



HAL
open science

Giving new life to out-of-print books: when publishers' and libraries' interests meet

Françoise Vandooren, Cécile Gass

► To cite this version:

Françoise Vandooren, Cécile Gass. Giving new life to out-of-print books: when publishers' and libraries' interests meet. *Learned Publishing*, 2008, 21 (3), pp.187-192. sic_00289985

HAL Id: sic_00289985

https://archivesic.ccsd.cnrs.fr/sic_00289985

Submitted on 24 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Giving new life to out-of-print books: when publishers' and libraries' interests meet

Françoise VANDOOREN and Cécile GASS

Université Libre de Bruxelles – Libraries

© Françoise Vandooren and Cécile Gass 2008

Published in LEARNED PUBLISHING VOL. 2 1 NO.3 JULY 2008

ABSTRACT. The Library of the Université Libre de Bruxelles and Editions de l'Université de Bruxelles (EUB), the University's publishing imprint, have recently agreed to collaborate to provide free online access to recent out-of-print books published by EUB. The e-books are available on the Digithèque website, a collection of digital copies of printed books created by the Library. This initiative is valuable for the scientific community and the general public who can freely access the books online, for the authors whose books have been digitized and widely disseminated, and for the publisher whose collections become more visible on the Internet, thereby generating more traffic on its website and potentially increasing sales of its other books. Around 20 books have been made available online so far. This article describes the context of the agreement, how the collaboration operates, the options of file conversion vs. book scanning, issues relating to copyright and users' rights, how access is provided to the digital copies, and future collaborative projects of the Library and EUB.

Introduction

Universities have the great advantage that they bring together all the different types of actors involved in the scholarly publication process: authors, reviewers, publishers, libraries, readers. When they join together and think about their common interests, this can lead to improvements in access to publications for the whole scientific community. This is what happened at the Université Libre de Bruxelles (ULB), when the library and Editions de l'Université de Bruxelles (EUB), the University's publishing imprint, decided to collaborate to provide online access to recent out-of-print books.

EUB publishes research monographs and university textbooks by authors from the European Union; it publishes in some 20 domains, including law, economics, political science, philosophy, history and geography. The publications are organized in 18 collections, each one dealing with some specific subject areas – social and cultural anthropology, philosophy and society, spirituality and free thinking, etc. Most collections are directed by a professor of the ULB; each book manuscript is normally submitted to at least two referees. Around 20 books are published annually, and these are sold through both specialist and general bookshops. The publishing costs are defrayed by a combination of sales revenue, public funding and subsidy by the university. EUB is a very small operation (three full-time employees), which has not yet turned to electronic publishing.

However, at the end of 2005 the ULB Library started digitizing ancient books from its collections and providing free online access to the digital copies through its Digithèque (digital library) website¹. The books selected for scanning by the library date from the late 19th century and early 20th century; are in bad condition but still very much used or of high intellectual value; and are out of copyright and out of print. Books are considered out of copyright if their authors have been dead at least 70 years; they can reasonably be considered out of print by checking specialized databases, like Electre, and online libraries². The objective of the Digithèque is to make digital copies widely available, while preserving the original printed books.

In order to meet the demand of the scientific community for continued access to more recent books which are out of print, and relying on the existing technological expertise of the digital library service,

EUB and the Library agreed to include in the Digithèque digital copies of out-of-print books published by the EUB. Around 20 books published by EUB between 1974 and 2004, relating to political science, sociology, European law, history, history of religions, statistics and the environment, are currently freely available online. This initiative was very much welcomed by the university community as it brings advantages to those concerned: to readers, both in the scientific community and in the general public, who can freely access the books online; to authors, whose books become available worldwide through the Internet – these authors can then provide their students and colleagues with links to the online copy, the books can be disseminated through the network, read online and indexed by search engines, thereby potentially increasing the number of citations and impact and enhancing the author's profile and scholarly career. To the publisher, it brings the advantage of making its books and collections more visible on the Internet, thereby generating more traffic on its website and potentially increasing sales of its other books³.

How it works

EUB selects the books to be integrated in the Digithèque, the principal criterion being that it is out of print. Priority is given to the most recent books. EUB provides the Library with a PDF copy, if there is one, as well as a paper copy of each title. If there is no digital file available, the Library scans the book (see Figure 1), applying the technical requirements and the workflow of the Digithèque scanning activities: digitization is performed with full colour acquisition, at 250 or 350 dpi for special characters or schemes; the resulting images are restored (removing the black outline and the fingerprints, and straightening the curved text); the resulting full-colour TIFF files are named following the Digithèque's naming convention and archived; then the colour files are converted into black & white (binary code) to reduce their size. The Library produces the final digital PDF copies, adding a scanned image of the book cover, a bookmarked table of contents, enabling users to go directly to a given chapter of the book, and information about copyright and users' rights. The digital copies are stored on the Digithèque server, which provides a permanent URL for each file, i.e. a URL that the institution undertakes not to change for.

The establishment of this project required a total of around 220 hours of labour, as follows:

- trials to establish feasibility of migrating the existing digital files into current readable format: 11 hours;
- complete digitization process, from scanning through to enhanced PDF file creation, to uploading the files onto the server, to completing the catalogue of bibliographic records: 118 hours;
- website creation and communication: 80 hours;
- workflow follow-up, i.e. following up digitization and all other operations (PDF creation, update of the catalogue and of the website, etc.) : 12 hours.

No additional funding was specifically allocated to this initiative; both EUB and the Library are utilizing existing human and technological resources, the cost of which it was not possible to estimate.

File conversion or scanning?

At the very start of the project, EUB provided PDF text files for some books, floppy disks containing a variety of file formats⁴ for others, and in the remainder of cases printed copies only. The Library spent some 11 hours trying to read these disks and convert the files into a manageable format with minimum loss⁵; it was felt that this work would be worthwhile, in order to obtain fully searchable and manipulable text files, rather than scanned image files. However, the experiment was not a success; at best, this conversion produced partially readable text files with seriously disrupted layout, which had to be completed and fully reformatted.

Since funding and human resources were not available to clean up these files, scanning the books turned out to be more efficient; the Digithèque can produce a digital PDF copy, in image mode, of a 300-page book in around six hours. Thus the EUB books were added to the list of works to be scanned by the Digithèque. Attempts to use optical character recognition (OCR) on the scanned copies were abandoned because, in addition to the few errors which this automatic recognition process inevitably produces, the same layout problems occurred.

Copyright and users' rights

EUB's publishing contracts require authors to transfer their copyright. However, the contracts for the books to be included in the Digithèque did not explicitly mention any form of digital publication. The EUB therefore contacted the authors of the selected titles to seek their agreement for inclusion of their book online, free of charge, in the Digithèque. The details were given in a contract amendment. All the authors agreed to the new terms, and indeed were very enthusiastic about open access digital dissemination. Where a book chapter had been originally published as an article in a journal to which the author had transferred copyright, and included in the print book by permission of the journal publisher, the EUB assumed it could not be included in the online book version; in such instances a page specifying where the content may be found replaces the missing chapter in the digital copy.

Although the authors are required to grant digital publication rights to EUB, it was agreed that the online version of the book would be made freely available and that the users' rights would be as defined for the other items included in the Digithèque. Users' rights are specified on the Digithèque website, as well as at the end of each digital copy⁶: the digital documents can be used for research, teaching and private purposes as long as the references are fully cited. For any other use, permission must be requested from EUB. Deep links, providing direct access to a particular digital document, are authorized provided that the ULB is clearly identified as being the access provider.

Authorized uses include: downloads, copies and storage of the whole digital document (other than deposit in another database, which is prohibited); exact facsimiles, prints and photocopies; and copying and pasting of text.

Maximizing visibility and access

In order to maximize the visibility and accessibility of the online books, various cataloguing options were considered:

- The books were already recorded in the Library online catalogue; the permanent URL of each digital copy has been added in the corresponding book record in the catalogue. Since the library catalogue is compliant with standard communication protocols (e.g. OAI-PMH and Z39.50), harvesting and federated search engines that are using these protocols⁷ can search the metadata from the ULB catalogue and provide links to the ULB book records. This therefore enhances the efficiency of the new digital documents search and discovery.
- A dedicated section was created on the Digithèque website to promote the digital book collections of EUB⁸, listing the titles, authors, and subject areas (see Figure 2). The website is indexed by general search engines such as Google, which again increases the visibility of the digital books⁹.
- Finally, the digitized books were reinstated in the EUB online catalogue¹⁰, which gives the URL of the digital copies.

In addition, the EUB online catalogue was added to the Library link server. A link server provides users with context-sensitive links to books, journals, and other library defined resources. It means that in this case, the user is redirected to the EUB website when looking for other books from the same author. All EUB had to do was to enhance its website so that it could handle OpenURL requests, i.e. the standard communication protocol on which link servers are based.

Conclusion: building on complementary skills

Following the recommendations of a study¹¹ commissioned by DG Research–European Commission, and to which some ULB researchers contributed, the ULB engaged in initiatives supporting open access to scientific information. The rector of the ULB signed the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities¹², which specifies that scientific literature should be freely available to any user connected to the Internet, and the collaborative agreement between EUB and the Library is one way in which this declaration is put into practice.

In addition to the Digithèque, other open access initiatives of ULB include BICTEL/e (the Directory of ULB electronic theses and dissertations)¹³ and its Institutional Repository, currently under development, which aims to collect and provide online open access to publications and scientific works by ULB professors and researchers¹⁴.

As well as continuing to upload digital copies of EUB out-of-print books to the Digithèque on a regular basis, EUB and the Library have already agreed on a number of future projects:

- Digitizing and providing online open access through the Digithèque to back issues of three EUB book series in history, 57 volumes in all (*Etudes sur le XVIII^e siècle*, *Problèmes d'histoire du christianisme*, *Problèmes d'histoire des religions*); there will be a 'moving wall' of seven years, i.e. the time period between the last issue available in the Digithèque and the current year. In this instance the digital files will undergo optical character recognition, which will then enable full-text search¹⁵. Specific funding has been allocated to this project by the Research Incentive Fund of the University, primarily to cover the cost of scanning (by working students)
- Publishing electronic-only university-level and research-oriented books. The Library is currently investigating the long-term preservation issues related to the contents of both the Digithèque and the Institutional Repository. An electronic archiving process will have to be set up¹⁶, which will automatically be applied to the EUB digital titles included in the Digithèque.

As was pointed out in a recent ARL publication¹⁷, although university presses know what needs to be done, they often lack the financial capital, technical staff and technological skills to engage in electronic publishing. By collaborating with libraries, which can provide the tools and infrastructure to support new forms of publishing, they can bring together the unique skills and experience of each, to engage in a common strategy to develop new open access electronic products.

References

1. More information: <http://digitheque.ulb.ac.be>.
2. So far, we do not digitize books that are under copyright or that are still published. However, we face problems with orphan books, i.e. works for which rights-holders cannot be identified or located. We look forward to solutions to this issue at a European level. See the European Digital Library Initiative High Level Expert Group – Copyright Subgroup report in 2007.
http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=3366
3. Some Publishers Warm to Google Book Search, *Chronicle of Higher Education*, 5 June 2007,
<http://chronicle.com/wiredcampus/index.php?id=2122>
4. In 30 years of existence, EUB have changed their publishing software at least twice, migrated from the Macintosh to the Windows environment, and followed the software versioning.
5. Faced with the same issue, one may find the following links useful: PRONOM database (<http://www.nationalarchives.gov.uk/PRONOM/BasicSearch/proBasicSearch.aspx?status=new>), Adobe forum (<http://www.adobe.com/cfusion/knowledgebase>) and the w3.org website (http://www.w3.org/Tools/Word_proc_filters.html).
6. Note that the copyright status of the book is indicated on the front page of the digital copy, with a hyperlink to the users rights placed at the end.
7. See, for instance, OAlster www.oaister.org and federated search engines like MetaLib from ExLibris.

8. <http://digitheque.ulb.ac.be/fr/digitheque-des-editions-de-luniversite-de-bruxelles/index.html>.
9. The Digithèque was also advertised in some Wikipedia records, and on some mailing lists that are expected to increase traffic.
10. Editions de l'Université de Bruxelles: <http://www.editions-universite-bruxelles.be/>
11. Study on the Economic and Technical Evolution of the Scientific Publication Markets in Europe. M. Dewatripont *et al.*, European Communities Publication Office, January 2006.
http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf
12. The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
<http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
13. <http://www.bib.ulb.ac.be/fr/bibliotheque-electronique/theses-bictele/index.html>
14. <http://www.bib.ulb.ac.be/fr/bibliotheque-electronique/depot-institutionnel/index.html>
15. Recognised text would be automatically enhanced by specially developed dedicated software, enabling the book layout to be reproduced more accurately, and correcting most of the remaining mistakes.
16. A state-of-the-art survey is under way.
17. ARL A Bimonthly Report on Research Library Issues and Actions from ARL, CNI and SPARC. 252/253 Special Double Issue on University Publishing, June/August 2007.

Françoise Vandooren and Cécile Gass
Université Libre de Bruxelles – Libraries
Avenue F. Roosevelt 50 – CP 180
1050 Bruxelles, Belgium
Email: françoise.vandooren@ulb.ac.be ;
cqass@admin.ulb.ac.be