Collaboration is Key – The importance of collaboration and partnership in digitisation projects

Yves Maurer
IT Department / Digitisation Services, Bibliothèque nationale de Luxembourg, Luxembourg.
yves.maurer@bnl.etat.lu

Claus Gravenhorst
Strategic Initiatives, CCS Content Conversion Specialists GmbH, Hamburg, Germany.
c.gravenhorst@content-conversion.com

Abstract:

In 2003 the National library of Luxembourg (BnL) started with its first digitisation projects in image-only mode. Very quickly the BnL realized that full text search was expected by users and so the library started using METS/ALTO in 2006 during a first trial with CCS. The digitisation project is annual, with an annual budget and an annual European tender. This restrictive working mode is necessary because of budget issues but has also helped the BnL by allowing the iterative improvement of processes each year.

From 2000 – 2003 CCS Content Conversion Specialists GmbH (CCS) collaborated in the EU funded METAe project with library and university partners. The project aimed to develop a comprehensive software product for the conversion of printed historical documents. CCS provided its software technology as well as expertise and learned from the project partners about their business and requirements regarding digitisation, conversion, preservation and access. Through the fruitful collaboration with libraries and universities this project represents the model a successful partnership. Since 2005 CCS has been engaged in various mid- and large-scale projects. It became obvious that both service provider and institution should be open to learn from each other and share knowhow and experience to lead the projects to success.

The BnL has always thought that it's important to share and collaborate with others. So, just as the BnL is a grateful recipient of knowledge, experience and tools from companies and the library community, it also open-sourced its newspaper viewer and publishes its tender documents for the inspiration of others. Good collaboration improves our processes and our results, which produces better services for users. It also helps to find solutions to unforeseen problems. With this paper the BnL and CCS would like to share with IFLA members the story of their collaboration.

Keywords: collaboration, newspaper digitisation, newspaper conversion

Collaboration is Key – The importance of collaboration and partnership in digitisation projects

The National library of Luxembourg traces its origins to 1798 as the central library of the "Département des forêts" during the occupation of present-day Luxembourg by the French revolutionary troops. During the 19th century and until the 1960s it remained a small institution and even nowadays it's one of the smaller national libraries. Besides having the patrimonial mission to collect and preserve the printed heritage of Luxembourg, the BnL is also a public library and an encyclopaedic and scientific library. Two-thirds of the holdings are non-patrimonial material in all subject areas.

As the biggest library in Luxembourg, and the first one to use an Automated Library System, the BnL has been at the forefront of library automation in Luxembourg. This partly explains why it is responsible for the national library network and why it provides Integrated Library System (ILS) services for 66 other libraries throughout Luxembourg. The national Consortium for acquisition and access to licensed content is run by the BnL, where its partners are the University of Luxembourg as well as three public research centres. These types of networks enable institutions to put resources in common to solve common problems and are a very successful example of institutional collaboration.

Indeed, the BnL implemented its first computerised cataloguing system with the help of RERO (REseau ROmand) which is a Swiss library network. Collaboration with Swiss institutions comes naturally for Luxembourg, because the issues with a multilingual environment are similar in both countries. When the BnL switched over to a different ILS in 2000, it concluded a service contract with the IDS (Informationsverbund Deutschschweiz) where again it could take advantage of the experience the IDS members had with the software and its configuration in a multilingual setting.

In each setting where the BnL partners with other institutions to learn from their experience, it also tries to give something back. The BnL was an early and active member of Europeana and has led the Europeana Connect Licensing Workpackage. In the Europeana Newspapers project[1], the BnL is an associated partner and actively shares its experiences with newspaper digitisation in METS[2] and ALTO[3].

Collaboration between people, institutions and companies allows finding solutions to complex problems that would go beyond the expertise of a single entity. Many of humanity's big projects like the great pyramids, the moon flight or the development of information technology are in fact stories of collaboration. We have luckily evolved from collaborating on moving huge blocks of stone, enforced by an overseer with a whip to collaborating in public-private development and research contracts like the NASA did with private American companies for space flight. In IT, collaboration happens most between private companies faced with questions of interoperability, standards and common infrastructure, such as e.g. Open Source projects financed by commercial companies that are natural competitors like IBM and Oracle who both actively contribute to Linux. But the basic tenet that other entities

are more specialised in a certain field than yourself and hence can help you to develop a better overall solution, has remained the same.

Beyond the expertise sharing, collaboration also allows entities to consider more viewpoints and get a wider variety of thoughtful analysis. As one of the co-founders of Sun Microsystems eloquently put it: "No matter who you are, most of the smartest people work for someone else". In order for collaboration to be successful, the institution or company has to be willing to recognise this basic tenet and be humble enough to admit that it doesn't know everything about the problem and that partners have constructive value to add. Admitting this in a given context is a first sign of trust between partners. For a fruitful collaboration this trust has to be well tended to. Clear, open communication is the basis for that.

In order to be open and clear about its requirements, the digitisation department of the BnL has developed a detailed requirements document that explains most of the processing steps for digitising newspapers in detail. It's partly a procedural requirement due to the nature of the European tender process it's required to follow, but it's also a great basis for the BnL and the service provider to exchange views and solve problems arising during the production. Since 2007 when the first newspaper tender was launched and 2014 when the latest one is in the works, these requirements have undergone numerous clarifications and iterative improvements thanks to the BnL's experience from previous projects and a constructive dialogue with service providers.

After a first wave of newspaper digitisation in the early 2000s in image-only mode, the library wanted to use OCR to enhance the discovery and usability of the material. The outcomes of the METAe project looked very promising, since they addressed more than just OCR, with the ALTO format, by also adding metadata about the scanning process and descriptive metadata about articles and sections using METS. A first prototype with CCS was successful and during that period, the BnL learnt a lot about the METS/ALTO digitisation process from CCS. It was clear from the start that while CCS had the technical expertise to produce METS/ALTO, the BnL had to decide on its markup policies which guide what exactly is an article, a section, a supplement etc. A big effort was launched by the BnL to do manual mark-up of a large sample of newspaper issues to identify the general mark-up policy.

Collaboration between the BnL and CCS has gone beyond only newspaper digitisation projects. When the BnL decided to digitise scientific serials, with a very different layout than newspapers, discussions with CCS were helpful to determine the page elements that should be detected and the way that e.g. printed table of contents could be treated. Requirements are always a compromise between wanted functionality, feasibility and effort required. Since the library doesn't do in-house digitisation on the scale as specialised service providers, getting their feedback on feasibility is very valuable. In a recent pilot project for digitising historical statistical material, CCS proposed a very innovative use of METS/ALTO, with the embedding of Excel tables linked to the pages where the historical data series appear.

Since digitising and producing METS/ALTO is only the first part of the process of making historic newspapers available to patrons in a new form, the visualisation of these METS/ALTO files has also been a concern of both the BnL and CCS. Here, at the request of the BnL, CCS collaborated with the ExLibris, the company developing the Digital Asset Management system used by the BnL in order to achieve interoperability. Later, when CCS was exploring options to develop their own display solution, they consulted the BnL to get its

opinion as an active user of a METS/ALTO-capable system. As the BnL developed its own viewer for METS/ALTO, it decided to open-source it and to make it available to the community under the permissive BSD license:
https://sourceforge.net/projects/bnlviewer/
This viewer has already been used by the National Library of Latvia's periodika.lv project and is undergoing active development.

Since the digitised newspapers have been online, the interest from research groups at universities has been growing. The BnL has developed a custom access solution for one group which investigated the 100 years of publication of the "Tageblatt" one of main newspapers in Luxembourg. It also makes its layout and OCR data available on request to projects like SDiv (Standardization in Diversity: The case of German in Luxembourg 1795-1920) which aims to uncover the historic standardisation of German or the HERA-funded ASYMENC project (Asymmetrical Encounters: E-Humanity Approaches to Reference Cultures in Europe, 1815-1992) which explores the influence of larger, reference cultures on smaller countries like the Netherlands, Belgium and Luxembourg. In turn, the BnL wants to learn from these projects about new methods in which to index and present the content it is digitising.

CCS Content Conversion Specialists GmbH (CCS) is a Hamburg based German Technology Company developing market-leading software and hardware for the creation and display of digital collections. The conversion software docWorks serves small-, medium- and large-scale digitisation projects, and is used by the most renowned national and university libraries, publishers, service and content providers. CCS is active in the field of software and workflow automation and is a member of the two important XML-standardizing committees METS and ALTO.

With regard to newspaper digitisation and conversion to XML content providers such as brightsolid (UK), Library of Congress (USA) and various national libraries in Europe, Australia and Asia are using docWorks in-house. But also service providers like Contentra (India), Digital Divide Data (USA, Asia) and Leta (Latvia) make use of this technology to process newspapers for outsourced newspaper digitisation and conversion projects.

From 2000 – 2003 CCS collaborated with 12 library and university partners in the EU funded METAe project[4]. The project aimed to develop a comprehensive software product for the digitisation and conversion of printed historical documents. As commercial partner CCS provided its software technology and digitisation/conversion expertise and learned from the project partners about their business and requirements regarding digitisation, preservation and access. Through the fruitful collaboration with libraries and universities this project represents the model of a successful partnership.

Since 2012 CCS has been active in the Europeana newspapers project and collaborating with 18 project partners. The goal is to convert 10 million newspaper pages to full text and to aggregate 18 million historic newspaper pages for Europeana and The European Library.
CCS has been providing its expertise and technology to set up and operate a mass digitisation workflow to create high quality structured content from 2 million scanned newspaper pages. The distributed workflow enables the contribution of project partners (content providers) to the integrated quality assurance process. All partners have one common goal and are actively sharing their know-how and expertise.

In the past CCS has been engaged in various mid- and large-scale digitisation/conversion projects. By gaining more experience and improving product and project management it became obvious that both service provider and institution should be open to learn from each other and share experience to lead the projects to a success.

Both cultural heritage institutions and commercial product/service providers have their expectations regarding openness and quality of communication. When approached by commercial providers, institutions often have complaints like "Don't presume you know my business and your solution fits! Get to know me first!". This seems to be a major issue for many institutions. They miss the willingness of really getting to know the customer, to empathize with him and to have a genuine interest in exploring and understanding his situation. Institutions would like commercial providers ask the important and correct questions, and above all to listen to the institution's representative.

For both sides trust is the necessary basis for collaboration and partnership. By showing integrity, honesty and transparency a relationship has to be built by both partners. Once established, it is important assume ownership, in order to maintain the relationship and guide the partner along the way.

Even if these basic rules are adhered to, the industry still complains about the fact that institutions tend to be reserved when approached by commercial suppliers. In these cases, the willingness to be open and build a relationship seems to be missing and in consequence, there is no way the customer's situation can be understood and suitably advice provided. Establishing a relationship already fails at the beginning and we often don't get to know the reasons for the lack of openness.

For successful collaboration the following basic rules should be followed:

- Listen and learn about others
- Show respect for others and their opinions
- Be open and transparent, share knowledge and experience
- Be honest and communicate clearly, concretely and correctly
- Build and earn trust

Lessons learnt from our various projects and contacts to institutions show that trust is not a given – it must be developed and earned. Both sides must be open to sharing knowledge and experience. It's also important to learn from previous projects and put improvements on the agenda. For clear communication and project control, dedicated project managers with clear responsibilities should be in place on both sides. A web-based collaboration platform enables fast and transparent communication and is a must at least for mid- and large-scale projects. Problems shouldn't be used to blame one other; they occur in every project and just need to be solved. It's essential to deal with problems and issues in an open, respectful and professional manner. Last but not least, pay tribute to objectives and milestones achieved and celebrate the end of successful projects!

CCS's collaboration with the BnL started in 2005. CCS and the BnL started off by sharing their knowhow and experience regarding historical newspapers and the digitisation/conversion to highly structured digital objects, in several face to face meetings. The BnL gave valuable information regarding the required markup and tagging to create administrative and structural metadata for newspaper issues while CCS shared details about

the conversion technology and the workflow to process bounded newspaper volumes and structure issues down to article level. Giving detailed information about the XML standards METS and ALTO for the description of the digital object and recording the metadata was also well received by the BnL. For both the BnL and CCS this was an important first step to building the relationship and earning each other's trust.

Having established the relationship, CCS entered into a first newspaper conversion project with the BnL in 2006. The goal was the processing of 55.000 newspaper pages on article level with the required output format METS/ALTO.

In 2007 the BnL launched a tender for the digitisation and conversion of 155.000 newspaper pages and contracted a service provider to scan the bound volumes and deliver article-based METS/ALTO output. For the conversion part to XML, CCS acted as a sub-contractor. There were many issues with scanning quality and respective rework, the BnL had to talk to two different providers and it was difficult for CSS to keep up the quality of the relationship and collaboration. In 2010 CCS had been contracted as a principal supplier for scanning and conversion of a newspaper collection with 125.000 pages. Scanning was provided by a subcontractor, but as the general contractor CCS was able to manage the project based on clear and fast communication. Both CCS and the BnL solved problems in a respectful and timely manner.

Another tender followed two years later: The BnL contracted CCS again in 2012 to process another newspaper collection of 125.000 newspaper pages (19th/20th century). Since there were still some quality issues in the 2010 project with scanning quality, CCS decided to provide both scanning and conversion to achieve the highest quality possible and avoid unnecessary rework. An important part of the project's kick-off meeting was a review of prior projects as well as defining the criteria for improved scanning quality. As with the other projects the BnL nominated a dedicated and experienced project manager, which makes it easy to have clear and effective communication. Both sides were and still are committed to the project goal and to handling problems in an open, appreciative and frank manner.

The BnL provided a very clear and detailed project specification, which saved a lot of time and consulting effort at the beginning of the project. To enable clear communication with a high level of transparency, the web-based MS SharePoint was and is being used as the project's collaboration platform. As much information as possible is shared via that platform such as delivery lists, details about redeliveries, clarification of open issues, BnL quality reports on CCS deliveries (batches), hard disk drive tracking list for traceability of shipped items as well as shared documents like specifications, the project time schedule (Gantt chart with tasks and milestones), meeting notes, samples, change requests etc.

The project's end is scheduled for Q3/2014 and a final face-to-face meeting is planned in order to conclude the project with a review and to honour the work carried out by both partners.

For over 20 years, CCS has been developing solutions for the digitisation of valuable library holdings and has been engaged in research projects as well as various outsourced digitisation and conversion projects. We have learned to listen to cultural heritage institutions and understand their business and requirements. We always shared our know-how and expertise and picked up a lot of valuable pointers from the cultural heritage community. Looking back on our past and current activities we can state that successful projects need fruitful

partnerships with powerful collaboration. As both the BnL and CCS are following the rules for successful collaboration, the relationship has reached a high level of quality and trust.

When the BnL subcontracts with private companies it's not only about getting a job done, but also to get a share from the domain expertise that companies build up when they work in a specialised field. So the BnL is always open to listen and try to understand our partner's point of view and also to share again the knowledge it accumulated through previous work. Confrontational relationships based on narrow reading of contractual obligations rarely produce satisfactory results. A collaborative approach lets trust develop and is often the key to success.

---

[1] Europeana Newspapers Project, 2012 – 2015, http://www.europeana-newspapers.eu/

[2] **M**etadata **E**ncoding and **T**ransmission **S**tandard, http://www.loc.gov/standards/mets/

[3] **A**nalyzed **L**ayout and **T**ext **O**bject, http://www.loc.gov/standards/alto/

[4] Metadata Engine Project, http://meta-e.aib.uni-linz.ac.at/