German ones, they only supplement them. They thus place huge pressure on our budgets.

eBooks have not become as well established within our branch as in other sciences. But that wave is hitting us hard at the moment, with many publishers now offering new or out of print books as eBooks, too.

For years the publisher Beck used to be the biggest player among several other competitors. Over the last years, Beck has purchased some of these rivals and thus increased its dominance in this specialized market.

5. CONCLUSION

Yes, there are law libraries in Germany. And there are people who work in them. But there are surprisingly few people all over Germany who can or should be called law librarians. Those few professionals who fulfil the criteria face the challenges outlined above, but it must be said that most of these are challenges that the allocation of sufficient resources to libraries could resolve. In this respect, there is little we can do as law librarians except lobby our sponsors for money and the legislator to change the copyright law.

Biography

Christian Wolf graduated in economics and law inter alia at Aberystwyth University. He qualified as a senior librarian from 2003–2005 and is currently the law librarian at the Philipps-Universität in Marburg, Germany.

Retrospective Digitisation of Legal Sources in Germany

Abstract: This article, written by Ivo Vogel and Elisabeth Schrecklinger, deals with the efforts of German libraries to digitise historical legal sources and make them publicly available. Although the main focus is on two selected libraries, a general overview is included. Commercial products are not considered since their contents are likely to become increasingly less relevant due to German law libraries' own initiatives. More attention is paid to problems that have been identified during the implementation of digitisation projects such as the recording of full texts is highlighted. The retrieval of digitised legal materials and, eventually, the digitisation of historical legal gazettes or parliamentary literature is also discussed. This contribution focuses exclusively on the retrospective digitisation of historical legal materials.
Keywords: law libraries; legal information; digitisation; Germany

FROM CATALOGUE ENRICHMENT TO MASS DIGITISATION

A few projects started the extensive digitisation wave of legal content in Germany. Two libraries were particularly prominent in that sphere: The Max Planck Institute for European Legal History (Max-Planck-Institut für europäische Rechtsgeschichte) in Frankfurt/Main[1] and the Berlin State Library – Prussian Cultural Heritage (Staatsbibliothek zu Berlin – Preussischer Kulturbesitz)[2] with their unique law collections. Regarding these projects a very different quality and quantity is to be considered, as well as the progress that has been made in this field during the last 10–15 years.

105

Ivo Vogel and Elisabeth Schrecklinger

MAX PLANCK INSTITUTE FOR EUROPEAN LEGAL HISTORY PROJECTS

The Max Planck Institute for European Legal History project "Juristische Dissertationen des 16. – 18. Jahrhunderts aus Universitäten des Alten Reichs"[3], which was initiated in 1998, covered approximately 73,000 legal dissertations of the 16th up to the 18th century from the area of the old empire. First, all writings were indexed systematically in the library catalogue. In addition to the cataloguing, special key pages were digitised in order to offer genuine added value for research. After completion of the project more than 92,000 title pages and dedication pages were accessible as high-quality colour images. Thus, this project can be seen more as catalogue enrichment than as a full digitisation. Links to access the digitised materials were created exclusively via the Max Planck Institute's catalogue, so that it is not possible to use those resources beyond the catalogue. This fact may cause individual usability conflicts. In the context of recent mass digitisation all dissertation texts might be progressively converted into digital format.

The project "Literatur zur Geschichte des deutschen, österreichischen und schweizerischen Privat- und Prozessrechts des 19. Jahrhunderts" ("Literature on the History of the German, Austrian and Swiss Private and Procedural Law of the 19th Century") is a major step towards mass digitisation. With funding from the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG) this substantial collection could have been digitised and thus made accessible to a wider scientific community. Overall, 4,316 books with approximately 1.350.000 pages were digitised from 1997 to 2002. This project meets the high demand for relevant literature in one of the preferred research fields of the newer legal history, namely the development of the private and civil procedural law of the 19th century. The Max Planck Institute's intention was to emphasise their support to this field of research by digitising those sources. Due to its broad topical spectrum this digital collection might not only be brought to the attention of legal historians. Books on inheritance and family law or on labour and social law can be as useful for social historians as works on trade and commercial law for economic historians. Most of the works are represented in several editions, which document the development and change of legal views. The books which were digitised within the project are directly accessible on-line. By recording the full text of the works' contents an indexing level was created which exceeds the existing conventional library indexing. However, again full text recognition is not realised yet.

COOPERATION IN DIGITISATION OF HISTORICAL LAW JOURNALS

The cooperative project "Juristische Zeitschriften 1703 – 1830" ("Legal Journals 1703–1830")[4] between the Max Planck Institute for European Legal History and the Scientific Information Service for International and Interdisciplinary Legal Research of the Berlin State Library has a very special character and marks a big step forward. The materials will not be selected; rather, all journals from the relevant period will be digitised. The basis for this project is Joachim Kirchner's work "Bibliographie der Zeitschriften des deutschen Sprachgebietes bis 1900" ("Bibliography of Journals of the German-speaking region up to 1900")[5]. Kirchner's bibliography will help to identify the appropriate titles. The journals which are to be digitised do not necessarily represent the holdings of one individual library. The larger aim of the cooperation between the Max Planck Institute and the Berlin State Library is to overcome institutional barriers and to build a truly comprehensive digital collection.

The journals of all German-speaking territories and all legal disciplines will be digitised within the project and thus made accessible to scientists in a modern, digital environment. When the project was launched 31 of the total 248 titles were already digitised; the overall plan is to digitise 217 journals in their entirety. However, at the moment only 182 titles are processed within the project as these are available at the project libraries. In the long term a completion of the digital collection should be achieved. The purpose of the project is, in particular, the digitisation of the journals and the formal cataloguing of individual contributions by using metadata and substructural indexing; continuing the Max Planck Institute for European Legal History project "Digitalisierung juristischer Zeitschriften des 19. Jahrhunderts" ("Digitisation of Legal Journals of the 19th Century")[6], which was carried out from 2002 to 2006. The technical project parameters are based on the DFG Practical Guidelines on Digitisation (Praxisregeln der Deutschen Forschungsgemeinschaft "Digitalisierung")[7]. On the one hand, this includes a colour digitisation with Tiff-files as digital masters in order to generate web files as well as an indexing of descriptive and structural metadata in MODS/METS format. In cooperation with the Max Planck Digital Library (MPDL) the digitised journals are provided on the web as an "eSciDoc Solution". The GWDG (Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen) provides the web servers, which are nessessary for the presentation of the digital journals' collection on the web as well as the network infrastructure. They also ensure the long-term digital preservation of the data.

BERLIN STATE LIBRARY PROJECTS

The second major institution that plays an essential role in digitising legal literature is Berlin State Library. In 2001 they started their activities in this field with a highly ambitious project, supported by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG). The core of the project "Preußische Rechtsquellen Digital" ("Prussian legal sources – digital")[8] was the comprehensive compilation of two Prussian legal collections of 1298 up to 1810: Corpus Constitutionum

106

Marchicarum (CCM) and Novum Corpus Constitutionum Prussico-Brandenburgensium Praecipue Marchicarum (NCC). Although these two volumes were published separately they needed to be digitised as a whole as they are considered in relation to each other. The CCM and NCC can both be seen as very heterogeneous materials. These sources are primarily text-based, but double column arranged, set in blackletter, and are partly accompanied by short comments. Moreover, the works contain multi-page illustrations (e.g. of diverse coins or tools) as well as large portions of tables and legal forms. More than 25,000 pages in folio format were scanned as bitonal images with a resolution of 600 dpi in TIFF 6.0 format. Within this process, technical and bibliographic data were stored in the TIFF-Headers. Finally, metadata and sub-structural indexing in XML was added to the collections by taking into account the Text Encoding Initiative (TEI) document type definitions.

Serious problems have been detected, particularly while recording individual documents. Due to the works' strict hierarchical order (particularly in the CCM), individual volumes have numerous paginations, which make their use more difficult. For instance, any column information of the collections needed to be standardised. Unique referencing of the pages was achieved by using a concordance between the page key and the associated image. Since summer 2002, the images (25,120) and successively also the sub-structutal data has been available on the web. This presentation has been made possible with the help of a javascript-based leafing tool. The current version makes it possible to leaf through the images as well as providing specific access to individual edicts by navigating through the contents. The collection of source material has been bibliographically and formally portioned into 32 (CCM), and 62 (NCC) sections according to volumes and issues of the printed version. Also, information from 9.525 edict contents and chronological registers were grouped and can be searched through the browser. The respective column information has been linked to the relevant image and leads directly to the requested page.

An essential project aim was to be linked to other projects, particularly to the Dictionary of Historical German Legal Terms (Deutsches Rechtswörterbuch der Heidelberger Akademie der Wissenschaften – DRW).[9]

This dictionary of the Germanic legal language refers to many partly digitised sources. Therefore, the digitally processed works CCM and NCC have been significant and often cited sources of the dictionary. As a result, the Dictionary of Historical German Legal Terms contains more than 1,400 references to State Library digital content. A link leads directly to the page with the cited references and ensures a connection between the two projects.

Regarding their web presentation, the Prussian Legal Sources are not up-to-date. Therefore, web presentation and search functionalities need to be improved. A re-launch needs to be realised in order to enable users to search for individual title key words or subject headings in the register under several search aspects. Further, a combined search for date of publication of the relevant legal source should be offered. A full-text search, including all criteria, would complete the search facilities mentioned above. The Berlin State Library faces a big challenge in realising all those new features. More precisely, retrieval difficulties may occur due to historically founded variations of German orthography, since the sources are represented in their spelling (e.g.: Münze [Müntze], Ernte [Erndte], Strafe [Straffe] and Holz [Holtz]). Several programs that are able to recognise various historic German spellings will be tested and developed in order to solve the existing problems and to provide successful search results.

The Berlin State Library has improved and modernised its procedures with regard to mass digitisation. The digitisation project "German Territorial Law from 1801–1900"[10], also supported by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG), focuses on a core of the State Library's legal collection, which includes the world's most comprehensive corpus of literature on German territorial law, with a high proportion of rare and unique resources. The collection consists of about 12,500 volumes and approximately 2.5 million pages. 'German Territorial Law' is a Berlin State Library pilot project. It has been the first large digitisation project, that is working with an external service provider. The project needs to handle two different workflows. Books of a favourable conservation state are being digitised by the external firm whereas works in a bad conservation state are being digitised in-house. All processes, which are necessary to ensure a good project organisation, are being documented by the project team. As a result, coming digitisation projects will be able to profit from the 'Territorial Law' project's experience. This could be a considerable advantage, since numerous special questions emerge in a project with the complexity of the 'Territorial Law'. In the long term all relevant parts of public domain resources of legal literature are to be converted into digital format. The digitised materials are being made freely accessible in the State Library's digital collection[11]. They are navigable through a metadata and sub-structural indexing, which is added to the images before they are presented on the internet. Thereby direct access will be created meeting the requirements of web-based, scientific working. The image quality of the digitised material will allow automatic character recognition (OCR). This feature is planned for the time after completion of digitisation. At the moment, 'Territorial Law' is improving its web presentation. Among others, several new features will make the use of the digitised materials easier.

STRATEGIC AND TECHNICAL ISSUES

Obviously, the digitisation projects described above have developed in very disparate ways as regards their technical conditions. There is a pressing need to create common standards. The DFG Practical Guidelines on Digitisation (Praxisregeln der Deutschen Forschungs-gemeinschaft

"Digitalisierung") may serve as guide towards realising this goal. The DFG expects that several projects will make concentrated efforts to ensure a high degree of recognition and an intensive use of the resources provided. Resources should be integrated into existing or emerging specific portals or catalogues. Now, each project guarantees an individual view at the existing digital collections themselves; for example, the zvdd (Zentralverzeichnis für Digitalisierte Drucke)[12] serves as a central portal for accessing digital content of printed publications of the 15th century up to now. However, there is a high need to improve several aspects such as universality, completeness, timeliness and expert and material based access. It is a pity that several library catalogues do not provide enough records of freely accessible digital sources, which is why users cannot be sure of finding relevant digital content, depending on the catalogue they search. But it is precisely these digitised publications that need to be displayed in library catalogues or discovery systems since library users are used to performing parallel searching. Therefore, every German library network should import digital content and make it available in local catalogues (at least those libraries which hold a physical copy should do so). Furthermore, digitisation projects should be taken into account and included when building discovery systems.

The Deutsche Digitale Bibliothek[13] marks a bright spot at the national level. It is dedicated to creating unlimited access to Germany's cultural and scientific heritage by establishing a network of all German culture and science institutions' digital content. However, so far this platform faces the same difficulties as the zvdd.

Looking at the projects discussed, another serious problem becomes clearly apparent, namely OCR recognition. Significant OCR (Optical Character Recognition) successes have so far only been achieved with works from the 19th century onwards as well as with works in blackletter from the second half of the 19th century onwards.

There are various aspects that make OCR recognition a serious challenge for some time to come:

Historical publications contain old typefaces that cannot be recognised by standard OCR programmes while modern computer types do not offer suitable equivalents. Moreover, the quality of the letters themselves is often bad (e.g. incomplete characters, mixed with dirt, handwritten comments or scan noise). When these documents were first published, there was no common spelling, so that very different kinds of writing can be observed. Scans of these documents are mostly inadequate. The most frequent sources of error are curled paper, nested layouts, curved lines etc. Historical books often show layout structures different from those of modern texts. Special algorithms, which were developed for the recognition of modern characters, do not produce satisfactory results when working with old documents. Additionally, projects need to decide how to handle books which contain dusty pages. A reliable method is needed to remove dust from those books, in order to achieve reasonable OCR results. The Berlin

State Library is already working with some external service providers, who are responsible for removing dust from respective works.

A practical alternative to OCR may be found in direct recording[14], which offers two methods used to transcribe texts: the single-key and the double-key method. In the latter case, a text will be recorded twice before the variations of both texts will be compared and filtered out automatically. This method ensures a nearly correct recording of 99,997%. Actually, an accuracy below 99,5 % can be considered worthless when recording manually, since every hundredth letter would be misspelled (one mistake in a line).

Direct recording might not be the best choice for mass digitisation projects, since it is particularly costly and requires a lot of human resources. However, there is a pressing need for all past and present digitisation projects to catch up on a practical method of character recognition. For this purpose, a national strategy is needed.

The pilot project on OCR application used in the digitisation of funeral scripts at Berlin State Library can be seen as a significant improvement in this field[15]. The main goal of this project was to explore ways and means of a programme for automatic character recognition of early modern prints from the German-speaking region. Within the project, two different OCR programmes were tested thoroughly. These programmes were then examined using standardised and comparable criteria. In the next step, they were assessed on OCR quality and configuration or optimisation opportunities.

HISTORIC LEGISLATIVE TEXTS AND PARLIAMENTARY PAPERS

At the end of this paper, mention should be made of some types of material that are highly relevant for legal historic research, namely legal gazettes and parliamentary papers. The present conditions to research historical German legal gazettes and parliamentary papers are not ideal, although these are public domain materials under German copyright law. Some projects in public institutions have started to deal with this issue (e.g. the project "Verhandlungen des Deutschen Reichstags" – "Session Reports of the German Reichstag")[16]. However, a nationwide and coordinated policy for digitisation is not yet in place. Instead, German libraries need to make available additional resources to acquire historical legislative texts and parliamentary papers conventionally. Therefore, there is a need to find an overall solution in this concern. This can probably only be achieved efficiently under the coordination of the Deutsche Digitale Bibliothek cooperating with the law libraries.

Other projects such as "Thüringen Legislativ Exekutiv"[17] seem to be quite confidence-inspiring but are very much a local product. The Austrian National Library project "Alex"[18] provides legal gazettes of the Third Reich, the Federal Republic, and of the provinces, as well as

stenographic protocols and thus may serve as a benchmark. Moreover, a chronological and formal structure, a reference search and a search in legal gazettes is provided.

CONCLUSION

Overall, German law libraries are heading in the right direction to realise retro-digitisation of legal materials in order to make them available to the world's legal (historical) research. Numerous small digitisation projects could not be mentioned in this article, although a number of legal materials are being digitised retrospectively.

Google's project to digitise large library holdings in co-operation with libraries is rather unsystematic and does not only focus on law. The large law libraries in Germany should set up a consortium, that would coordinate all projects, dealing with the digitisation of German law materials. This network could contribute to the final goal to digitise German law materials in their entirety. Accompanying services, such as digitisation on demand, should complement this process. The Scientific Information Service for International and Interdisciplinary Legal Research of Berlin State Library will set up such a service in a timely manner, supported by the German Research Foundation.

Footnotes

[1] The digital library of the Max Planck Institute for European Legal History offers the title pages and dedications of more than 72,000 legal dissertations of the 16th–18th centuries, nearly 450 titles, eg. German legal literature from the 15th to the 16th century as digital facsimiles or partly in full text, approximately 20,000 pages of Legal sources of the Holy Roman Empire (digital facsimiles, table of contents as searchable full text), first results of the current digitisation project 'Legal journals from 1703 to 1830' and the 'Legal journals from 1800 to 191' collection of 75 journals containing 1.320 fully digitized pages volumes as well as a collection of 4316 fully-digitised volumes of 'Source literature on German, Austrian and Swiss private and civil procedural law of the 19th century'.

[2] 'Fachinformationsdienst Recht' <http://staatsbibliothek-berlin.de/recherche/fachgebiete/rechtswissenschaft/> accessed 26 March 2014

[3] 'Juristische Dissertationen des 16. – 18. Jahrhunderts aus Universitäten des Alten Reichs' <http://dlib-diss.mpier.mpg.de/> accessed 26 March 2014

[4] 'Juristische Zeitschriften 1703–1830' <http://www.rg.mpg.de/de/virtuellerlesesaal/zeitschriften1703-1830.cfm#i324> accessed 26 March 2014

[5] Joachim Kirchner, Bibliographie der Zeitschriften des deutschen Sprachgebietes bis 1900' vol 1 'Die Zeitschriften des deutschen Sprachgebietes von den Anfängen bis 1830', p. 140–152 (nr. 2486 – 2709) and p. 365 (1969 Hiersemann)

[6] 'Legal journals of the 19th century' <http://www.rg.mpg.de/en/virtuellerlesesaal/zeitschriften1800-1918.cfm> accessed 26 March 2014

[7] 'DFG Practical Guidelines on Digitisation' <http://www.dfg.de/formulare/12_151/12_151_en.pdf> accessed 26 March 2014

[8] 'Preußische Rechtsquellen Digital' <http://web-archiv.staatsbibliothek-berlin.de/altedrucke.staatsbibliothek-berlin.de/Rechtsquellen/> accessed 26 March 2014

[9] 'The "Deutsches Rechtswörterbuch" – Dictionary of Historical German Legal Terms' <http://www.rzuser.uni-heidelberg.de/~cd2/drw/drw_english.htm> accessed 26 March 2014

[10] 'German Territorial Law' <http://staatsbibliothek-berlin.de/en/about-the-library/projekte/deutsches-territorialrecht/> accessed 26 March 2014

[11] 'Digitalisierte Sammlungen der Staatsbibliothek zu Berlin' <http://digital-beta.staatsbibliothek-berlin.de/suche/?DC=rechtswissenschaft> accessed 26 March 2014

[12] 'Zentralverzeichnis für Digitalisierte Drucke' <http://www.zvdd.de/en/start/> accessed 26 March 2014

[13] 'Deutsche Digitale Bibliothek' <https://www.deutsche-digitale-bibliothek.de/> accessed 26 March 2014

[14] Example for direct recording: 'DRQEdit' <http://drw-www.adw.uni-heidelberg.de/drqedit/> accessed 26 March 2014

[15] 'Pilotprojekt zum OCR-Einsatz bei der Digitalisierung der Funeralschriften der Staatsbibliothek zu Berlin' <http://staatsbibliothek-berlin.de/die-staatsbibliothek/abteilungen/historische-drucke/projekte/funeralschriften/> accessed 26 March 2014

[16] 'Session Reports of the German Reichstag' <http://www.reichstagsprotokolle.de/en_index.html> accessed 26 March 2014

[17] 'Thüringen Legislativ & Exekutiv' <http://www.urmel-dl.de/Projekte/LegislativundExekutiv.html> accessed 26 March 2014

[18] 'ALEX – Historische Rechts- und Gesetzestexte Online' <http://alex.onb.ac.at/index.htm> accessed 26 March 2014

Biographies

Ivo Vogel is a lawyer and chief law librarian at Berlin State Library (Scientific Information Service for international and interdisciplinary legal research). Since 2010 he has been a board member of the International Association of Law Libraries.

Elisabeth Schrecklinger has been working for the Berlin State Library digitisation project "German Territorial Law" 1801 – 1900" since July 2012.