most of the library services for the blind and visually impaired worldwide are members of the DAISY Consortium. The world map below shows all the countries that have full or associated member organizations in the Consortium, see Fig. 4.

The goal of the DAISY Consortium is:

> Our vision is that all published information, at time of release to the general population, be available in an accessible, highly functional, feature rich format and at no greater cost, to persons with print disabilities (*www.daisy.org*).

4. Technology and standards

The DAISY Consortium is developing standards and technology for DTB (Digital Talking Books). DTB is a XML based format for accessible multimedia publishing.

Books, journals and a lot of other information can be marked up in the DAISY format. By using converting software, it is possible to translate a DAISY document into different accessible formats like Braille, large print or synthetic speech.

The following standards have been developed:

I 1996 DAISY Standard V. 1 (Proto type, Proprietary).
II 2002 DAISY Standard V. 2.02 (XHTML, talking book with TOC navigation).
III 2004 DAISY NISO Standard Z3986 (XML, schoolbooks and scientific publications).

In September 2007 IDPF (International Digital Publishing Forum) published the OPS V. 2.0 specification, which includes NCX and DTBook (part of Z3986).

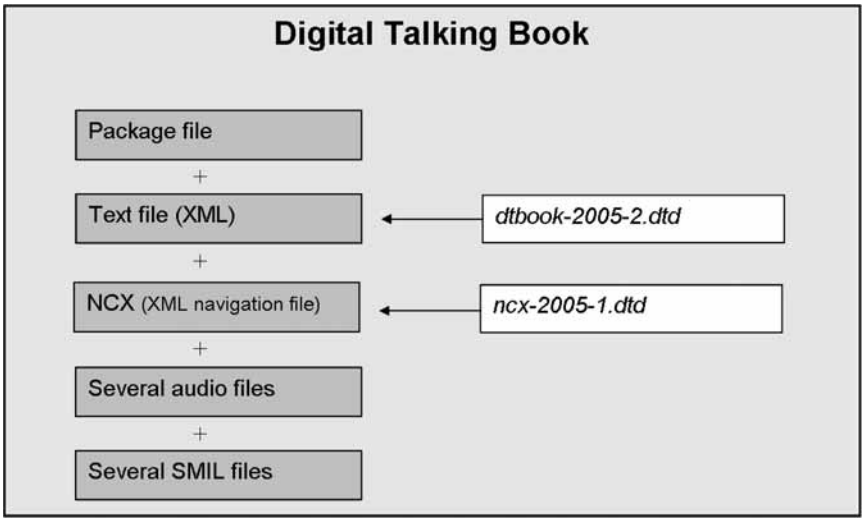

Fig. 5. NISO standard Z3986.

For more information see: *www.idpf.org*.

In November 2007 Microsoft announced a collaboration project with the DAISY Consortium. The name of the project is "Save as DAISY XML". Open software will be developed, to make it possible to store Microsoft office documents directly into the DAISY format.

For more information see: *http://www.microsoft.com/presspass/features/2007/nov07/11-13daisy.mspx*.

4.1. DAISY/NISO Standard Z3986

The DAISY V. 3 specification is a formal standard of the American standard organisation. Officially, the ANSI/NISO Z39.86 specifications for the Digital Talking Book.

Abstract:

> This standard defines the format and content of the electronic file set that comprises a digital talking book (DTB) and establishes a limited set of requirements for DTB playback devices. It uses established and new specifications to delineate the structure of DTBs whose content can range from XML text only to text with corresponding spoken audio, to audio with little or no text. DTBs are designed to make print material accessible and navigable for blind or otherwise print-disabled persons.

The picture shows general file structure of a DTB and differences of DAISY V. 2.02 and DAISY/NISO Z3986, see Fig. 5.

5. Projects

The following projects have been established by the Consortium:

- DAISY for all

  - A support project to establish DAISY into developing countries;
- Braille in DAISY
  - To develop facilities to translate content by using the Z3986 standard into different Braille formats;
- MathML modular extension for Z3986
  - To include markup for mathematical expressions in an accessible format;
- DAISY and SVG
  - To include appropriate markup for structured graphical information into the Z3986 standard;
- Urakaawa Project
  - An open source API has been created, which can be used for Z3986 standard applications development;
- Online distribution
  - API and facilities will be developed to standardize online distribution processes for DAISY content worldwide;
- DAISY Pipeline
  - A Multi-Format-Converter is available to translate documents into the DAISY format.

6. Content and applications

Currently the DAISY V. 2.02 (talking books with navigation for searching and browsing into books) is in common use by the library services for the blind:

- More than 200,000 DAISY titles are available worldwide;
- By 2010 more than 30,000 titles will be available in Germany.

The advantages of the DAISY V. 2.02 books are:

- One book fits on one media;
- 40 hours audio on one CD;
- Users can search and browse the table of contents;
- Text and Audio synchronization is established by using SMIL.

The more complex books are provided in DAISY V. 3 Z3986.
DAISY V. 3. content is already being produced in the following projects:

- American publishers providing data in NIMAS format, a subset of Z3986.
- *www.bookshare.org* provides access to more than 50,000 DAISY documents in the USA only to users who have reading disabilities.

Fig. 6. DAISY player.

– *www.Braillenet.fr* and French publishers collaborate to translate information into Braille and other accessible formats.

7. Playback

It is possible to play DAISY books on special players, on a PC and on standard MP3 or DVD-Players.

The special Players, PTN1, Victor Reader Stream and Milestone, are designed for the needs of people who cannot read displays and have only limited computer skills, see Fig. 6.

Software is: DAISY-Leser, AMIS and TpB. They can be used free of charge. They can be used to play the audio content, navigate into the book structure and to read the text of the book on the screen or on a refreshable Braille-display.

DAISY-Leser, playback software developed in DZB, can be downloaded from: *www.daisyleser.de*.

A current project at DZB is the development of DAISY playback software on mobile phones. DAISY2go is a player that will run on smart phones with the Simyon operating system.

Standard DVD and MP3 Players also play DAISY books, but they do not support the advanced DAISY features like the navigation and browsing capabilities. Usually these devices can only be used by reading out the information from the display. Blind people do not have access to these off the shelf products.

8. Example

Figure 7 shows an example of a DAISY book.

9. Cross-publishing

The great challenge of DAISY is that the format can be used for cross-publishing. By using the DAISY data as a single source file information can be translated into different formats that are suitable for the users, see Fig. 8.

An XML document provided by a publisher can be semi-automatically translated to Braille, large print or synthetic speech. The publisher can also use the XML document for different types of online and print publications.

Fig. 7. Example of a DAISY book.

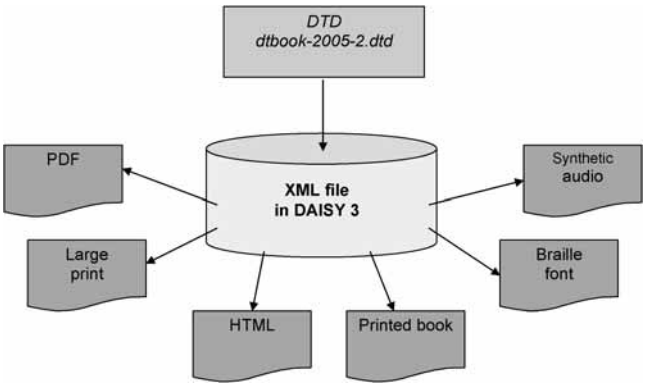

Fig. 8. Cross-publishing.

Information Services & Use 28 (2008) 159–161
DOI 10.3233/ISU-2008-0561
IOS Press

Publishing in the digital age: Challenges and requirements for authors and users of SpringerLink

Gertraud Griepke
Vice President SpringerLink Operations, Springer-Verlag, GmbH, Tiergartenstr. 17, 69121 Heidelberg, Germany
Tel.: +49 6221 487 8457; Fax: +49 6221 487 8364
E-mail: gertraud.griepke@springer.com
URL: www.springerlink.com

1. Introduction

The use of technology to support the digital production, distribution and storage of scientific content in full-text formats poses challenges for authors and publishers as well as for the librarians and scientists who access the materials. While scientific publishers still have to produce high-quality print products, they have had to develop new production systems that are now able to produce book and journal pages in both print and digital form.

This paper will discuss how Springer organizes its digital production and distribution systems and some of the challenges we will face in the future.

2. From the manuscript to the final paper – tagging

All of the world's scientific publishers have agreed on a standard set of XML tags, called metadata, to identify all of the various data elements – journal title, ISSN or ISBN for books, volume and issue number etc. – that can be used in scientific books and journals. Authors can insert some of these tags as they compose their works, but the publishers insert most of the tags.

Graphics are also identified and linked with the appropriate part of the work. Standards for tagging and handling supplementary materials like videos, chemical structures and animation are still in the discussion stage among publishers and other interested parties within the publishing community.

2.1. HTML or PDF format?

The other issue that is being debated by the scientific community for journal articles is how the content will be displayed in electronic form. Most publishers and search engine vendors prefer HTML but user surveys indicate that the scientists who will read the articles in electronic form or as printouts prefer

the PDF format by wide margins. There are strong arguments for both, and it is too early to tell which format will be the final choice.

3. From the manuscript to the final paper – publication

A Digital Object Identifier (DOI) is added to each article after copyediting and tagging. The DOI is a persistent electronic identifier that can connect the paper to other publications through cross-reference links. More than 25 million DOIs have been assigned since the program started in 2000.

The actual typesetting process is electronic, and the online version of the paper is published first. Corrections after publication are always in the form of ERRATA, as content cannot be removed from the database of digital articles.

4. The universe of online publishing

Springer's electronic content is distributed through a data delivery system to its own website, springerlink.com, and parts of the content are distributed to approximately 100 different parties, which include bibliographic services and some large library consortia that store content on their own sites.

Springer supports and promotes its online content through springer.com, an extensive website that includes catalogs, promotional materials, and information for booksellers and librarians. There are tools for journal authors to communicate with journal editors and other helpful functions for customers to access.

4.1. SpringerLink

All of Springer's content that is in electronic form is available online at springerlink.com. Individuals and institutions can subscribe. A subscription includes a comprehensive search capability and, of course, extensive links to electronic journals in other publishers' databases.

Libraries have additional requirements that include individual usage statistics within their institutions. Springer provides these statistics in the COUNTER format, which libraries use to compare usage across multiple publishers.

Finally, SpringerLink is the website that searchers arrive at when they come across a reference to a Springer online journal. These search queries come from a number of different websites, but the Google portal had the highest referral rates of any other portal last year.

5. Long-term outlook

In the future we can expect to find increasing amounts of the world's literature being published in electronic form, and sometimes exclusively in electronic form. Searching will be even more intuitive and will be facilitated by semantic links; that is, users will be able to search for similar content. They will also be able to use mathematical formulae to compute graphs and tables, and chemical formulae will be searchable by structure.

In order to make all of this happen, publishers will have to make sure that the content we prepare for publication today will be ready for the enhanced storage, retrieval and searching techniques of the future.


SpringerLink
Content Types Subject Collections
Back to: All Search Results
Journal
Athens Authentication Point
Recognized as:
Springer (200-76-474)
Welcome!
To use the personalized features of this site, please log in or register.
If you have forgotten your username or password, we can help.
My SpringerLink
Marked Items
Alerts
Order History
Saved Items
All
Favorites
Journal of the American Oil Chemists' Society
Publisher Springer Berlin / Heidelberg
ISSN 0003-021X (Print) 1558-9331 (Online)
Subject Collection Chemistry and Materials Science
SpringerLink Date Thursday, November 09, 2006
Add to marked items
Add to saved items
Register for TOC Alerts
Recommend this journal
Editorial Board
Submissions
About This Journal
Biosolids Treatment Processes
Wang, Lawrence K. , Shammas, Nazih K. , Hung, Yung-T
Disable Highlighting
Editorial View
Expanded List View
Condensed List View
59 Volumes
First Page | Next Page
Current-82
Access to all content
Access to some content
Access to no content
Volume 84
Number 4 / April, 2007 315-404
Number 3 / March, 2007 205-314
Number 2 / February, 2007 107-204
Number 1 / January, 2007 1-105
Volume 83
Find more
Within all content
Within this journal
Starts With
Content Status
Online First™ (10)
In Publication (6,790)
SpringerLink Date
In the last week (5)
In the last two weeks (6
In the last month (522)

Fig. 1.

Information Services & Use 28 (2008) 163–167
DOI 10.3233/ISU-2008-0566
IOS Press

APE 2008: Pre-conference day – The purpose of publishing, education and training course for younger academic publishers. Tutorial 2: Providing readers with the Google experience (user experience)

Brill's publications on Google's BookSearch platform

Presentation: Dr. Matthias Wahls, Manager Business Development, Brill, Leiden, The Netherlands

1. Introduction

In this paper, Brill's participation as "early adopter" in the GooglePrint initiative, as it was called at that time, launched at the Frankfurt bookfair in 2004, is outlined, and first results and gained insights are briefly discussed in due course. Facts, like traffic and usage data, are given and analyzed. An overview of benefits for Brill publication readers, authors, and the publisher itself are briefly reviewed and used for a brief discussion of assumed future needs of current and future user generations of Google's book search tool, and its upcoming competitors.

2. About Brill

Founded in 1683 in Leiden, The Netherlands, Brill is a leading international academic publisher in the fields of Ancient Near East and Egypt, Middle East and Islamic Studies, Asian Studies (including the Hotei imprint), Classical Studies, Medieval and Early Modern Studies, Biblical and Religious Studies, Social Sciences, Biology, Science (the VSP imprint), Human Rights and Public International Law (the Martinus Nijhoff imprint). With offices in Leiden and Boston (MA), USA, Brill today publishes more than 100 journals and about 500 new books and reference works each year. Almost all publications released by Brill nowadays are written in the English language, with the occasional exception of a work in the French or German language.

In addition, Brill is a major provider of primary source materials online and on microform for researchers in the humanities and social sciences through its imprint IDC Publishers. In principle, all publications are available in print or microform, as well as in electronic form.

3. E-publishing @ Brill

3.1. Overview

From the subscription year 2000 onwards, literally all Brill journals have been made electronically available, through the Ingenta platform [1]. Also in 2000, Brill started to offer selected book titles as e-books through some selected vendors, who are primarily focusing on the institutional (librarian's) market, like Netlibrary [2] and e-brary [3], in addition to some others. From 2005 onwards a few hundred books, primarily Humanities titles, have become available on Questia's platform [4], which actually marks the first Brill experiment of offering e-books to individuals. In 2005 Brill launched its reference platform called Brillonline [5], through which many of Brill's Major Reference Works and Encyclopedias are made available electronically. In 2005, partnerships with both Amazon [6] and Google [7] have been entered into, to allow indexing of most of Brill's book titles through Amazon's "Search Inside the Book" and "GooglePrint, also named Google's BookSearch (GBS)" program, respectively. Both services, although slightly different in their purpose and market approach, do have in common that their primary market certainly is that of the individuals.

Finally, it is worth to mention that since 2005, some selected IDC microfilm collections have been made available electronically through IDC's digilib application [8].

3.2. Traffic to HP

Within the context of this paper, it is important to understand the various ways Brill customers (primarily authors, readers and librarians) approach ("find") Brill's homepage: www.brill.nl.

About half of all web-visits on brill.nl enter the publisher's homepage through either Google (40%) or Yahoo! (10%), while GoogleScholar accounts for another approximate 10%. Brill's participation on Google BookSearch accounts for additional traffic to the site, currently at a single digit % level. Navigation to the site directly is undertaken by approximately 40% of all visitors. That figure is certainly remarkable and confirms Brill's strong brand recognition.

3.3. Brill publications on Google's BookSearch platform

Since summer 2005, about 6000 Brill titles (under the imprints Brill, MNP and VSP, respectively) have been uploaded to the GBS platform [9]. From October 2007, a workflow has been in place to allow the regular uploading of all new book titles, immediately after publication. For all these titles, cross-links are available between the individual title's product site on Brill.nl and GBS, respectively, and an extended choice of purchase options for the customer within the Google environment.

Additionally, a huge amount of currently commercially non-available (except for any occasional 2nd hand opportunities), so-called out-of-print titles have been uploaded to the GBS. A significant amount of these titles have been published in the 19th century, under the Brill imprint, which itself is in used since 1849/1850. These "archive" titles so far lack cross-link facilities between their own product homepages at either brill.nl or their GBS site, as well as purchase opportunities. Brill, however, is evaluating,

primarily from feedback received from customers, market opportunities to republish selected titles from this list in the future.

E-access and POD facilities, which have been expected to be launched by Google for quite some time already would, if coupled to the initial GBS indexing platform, allow Brill (like all other publishers) to keep publications (digitally) available to customers relatively easily as well as long term. Such approach would represent an alternative to serve existing and future archive needs, both of the community and the publisher itself, as libraries would not be involved in such a set-up.

4. Discussion and conclusion

4.1. Impact

Up until the end of 2007, all book titles Brill has uploaded to the GBS platform generated approximately 4.5 million book-visits so far, of which about 25% (1.2 million) have been viewed books and have generated a total of 14.4 million page views. This calculates an average of 10–11 page-views per session/visit. About 3.5% (+40,000) of all those customers who have viewed pages of Brill publications on the GBS platform have clicked onto the "Buy this Book" link on the GBS product page of individual Brill books, to inquire for further product information, e.g. purchase opportunities. About one third of all BTB clickers have done so by approaching the product page on brill.nl.

Although final evidence is missing, it is strongly believed that the increased visibility of in particular deep backlist titles has supported the sales figures of individual titles, in line with the 'long tail' theory.

4.2. Benefits

The overall traffic and usage of Brill's publications within the GBS environment confirm that the great majority of Brill's book publications remain academically relevant for a long period of time. This effect is certainly caused by the nature of the academic areas of research typically served by Brill's publications, e.g. the Humanities. A large amount of those Brill books which are showing these huge traffic figures have not been made electronically accessible by the publisher in the past for a variety of reasons. Through GBS, these publications became visible for the first time on the internet. This looks likely to be a great service in particular also for those researchers who lack access to these works through an academic library, for example. As an additional benefit, it is in essence for the first time possible to gain a quantified picture of traffic and usage for academic books, from the publisher's point of view.

4.3. Conclusion

Usage and traffic data gained so far through the GBS platform undoubtedly indicate that the great majority of Brill titles attract interest from a presumably large public.

From these sheer figures, it is concluded that the user experience(s) must be predominantly positive. One has to divide users into at least three different categories in this context:

1. *Reading customers* all over the globe can find publications in their fields of interest relatively easily, check limited portions of the text for relevance, and are enabled to decide whether to purchase such a title, from various vendors, quickly and almost hassle-free. In pre-internet times, such a discovery, if possible at all, would have taken days, if not weeks.

2. *Authors* are benefiting as well, as they are gaining attention for their own works from a much broader audience through participation in GBS.
3. *Publishers*, last but not least, as the representative of their authors, benefit from the increased attention a title can gain through participation in GBS in at least three ways: (a) by becoming an even more attractive choice for authors as their preferred publisher, (b) by generating additional sales, and (c) by being able to digitize the book backlist titles through Google at a very limited cost.

If monetizing tools, such as the earlier mentioned e-access and POD facilities would be made available within the GBS environment, an additional benefit would arise for publishers, certainly for the smaller ones, namely that (deep) BL titles could remain "in print" perpetually at little, if any, cost to the publishers. Such a solution would eventually be to the benefit of all three major stakeholders, namely readers, authors and publishers.

One can safely state that the participation in the GBS program with all its benefits mentioned in this paper so far *supports the publisher Brill in its mission* to allow the best possible access and dissemination to its published works. Whether this happens through the printed or electronic version of any given publications is of no relevance in principle, as long as all different customer demands are served appropriately. Google's book search platform enables Brill to reach new groups of customers, of which many in the pre-internet times most likely were unable to access Brill books.

Therefore, the internet, which has been available to the mass audience on a broad scale for no more than 15 years, has significantly changed academic (book) publishing in a relatively short period. Google's book search program, launched in 2004, i.e. approximately 10 years after internet became a widely used medium, constitutes a milestone for the academic book publishing business. Academic books – like journals – nowadays can reach a much greater number of its potential readers, everywhere and any time, and with little effort.

5. Outlook

Three and a half years after the introduction of Google's BookSearch, and at a time when Microsoft's Live Book Search tool, and the Libreka! Initiative of the Deutsche Boersenverein, among others, have just been launched, it may be safely stated that Google has greatly influenced the (academic) book market with its formerly called GooglePrint initiative. Whether any of the mentioned alternatives or any others yet to come, or the whole Open Access Movement in general, will also influence the way academic books and other works are published, and if so to what extent, only time will tell. Academic publisher will have to monitor closely a number of related issues, such as Google's long term commitment to GBS and its related products, any financial implications, but clearly also all developments regarding any Semantic Web application.

So far, the internet in general and GBS in particular have made a great impact onto the academic book publishing business in a relative short period of time. GBS certainly represents a new and valuable tool to experience book publications.

The most important question, however, remains: Does this tool serve all needs of current and future users? Still, customers have to wait for access facilities (e-access, POD) to books on the GBS platform. In the academic field of publishing, however, readers are often prepared to pay considerable amounts for 'must have' knowledge. The current GBS approach unfortunately does not answer all customer needs in

that respect. Any of the upcoming alternatives may fill in this gap eventually, but that is speculation at this moment.

Regarding any of the literature search habits and principal needs of the upcoming and future generations, one can only speculate. Already the current generation of 'young researcher' considers almost only material which is available and accessible online as existing. Therefore, a great task remains for the content providers (publishers) and the search institutions (like Google, Microsoft and Yahoo!), namely: how to make the content, which has been digitized at considerable costs, fully available, while respecting all (copy-)rights of all stakeholders involved? As customers these days tend not to accept second best solutions, all stakeholders involved will have to offer acceptable tools to allow access to academic book publications as soon as possible.

Acknowledgements

The author of this paper would like to take the opportunity to kindly thank Ms. Laura Driussi, Mr. Jason Hanley and Mr. Jens Redmer of Google Inc. for their support during the implementation of GooglePrint/GBS at Brill Publishers in Leiden and for the fruitful co-operation over the past years.

References

[1] www.ingentaconnect.com.
[2] www.netlibrary.com.
[3] www.ebrary.com.
[4] www.questia.com.
[5] www.brillonline.nl.
[6] www.amazon.com.
[7] www.google.com.
[8] www.idc.nl.
[9] www.books.google.com.

< APE 2008

The International Conference "Academic Publishing in Europe"

QUALITY & PUBLISHING

Sponsors

European Commission, Research Directorate-General

Akademische Verlagsgesellschaft Aka GmbH

The Electronic Publishing Working Group (AKEP)

ALPSP

ARIES

German Association of Publishers and Booksellers (Börsenverein)

German Association of Information Theory and Practice (DGI)

Ebsco Information Services

Elsevier

Federation of European Publishers (FEP)

Frankfurt Book Fair

International Federation of Library Associations and Institutions (IFLA)

IOS Press

IPA

Klopotek & Partners

Lightning Source

MetaPress - Electronic Content Management and Hosting

Royal Brill

Springer Science + Business Media

The International Association of Scientific, Technical and Medical Publishers (STM)

Swets

Taylor & Francis Group

APE 2009

Academic Publishing in Europe: The Impact of Publishing

20-21 January 2009 – Academy of Sciences, Berlin

This fourth APE 2009 Conference is about **'the impact of publishing'** in a broad perspective, varying from usage and users, peer reviewing and quality assurance, branding, funding and investing, business models, new types of information, enabling technologies, search engines, to research excellence, dissemination, access and sharing of knowledge.

APE Conferences encourage the debate about the future of scientific publications, information dissemination and access to scientific results. They offer an independent forum for 'open minds' with a free exchange of opinions and experiences between all stakeholders.

Participants are: academic, educational, scientific, technological, medical, legal and professional publishers, university presses, scientists, authors, editors, librarians, teachers, learned and professional societies and associations, funding agencies, politicians and policy makers, subscription agencies and booksellers, recruiting agencies and technology providers.

Aims and Scope: Academic Publishing and library activity face significant technical, financial and political challenges in the coming years. This may have major consequences for scholarly communication. Europe is the largest producer of scientific research and scholarly knowledge with a very strong publishing and library culture, with different scientific and professional cultures and a multitude of languages, but the acceptance, requirements and expectations are changing rapidly. Although the supportive role of academic organizations, societies and the private sector is subject to an intensive debate, there is a lot of uncertainty about current challenges and possibilities of the scholarly communication system, like new business models, the role of established players, funding schemes, quality and integrity of information, intellectual property, or life cycle and preservation of digital objects. Those issues are often discussed fiercely and emotionally, and there seems to be little interaction and knowledge exchange between the different players of the scholarly communication system, like librarians, publishing houses, university presses, scientists, funders and policy makers.

APE Conferences aim at a better understanding of scholarly communication through value-added publishing and the increasingly important role of relevant information in science and society. They provide an inventory and show the way ahead in a collaborative effort involving all stakeholders. The need for a wider dissemination and improved access, with new content like raw data, computer simulations, laboratory manuals and other background information has been identified. At the same time APE Conferences are open for a review of new enabling technologies that may cope with new requirements but also look at sustainability given the changes of behaviour. There is as yet no complete understanding of the economics of electronic publishing in the private and the public sector and the various functions in the value chains. Coherent usage indicators are needed to measure the actual use of scientific information as a basis for future funding in the public sector and investments in the private sector. These and many other topics will be addressed in January 2009.

APE 2009 will offer a very interesting program with enough room for discussions. We look forward to seeing you (again) in Berlin!

www.ape2009.eu **info@ape2009.eu**

INFORMATION SERVICES & USE

Aims and Scope

Information Services & Use is an information and information technology oriented publication with a wide scope of subject matters. International in terms of both audience and authorship, the journal aims at leaders in information management and applications in an attempt to keep them fully informed of fast-moving developments in fields such as: online systems, offline systems, electronic publishing, library automation, education and training, word processing and telecommunications. These areas are treated not only in general, but also in specific contexts; applications to business and scientific fields are sought so that a balanced view is offered to the reader.
"Information Services" are considered by the journal to comprise both the entities and the functions with which they are involved. These subjects are therefore considered as distinct from information uses which are concerned with service applications. Under this concept, the journal reports on intellectual processes for the organization of information, transformation techniques, media, economics and all varieties of support to services, including education and training. Education and training are also the essential elements for the other facet of the journal, the use of information. Information technology available, at the moment, is solutions waiting for problems.

Published quarterly 0167-5265/08/$17.00 Printed in the Netherlands

INFORMATION SERVICES & USE

Subscription information
Information Services & Use (ISSN 0167-5265) is published in one volume of four issues a year. The subscription price for 2008 (Volume 28) is EUR 350 (for electronic subscription) + EUR 36 p.h. = EUR 386 (US$ 503) (for print subscription or for a combined electronic and print subscription). The Euro price is definitive. The US dollar price is subject to exchange-rate fluctuations and is given only as a guide. 6% VAT is applicable for certain customers in the EU Countries. Subscriptions are accepted on a prepaid basis only, unless different terms have been previously agreed upon. Personal subscription rates and conditions, if applicable, are available upon request from the Publisher. Subscription orders can be entered only by calendar year (Jan.–Dec.) and should be sent to the Subscription Department of IOS Press, or to your usual subscription agent. Postage and handling charges include printed airmail delivery to countries outside Europe. Claims for missing issues must be made within six months of our publication (mailing) date, otherwise such claims cannot be honoured free of charge.

Instructions to authors
For detailed instructions please refer to the Authors Corner on our Web site www.iospress.nl.

Submission of manuscripts: Authors are requested to submit their manuscript either (i) per e-mail to Johan van Halm; e-mail: johanvanhalm@cs.com or (ii) as hardcopy with an electronic file on cd-rom to either one of the Editors-in-Chief.

Preparation of manuscripts:

1. Manuscripts must be written in English. Authors whose native language is not English are recommended to seek the advice of a native English speaker, if possible, before submitting their manuscripts.
2. Manuscripts should be typed on one side of the paper only, with wide margins and double spacing throughout. For the electronic file of the text you may use any standard word processor. Do not use page layout software.
3. The title page should contain (i) a title, (ii) author's name(s), (iii) affiliation(s), (iv) an abstract, (v) a complete *correspondence address*, including a telephone number, a fax number, and an e-mail address.
4. Each *table* should be provided on a separate page. Tables should not be included in the text.
5. *Figures* should (i) be numbered, (ii) not be included in the text, (iii) be reducible, (iv) have their captions typed on a separate page. Please note that *color figures* must be paid for. The digital files should (i) have a minimum resolution of 600 dpi, (ii) preferably be saved in eps or tif, (iii) NOT been taken from the Internet.
6. *Footnotes* should be kept to a minimum, and they should be provided all together on a separate page.
7. *References* should be listed alphabetically in the following style:

   [1] R.E. Jacobson, Approaches to total quality for the assessment of imaging systems, *Information Services & Use* **13**(3) (1993), 235–246.
   [2] K. Kocho, Multimedia networking: roadblocks and solutions, in: *Multimedia Communications. Forging the Link: Market – Technology – Policy*, D. Phillips and P. Desrochers, eds, Proceedings of the Multimedia Communications '93 Conference, IOS Press, Amsterdam, 1993.
   [3] J.N. Srivastava, *A Survey of Statistical Design and Linear Models*, North-Holland, Amsterdam, 1975.

Proofs: The corresponding author is asked to check the galley proofs. Corrections other than printer's errors, however, should be avoided. Costs arising from such corrections will be charged to the authors.

Reprints: For each contribution the corresponding author will receive one copy of the issue free of charge unless otherwise stated. An order form for reprints, additional issues, or a pdf file will be provided along with the galley proofs.

Copyright: Authors submitting a manuscript do so on the understanding that they have read and agreed to the terms of the IOS Press Author Copyright Agreement on our web site: www.iospress.nl.

Publisher

IOS Press
Nieuwe Hemweg 6B
1013 BG Amsterdam
The Netherlands

Tel.: +31 20 688 33 55
Fax: +31 20 687 00 19

E-mail:
General information: info@iospress.nl
Subscription Department: order@iospress.nl
Advertising Department: market@iospress.nl
Desk Editorial Department: editorial@iospress.nl

Internet:
www.iospress.nl