

## **Report on my professional visit at British Library, London and Bodleian Digital Library Systems & Services Department, Oxford in December 2014**

**(supported by BII grant)**

As I have been involved in the organization of a digitisation project on German Territorial Law in the 19<sup>th</sup> century at the Staatsbibliothek zu Berlin, my professional visit was aimed at exchanging with my British colleagues to get an idea of their digitisation activities and their project implementation. Moreover, I wished to speak to my colleagues, so that we might be able to discuss and exchange information about any emerging project issues, which may face us in the future.

### The British Library

From 8<sup>th</sup> to 12<sup>th</sup> December I visited the British Library in London, where my colleague Rossitza welcomed me with an introduction in the most significant digitization activities of the BL. She also gave me an excellent overview about the digitised collections on the BL website. Links to any important websites will be provided below or come with the project descriptions.

Indeed, all digitization projects at the BL are externally funded, and there are numerous collaborations with very different foundations on an international level. This fact brings great opportunities to the library, and it enables them to implement a wide range of projects. Also, the funding route (as in the case of the BL and Qatar Foundation partnership) allows the library to design their projects in a bespoke way. This means, that funding sometimes includes i.e. financial help to equip a project with its own digitization studio, or - as a project - to be able to come up with new, innovative ideas, and bring them into the implementation process. I will talk about what this can be in detail, later on.

In the following, I will describe those projects, which I visited. I will also name those projects, which I did not visit and give their URL, at least.

Europeana 1914-1918 and Europeana Newspapers:

<http://blog.britishnewspaperarchive.co.uk/2014/11/18/9-million-newspaper-pages-are-now-online/>

The European Library:

<http://www.theeuropeanlibrary.org/tel4/newspapers/issue/3000095266893#>

Manuscripts:

[http://searcharchives.bl.uk/primo\\_library/libweb/action/search.do?dscnt=1&fromLogin=true&dstmp=1418051255568&vid=IAMS\\_VU2&fromLogin=true&fromLogin=true](http://searcharchives.bl.uk/primo_library/libweb/action/search.do?dscnt=1&fromLogin=true&dstmp=1418051255568&vid=IAMS_VU2&fromLogin=true&fromLogin=true)

### Google Books Digitisation

The British Library as well as the Bodleian Digital Library Systems & Services Department, have been implementing joint projects with Google, to digitise large numbers of books. Google co-operations have considerable advantages for the libraries, as Google meets the costs for both shipping and digitisation. After completion of digitisation, Google provides the files of the digitised materials on an on-line platform. The Partner Libraries receive copies of the files to preserve in their repositories and display on the Library's own viewer. A lot of collaborative staff effort within the British Library goes into selecting and preparing the books for digitisation and monitoring the download of the files.

Digitised by Google (example):

[http://books.google.co.uk/books?id=Xa5bAAAACAAJ&printsec=frontcover&source=gbg\\_summary\\_r&cad=0#v=onepage&q&f=false](http://books.google.co.uk/books?id=Xa5bAAAACAAJ&printsec=frontcover&source=gbg_summary_r&cad=0#v=onepage&q&f=false)

## International Dunhuang Project

<http://idp.bl.uk/>

IDP is an international collaboration of different cultural heritage institutions, dealing with the digitization of any kinds of literature and materials relating to Dunhuang and archaeological sites of the Eastern Silk Road. IDP London, located at the British Library and established in 1994, acts as the directorate of IDP. It maintains the IDP database and web site and carries out fundraising. It coordinates the conservation, cataloguing, digitisation and research of the Central Asian manuscripts in the British Library collections in liaison with British Library conservators and curators as well as scholars worldwide. It has its own digitisation studio using the latest digital backs on high-format cameras, as well as staff specialising in conservation, cataloguing, research and education. It is largely externally funded.

IDP London also hosts the images and metadata of other collaborating institutions which do not have their own IDP website, such as the British Museum and Princeton University. (from:<http://idp.bl.uk/pages/about.a4d>)

## The British Library Qatar Foundation Partnership

<http://www.qdl.qa/en>

The Qatar Digital Library (QDL) is making a vast archive featuring the cultural and historical heritage of the Gulf and wider region freely available on-line for the first time. It includes archives, maps, manuscripts, sound recordings, photographs [...], complete with contextualised explanatory notes and links, in both English and Arabic. The QDL has been developed as part of a 10-year Memorandum of Understanding on Partnerships between the Qatar

Foundation, the Qatar National Library and The British Library. (from: <http://www.qdl.qa/en/about>)

The [Qatar Digital Library](#) [...] will improve understanding of the Islamic world, Arabic cultural heritage and the modern history of the Gulf providing researchers around the world with the opportunity to perform groundbreaking research in subject areas such as the history of Gulf trade and politics, key individuals in the Gulf and the history of science in the Arabic-speaking world. (from: <http://www.bl.uk/qatar/>)

As regards copyright questions for the digitisation, this project has been of particular significance. There is a copyright team attached to the project, as their materials cover a large period of time. The latest documents can be considered of the very recent time. The copyright team is busy writing letters to the copyright holders in order to clear the rights. For that reason, any volumes must be examined carefully to identify the rights relating to individual documents.

Endangered Archives Programme

<http://eap.bl.uk/>

### **The Threat to Archives**

**“Documentary heritage reflects the diversity of languages, peoples and cultures. It is the mirror of the world and its memory. But this memory is fragile. Every day, irreplaceable parts of this memory disappear forever.”**

UNESCO Memory of the World Programme.

(from: <http://eap.bl.uk/pages/threat.html>)

EAP is a world-wide BL-project, which is funded by Arcadia.  
<http://www.arcadiahfund.org.uk/about-arcadia/about-arcadia.aspx>

With this project, the British Library wants to make a contribution to the preservation of archival material that is in danger of destruction, neglect or physical deterioration world-wide. It enables countries to digitise their national archives and collections. Particularly, it is targeted to such countries, which would not be able to carry out such digitisation projects by their own efforts.

The digitisation must take place in the origin country that holds the collections. In case of a lack of expertise, training can be provided prior to the actual project start. After the completion of digitisation, hard drives with the digitised material must be sent to the British Library, which cares for the importing of the data into BL's systems. The Library preserves the digital files and currently gives access to them via a dedicated website. The project receives minimal descriptive or structural metadata. This is not a common practice in the British Library digitisation, in general.

Overall, I feel deep appreciation for the Endangered Archives Project. It helps countries not only to digitise their national literature, but also makes a significant contribution of developing aid.

### BL Labs

British Library Labs is an initiative funded by the Andrew W. Mellon Foundation (<http://www.mellon.org>) that invites researchers and developers to work with the British Library and its digital collections to address important research questions. (from: <http://labs.bl.uk/>) This externally funded initiative gives the Library the opportunity to understand how users interact with both digital born and digitized content and will inform decisions about new services that will

support user engagement with the Library's digital collections and digital research practices. The BL Labs targets the researcher community, in order to find out, how they would like to work with the digitised content. Evidence shows that users no longer want to consult individual items through the Library catalogues but discover whole collections, including full text files and images within books, and export any freely reusable content for the purpose of text mining and other digital analysis. Projects can be designed in a bespoke way by gaining feedback from professionals. This can happen in many ways, i.e. in organizing talks, where digitisation can be discussed against different backgrounds (digital conversations). Moreover, BL Labs carries out competitions. One of the library's key findings in this field is that it is important to researchers to have high resolution digitised images accessible that are OCRd. However, substantially less attention is given to the addition of any structural data to the digital copies. The level of structural metadata for British Library digitisation projects varies depending on the size and resource dedicated to the project. For more information on the important and innovative work of BL Labs see the Labs website (link see above).

## IT

Cooperation procedures between the projects and the IT department are often rather limited, and it mainly involves storage considerations. The BL's IT department has developed share points for some of the major projects. So, if a project collaborates with an external partner, both can access a common share point on-line. This is a good way for all sides to consider the current project status at all times, and it helps to avoid vast communication with external partners, as well as within the projects. Currently the Library has projects that are reviewing and developing a new strategic end-to-end digitisation project.

## Networking activities

Digital Preservation Coalition:

<http://www.dpconline.org/events/previous-events/1340-current-trends-and-future-directions-for-digital-imaging-in-libraries-and-archives>

IIIF (International Image Interoperability Framework)

<http://iiif.io/>

IIIF is a world-wide digitisation network of cultural heritage institutions. Its goals are:

- To give scholars an unprecedented level of uniform and rich access to image-based resources hosted around the world.
- To define a set of common application programming interfaces that support interoperability between image repositories.
- To develop, cultivate and document shared technologies, such as image servers and web clients, that provide a world-class user experience in viewing, comparing, manipulating and annotating images. (from: <http://iiif.io/about.html>)
- The British Library is implementing a new Viewer for its digitized content which will be IIIF compliant.

For more detailed descriptions on the use of IIIF, please see the IIIF website.

## Wellcome Trust Library, London

During my London visit, I also met the Digitisation Programme Manager at the Wellcome Trust Library. Like the Staatsbibliothek zu Berlin, the Wellcome Trust Library has been using Goobi as their workflow software.

The Library has been implementing a cooperation project with the Internet Archive (<https://archive.org/index.php>). The practice in this project is to harvest the requested internet archive's files, in order to import them automatically via Goobi into their own presentation. Eventually, the digitised materials exist twice. However, the Wellcome Library prefers this practice rather than to create links, which would lead to the Internet Archive's presentation. Like in several other BL projects, which I have visited, the long-term preservation is left to the ingest coordinator.

At the Wellcome Library, I had the opportunity to cast a glance at the OCRd collections. OCR is being done in different ways. Within the internet archive project, the digitised materials can be OCRd with the IA's OCR software, which is Abbyy. Another way is to do OCR through Goobi with Intranada (<http://www.intranda.com/en/digi/intranda-ocr-service/intranda-ocr-service-overview/>). However, all the OCR is being done via Abbyy. OCR procedures have been working very successful, even with books, which contain different kinds of characters. By using the OCRd collection it is possible to search one special term within a volume, or to highlight passages. Here is an example for a successful OCRd work at the Wellcome Trust Library:

<http://wellcomelibrary.org/moh/report/b19821955/39%20-%20?asi=0&ai=39&h=war&z=-0.5851%2C0.2348%2C1.9361%2C0.8312#?asi=0&ai=0&z=-1.366%2C0.0594%2C3.1339%2C1.4304>

Bodleian Digital Library Systems & Services Department

The following week - from 15<sup>th</sup> – 19<sup>th</sup> December 2014 - I visited the Digital Library Systems & Services Department of the Bodleian Libraries in Oxford.

<http://www.bodleian.ox.ac.uk/bdlss>

One of the Bodleian Digital Library Systems & Services Department's strategies to find new digitization collaborations is to review their own collections and to check, which libraries might have similar collections, which could complement their own holdings in a reasonable way. Thereby, a significant, new digitization project is just emerging in cooperation with the Vatican Library. <https://www.vatlib.it/home.php>

The Bodleian Libraries found a comprehensive corpus of religious literature from the 12<sup>th</sup> and 13<sup>th</sup> century in their holdings, which can be complemented with respective holdings of the Vatican Library in Rome very well. During my Oxford visit, the colleague of the Vatican Library also visited the Oxford colleagues, and I had the pleasure to join this meeting. The Italian colleague has been advised by a British Library colleague, which also took part in the meeting. We had been given a presentation of the current project status in Rome by the Vatican librarian. The project has been using Goobi as their workflow software. The two project partners plan to create a common on-line platform to provide the content, which shall be independent from the Bodleian or Vatican presentations. The overall aim in the collaboration between the Digital Library Systems & Services Department of the Bodleian Libraries in Oxford and the the Biblioteca Apostolica Vaticana (Vatican Library) is to opening up their repositories of ancient texts for the public. This project partnership is funded by the Polonsky Foundation. <http://bav.bodleian.ox.ac.uk/>

*Note:* The British Library also has a Polonsky Foundation funded project that is digitizing 1250 Hebrew Manuscripts from the Library's collection.

The colleagues in Oxford are endeavouring to find U.S. foundations to support their digitization activities, as well. With most of the American grants, it is possible to implement projects much more individual, as there are less requirements and less monitoring, as the U.S. see themselves mostly as sponsors.

An overview of all current digital projects in Oxford can be found at <http://www.bodleian.ox.ac.uk/bdlss/digital-projects/current-projects>.

## OCR

Like in London, it is as well a common practice in Oxford to include scholars more and more into the projects, in order to assess their working interests. Similarly as the BL Labs project, the Oxford colleagues think of new ideas, how to provide the digitized content for an ideal use including OCR i.e. searchable content. For instance, users could directly get involved into the actual OCR process, whilst a project would create OCR crowdsourcing features on its website, which would enable them to make any enhancements in a work as needed. A respective announcement could be made onto the project website. Such a practice could be conceivable, provided that the community would be willing to do OCR work on their own.

In the follow-up to the presentation of my own project and digitization activities in Germany, which I gave to the Oxford colleagues, they came up with some interesting and useful OCR ideas for my Berlin digitization project

work within the Territorial Law project. For instance, they suggested building a feature, which could enable users to search a special term and to find out, how this term could have changed or if there could have been a different use of it in different areas.

### After my professional visit in England...

...and some Christmas holidays, I went back to Berlin being greatly inspired by what I saw and heard about the British digitisation activities and with lots of ideas. I am most grateful for this exchange of experience and to have met lots of wonderful colleagues, who gave me such a warm and friendly welcome, and shared their best practice with me. I had been completely unable to cover all aspects of my UK visit in this report. However, I have sought to provide a helpful overview.

by Elisabeth Schrecklinger

Staatsbibliothek zu Berlin – Preußischer Kulturbesitz

Digitisation project "German Territorial Law 1801 - 1900"