Mass Digitization for Research and Study: The Digitization Strategy of the Bavarian State Library

Klaus Ceynowa
Bavarian State Library
Munich, Germany

From the "Hybrid" to the "Digital" Library

The relevant infrastructure for the research, distribution and use of information today is the internet. Researchers and students expect to satisfy their need for information promptly and comprehensively via the internet. Since the variety of web-based offers meets this demand increasingly frequently and comfortably, users have less and less need to use conventional information sources and resources. At least for the generation born after 1981, the so-called "millenials", who grew up with user-friendly web technology and for whom - primarily due to the ubiquitous availability of high-speed broadband networks - immediate access to digital offers and services, independent of time and place, has become a matter of course, the following is true today: Information has to be accessible easily, at a mouse click, on the internet, otherwise it is no longer even taken notice of. This generation is currently studying at university or college and will shortly be playing a decisive role in shaping the forms of generating, communicating and acquiring information in research, teaching, business and career.

Current surveys and opinion polls clearly point out the resulting challenges for libraries. In the 2006 OCLC-Report "College Students' Perceptions of

Libraries and Information Resources" 39% of the responding college students said that, preferring the use of the internet, they avail themselves of the services of their university library considerably less frequently.[1] In reply to the question which type of research suited their habits best ("fits perfectly with your lifestyle"), 64% of respondents mentioned using search engines, however only 30% mentioned the "online library" and only 24% the "physical" library.[2] In line with this, 72% of the college students prefer search engines when searching information, but only 14% prefer the "real" library, and 10% prefer the electronic library.[3] As is shown by the survey compiled in 2006 by PEW Internet & American Life Project titled "The Internet as a Resource for News and Information about Science", this information-using behaviour is by no means applicable only in the fields of structured factual questions or pragmatically motivated information access (such as e. g. preparing for an exam), but likewise in the field of more sophisticated and specialised information demand. Three scientific disciplines were picked out - stem-cell research, climate change and the origin of life. A representative number of Americans were asked where they would go first in order to learn more (!) about these specific scientific topics. Also with the aim of acquiring more in-depth knowledge, 67% (stem-cell research), 59% (climate change) and 42% (origin of life) of the respondents would resort to the internet, whereas only 11, 12 and 19% mention the library as their preferred information source.[4] The study "Researchers' Use of Academic Libraries and their Services" published in 2007 by RIN (Research Information Network) and CURL (Consortium of Research Libraries) among other things analyses the importance of the "real", physical library for scientific and scholarly users. Scientists and scholars of different

---

[1] Cf. College Students' Perceptions of Libraries and Information Resources, 2007, p. 61 (3-19) under http://www.oclc.org/reports/perceptionscollege.htm
and Perceptions of Libraries and Information Resources (2005) under http://www.oclc.org/reports/perceptionscollege.htm.
[2] Cf. College Students' Perceptions of Libraries and Information Resources, 2007, p. 62 et seq. (3-20 et seq.) under http://www.oclc.org/reports/perceptionscollege.htm
[3] Cf. College Students' Perceptions of Libraries and Information Resources, 2007, p. 23 (1-11) under http://www.oclc.org/reports/perceptionscollege.htm
[4] Cf. http://www.pewinternet.org/pdfs/PIP_Exploratorium_Science.pdf, p. 21.

disciplines were asked to compare the intensity of their library visits in the year 2006 to that of the year 2001, and to make a prognosis for 2011.[5] Regardless of the field of study the result is very unambiguous: Whereas in the year 2001 40% of the responding scientists and scholars still used to visit the library at least once per week, in 2006 their number had diminished to 22.5%, and according to the respondents' own prognosis in the year 2011 this number will have shrunk to 18.5%. A particularly significant drop is to be expected in the fields of life sciences with merely 9.5% and physical sciences with 10% predicted regular on-the-spot users in the year 2011. However, also regarding the social sciences, only a very limited demand for local library services (19%) is to be expected in 2011. Only in the field of arts and humanities 40% of the respondents plan to keep visiting the library at least once per week in 2011.

If the libraries wish to maintain their position as the hubs of the academic and general public information infrastructure also in the future, they have to act on their conclusions drawn from the foreseeable changes in the patterns of use. In the long term, this means that the trend is leading away from the well-known model of the "hybrid library" with peacefully coexisting printed and electronic media, and towards the model of a genuinely "digital library" with a predominantly net-based offer of contents and associated services. It is not a contradiction that parallel to this development, also "real" libraries (if not all libraries) will continue to exist as centres of communication and cultural exchange, and as guardians of heritage.

International Research Library on the Internet

The challenges posed by the model of a primarily - or even exclusively - web-based information infrastructure differ in dependence on the type of library and the users to be serviced. For the Bavarian State Library as a great European universal library and one of the world's leading international research libraries, its collections acquired systematically in

---

[5] Cf. http://www.rin.ac.uk/files/libraries-report-2007.pdf, p. 19 et seq.

the course of 450 years represent the decisive "distinguishing feature" at the centre of its "business model" for the digital age. These collections, which encompass a large part of the written cultural heritage of the occident, and which are continuously further extended, are a unique resource for academic research and study. With 91,000 manuscripts, the Bavarian State Library holds one of the five largest collections worldwide, and with its collection of 20,000 incunabula it even holds the world's largest collection of incunabula. With 130,000 prints of the 16th century, it owns Germany's largest collection of works from this period. As for its holdings of periodicals - about 49,000 titles, a growing part of which is also available in electronic form - it is outranked in Europe only by the British Library.

Due to its international orientation the Bavarian State Library, in contrast for example to a university library, does not have a locally defined "primary user group". Rather, its services address researchers, students and information seekers on an international level. The demand is of a correspondingly international nature, as is shown by the intensive use by foreign academics and the great demand by document delivery services (about 500,000 orders processed in 2007).

Insofar the internet represents the natural medium for the Bavarian State Library, for it allows everybody to access information, no matter when or where. Consequently one important strategic goal of the Bavarian State Library is to digitise its unique holdings as fast as possible and - very pragmatically - to make them accessible to the world, as far as possible under the given legal and technological conditions.

The Bavarian State Library has been pursuing the goal of a comprehensive digitisation of its holdings since 1997, the year of foundation of the Munich Digitisation Centre (MDZ - Münchener Digitalisierungszentrum), which was established with the financial support of the German Research Foundation. This centre is seamlessly integrated in the work processes of the library in the form of the "Division Digital Library" (Referat Digitale Bibliothek) today. Since its foundation it has carried out more than 80 projects mostly

funded by third parties, the majority of them representing cooperative projects with university and non-university research. In addition to developing innovative technologies and workflows the Bavarian State Library focuses on digitising its manuscripts and Bavarica, as well as the collections of the DFG-supported areas of collection emphasis, in particular history and Eastern Europe. By the end of 2007 over 24,000 titles with a data volume of 50 terabyte altogether had been scanned and processed for posting on the internet. The "digitisation line" of the Bavarian State Library is currently equipped with two high-resolution book scanners, two so-called "Graz camera tables" for the gentle digitisation of the most valuable manuscripts and prints, and three fully automatic scanning robots with a maximum throughput of 1,800 pages per hour.

The know how aggregated at the MDZ now puts the Bavarian State Library in a position to realise its comprehensive digitisation strategy in a variety of suitable ways, given the diachronic structure of its collections and the different types of materials used. With the continued financial support by the German Research Foundation the manuscripts, incunabula and historical prints of the 16th century, the music materials and maps will be digitised "in house" using the digitisation line of the MDZ. In a current, DFG-financed project for the digitisation of the 38,000 German-language prints of the 16th century held by the Bavarian State Library, 7.5 million pages are scanned by means of scanning robots within a period of only two years.

The current support guidelines of the DFG, in particular concerning national licenses for current periodicals and the digitisation of materials belonging to areas of collection emphasis[6], furthermore allow the increasingly comprehensive provision also of works that are subject to copyright protection. Already in 2008, again supported by the DFG, the Bavarian State Library is going to conclude contracts with renowned

[6] Concerning the DFG support programme "Nationallizenzen" (national licenses) see http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/projektfoerderung/foerderziele/nationallizenzen.html and http://www.nationallizenzen.de. Concerning the programme "Digitalisierung der DFG-Sondersammelgebiete" (digitisation of the DFG areas of collection emphasis) see http://www.dfg.de/forschungsfoerderung/formulare/download/12_154.pdf

publishing houses in the fields of humanities and social studies, allowing the digitisation and open-access provision of the publishing programmes frequently up to the year of publication 2004 (!). The compensation payments will be manageable, and a comfortable moving-wall solution will provide for the continuation of these digitisation programmes.

"Boutique Digitisation" versus Mass Digitisation

Regarding the collection segment from the 17th to the 19th century, flanked by the historical collections of the manuscripts, incunabula and early prints on the one hand, and the copyright-protected collections of the 20th century and the new publications - in printed or electronic form - on the other hand, the Bavarian State Library has taken the path of a public-private partnership. In the spring of 2007 the Bavarian State Library and the world's leading internet search-engine provider Google have signed a widely noticed cooperation contract of a duration of several years. Within the framework of this agreement Google will scan the copyright-free collections of the library from the 17th to the 19th century - encompassing far more than one million books and bound volumes of periodicals. The digital copies will be made available on the internet for worldwide use, both via the web site of the Bavarian State Library (http://www.bsb-muenchen.de) and via Google Book Search (http://books.google.com) und the general Google web search. The cooperation with Google will make a substantial part of the European written cultural heritage - hundreds of thousands of literary and scholarly source works in many languages, as well as a comprehensive, systematically built corpus of historical research literature - available at a mouse click for every user with an internet access. In addition to the Bavarian State Library currently a whole group of renowned libraries, among others the university libraries of Stanford, Michigan, Harvard and Princeton, the New York Public Library and the Bodleian Library of Oxford University take part in the Google library project.

Which were the reasons that convinced the Bavarian State Library to commit itself to this partnership, which is unique in scope in Europe, with Google? The projects of the Munich Digitisation Centre, similar to the activities of other large German libraries, usually follow a subject-related or material-specific approach directed at relatively small text collections, which are of special research interest or which are unique as regards their collection profile. Ronald Milne of the British Library has called these projects "boutique digitisation projects". In contrast, the cooperation between the Bavarian State Library and Google means that for the first time in the history of German libraries a digitisation project of an industrial scale, a true mass digitisation project, is tackled technically and logistically. A "selection" of books takes place only according to their fitness for scanning from a conservational perspective, and according to certain requirements of size and volume posed by Google's proprietary scanning technology.

The funds necessary for a mass digitisation project designed to encompass more than a million books, especially if this project is to be carried through within a period of a little more than five years, can only be provided within the framework of a public-private partnership. The current publications and recommendations for example by the European Commission and the European Council for the "Digitisation and Online Accessibility of Cultural Material and Digital Preservation" express rather clearly that the European Union will not contribute to funding the operative expenses of the mass digitisation of cultural material, but considers this a task of the member states. Also the current initiatives on a German federal and state level for developing a "German Digital Library" by way of contribution to a "European Digital Library (Europeana)" are still largely undetermined as to who is going to shoulder the financial burden of a comprehensive digitisation of the German cultural heritage in libraries, archives and museums. In view of the immense funding requirements, the European Commission explicitly points out the necessity

of public-private partnerships, in order to cope with the task of digitising the respective national cultural heritage.[7]

„Google Digital Copy" and „Library Digital Copy"

The decisive advantage for the Bavarian State Library in cooperating with Google lies in that Google absorbs the complete digitisation cost. By way of compensation, Google keeps a "Google digital copy" for integration in the Google Book Search and the general Google web search. The Bavarian State Library for its part obtains a "library digital copy" for integration in its various internet offers. As regulated by corporate law, and as is customary in public-private partnerships, the details of the contract are confidential. The central point of the contract concluded with Google is: Through obtaining a digital copy of the data generated by Google, the Bavarian State Library acquires these data "physically" and thereby permanently. The "library digital copy" may be made available within the framework of the online catalogue (OPAC) and the web offers, for example in the form of research collections selected according to subject areas and processed especially for scholarly research.

The digitised books can be browsed or leafed through virtually by means of comfortable navigation instruments. What is more, due to progressive optical character recognition, the majority of the digitised works will be searchable in full-text mode. Since the digitisation project encompasses exclusively copyright-free holdings of the Bavarian State Library, the user also has the option to download or copy the complete work for scholarly purposes. The provision of the digital copies in the form of full-text indexed research collections, catalogued with differentiated structural data and web 2.0 functions, will create entirely new possibilities of working, in particular for the areas of humanities, cultural studies and social studies. This means that the Bavarian State Library does not only set great value by content, but also by context.

---

[7] Cf. Official Journal of the European Union, 2006/C 297/01, Council Conclusions on the Digitisation and Online Accessibility of Cultural Material and Digital Preservation, http://ec.europa.eu/information_society/activities/digital_libraries/doc/culture_council/council_conclusions_nov_2006/en.pdf

The metadata of the digitised collections furthermore allow the Bavarian State Library to integrate the "library digital copy" in regional, national and international portals and services without any restrictions. The holdings digitised by Google can thus for example be integrated in services such as the "Bayerische Landesbibliothek Online", Bavaria's arts and humanities portal, or the planned "European Digital Library", if necessary upon prior selection according to subject or material. In this context it should be stressed that the contract with Google is non-exclusive. In the light of new technologies developed in the future, special requirements of use or a change in the strategic framework conditions it is thus left up to the Bavarian State Library to produce new digital copies also of such titles already scanned by Google.

It goes without saying that Google as a stock-exchange listed corporation does not digitise the complete copyright-free holdings of the Bavarian State Library for altruistic reasons. Given the ranking of the Google Book Search in the complete Google portfolio, it is obviously not intended to directly market the "Google digital copy". This does not make much sense in view of the fact that the library may offer the identical "library digital copy" free of charge via its web offers. The "Google digital copy" can be accessed via the general Google web search or via the special web service of Google Book Search. At least for the latter no online advertising is planned so far to accompany the search results.

Nonetheless, the library project provides Google with an indirect, yet significant competitive advantage. The full-text digital copies add an enormous amount of content to the Google search index, which in the end results in an optimised search, more users, more clicks on the advertisements and thereby an increase in profit. Generally, the variety of services - Google Earth, Google Scholar, Google Mail and also Google Book Search - surrounding the web search engine representing the Google core product seem to primarily fulfill the function of making Google appear the more attractive service in comparison to competing search engines, which are always "just a mouse click away". „Users like Google better" –

according to this principle the Google Book Search also makes a contribution to making the search engine more attractive for online advertisers.

Conservation of Collections and Long-Term Archiving

As the central state and repository library of the Free State of Bavaria, the Bavarian State Library has to guarantee the long-term usability of its holdings. The collections of the Bavarian State Library of the 19th century are seriously threatened by paper disintegration. The digitisation by Google thus serves also the purpose of collection conservation, since by preparing digital copies at least the information contained in the threatened books can be saved. Since the extent of the damage is increasing rapidly, time is of the essence. The digitisation represents a special conservational protection also for those parts of the collections that are not threatened by disintegration, since in the future readers will normally use the digital copy, whereas the original will only have to be consulted for special research purposes. In line with its digitisation strategy the Bavarian State Library will in the future generally halt the production of microfilms of threatened library holdings. Digital copies will be produced instead, which, owing to their conditions of use, have already developed into the far more attractive secondary form. The Bavarian State Library sets about the challenging task of safeguarding the digitised holdings in the long run together with the technically superbly equipped Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities.[8] Several large-scale projects supported by third-party funds are currently pursued in this field. In support of all the process steps of its digitisation projects the Bavarian State Library has developed an electronic publishing system with elaborate functions of a logistics server, which is designed particularly for "industrial mass digitisation" with several thousand volumes to be "moved" per week, and which supports all

---

[8] Bernd Reiner: Sicherung des Weltkulturerbes am Leibniz-Rechenzentrum, in: Akademie Aktuell. Zeitschrift der Bayerischen Akademie der Wissenschaften, 2007, no. 22, p. 18-22.

processing steps from retrieving the volumes to their web presentation, on to their long-term archiving.

Conservational Fitness and Quality Standards

The Bavarian State Library issued an invitation to tender for the planned mass digitisation of its copyright-free holdings in the Supplement to the Official Journal of the European Union (Tenders Electronic Daily), advertising the project as a service concession in a transparent and non-discriminating procedure in the form of an "invitation to participate in the negotiation procedure". The thus concluded cooperation agreement with Google is subject to the obligation of secrecy that is customary for contracts between public institutions and commercial enterprises, meaning that no information must be given to third parties regarding the technical and organisational details of the contract, the exact scope of the holdings intended for digitisation and the exact duration of the project. This circumstance has led to critical enquiries, primarily regarding the problem of fulfilling the conservational and qualitative requirements that library digitisation projects are usually subject to.

It has to be stressed in this context that the conservational criteria underlying the decision which books are in a sufficiently fit state for digitisation and which are not, are as a rule established by the Bavarian State Library and Google together. In cases of doubt the final decision is always taken by the library. Through the consistent involvement of the internationally renowned "Institute for the Restoration of Books and Manuscripts" of the Bavarian State Library it is ensured that no book is handed over to Google which could be damaged during scanning using Google's technology. Also the quality standards agreed with Google are within the tolerance range that is customary for third-party funded digitisation projects. It is advantageous in this context that the Bavarian State Library does not belong to the first participants of the library project initiated by Google in 2004, but as a "late entrant" can profit from the quality checks continuously optimised by Google.

Libraries and the Internet Economy

In the light of the Google library project, in which currently already 26 large libraries are participating with a total of approximately 30 million books to be scanned, the fundamental question arises whether this enormous project does not sound the death knell for libraries in the long run, in particular since a growing part of the new literature is offered in digital form by the publishing houses, as E-journals and E-books. Will the readers still use the reading rooms, the lending desks and the special collections of the libraries, if great parts of their holdings can easily be retrieved online at every internet workstation? Given that the reading rooms of the Bavarian State Library are frequently extremely crowded - in 2007 more than one million visitors were counted - and that the new opening hours until midnight were embraced enthusiastically, there is obviously no need to worry about this, not even in the long run. Rather, the current development can be characterised as a renaissance of libraries, which are enjoying a continuous increase in use as places of cultural and scholarly exchange and concentrated learning, even though increasingly comprehensive parts of their information offer are made available online.

Sometimes it is also doubted whether the "library digital copy" left at the free disposal of the Bavarian State Library will still be in demand after all, given that the identical "Google digital copy" can be retrieved using the most popular search engine of the world. From the perspective of the Bavarian State Library also this doubt is unfounded. The library is in fact not unhappy at all that its copyright-free collections will in the future be retrievable also via Google, the world's most frequently used web search engine. After all, this corresponds to the basic mission of every library: To join people and knowledge. The offers of Google Book Search and the Bavarian State Library therefore have to be understood as complementary. Ultimately, Google is striving to add additional content to its search index that is not at the disposal of its competitors, in an

attempt to consolidate its market leadership in the online-advertising business. The processing of extensive digital text collections for specific scholarly interests of use, and their embedding in web-based research and learning environments, which will be characteristic for the future offer of the "library digital copy" by the Bavarian State Library, does at least not represent the core business of Google. It can therefore be assumed that instead of competing with each other, rather two services and interests of use with a different focus will coexist.

A prognosis of the potential demand of the future mass offer of digitised research literature by the library can be made with the aid of the so-called "long-tail" phenomenon of the internet economy. The term long tail designates a curve for the sale of goods which drops steeply at the start and then slowly progresses toward zero. At the top of the curve are the products that are in strong demand (in the case of DVDs for example these are the latest blockbusters), at the end of the curve are the only marginally sought after niche products (for example a subgenre of Japanese anime films). In his much-noticed book "The Long Tail. Why the future of business is selling less of more" Chris Anderson, head editor of the US periodical "Wired" has shown by means of a multitude of examples that this curve is no longer universally applicable to internet trading.[9] Online traders such as Amazon, iTunes or Netflix owe up to one half of their turnover not to "hits" or "bestsellers", but to perceived niche products from the continuously flattening long tail. This circumstance is owed to the fact that in the internet the geographic boundaries of conventional, local and regional retail trading do not exist. In the web, goods can be offered globally and 24 hours per day, so that also such products which are intended for a special group of customers and not for the mass market meet with sufficient demand. Products that are regarded as "shelf warmers" in a conventional retail outlet or that are not even included in the line of goods, can be sought after by thousands of customers in online trading.

---

[9] Chris Anderson: The Long Tail. Why the Future of Business is Selling Less of More. New York 2006

Large universal libraries have conducted a long-tail business already under the conditions of the analogous age - consciously disregarding the dictate of economic usability -, by providing "product storages" for a demand that is usually rather special and unforeseeable regarding time and scope. In the digital age this offer now meets with exactly the right distribution channel to reach a demand potential that is spread throughout the world, but which is nonetheless of a significant total volume. This becomes obvious already when comprehensive retro-conversion projects render the metadata of specific collection segments globally accessible via online catalogues. After the conversion of the so-called "Quarto Catalogue" of the Bavarian State Library, containing the collections from 1841 to 1952, was concluded at the end of 2005, the demand for these titles within the framework of local lending and document delivery could be increased by more than 20 %. If in the future the digital full-text copies of increasingly comprehensive collection segments can be accessed directly via the catalogue entries, a significant increase in use stands to be expected in comparison to the conditions of analogous information provision. In this case the demand follows the offer. The unique holdings of the library, which had always been intended for use on an international scale, can now finally reach their clients, who had existed all along, but, for being spread all over the world, could not be addressed efficiently so far.

On the whole, the basic idea of the digitisation strategy of the Bavarian State Library is to contribute to the core mission of the library: To preserve the complete unique holdings for future generations and to simultaneously optimise access for the present generation. The "market place" in which libraries of future viability have to fulfill this mission is the web, which decisively shapes the work processes in science and research. Today libraries can no longer be confident that the users will adapt their workflows to the library's structures, but rather have to integrate their services and offers in the network-based work processes of the users. The strategy of the comprehensive digitisation of its holdings enables the

Bavarian State Library to bring its self-conception as an international research library to bear in the internet world in an appropriate manner.

Dr. Klaus Ceynowa
Deputy Director General
Bavarian State Library, Munich, Germany
ceynowa@bsb-muenchen.de

08-05-15