

Access to European Grey Literature

Joachim Schöpfel

▶ To cite this version:

Joachim Schöpfel. Access to European Grey Literature. Grey Literature Repositories, Radim Bacuvcik VeRBuM, pp.20-33, 2010. sic_00597798

HAL Id: sic_00597798 https://archivesic.ccsd.cnrs.fr/sic_00597798v1

Submitted on 2 Jun 2011

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.

Chapter 2

Access to European Grey Literature

Joachim Schöpfel

Collecting grey literature remains a challenge to library and information science (LIS) professionals. Grey items such as reports, proceedings, or working papers cannot be purchased or bought like journals and books. There is no special agency or supplier for grey materials.

Buying information is part of the traditional library role, together with gateway and archive functions. In line with the economic definition of grey literature, "material that usually is available through specialized channels and may not enter normal channels (...) of (...) distribution", one comes to understand that a systematic collection of grey literature calls on specific attention, competency, and procedures.

2.1 Networking

Frequently, grey holdings are the result of a patient and long-term investment in professional contacts and networking. Networking means sharing information about grey content, resource discovery and acquisition channels with other LIS professionals. Two recent personal Web 2.0 initiatives of sharing insider knowledge on biomedical grey literature are Barrera's personal Netvibes page with RSS files and Giustini's report *Finding the Hard to Finds* on grey holdings and search strategies².

Yet, LIS professionals started collaborative "grey work" many years before Tim O'Reilly invented the Web 2.0. In 1980, EU scientific information centres established the "System for Information on Grey Literature in Europe" (SIGLE database) to provide access to European grey literature and to improve bibliographic coverage. After initial funding by the Commission of the European Communities, the national centres formed a non-profit network for the acquisition, identification and dissemination of grey literature called "European Association for Grey Literature Exploitation" (EAGLE). In this network, each national centre for scientific and technical information (STI) held the national grey collection or at least, guaranteed the document supply of distributed holdings.

¹ http://www.netvibes.com/ernestobarrera#Grey_literature

 $^{^2\,}http://www.slideshare.net/giustinid/finding-the-hard-to-finds-searching-forgrey-gray-literature-2010$

In 2005, a general assembly resolved to liquidate EAGLE because its organisational structure and business model were unable to cope with Internet technology and the Google generation, e.g. SIGLE offered no solution for online cataloguing, metadata harvesting, links to full-text and other resources (Schöpfel et al. 2007).

But the same 2005 EAGLE general assembly decided unanimously to preserve the European co-operation for grey literature and to transform the 1980 model into a sustainable network in the emerging environment of open access to scientific information, especially in the context of the 2003 Berlin Declaration.

The first step was to archive the SIGLE records in an open and freely searchable database, compliant with the OAI metadata harvesting protocol. The French INIST (CNRS) developed OpenSIGLE³ based on MIT software (DSpace) and loaded most of the SIGLE records in a simplified XML format (Farace et al., 2008). The second (future) step should be the federation of European open access projects for grey literature in order to (re)establish a gateway to European grey literature.

Other grey networks may exist on a regional or community level. For instance, in the early 90s the French government operated an interdepartmental group (LIGRIA) for the management of official grey literature.

What can be learnt from these and other initiatives? Most LIS professionals in charge of grey collections are interested in collaborative work. Librarians like networking. Cultural mediation is collective, not solitary. But "network attitude" is not enough. Efficient networking needs experience and competency, common objectives, well-organized governance structures and a sustainable business model. This may explain why networking sometimes remains an individual (personal) rather than an institutional affair.

2.2 Mandatory policies

Some collections of grey literature are the expression of a clear mandatory policy, e.g. result from explicit national or regional decisions. These may be scientific, cultural and/or political decisions, for instance to guarantee preservation of and access to specific contents, or to contribute to the construction of cultural (scientific) heritage.

The first case is a kind of legal deposit of grey items. One of the three special scientific libraries in Germany, the German National Library of Science and Technology in Hanover (TIB) celebrated its 50th 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". 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. In 2010, TIB holds more than 210.000 print and 30.000 digital German research reports on engineering or natural sciences. Each month, around 200 new electronic and 500 print reports are added. TIB is the deposit library of the (digital) final project reports funded by the Federal Ministry of Research. Since 1996, any

-

³ http://opensigle.inist.fr

research project has to deliver its final report to the TIB both as a free, printed copy, and on an electronic storage medium (see also Meyer, 2009). –

A quite different model of mandatory policy is the distributed collection of Ph.D. theses by academic libraries, with a central access point. The French government published in 1985 a decree that regulated and improved the deposit and dissemination of doctoral theses. The local library stores the document but the record is part of the French national union catalogue SUDOC that allows for ordering and delivering of print copies (Paillassard et al., 2007). This distributed model is different from a central storage facility such as the Moscow Scientific and Technical Information Centre (VNTIC) that receives copies of Russian PhD theses and holds 500,000 theses since 1982.

A third model are bi- or multilateral agreements for the acquisition and dissemination of grey items in the context of a national STI policy. Again, a French example may illustrate this model: one part of the significant INIST holdings of French Ph.D. theses and scientific reports are related to settlements with the Ministry of Higher Education (theses) or publishing bodies (research organisations, ministries). Another example: the Danish Royal Library was a depository library until July 2002 for the Council of Europe, United Nations, NATO, OECD, UNESCO and other international organisations.

Especially when accompanied by public funding, these explicit mandates allows for long-term collection and preservation of specific grey items. Sometimes they will also facilitate digitization projects (scientific heritage). The problem with these public mandates and agreements is that they may be under-funded and/or of limited duration, with a risk of incomplete and disrupted holdings.

2.3 Acquisition strategy

Defining a coherent acquisition policy is a crucial part of a library's function. Often, this policy will reflect patrons' needs and suggestions, subject choices and budget structure. Some libraries may also develop a specific acquisition strategy with regards to grey literature, independently from an external mandate. Over the time, such an institutional ("intrinsic") strategy may generate exceptional holdings.

Probably one of the most famous holdings of this kind is the Boston Spa conferences collection of the British Library with around 450,000 items. "British Library holds one of the most comprehensive and easily accessible collections of English language conferences in the world. (...) The British Library believes holding the material is only part of the process to enable access, and has developed various products to aid the user in locating this material" (Tillett & Newbold, 2006). This holding reveals an explicit choice to collect all English-speaking scientific conference proceedings. During the golden age of document supply, LIS professionals and customers knew that Boston Spa possessed (nearly) all international scientific conferences.

Another distinct area of the collecting and focus maintained at Boston Spa are scientific and technical reports, from several thousands public and private sector British, American and international sources such as the NASA, the UN Food and Agriculture Organization or the European Space Agency. The 2008-2009 British Library annual report mentions 10,5 million reports on microform while the website provides the figure of 4,9 million unrestricted reports available for public use.

As a complement to their mandate for German reports, the TIB invests in the systematic collection of foreign scientific reports, especially by the National Technical Information Service (NTIS) of the US Department of Commerce or the NASA but also from an important number of European institutions (nearly 2 millions).

These two examples show, too, that up to now we cannot speak of a formal European coordination of national or regional grey collections. National or institutional considerations prevail, e.g. preservation strategies, national independency, bilateral agreements etc.

Somewhere on the crossroad between mandate and institutional choice we find holdings of theses and dissertations. Usually, academic libraries are mandated to hold theses from their own university; at the same time, they collect more or less systematically theses from other universities, following their own rules and criteria (disciplines, subjects, institutions...). At first sight, this does not make any difference. But when libraries evaluate and weed their collections, they maintain the "local collection" and discard the rest.

2.4 From print to digital collections

Since the invention and success of the web, libraries leave the Gutenberg galaxy with its millions of print items. They do it in two ways. They convert their print holdings into digital collections, and they collect and archive born digital material.

Grey literature enters the two circuits. Significant retro-digitization projects were launched for PhD theses in print format or on microforms. The British Library digitizes theses from UK universities for the new EthOS-portal⁴ that integrates free access to electronic theses and dissertations (ETDs) harvested from open repositories with supply from theses digitized on demand.

In France, the national reproduction centre for theses ANRT⁵ develops a capacity for digitisation for its service "Thèses à la carte". The German digitization projects funded by the DFG⁶ contain namely early primary sources (cultural heritage) but also, scientific publications (manuscripts, journals etc.). In the UK, the JISC invests since 2003 in digitizing content from special collections⁷ like the approximately 600 volumes of historical population reports (census reports) hold by the University of Essex and 10,000 theses for EthOS.

The list of European digitization programs for grey literature is long, and we could add the retro-digitization programs by the TU Delft for their E-thesis pilot or by the University of Uppsala for more than 11,000 theses submitted in the 18th and 19th century, the Catalan electronic theses and dissertations network TDX, or the Poznan PSNC Digital Libraries Team project coordination activity.

In comparison, less digitization programs are scheduled for reports, conference proceedings or other forms of grey literature. Also, we didn't hear from important European initiatives that could match with the US digitization project of DOE report

⁵ http://www.anrtheses.com.fr/

⁷ JISC digitisation and e-content program http://www.jisc.ac.uk/digitisation

⁴ http://ethos.bl.uk/

⁶ http://gepris.dfg.de

collections or the OSTI collaboration with other sites (FERMI, LANL etc.)⁸. Three recent initiatives in France involve BRGM reports for an earth sciences portal, the LARA platform for scientific reports from different institutions⁹ and the mathematics archive NUMDAM with 29 seminaries from 1948 to 2007¹⁰.

These programs share three common features, e.g. clearly identified grey collections, scientific heritage character and low coordination with other initiatives.

Sometimes, digitized grey items are mingled up with born digital material and/or living e-collections. One (but not the only) example is the French national repository for ETDs, TEL, with more than 10,000 recent theses (2005-2010) but nearly 2,000 digitized PhD theses published in 1990 or before.

Another site, the UK Centre for Environmental Data Archival (CEDA) based at the STFC Rutherford Appleton Laboratory, is for grey literature primarily concerning Earth observation and the atmospheric sciences¹¹. Apparently, all CEDA items – more than 600 - are born digital. Other European repositories with grey digitized and/or born digital material can be found in the OpenDOAR directory: from the 776 registered sites (March 2010), 54% hold ETDs, 42% unpublished reports and working papers, and 40% conference and workshop papers. In France, ³/₄ of the OA repositories contain grey literature.

2.5 The open access principle

The OpenDOAR directory highlights the fundamental impact the shift from print to digital holdings produced on the underlying business model and the distribution channels of grey items.

Grey literature is defined through its non-commercial dissemination channels. With the development of the open archive (OA) initiative, grey documents quite logically took their place in these new repositories, especially in institutional repositories (Schöpfel et al., 2009) but also in subject-based or other types of open archives.

A longitudinal survey 2005-2009 describes how five international STI centres adopt a strategy of open access publishing, in different environments, with different objectives, and with more or less success (Boukacem-Zeghmouri & Schöpfel, 2006; Schöpfel & Prost, 2009). 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 a significant part of grey items. This special material is by definition part of the long tail – a lot of items with a low demand.

Luzi et al. (2008) describe the preparation of an institutional repository by the Italian National Research Council (CNR); at least one third of the deposits in existing CNR OA sites are grey items (reports, oral presentation, theses etc.).

The open archive may be the best solution for this kind of "stuff" because of low acquisition, management, conservation and supply costs. Yet, this remains an

11 http://cedadocs.badc.rl.ac.uk/

⁸ See for instance Fermi's 1947 report on the Future of Atomic Energy available at http://www.osti.gov/accomplishments/documents/fullText/ACC0043.pdf

⁹ http://lara.inist.fr/ see also Stock et al. (2006)

¹⁰ http://www.numdam.org/

assumption without empirical support because there is no economic or financial evidence so far as we know.

In the next future, will all grey documents be available on OA web sites? In spite of the Willinsky (2006) claim that "open access is a public good" and that "commitment to scholarly work carries with it a responsibility to circulate that work as widely as possible", a significant part of grey material probably may not enter the open archive landscape – because of lacking interest or budget for digitization of older print materials or restricted access, or because the items are already available on other websites (personal pages, institutional sites with links to PDFs etc.) but not well indexed.

Nevertheless, the proportion of "grey" documents published on the Web continues to increase. This development is closely linked to the production of grey literature in e-environments, as well as to retrospective activities leading to republication. The Internet will encourage a greater diversity in the types of "grey" resources available (raw research results, notes and personal comments, lectures, newsletters, product catalogs, etc.).

New technologies of information and communication facilitate resource archiving in general, and there is strong incentives from the OA movement. Nevertheless, the question of "who should archive what, where, when, and for how long" remains largely unanswered. Aware of information policy and the concomitant financial aspects involved, answers are rather urgently needed, even if they were to now address only part of grey literature resources (Schöpfel & Farace, 2010).

2.6 From collection to gateway

"A library is a collection of sources, resources, and services (...); it is organized for use and maintained (...)" Can an open repository be called a collection? Is it part of the library collections?

Probably, these are yesterday questions. In the coming Google era, do we really need well-structured and maintained grey collections? Or do we need tools to search, retrieve and access grey items? Can we imagine grey "collections" as a kind of global grid?

Perhaps there is no answer yet. Instead, we would like to draw the reader's attention to some recent developments, products and services.

A couple of years ago, the Royal Netherlands Academy of Arts and Sciences (KNAW) stopped all acquisition activity of the Institute for Scientific Information Services (NIWI), formerly one of the major document suppliers. Instead, the KNAW invested in the creation of a new gateway called NARCIS that gives access to OA publications from Dutch universities and research institutes, datasets, descriptions of research projects, institutes and researchers, and research news. In this environment, the borderline between "grey" and "white" (commercial) literature becomes increasingly indistinct.

The National Documentation Centre EKT at Athens maintains the Hellenic Dissertations database linking to 13,000 theses hold by Greek universities.

The ETH Zurich library portal provides access to 2,1 million reports from other libraries, databases and search engines.

¹² "Library" Wikipedia, The Free Encyclopedia (accessed March 13, 2010).

The Irish Virtual Research Library and Archive is meant to realize the latent potential of archival collections within the University College of Dublin.

The DiVA portal¹³ gives access to 270,000 research publications and student theses written at 27 Swedish and Norwegian universities and colleges of higher education; 44% of the content is grey.

Scirus, Elsevier's free academic search engine, indexes more than 30 OA repositories called "preferred Web sources" that include European institutions and grey items.

The DART-Europe E-theses portal¹⁴ provides access to more than 130,000 full-text research theses from 233 Universities sourced from 16 European countries (March 2010). DART-Europe, a partnership of research libraries and library consortia, is the European Working Group of the Networked Digital Library of Theses and Dissertations (NDLTD, access to nearly 750,000 ETDs).

Another recent tool for resource discovery is the German PUMA project for the management of academic publishing ¹⁵.

These are but some illustrations. It is impossible to propose an exhaustive list of all European initiatives. The common point is that the notion of collection has been replaced by the concept of access. This gateway function often stays with the library; but other players enter the scene, such as publishers, search engines, computing centres etc. These new players never managed library holdings; the accent is on selection, dissemination, access, not on preservation and organization.

Roosendaal et al. (2010) depict very clearly the dynamics of this new publishing paradigm and the underlying business model. The advantage is obvious: a critical mass of information, a single access point, powerful search and selection tools, nomad access.

Some problems have been listed by Stock (2007) in her study on European ETDs in open repositories: partial or restricted access to the full text, records without full-text, missing or incomplete metadata, language barriers. Other problems are lacking standards and interoperability.

So far as we can see today, searching and collecting grey literature will not become as straightforward as it is for journals and books in the traditional publishing sector. New tools for collecting, depositing, and archiving does not make grey literature less ephemeral and volatile than in the past. Our research indicates that until an organization formulates a policy on grey literature backed by budget appropriations, the implementation of technology cannot be guaranteed and thus the environment in which grey literature has coexisted in the past will remain unstable in the likely future (Schöpfel & Farace, 2010).

2.7 From library to eScience

The research environment is changing and becomes more and more data-driven, with growing needs related to data acquisition, storage, processing, management, mining etc. New data integration services are already emerging, transforming data discovery on the web from lists of search results into tools that compute answers to

14 http://www.dart-europe.eu/

¹³ http://www.diva-portal.org/

¹⁵ http://puma.uni-kassel.de/

structured questions (Fry et al., 2009) but as Osswald (2008) points out, so far scientific libraries have not played an important role – if at all – in e-Science projects implemented in the EU.

Access to research results (both publications and data) is "the last key ingredient of the research infrastructure (...). Thus the e-Science revolution will put libraries and repositories centre stage in the development of the next generation research infrastructure." (Hey & Hey, 2006).

Portals like NARCIS already include datasets in the resource offerings. With regards to the most recent developments of academic publishing (dynamic publications, 3D illustrations, primary datasets embedded in journals etc.), this is quite natural. But what is the color of datasets? Are they part of scientific literature? Or will they replace, at least partially, scientific publishing?

Today, the "article of the future" concept is in the center of scientific and professional debate. Commercial publishers (and software companies) invest heavily in advanced editing software in order to integrate data and publication.

What about grey literature in this data-intensive research environment also called "fourth paradigm" (Hey et al., 2009)? While inclusion of raw data is a relatively new functionality for journals, supplementary materials - CD-ROMS, maps, tables or voluminous data appendices – often have accompanied theses and reports already in the past.

Grey literature provides raw material for data mining and scientific alert services. For instance, scanning pharmacological conference announcements and abstracts allows for economic intelligence (industrial trends analysis etc.); exploitation of state of the art sections and bibliographies of PhD theses contributes to scientometrics.

The question is NOT if grey literature has to do with eScience but how scientific data in theses, reports, communications, working papers etc. should best be valorized. One solution is the creation of powerful data repositories by the scientific communities and their libraries, and the development of new data publishing models. Osswald (2008) warns that libraries may lose an important part of their tasks within the research community if they don't try to gain a role in eScience projects. The risk is real. But some recent initiatives provide evidence that libraries become part of the emerging scientific cyberinfrastructure. The most promising European project actually seems DataCite¹⁷ that "promotes data sharing, increased access, and better protection of research investment".

As a first step, DataCite will promote the use of Digital Object Identifiers (DOI) for datasets. The next step should be the interconnection between open data and publication archives, by the scientific communities and institutions ¹⁸ – if they want to limit control of research results by commercial publishers and global information companies. What about unique identifiers for grey objects?

Anyway, datasets challenge the certification and preservation function of publishers and libraries. Maybe, their real place is outside of commercial distribution channels and not in the "article of the future". Most scientists don't want to store data with the publishers. Tomorrow, perhaps we will not have one but many NARCIS information systems, and perhaps we will have, too, a unique gateway to

¹⁶ See Elsevier's "Article of the Future" initiative http://beta.cell.com/

¹⁷ Project leader: TIB Hannover; http://www.datacite.org/

¹⁸ For instance, in a current research information system (CRIS) environment.

access and connect Dutch, UK, German, Swedish and Czech datasets and publications. Let's dream.

2.8 The future of grey collections

This chapter tried to provide an idea of the richness and dynamics of European grey literature. Of course, it is impossible on some page to list all significant collections, such as the special collection of more than 60,000 rare publications and samizdat literature hold by the Jagiellonian library at Cracow or the 15,000 digital maps at the Institut Cartogràfic de Catalunya ICC at Barcelona. The reader will find links to more resources on the websites of different LIS networks, e.g. GreyNet¹⁹ or LIBER²⁰ with a forum on digitisation, resource discovery, heritage collections and preservation.

In the ongoing discussion on new business models of academic publishing, eScience and open access to public research results, non-commercial distribution channels will continue to play a central role as vectors of scientific communication, alongside commercial publishing.

Scientists ask for trustworthy information. The ClimateGate discussion²¹ showed that the question of quality needs attention. Of course, one part of grey literature has some kind of quality label (Ph.D. jury, scientific committee selection of communications etc.). What about the rest? How can libraries guarantee high quality for grey items? Can they? Will Web2.0 item tagging become an alternative to review and selection?

Another question is about impact and usage. In the past, impact metrics were limited to citations and journals. Today, usage metrics offer new opportunities to measure impact of a large scale of digital resources, also on the individual item level. Tomorrow, these metrics will provide additional information on quality and popularity to the end user.

Open archives will offer more appropriate services and functions for at least some segments of grey literature if not for all. But bibliographic control of grey literature will remain problematic despite the trend toward standardization of digital documents. And the libraries, together with their scientific communities, need to find new forms for the fundamental functions of scientific publishing, applied to open repositories, non-commercial items and datasets.

A very last remark: the article is on European grey literature, and the author is deeply attached to the European idea. But the philosophy and technology of Internet pay no attention to frontiers, nations and supranational structures. The problem is with language barriers, metadata and formats; the chapter ends with a plaidoyer for interoperability and standards.

20 http://www.libereurope.eu/

¹⁹ http://www.greynet.org/

²¹ The mistake on glacier melting introduced in the 2007 UN Intergovernmental Panel on Climate Change (IPCC) report.

Bibliography

- C. Boukacem-Zeghmouri & J. Schöpfel (2006). `Document supply and open access: an international survey on grey literature'. *Interlending & Document Supply* **34**(3):96-104.
- D. Farace, et al. (2008). `OpenSIGLE, Home to GreyNet's Research Community and its Grey Literature Collections: Initial Results and a Project Proposal'. In *GL10 Conference Proceedings. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society. Amsterdam, 8-9 December 2008.*
- J. Fry, et al. (2009). 'Open science in e-science: contingency or policy?'. *Journal of Documentation* **65**(1):6-32.
- T. Hey & J. Hey (2006). e-Science and its implications for the library community'. *Library Hi Tech* **24**(4):515-528.
- T. Hey, et al. (eds.) (2009). *The Fourth Paradigm. Data-Intensive Scientific Discovery*. Microsoft Corporation.
- D. Luzi, et al. (2008). 'Towards an Institutional Repository of the Italian National Research Council: A survey on Open Access experiences'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society*, 8-9 December 2008.
- T. Meyer (2009). *Die Zentralen Fachbibliotheken und ihre Rolle für die Fachinformation in Deutschland*, vol. 248. Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin.
- A. Osswald (2008). `E-science and information services: a missing link in the context of digital libraries'. *Online Information Review* **32**(4):516-523.
- P. Paillassard, et al. (2007). 'Dissemination and preservation of French print and electronic theses'. *The Grey Journal* **3**(2):77-93.
- H. E. Roosendaal, et al. (2010). *Scientific Publishing: From vanity to strategy*. Chandos Publishing.
- J. Schöpfel, et al. (2007). `From SIGLE to OpenSIGLE and Beyond: An In-depth Look at Resource Migration in the European Context'. *The Grey Journal* **3**(1):45-51.
- J. Schöpfel & H. Prost (2009). `Document supply of grey literature and open access: an update'. *Interlending & Document Supply* **37**(4):181-191.
- J. Schöpfel, et al. (2009). Usage of grey literature in open archives'. In GL11 Conference Proceedings. Eleventh International Conference on Grey Literature: The Grey Mosaic: Piecing It All Together. Washington D.C., 14-15 December 2009.
- J. Schöpfel & D. J. Farace (2010). 'Grey Literature'. In M. J. Bates & M. N. Maack (eds.), *Encyclopedia of Library and Information Sciences, Third Edition*, pp. 2029-2039. CRC Press.
- C. Stock, et al. (2006). `LARA—Open access to scientific and technical reports'. *Publishing Research Quarterly* **22**(1):42-51.
- C. Stock (2007). 'Open access to full text and ETDs in Europe: improving accessibility through the choice of language?'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, 10-11 December 2007.*
- S. Tillett & E. Newbold (2006). `Grey literature at The British Library: revealing a hidden resource'. *Interlending & Document Supply* **34**(2):70-73.
- J. Willinsky (2005). *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)*. The MIT Press.