

# Document supply of grey literature and open access: an update

*Joachim Schöpfel*

*Hélène Prost*

## Abstract

**Purpose:** The article investigates the impact of the open archive initiative on the document supply of grey literature.

**Approach:** The article is based on a comparative survey of five major scientific and technical information centres: The British Library (UK), CISTI (Canada), INIST-CNRS (France), KISTI (SouthKorea) and TIB Hannover (Germany).

**Findings:** All major document suppliers are more or less deeply involved in the open archive movement, and this involvement has an obvious impact on the policy of acquisition, archiving and supply of grey literature (dissertations, reports, conferences etc.).

**Originality:** The article is a follow-up study of our survey published in 2006.

## Keywords:

Grey literature, scientific and technical information, document supply, open archive initiative (OAI), open access, institutional repositories, e-Science, STI centres

**Paper type:** research

## Introduction

In 2005, we conducted a survey on open access (OA) projects and the document supply of grey literature, based on data collected from five major scientific and technical information (STI) centres<sup>1</sup>, (Boukacem-Zeghmouri & Schöpfel, 2006). Our main findings were:

The STI centres placed special emphasis on grey literature and had important grey collections, especially of conference proceedings, technical reports and dissertations. Nevertheless, the relative part of grey document supply differed with INIST and the British Library having a grey document supply of up to 5%, and CISTI and TIB Hannover with more than 10%.

The “grey supply” generally followed the overall downward trend of document supply overall. Nevertheless, in two STI centres (CISTI and TIB) the supply of grey literature slightly increased.

Due to their public mission, all institutions were interested and involved in open access projects. Their specific involvement depended on the integration into the national information market and institutional environment (higher education, research communities) but also in financial and human resources. One part of these open access projects was related to traditional grey literature.

This involvement did not impact their traditional functioning and activities in any significant way. We observed little impact of OA on acquisition policy or service

---

<sup>1</sup> The British Library (UK), CISTI (Canada), INIST-CNRS (France), KISTI (South Korea) and TIB Hannover (Germany).

development. Only a few changes were noted in the information systems on which the supply of grey literature is based, or in the bibliographic control.

Our 2005 survey showed a great diversity between the document suppliers on a topical issue which directly concerns them. We suggested that access to grey literature in an electronic context may have greater economic potential than in the traditional paper era and that the commitment of the document suppliers in the domain of open access to the distribution of grey literature may be a strategic means of establishing their position in the broader scientific and technical information market.

Four years later, our intention is to provide more evidence on the relationship between the OA movement, document supply and grey literature. Since 2005, the OA movement has steadily developed. According to the statistics from the two main OA directories [1], the number of OA journals and repositories increases at 20-30% per year. In May 2009, the cited directories included more than 4,200 journals and 1,400 repositories.

Compared to 2005, open access has become a central part of the scientific and technical information market, offering free and seamless dissemination of 10-20% of current scientific production and challenging the subscription-based business model of academic publishing.

In the past, STI centres played a central role in the value chain for print journal publishing. Today, they face two threats: the developing e-commerce from the main academic publishers with growing disintermediation, and the open access movement with community-based direct communication between scholars.

Our conviction is NOT that these threats will necessarily destroy STI centres' service offerings<sup>2</sup> but that they will deeply affect their functioning and strategies. A recent report on nine American universities<sup>3</sup> showed that on the user side "institutional repositories (IR) and open access (OA) materials (may not) have ... substantially impacted interlibrary loan services ...". Most of the participants report the same or an increased volume of business", (Kelsey, 2009). Nevertheless, they also report rather low use of "commercial suppliers"<sup>4</sup>. So what about these service providers?

The head of sales and marketing at the British Library stated one year ago "(... the last five to seven years have been a roller-coaster ride for our document supply service", (Pfleger, 2008). Is this true for all STI centres?

As in our 2005 survey, our 2009 focus is on grey literature<sup>5</sup> and document supply. Why this focus on grey literature? Because of the importance of grey resources for scientific research and teaching, all major public document suppliers invest in collections and delivery services for theses, conference proceedings, reports and unpublished working papers. These special collections are costly and "grey supply" is often more expensive than article supply.

On the other hand, grey literature represents a significant part of the content of institutional and other repositories and is more and more freely available on the Web but not always easy to find. Therefore our underlying assumption is that grey literature may be a sensitive indicator for the evolving strategy of STI centres.

---

<sup>2</sup> Even if some document suppliers already disappeared (eg. NIWI the Netherlands Institute for Scientific Information which is an Institute of the KNAW, the Royal Netherlands Academy of Arts and Sciences)) or restricted their service (eg. TU Delft).

<sup>3</sup> University of Texas at Arlington, Tulane University, University of Minnesota, Indiana University-Purdue University Indianapolis, Brigham Young University, University of Tennessee, Colorado State University, Oberlin College and Stony Brook University.

<sup>4</sup> Among these "commercial suppliers" are Ingenta, the British Library, CISTI, NTIS, InfoTrieve and the National Library of Medicine.

<sup>5</sup> "Grey literature (is) outside of the realm of commercial publishers, and (...) rather ephemeral—often poorly controlled by catalogs, databases, and bibliographies" (Schöpfel & Farace, 2009).

## Questions and methodology

Our follow-up study reproduces largely the methodology of the initial survey (Boukacem-Zeghmouri & Schöpfel, 2006). The survey sample remains the same as in the initial study:

The *British Library* (BL) [2]

The NRC *Canada Institute of Scientific and Technical Information* (CISTI) [3]

The French CNRS *Institut de l'Information Scientifique et Technique* (INIST) [4]

The *Korean Institute of Scientific and Technical Information* (KISTI) [5]

The *German National Library of Science and Technology* at Hannover (TIB) [6]

These traditional document suppliers have in common a public mission to collect, preserve, archive and disseminate scientific information through a non-profit ILL and document supply service that is based on a mixed economic model with their income supplied both by public funding and their customers' fees. ILL and document supply networks without holdings and corporate, profit-based suppliers are excluded from the sample. The data collection can be described in the following way:

(a) We searched for open source information about the development, services and projects of the sample on the institutional websites, in activity reports and published articles.

(b) We asked each institution for information on the following topics:

1. Figures on their grey document supply and ILL in 2008.
2. Comparison of these figures to the overall supply and ILL (%).
3. The recent evolution compared to previous years.
4. Their projects in the area of grey literature.
5. Their open access projects.
6. The impact on the collection of grey literature.
7. The impact on document supply (service offer, pricing).
8. The impact on the bibliographic control of grey literature (cataloguing, record data).
9. The impact on the information system.

(c) We communicated the data synthesis to the institutions for comments and validation.

The results are analysed in two ways: a comparison between the five institutions, and a comparison with the results published in 2006.

## Results

In the following, we present the data and information for each STI centres.

### *The British Library*

The British Library's grey holdings – mainly dissertations, reports and conference proceedings – are extremely rich with 10.3 million reports in microforms, 13.7 million patents specifications, 164,265 theses, 4.3 million cartographic items etc, (British Library, 2008). The British Library was the most important national input centre in the European EAGLE network. Recently, the British Library “has been charged with achieving cost recovery for its document supply operation within two years – by March 2011”, (Prowse, 2009) and will work on a sustainable business model for the document supply service through increased efficiency, reduced costs and improved productivity (see British Library, 2008).

**Grey document supply and ILL in 2008:** We have no updated data but from the most recent figures, we can estimate that BLDSO received around 70,000 requests for grey

literature in 2008. The satisfaction rate for supplying grey literature was 85% in 2003.. (Boukacem-Zeghmouri & Schöpfel, 2006)

**Comparison with the overall supply:** This probably represents 5% of the total items supplied in 2008.

**Evolution:** Since 1998/1999 the British Library has experienced a significant decrease in remote document supply (RDS). Even if the exact number of requests is no longer published in the annual reports (“commercially sensitive”, see Prowse, 2006), the decline can be estimated at 10% per year. The face of document supply changed: “In 2002, over 50 % of material demanded was for papers published in the last two years. In 2007 this half-life had moved to five years and the spread across publishers has increased” (Pfleger, 2008). Pfleger reports 1.6 million requests for 2007.

**Projects in the area of grey literature and OA projects:** The British Library contributes to OA projects mainly through the Research Support Libraries Programme (RSLP), which indicates an improved relationship with the UK higher education community. The EthOS system provides improved access to UK theses. EThOS [7] is an open access repository providing electronic access to UK doctoral theses (immediate access to the full text of 12,000 theses, growing to 100,000 theses within three years and an option to request digitisation from 250,000 paper-based theses). The system was developed by the EThOS partnership comprising 90 UK Higher Education Institutions and the British Library with funding from the Joint Information Systems Committee (JISC), Research Libraries UK and partners. The service was launched in January 2009 and is currently in beta version (see Prowse, 2009).

The BL is developing two new subject-focused websites that will encourage publishers of research reports, working papers and other grey literature to deposit their print and/or electronic material. The first of these websites focuses on the Olympics, a topic which is attracting increasing research interest in the run-up to London 2012. It aims to provide a hub of information about resources for the study of the Olympics. It will be launched in summer 2009. The website includes a section on the print and digital legacy of the Olympics and the need to ensure material is not lost forever, plus a form for publishers to contact the BL in order to deposit print and digital material.

The second website focuses on management and business studies (MBS). Its target audience is academics and senior practitioners with an interest in the latest management research. It aims to bring together the BL print and digital collections for this subject area in an interactive Web 2.0 environment. The main professional bodies and learned societies for this subject-area in the UK are working with the BL to promote the new website to their members. The added value lies in bringing together in one place so much high quality content for this subject, together with user-generated content and subject expertise. In terms of grey literature, the BL is targeting 30 key UK publishers whose outputs are considered by users in this subject-area to be high quality. The BL approaches the publisher to actively encourage them to supply print and digital material in a timely manner, and if they are agreeable, they obtain permission to harvest and republish their digital material on the MBS Portal as well as adding it to the British Library's Digital Library Store for long-term preservation. It will have a UK focus and be launched in October 2010.

UK PubMedCentral [8] was launched in 2007 and provides a stable, permanent, and free-to-access online digital archive of full-text, peer-reviewed research publications. It is based on PubMed Central (PMC), the US National Institutes of Health (NIH) free digital archive of biomedical and life sciences journal literature, and is part of the network of PMC international repositories,[9]. Through UK PubMed Central the BL aims to provide a freely accessible, UK-based archive of biomedical and health research findings. The BL is leading a partnership to host, manage and develop UKPMC on behalf of the Funders, [10]. Its ambition is to

become the information resource of choice for the UK biomedical and health research communities. Current developments of UKPMC include:

- Establishing a comprehensive, sustainable repository for UK-funded research outputs.
- Improving information retrieval and knowledge discovery through the development of text and data-mining solutions.
- Providing access to additional content that integrates seamlessly into the UK PubMed Central website.
- Creating comprehensive analysis and reporting tools for researchers and funders to inform strategy and policy making.

The British Library is leading a work-package on additional content. The work-package will identify relevant content and make it discoverable through UKPMC. Much of the additional content being identified is grey literature including, clinical guidelines, single-issue research reports and theses.<sup>6</sup>

Other projects at the edge of grey and OA are the selective archiving of websites, the archival sound recordings project and the digitisation of selected collections (“hidden treasures”) in the context of the UK Digital Preservation Coalition. Some “Co-operation and Partnership Programmes” are focused on the preservation, digitisation and display of national heritage resources and non traditional items: for example, legal materials, company reports, official publications. The 19<sup>th</sup> century British Newspapers website, launched in October 2008, makes over two million searchable pages of historic newspapers available online to the UK’s Higher and Further Education communities, [11].

**The impact on the collection of grey literature:** No updated information available.

**The impact on document supply:** “With UK PubMed Central and EThOS the British Library will be making material freely available that would previously have had to be obtained via RDS. That seems to be the way that much RDS has been going. Previously it was quite expensive, took a while and had to be done via an intermediary; increasingly the documents traditionally obtained via RDS are free and available directly to users immediately. It is an interesting turnaround is it not?”, (Prowse, 2007).

**The impact on the bibliographic control of grey literature:** The British Library is rethinking its approach to catalogues, (see Brazier, 2007). Topics include:

- i. better integration into new resource discovery services via a single entry point and with links to RDS, definition of baseline quality,
- ii. enrichment of records,
- iii. integration with digital libraries,
- iv. cataloguing of non-textual material (sound recordings),
- v. development of interactive Web 2.0 services.

**The impact on the information system:** Open access to scientific information is part of the demands and expectations mentioned in the British Library’s Strategy 2008–2011 [12]. The impact on the library’s information system will be multiple: improved digital storage capacities, a new integrated archives and manuscripts system, a new resource discovery system with integrated Web 2.0 functionalities, a central repository system for UK Higher Education research, new services for the UK e-infrastructure (virtual research environments, storage and access to datasets etc.), development of the library system with improved digital rights and policy management, etc.

*Acknowledgments to Elizabeth Newbold for providing data and information on document supply and open access to grey literature.*

---

<sup>6</sup> Further information on the development of UKPMC are available at: [http://ukpmc.ac.uk/ppmc-localhtml/future\\_plans.html](http://ukpmc.ac.uk/ppmc-localhtml/future_plans.html)

## ***CISTI***

CISTI (or NRC-CISTI) is part of the Canadian National Research Council (CNRC) and Canada's National Science Library. It is not only one of the largest scientific and technical libraries in North America, but also one of most important and appreciated document suppliers with a global activity. It is also the largest publisher of scientific books and journals in Canada, (Ireland, 2008). Its holdings contain over 50,000 scientific journal titles, more than 800,000 books, conference proceedings and technical reports and two million technical reports in print and on microfiche. CISTI makes a special effort to locate, purchase and catalogue conference proceedings from around the world, its collection of published scientific conference proceedings is one of the best in the world.

**Grey document supply in 2008:** In 2008 CISTI received 45,261 requests for grey literature, most of them for conferences proceedings.

**Comparison of these figures to the overall supply:** Requests for grey literature in 2008 accounted for 9.2 % of the 490,000 requests received by CISTI.

**Evolution:** Since 2004, the supply of grey documents has decreased by 27% (16% from 2007 to 2008). The total number of grey literature requests has declined in proportion to the decrease in overall requests remaining at 9-10% of the total. From 2001 to 2008, the average number of requests fell from more than 4,000 to 2,000 per day, (Ireland, 2008).

**Projects in the area of grey literature and OA:** CISTI has launched a digital institutional repository for NRC publications (NPArc - the NRC Publications Archive). The new search interface "Discover" [13] contains more than 20 million STM article records but no grey material so far; nevertheless, the integration of metadata from reports or conference proceedings seems to be one of the future options.

In May 2009, the CISTI launched a new service called "Gateway to Scientific Data" that provides access to Canadian STM data sets from a broad range of disciplines [14] and to selected policies and best practices guiding data management and curation activities in Canada. [15]: for instance, in June 2009, CISTI links to 12 data providers in biological sciences, to 11 data providers in genomics and to 11 data providers in environmental sciences. The project is part of a national initiative in e-Science, Research Data Canada, [16] in favour of access to and preservation of primary research data from Canadian public research.

**The impact on the collection of grey literature:** The availability of grey literature on the Web does not seem to have affected CISTI activities.

**The impact on the information system:** The CISTI IT architecture is undergoing a fundamental change from primarily supporting a print-based remote document supply to supporting a digital library with gateway function and digital rights management. The drivers for this change originate more in the peer-reviewed publishing environment than from any OA or GL projects, (Ireland, 2008); e.g. the future IT system will include linking to the publishers' sites, linking from Google Scholar and OCLC, linking to the Copyright Clearance Centre, functionalities of e-commerce and Web 2.0. CISTI implemented a wiki called "CISTI Lab" [17] for collaborative test and evaluation of experimental and innovative services. The place of OA and GL in these new, mostly published article-based projects seems uncertain so far.

*Acknowledgments to Marsha Kaiserman and Michael Ireland for providing data and information on document supply and open access to grey literature.*

## ***INIST***

Founded in 1988, the Institute for Scientific and Technical Information is part of the French National Centre for Scientific Research (CNRS). Its mission is to collect, analyse and disseminate the results and findings of worldwide research in STM, social sciences and

humanities. The supply of copies of scientific and technical documents is part of its traditional activities, whether or not the documents are held at INIST.

The legal compliance of the INIST document supply service is guaranteed through agreements with the French copyright agency (CFC) and main academic publishers.

The INIST grey collections are significant: 69,000 conference proceedings, 155,000 theses and 65,000 scientific and technical reports. INIST is a central part in the French landscape of the collection and dissemination of grey public research documents.

**Grey document supply in 2008:** INIST received 11,707 requests for grey items (75% conference proceedings, 17% dissertations, 8% reports). 29% were delivered from INIST's own holdings, the rest (71%) was supplied through the INIST back-up network. The satisfaction rate was 94%. All items were delivered as print copies or through ARIEL with no supply of returnables. 49% of the grey items were requested by corporate (for-profit) customers.

**Comparison with the overall supply:** The supply of grey literature represented 3.5% of the overall activity. The satisfaction rate (94%) is lower than for the document supply of articles (97%). In 2008, 71% of the supplied grey items were from other French and foreign libraries. This means that the document supply of grey literature relies much more on the INIST national and international network than the overall supply : the INIST back-up network delivered 23% of all documents provided (see also Schöpfel and Gillet, 2007).

**Evolution:** Since 2005, the downward trend has slowed down and is currently about 5% per year. Between 2004 and 2008, the figures for grey literature supply (- 24%) decreased less than the overall activity (- 36%). The evolution of the major types of grey literature is various: while conference proceedings declined by 11% and reports by 7% since 2006, the demand for dissertations increased significantly by 51%.

**Projects in the area of grey literature:** From the five major projects in the area of grey literature mentioned in our 2005 survey, two have been completed (cataloguing of 14,000 French STM dissertations and the participation in the *Grey Literature International Steering Committee* (GLISC). Two other projects are related to the open access movement (see below).

**OA projects:** INIST continues its activities in the OA environment, with .:

- An OA platform for digital periodicals and conferences (I-Revues) based on the MIT *DSpace* and the French *LODEL* software.
- A website dedicated to OA issues with international news, reference texts and review articles (openaccess.inist.fr).
- A new initiative with European research libraries and information providers in order to establish a not-for-profit agency that will register research datasets and assign persistent identifiers (DOI).

Two OA projects are related to grey literature:

- The further development of the OA platform "LARA" for scientific reports based on the MIT *DSpace* software. The project includes the retro-digitisation of print reports and online deposit of digital material by the organisations that produce it. At present the LARA archive holds nearly 1,000 items.
- INIST successfully transformed the former SIGLE database into a *DSpace* open archive of European grey literature (OpenSIGLE, see Farace *et al.*, 2008). OpenSIGLE can be searched via Google and contains today more than 690,000 bibliographic records. INIST and TextRelease (Amsterdam) have started to add the full text of the GL conference proceedings (102 items so far). The next challenge will be the construction of a new international network (GL community) in order to maintain and improve the site and to foster the deposit of (and linking to) full text items.

**The impact on the collection of grey literature:** No updated information.

**The impact on document supply:** INIST offers traditional print and electronic delivery of PDF files with digital rights management (for DRM, see Gillet, 2007). URLs of freely available online resources are supplied. The evolution of INIST document supply will focus on article delivery, not on grey literature as such.

**The impact on the bibliographic control of grey literature:** Increased outsourcing of record production and metadata will increase visibility but decrease the bibliographic control of grey literature.

**The impact on the information system:** Since 2008, INIST has developed a new document supply system with e-commerce functionalities, further automation of request processing, improved linking to the backup library network, full integration of the electronic document supply and an augmented database (35 million items, mostly articles). The beta-version was successfully presented at the I-Expo 2009 exhibition in Paris. Additionally, INIST created a Web 2.0 version of the CNRS information portals. [18]

*Acknowledgments to Christiane Stock for providing data on the INIST holdings of grey literature.*

## **KISTI**

The Korea Institute of Science and Technology Information (KISTI) is a national information centre that has collected, analyzed and managed science and technology information resources comprehensively since 1962. KISTI also conducts research on technology, policies, and standardization concerning information management and dissemination. KISTI has contributed to the increase in national R&D productivity and industrial competitiveness through building the supercomputer infrastructure for the national R&D programmes.

KISTI builds essential Science and Technology databases on bibliographies, R&D reports and patents, and provides information services. It acquires Sci-Tech journals, e-journals, technical reports and conference proceedings. It is part of the Korea Knowledge Portal (Kim & Kwon, 2008). KISTI also supports online publishing through the ACOMS software, a system to manage e-journal workflows and to collect digital documents produced by more than 200 Korean academic societies, (Choi *et al.*, 2007).

**Grey document supply in 2008:** In 2008, KISTI supplied 6,300 copies from grey documents (reports 1,600, dissertations 1,300, conferences proceedings 3,400).

**Comparison with the overall supply:** Requests for grey document supply in 2008 accounted for 2.1% of the overall number of supplied items registered at KISTI.

**Evolution:** Compared to 2007, KISTI reports a decrease in the percentage of grey documents supply compared to from 2.9% to 2.1%, even though the number of grey documents remained higher than in previous years. The overall supply figure of 295,200 satisfied requests in 2008 is an increase of 14% since 2007.

**OA and GL projects:** Interest in open access and institutional repositories is increasing in Korea (Kim & Kwon, 2008) and the Korean scientific communities are pressing for a Korean repository for their publications, (Hwang *et al.*, 2006). KISTI contributes to the OA movement on different levels. Together with the Information Center for Physics Research (ICPS), it developed the first subject-based e-prints archive (Science Attic<sup>7</sup>, see Hwang and Choi, 2006). KISTI launched a current research information system, NTIS, for government funded research projects that provides access to research outputs (National Technical Information System<sup>8</sup>) and a multilingual electronic theses and dissertations (ETD) system as a

---

<sup>7</sup> <http://science-attic.org/>, nearly 600 items in June 2009

<sup>8</sup> <http://www.ntis.go.kr/>, more than 25,000 theses, patents and registrations in June 2009

part of the National Digital Science Links (NDSL).<sup>9</sup> KISTI also operates the KoreaScience portal [19] that provides access to Korean STI including OA items.

Two more recent OA projects are stOAI and the Korean National Repository. stOAI (Science & Technology Open Archives Initiative) is the OAI-based repository for international STI. Its objective is to improve access to OA journals for Korean scientists. Additionally, stOAI contains a wide variety of resources such as dissertations, trend analyses, Korean Patents, totalling 1,540,000 items in April, 2009. The Korean National Repository is a national project (2009-2012) for facilitating the OA movement and sharing knowledge and information. KISTI hosts the site and will create intellectual property (IP) for open access content from Korean journals, R&D results, research reports and dissertations; it also works on improving IP laws and regulations.

KISTI has been working on several grey literature projects since 2006. One major project is to develop an acquisition and distribution model for international grey literature, through usage assessment and evaluation of access conditions for foreign grey literature in Korea.

**Impact on collection, supply, bibliographic control or system:** No information.

*Acknowledgments to Sunae Lee for providing data and information on the document supply and open access to grey literature.*

### ***TIB Hannover***

One of the three special scientific libraries in Germany, the German National Library of Science and Technology in Hannover (TIB) celebrated its 50<sup>th</sup> anniversary in June 2009. TIB defines itself as a transfer centre for scientific knowledge; its task is “to comprehensively acquire and archive literature from around the world pertaining to engineering and the natural sciences”, [20]. TIB holds 5.4 million volumes, 3.4 million microforms, 18,300 subscriptions and 14.4 million patents. The library places a particular emphasis on the acquisition of grey literature (conference proceedings, research reports, standards and dissertations in print and digital format). Its grey holdings are unique in Germany. The TIB is closely linked to, but independent of, the Hannover university library (UB). It has developed a strategic partnership with the other German scientific libraries (ZBW and ZBMED) and scientific information centres (FIZ), and on the European level TIB promotes a decentralised network for scientific information, especially for research data (for more details, see Meyer, 2009).

**Grey document supply in 2008:** TIB received about 113,000 orders for grey literature, most of them for scientific and technical reports from its own holdings.

**Comparison with the overall supply:** The requests for grey documents represented 35% of the 322,000 requests in 2008.

**Evolution:** Document supply (excluding lending) has decreased since 2005 by 39%. There are two reasons: firstly changes in the German copyright law since January 2008 which mean that no electronic delivery is allowed outside of licence contracts, except for academic users and then only if there is no obvious pay-per-view-offer; secondly there is a very good offering of national licences for academic libraries [21]. The decrease of 15% in grey document supply was less significant. As a result, the proportion of grey material in document supply increased from 27% in 2005 to 35% in 2008.

**OA and GL projects:** TIB participates in a large number of projects, with a focus on electronic publications and the development of a digital library, especially for non-textual material and for long-term preservation. Some current projects are on grey literature and/or open access:

---

<sup>9</sup> <http://www.ndsl.kr/> NDSL provides one-stop service by offering information on 40,000 academic journals, 2,500 million theses and other materials.

- **Publication and Citation of Scientific Primary Data:** To make primary scientific data citeable as publications, several organisational and technical pre-conditions have to be met: quality control of the primary data set and the descriptive metadata, long-term availability of the published data in online repositories, a search function for data publications in library catalogues (e.g. GetInfo), access to the primary data with assignment of a persistent identifier and resolver system, for example a DOI [22] resolver. TIB has developed a platform for scientific data publishing and initiated a European network to establish a not-for-profit DOI registration agency that enables organisations to register research datasets and assign persistent identifiers to them (Brase, 2009).

- **PROBADO:** The use of complex, non-textual data and documents is becoming more and more important. However, today's digital libraries do not support these data optimally because they are based on the assumption that all documents can be described textually. The goal of PROBADO [23] is to develop tools and systems as well as create methods and workflows that allow academic libraries to treat non-textual documents in the same way as textual documents, e.g. 3D computer graphics (architecture) and music.

- **SCOAP<sup>3</sup>** (Sponsoring Consortium of Open Access in Particle Physics Publishing): SCOAP<sup>3</sup> will, for the first time, link quality and price, stimulating competition and enabling considerable medium- and long-term savings [24]. Each country will contribute according to its share of High Energy Physics publishing. In Germany there are three partners (Max-Planck-Society, DESY and TIB). TIB will develop and organize a model for the contribution of the German universities, (see Fournier, 2007).

**The impact on the collection of grey literature:** The mission of TIB is to collect all significant publications in its scientific domains (science and technology), at least those published in the most important languages. This objective also applies to grey literature. As these documents are more and more only available in a digital format, TIB tries to obtain authorization from the copyright owners to download the files to a local server and, if possible, to obtain or produce a print copy for archiving and usage. TIB may negotiate this permission individually or at the institutional level. For all the projects funded by the German Federal Ministry of Research, TIB is the deposit library for the digital project reports. These reports are freely available online through the service GetInfo. TIB plans (together with FIZ Karlsruhe) to store open access publications in its fields from researchers of the Leibniz-Gemeinschaft. These documents may also be published through commercial channels, (the green road). TIB's task is to make the literature available to its customers in the long term, an important task especially with regard to grey literature. TIB is also involved in restoration and long-term preservation projects.

**The impact on document supply:** The main impact comes from the new German copyright law (see above) but also from the offer of national licences not only for back files but also for current content; TIB is heavily involved in the negotiation process. While the document supply requests decrease, the access to licenced items is growing rapidly with three million downloads in 2008. Additionally, TIB continues to negotiate licences for pay-per-view-options (ppv), especially (but not exclusively) for the corporate sector. Currently more than one million articles are available via pay-per-view.. TIB's objective is to provide community-specific solutions for direct full text access, whether through national or local licences, ppv or open access, (see Rosemann, 2008).

**The impact on the bibliographic control of grey literature:** The collection and bibliographic control of grey literature is the same whether or not it is open access. TIB tries to obtain permission to download, store and archive the grey literature and to give access to customers for the long term. Irrespective of this, all grey documents in TIB are catalogued. In addition, TIB facilitates access by integrating metadata of grey literature into GetInfo.

Finally through ScanTOC, TIB facilitates access to more than 100,000 conference proceedings.

**The impact on the information system:** Since the launch in 2007 of the new IT system GetInfo<sup>10</sup>, TIB offers a customer-friendly interface based on the search engine “Lucene”. It is an integrated system for search and full-text options (print and electronic, licences and pay-per-view) with additional integrated filter mechanisms : It’s possible to limit a search to either an author, a document format, a document type, an year of publication, a language, a database or a publishing house. Searches are stored in a historical TIB will partner increasingly with the two other special scientific libraries in Germany.

*Acknowledgments to Irina Sens for providing data and information on the document supply and open access to grey literature.*

## Discussion

All STI centres were willing to participate in this survey and contributed with detailed and (partly) insider information. More facts were gathered from open sources (websites, blogs, newsletters, reports, conferences and journal articles). Short-living, abandoned or finished projects like CISTI’s e-book venture with Ingram Digital (*MyiLibrary*) or INIST’s project of a specific French and international dissertation service were ignored

Together, this information provides an interesting and contrasted image. Compared to our 2005 survey, the changes in the functioning and activities are significant.

### *Document supply – overall and grey*

Based on the survey results and open source information, the situation of the document supply services can be summarised as follows (table 1).

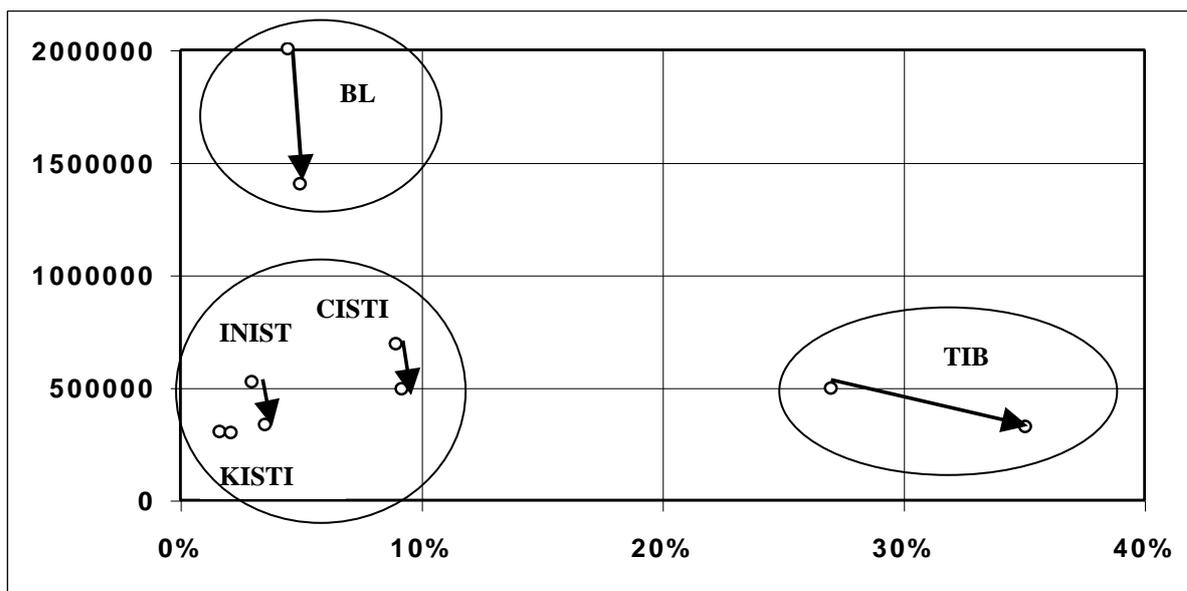
| Institution     | Document requests 2008 | Tendency since 2004 | % GL supply 2008 | GL tendency since 2004 | Main type of supplied GL |
|-----------------|------------------------|---------------------|------------------|------------------------|--------------------------|
| British Library | 1,440,000*             | ↓                   | 5%               | =                      | ND                       |
| CISTI           | 490,000                | ↓                   | 9%               | =                      | Conferences              |
| INIST           | 324,000                | ↓                   | 3,5%             | =                      | Conferences              |
| KISTI           | 328,000*               | ↑                   | 2,1%             | ↓                      | Conferences              |
| TIB             | 322,000                | ↓                   | 35%              | ↑                      | Reports                  |

**Table 1: Synthesis of 2008 document supply data (\* = best estimate)**

In 2008, the five STI centres received about 2.9 million document requests and supplied something like 2.45 million items (best estimate). Nearly 250,000 orders (or 9% of all requests) were for grey material – conferences, reports, theses etc. Compared to the 2004 figures, the “grey part” is slightly increasing.

However, the situation of the STI centres is not the same. The next figure illustrates some significant differences (figure 1).

<sup>10</sup> A common service of TIB and the German STI centres (FIZ Chemie, FIZ Karlsruhe, FIZ Technik). The beta version went online earlier this year.



**Figure 1: Evolution of document supply 2004-2008 (overall supply and % of grey items)**

In fact, we can distinguish three situations: (a) High-level supply, low proportion of grey material (the British Library). (b) Medium-level supply, low proportion of grey material, especially conference proceedings (CISTI, INIST, KISTI). (c) Medium-level supply, significant proportion of grey material, especially reports (TIB).

Except for KISTI, the relative importance of grey items is increasing, between 1% (the British Library, CISTI and INIST) and 8% (TIB). It may surprise that the main academic grey document type, the theses and dissertations, apparently do NOT contribute to this relative increase. Perhaps the explanation is that these documents are more and more available through institutional or other open access repositories.

### **Open access projects**

The following table shows the significant open access projects of the STI centres. Again, we can perceive considerable diversity (table 2).

| <b>Institution</b> | <b>Theses</b> | <b>Conferences</b> | <b>Reports</b> | <b>Others</b>   |
|--------------------|---------------|--------------------|----------------|---|
| British Library    | NR (EThOS)    |                    |                | NR (UK PubMed Central)                                      |
| CISTI              |               |                    |                | NR (NPArc )   |
| INIST              |               | OA (I-Revue)       | OA (LARA)      | OpenSIGLE   |
| KISTI              | NR            |                    |                | Journals (stOAI)  |
| TIB                |               |                    | NR             | Journals (SCOAP <sup>3</sup> ), non textual items (PROBADO) |

**Table 2: Synthesis of 2008 open access projects (NR national repository, OA open access)**

Apart from INIST, all centres are engaged in national repository projects. But while the British Library and KISTI are running projects in the field of theses and dissertations, CISTI collects publications from the Canadian National Research Centre (NRC), and TIB collects German public STM reports. The British Library also hosts the UK PubMed Central repository for UK medical publications.

Other open access initiatives are document specific (as INIST's LARA and I-Revues sites or TIB's PROBADO). Up to now, OpenSIGLE, probably the most important DSpace archive worldwide is restricted to former SIGLE records but may give free access to grey documents in the future. UK PubMed Central, stOAI or SCOAP<sup>3</sup> are clearly NOT about grey material, even if in the future this may change for stOAI and UK PubMed Central.

### ***Collection development***

Collection development and acquisition policy are linked to budget. We didn't ask for financial data but some of them are freely available on the web (annual reports etc.). Since 2005, we can discern three developments: (a) Major investments have been made for the development of access to online collections and digital archives, e.g. Big Deals with journal publishers and the acquisition of backfiles. (b) Development of grey collections is mostly linked to an open archive policy, e.g. through deposit of grey items, digitization of print documents and metadata harvesting. (c) The demand for grey items encourages networking with other libraries and document suppliers.

### ***Bibliographic control***

In the last few years, the surveyed institutions began to rethink their traditional cataloguing practice. On the agenda are new definitions of baseline quality, records enrichment, scanning tables of content, metadata harvesting, extension of cataloguing to non-textual material and outsourcing.

The institutions seem to share the same concerns: create records for all kind of available material in order to increase visibility of "hidden" resources, integrate the greatest number of records into portals and other discovery services, accelerate records production, and control the related costs. The price to pay will probably be increased heterogeneity and variable records quality.

### ***Information system***

Compared to our 2005 study, the impact on the information system is very significant. Some common features are : the development of e-commerce, the ongoing integration of information technology, the linking to other external services and software (networking, referencing, discovery tools), the improvement of storage capacities, and the adoption of a Web 2.0 philosophy (personalization, RSS syndication, interactivity etc.). For instance, INIST has just launched the beta version of its new STI portal BiblioCNRS based on Netvibes software. In order to improve the description of documents and non-textual items, the British Library will provide the opportunity for interactive resource tagging by the end-user.

STI centres are migrating from print document supply to digital libraries. As their access and gateway function becomes more importance than remote document supply, the principal reason for a new information system is improved discovery of digital resources (see for instance Ireland, 2008 for CISTI) and cost reduction through automation, resource sharing and networking or linking.

CISTI and INIST information systems are undergoing fundamental changes. The British Library and TIB continue to improve their software. The British Library is preparing a new DRM system while TIB carries on networking and IT integration with the two other national libraries in Germany.

### **Conclusion**

Taken together, these five STI centres satisfied in 2008 around 2.45 million requests. This is not a small figure, but compared to the access to online resources through digital libraries,

portals and open archives, it becomes subsidiary and nearly insignificant<sup>11</sup>. We are very far from the golden age of document supply ten years ago when the number of requests was more than double and this activity made the STM publishers somewhat nervous.

Decline, but also (potential) rise. Confronted with globalization, e-commerce, IP laws in favour of right-holders and the financial position of (some) information companies, the STI centres are in a paradoxical situation, at the same time strategic partners<sup>12</sup> and commercial competitors especially in the corporate STI market.

Of course, their strategic approach and perspective are different, shaped by their national context and organisational characteristics. Nevertheless, they seem to share some common features, perhaps even a kind of common culture of what to do, and how to do it. Our conclusion is on three topics, long tail, repositories, and e-Science.

### *Long tail*

“(…the last five to seven years have been a roller-coaster ride for our document supply service. Whilst we have delivered efficiencies and successfully identified and secured new areas of opportunity and income for the library the next three to five years are likely to bring challenges of a similar magnitude. Open Access, new publisher access models and archival digitisation will continue to make it easier and cheaper for researchers to find content at a granular level and as such our supply will become more focused on the long tail of publishing”, (Pfleger, 2008).

Mat Pfleger’s observation that document supply becomes a long tail service is consistent with our own findings: “Because of the growing access to core collections through online subscriptions, customers’ orders are shifting to the ‘long tail’ of scientific publications, to highly specialised niche and marginal titles” (Schöpfel and Gillet, 2007; see also Bador et al., 2007).

The British Library’s UK Research Reserve project reflects a “long tail strategy”, (Wright, 2008). In the first instance, the UKRR’s focus is on serials; however in the future, it may include other materials such as books and grey literature, (Shorley, 2009). This is logical and coherent: grey literature is difficult to find and obtain, and the demand is lower than for articles, so the economic challenge from commercial publishers is less important.

The survey, too, confirms this observation. The relative part of grey document supply increases and becomes really important when a STI centres owns and develops a unique (“monopolistic”) collection of highly valuable, grey research documents, such as TIB with its holdings of German public scientific reports.

“Give them what they want”. The long tail approach is the first answer to cope with the new reality. This may not be a sustainable business model, especially for collection-based establishments with flat or declining acquisition budgets<sup>13</sup>. But even in the new public management context with limited public spending and an obligation to provide a return on investment, providing information for research and developments remains a necessary function. Perhaps the British Library’s ongoing work on a new business model for document supply will provide a first answer to this question.

In the meanwhile, the supply of 2.45 million items continues to generate significant revenues and royalties, not enough to cover the total production costs and acquisition budgets but enough to justify investment in new products and technology.

---

<sup>11</sup> So far as we can see, no figures have been published on this topic. From our own French studies (see Schöpfel and Boukacem-Zeghmouri, 2008), the best estimate of the part of document supply compared to online access (download of items) is less than 2%.

<sup>12</sup> Cf. for instance the ICSTI *WorldWideScience Alliance* and the recent DOI initiative.

<sup>13</sup> This is the fundamental difference with corporate document suppliers such as InfoTrieve or ReprintsDesk whose business model is NOT based on collection acquisition.

## ***Repositories***

STI centres adopted a strategy of open access publishing, in different environments, with different objectives, and with more or less success. The total number of items freely available through their open repositories is difficult to estimate; it may be approximately 3.5 to 4 million items including at least 25% records without full-text access. This figure represents less than 10% of the content document suppliers potentially can deliver.

OA repositories and Remote Document Supply (RDS) services begin to live together rather than to compete. The STI centres are developing strategies for integrated and/or complementary service offerings, without antagonism between open access and commercial activities. OA, contrary to the Big Deals, appears to be compatible with RDS, at least it seems NOT a real threat so far. Perhaps this may change as OA content continues to grow and change..

Our conclusion is twofold: OA strategy is first of all a political choice. It seeks an answer to the challenge posed by the serials pricing crisis to the traditional value chain of scientific and technical information; it tries to reinforce the relationship with the (national) scientific communities, especially with the universities, and facilitate a return to the basics of scientific documentation, e.g. offering the best (cheapest, quickest) access to needed content.

The second conclusion relates to grey literature. This special material is by definition part of the long tail – a lot of items with a low demand. It may be that the open archive is the best solution for this kind of “stuff” because of limited acquisition, management, conservation and supply costs. Yet, this remains an assumption without empirical support because there is no economic or financial evidence so far as we know.

## ***e-Science***

Finally, all surveyed institutions invest in e-Science and e-infrastructure, as a new and complementary service activity (see table 3). These new projects underscore their national position and attract new revenues from the public funding bodies. The shared goal is to develop and improve access to scientific datasets on the Internet, especially from public research organisations and Higher Education.

| <b>Institution</b> | <b>e-Science projects</b>                |
|--------------------|--|
| British Library    | UK e-infrastructure, DOI                 |
| CISTI              | STM data portal                          |
| INIST              | Life sciences data e-infrastructure, DOI |
| KISTI              | NR STM data, infrastructure (grid)       |
| TIB                | DOI datasets registration agency         |

**Table 3: Synthesis of 2008 e-Science projects**

Even if the landscape is moving fast, we can distinguish three approaches: a contribution to the national e-infrastructure (the British Library, INIST and KISTI), the management and/or federation of datasets repositories (CISTI and KISTI), and the standardization of datasets processing and preservation (TIB).

The last project – the creation of a DOI registration agency for scientific datasets – is of special interest because it opens the way to international networking with other STI centres, including the British Library, CISTI and INIST.

Among our sample, KISTI occupies a privileged position, because of its experience with advanced scientific calculations, its outstanding technological capacities and major public

funding for Korean grid infrastructure implementation and middleware development (project K\*Grid).

These strategic initiatives again reveal political choices and perhaps also anticipation of the “time after RDS”. Even more than the OA projects, they require intimate knowledge of and enhanced integration into national and international scientific communities. Sometimes, the e-Science initiatives are linked to novel R&D activity. -

Our survey ends here. We have tried to give a snapshot of the situation and projects of some major traditional document suppliers in the emerging digital environment and STI market. What we have is not a still image but “streaming”. Compared to our 2005 findings, the transformation of these institutions into modern STI service providers is accelerating, with a great variety of projects, initiatives and plans but also a number of common features, with traditional services (photocopies, database production, ILL) coexisting with advanced and cutting edge technologies.

All changes bear costs. We didn’t ask about human resources, organisational structures and business models. We didn’t ask, either, how the STI centres were affected by the current financial crisis. Another update of this survey in some years could provide evidence on these topics, along with data on repositories, grey documents and document supply.

## **Bibliography**

- Bador, P., Boukacem-Zeghmouri, C., Lafouge, T., Prost, H. and Schöpfel, J. (2007), “Analysis of the customers of document supply in pharmacology: a case study from INIST in France (part 3)”. *Interlending & Document Supply*, Vol 35, No 3, pp. 138-144.
- Boukacem-Zeghmouri, C. and Schöpfel, J. (2006), “Document supply and open access: an international survey on grey literature”. *Interlending & Document Supply*, Vol 34, No 3, pp. 96-104.
- Brase, J. (2009), “Research data in library catalogues and the joint initiative of European technical libraries for data registration”. *Workshop Primary Data for Libraries. ICSTI winter meeting*, Paris, 2 March 2009.
- Brazier, C. (2007), “Resource Discovery at the British Library: New Strategic Directions”. *World Library and Information Congress: 73<sup>rd</sup> IFLA General Conference and Council*, 19-23 August 2007, Durban, South Africa.
- British Library (2008), *Bringing knowledge to life. Annual Report and Accounts 2007/08*. The British Library, London.
- Choi, M.S., Park, J.W., Jeong, H.S., Kang, M.Y. and Hahn, S.H. (2007), “KISTI-ACOMS: article contribution management system”. *Data Science Journal*, Vol 6., Suppl., 11 March 2007, pp. S155-S158.
- Farace, D., Frantzen, J., Schöpfel, J., Stock, C. and Henrot, N. (2008), “OpenSIGLE, Home to GreyNet’s Research Community and its Grey Literature Collections: Initial Results and a Project Proposal”. *GL10 Conference Proceedings. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*. Amsterdam, 8-9 December 2008.
- Fournier, J. (2007), “Information, Infrastructure, Involvement: The open access agenda in Germany”. *Berlin 5 Open Access. From practice to impact: Consequences of Knowledge dissemination*. Conference, Padova (Italy), 19-21 September 2007.
- Gillet, J. (2007), “Le service de livraison électronique sécurisé à l’INIST”. *World Library and Information Congress: 73<sup>rd</sup> IFLA General Conference and Council*, 19-23 August 2007, Durban, South Africa.
- Hwang, H. and Choi, H. (2006), “The Open Access Movement in Korea’s R & D Environment.” *World Library and Information Congress: 72<sup>nd</sup> IFLA General Conference and Council*, 20-24 August 2007, Seoul, Korea.

- Hwang, H., Choi, H. Seo, T.S. and Lee, S.S. (2006), "Patterns of Research Output produced by Scholarly Communities in South Korea." *International Journal of Grey Literature*, Vol 2, No 1, pp. 43-49.
- Ireland, M. (2008), "Nouvelles orientations en matière de fourniture d'information numérique dans l'environnement Web à l'Institut canadien d'information scientifique et technique du CNRC." *World Library and Information Congress: 74<sup>th</sup> IFLA General Conference and Council*, 10-14 August 2008, Québec, Canada.
- Kelsey, P. (ed.) (2009), *Profiles of Best Practices in Academic Library Interlibrary Loan*. Primary Research Group, New York.
- Kim, J. and Kwon, E.K. (2008), "Scholarly communication in Korea". In Xia, J. (Ed.), *Scholarly communication in China, Japan, Korea and Taiwan*. Chandos, Oxford.
- Meyer, T. (2009), *Die Zentralen Fachbibliotheken und ihre Rolle für die Fachinformation in Deutschland*. Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft Heft 248. Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, Berlin.
- Pfleger, M. (2007), "Delivering Access through Partnership". *10<sup>th</sup> ILDS Conference IFLA*, Singapore, 29-31 October 2007. <http://www.nlbconference.com/ilds/speakers-mat.htm>
- Pfleger, M. (2008), "The British Library: the changing face of document supply". *Interlending & Document Supply*, Vol 37, No 3, pp. 132-134.
- Prowse, S. (2006), "Recent developments in remote document supply (RDS) in the UK – 2". *Interlending & Document Supply*, Vol 34, No 3, pp. 125-131.
- Prowse, S. (2007), "Recent developments in remote document supply (RDS) in the UK – 3". *Interlending & Document Supply*, Vol 35, No 4, pp. 187-191.
- Prowse, S. (2008), "Recent developments in remote document supply (RDS) in the UK – 4". *Interlending & Document Supply*, Vol 37, No 2, pp. 100-104.
- Rosemann, U. (2008), "Product development and services provided by the German National Library of Science and Technology". *La biblioteca scientifica e tecnologica: servizi per l'informazione scientifica*. Conference, Roma, 17 April 2008.
- Schöpfel, J. and Gillet, J. (2007), "On document supply in the digital world". *Interlending & Document Supply*, Vol 35, No 4, pp. 195-204.
- Schöpfel, J. and Boukacem-Zeghmouri, C. (2008), "On the usage of e-journals in French universities". *Serials*, Vol 21, No 2, pp. 121-126.
- Schöpfel, J. and Farace, D.J. (2009), "Grey Literature". Bates, M.J. & Maack, M.N. (ed): *Encyclopaedia of Library and Information Sciences*. 3rd edition. Taylor & Francis, Abingdon. (forthcoming)
- Shorley, D. (2009), "UK Research Reserve (UKRR)". *Journées de l'ABES*, Montpellier, 27 Mai 2009.
- TIB/UB Hannover (2009), *Jahresbericht 2008*. TIB/UB, Hannover  
<http://www.tib-hannover.de/fileadmin/informationssysteme/jahresberichte/tibub-jahresbericht-2008.pdf>
- Woods, B. and Ireland, M. (2008), "eBook Loans: an e-twist on a classic interlending service". *Interlending & Document Supply*, Vol 36, No 2, pp. 105-115.
- Wright, N. and Crawford, J. (2008), "Supporting access to the UK's research collection: the UK Research Reserve project". *Interlending & Document Supply*, Vol 36, No 4, pp. 210-212.

## Web references

- [1] OpenDOAR available at <http://www.opendoar.org>  
DOAJ available at <http://www.doaj.org>

- [2] <http://www.bl.uk/>
- [3] <http://cisti-icist.nrc-cnrc.gc.ca/>
- [4] <http://www.inist.fr>
- [5] <http://www.kisti.re.kr/english/>
- [6] <http://www.tib.uni-hannover.de/>
- [7] <http://ethos.bl.uk> and <http://www.ethos.ac.uk/>
- [8] <http://www.ukpmc.ac.uk>.
- [9] <http://www.pubmedcentral.nih.gov/about/pmc.html>
- [10] <http://ukpmc.ac.uk/funders/>
- [11] <http://newspapers.bl.uk/blcs/>
- [12] <http://www.bl.uk/aboutus/stratpolprog/strategy0811/>
- [13] <http://discover-decouvrir.cisti-icist.nrc-cnrc.gc.ca/dcvr>
- [14] <http://cisti-icist.nrc-cnrc.gc.ca/eng/services/cisti/scientific-data/data-sets/index.html>
- [15] <http://cisti-icist.nrc-cnrc.gc.ca/eng/services/cisti/scientific-data/data-management-curation/index.html>
- [16] <http://data-donnees.gc.ca/>
- [17] <http://data-donnees.gc.ca/>
- [18] <http://www.bibliocnrs.fr/home.html>
- [19] <http://www.koreascience.or.kr/>
- [20] <http://www.tib-hannover.de/en/the-tib/overview-of-the-library/>
- [21] <http://www.nationallizenzen.de>
- [22] Digital Object Identifier, cf. <http://www.doi.org/>
- [23] <http://www.probado.de>
- [24] <http://scoap3.org/>

All web sites visited between May and July 2009.

### **The authors**

**Joachim Schöpfel** is Head of the Department of Information and Library Sciences at the University of Lille 3.

[joachim.schopfel@univ-lille3.fr](mailto:joachim.schopfel@univ-lille3.fr)

**Hélène Prost** is Information Professional in Social Sciences and Humanities at the French Institute for Scientific and Technical Information (CNRS).

[prost@inist.fr](mailto:prost@inist.fr)