



European Union
European Regional
Development Fund

Leverage from
the EU
2014–2020

Report on the benchmarking interviews in the Digital Open Memory project

Liisa Näpärä, PhD

Erno Liukkonen, MSc

2020

Contents

1	Introduction.....	1
2	Benchmarking.....	1
2.1	Interviews.....	2
2.2	Summary of the benchmarking data.....	3
3	The motives and stories behind building labs.....	4
4	The offering and operations of labs.....	5
4.1	Datasets.....	5
4.2	Tools.....	6
4.3	APIs.....	7
4.4	Collaboration with researchers.....	8
5	Funding and staff.....	9
6	Networking and communication.....	10
7	Facilities.....	12
8	Strategic tips for our lab.....	13
9	Summary.....	14
	Data.....	15
	References.....	16
	Appendix.....	i

1 Introduction

This report is part of the Digital Open Memory (DOM) project, funded by the European Regional Development Fund and its Leverage from the EU 2014–2020 programme. It discusses the library lab benchmarking that was conducted as part of the DOM project. The report's focus is on developing the research services at the National Library of Finland, but it also reflects the views of the first author, Liisa Näpärä, in trying to create an understanding of what the conceptual and concrete meanings of the term 'library lab' are. Because the author's background is in the humanities and anthropology, the interpretation of library labs and digital humanities offered in this report reflects this. The author wrote the interview questions and conducted the interviews together with Minna Kaukonen, the head of planning at the National Library of Finland. The technical sections concerning the lab benchmarking were written in collaboration with Erno Liukkonen, the DOM project's technical expert.

This report uses the name 'Louhos' (Mining Lab) to refer to the library lab that is being developed at the National Library of Finland and the term 'library lab' or 'lab' to refer to other hubs that bring together library professionals, digital information resources, researchers and research tools. Many library labs are multidisciplinary in nature and fall under digital humanities in national libraries or university libraries. According to various definitions, what makes a library lab is the interaction and physical space that connects data and researchers, but the labs have yet to establish a single model. All library labs are unique, and no uniform model can really be offered. Moreover, not everyone understands the concept of library lab in the same way. This report outlines how the people interviewed describe the lab in their own context.

Although library labs come in all shapes and sizes, they all share certain key features: digitalism, openness, innovativeness, accessibility, collaboration, multidisciplinary, user-centredness and research integrity. Although these keywords could be used to describe the values of modern science in general, labs are essentially meeting places where library professionals and researchers interact with data. Library labs can be broadly divided into three categories: national library labs, university library labs and public library labs. This report focuses solely on the first two categories and ignores the last, because the purpose of collecting and analysing data for the DOM project is to create a lab for the National Library of Finland to offer data-driven researcher services primarily to digital humanities scholars.

Much like the library lab, digital humanities is also a concept that entails relatively many definitions and operates somewhere at the intersection of the humanities and computer science. Different schools and researchers understand the concept of digital humanities somewhat differently. (Nyhan et al. 2012.)

2 Benchmarking

Benchmarking is essentially peer learning: it means learning from role models and organisations whom you aspire to emulate. The method was originally developed in Japan after the Second World War and it has mostly been applied in the corporate world (Holtzman et al. 2011). Many benchmarking guides therefore describe how to develop products or production processes, but the method is equally applicable to developing services, such as the researcher service concept at the National Library of Finland. In benchmarking, the operations of others are observed and their best practices are incorporated into the organisation's own strategy.

We began the library lab benchmarking process by exploring how labs are described online. Particularly important sources were lab websites and the Open a GLAM Lab book from 2019. In addition to online material, some of which was of very high quality, we wanted to hear from people who work in library labs because we wanted to achieve further clarity on what it is that the labs do. After all, no public website can offer the full picture, even if labs are known to be very open and outspoken about their work.

At the early stages of the benchmarking process, we contacted lab members who were already a part of our network and with whom Minna Kaukonen had been in contact. Our purpose was to start the benchmarking process at the Digital Humanities in the Nordic Countries conference in Riga, for which we had already set up meetings. A little more than a week before the conference, however, COVID-19 closed the borders and forced us to cancel our conference trip. We then changed our approach and started to increasingly conduct lab member interviews online and contact people who came up in different sources. We also trialled a short questionnaire seeking experiences from people working in labs, but because we mainly found our interviewees through other channels, we did not share the questionnaire very actively after its publication; we apparently had already succeeded in contacting the key people who were willing to share their experiences.

2.1 Interviews

In the spring and early summer of 2020, Minna Kaukonen and I conducted interviews with members from seven labs, 12 people in total. Individual interviews included a varying number of people: the least number being only one interviewer and one interviewee. Before the interview, we gave each interviewee a set of themed questions (see Appendix 1), which we had tailored specifically for each lab. At the time of the interviews, some of the interviewees did not yet consider their lab to be a lab, and our approach to the interviews was shaped by our thoughts and assumptions of what the labs do. We also did not rigidly follow our list of questions, so the interviews were semi-structured thematic interviews in which the questions were used to loosely guide the interview.

We conducted the interviews on Zoom and asked the participants for permission to record them for the DOM project's internal use. The length of the interviews varied from half an hour to a little over an hour, depending on the interviewees' schedules and the amount of information we had acquired about the lab before the interview. We transcribed the interviews using clean verbatim transcription, i.e., only including what was relevant. This made analysing the interview data faster and easier, although the transcription work itself was nevertheless time-consuming. We informed the interviewees that they would have the chance to read our report about the interviews later and to decide if they wished to be identified by name. If possible, we will also make the report available to all members of the library lab network more generally and share the results within the community.

Because the lab community is small, we will refer to the interviews by the lab's name. However, as many labs are centred around one or two people, it is more than likely that people working in labs will be able to recognise one another and identify the individuals who took part in each interview. We cannot therefore guarantee anonymity, which is common when studying and observing a small community, but anonymising individual interviewees gives the participants at least some degree of privacy outside the lab community.

All the labs included in the benchmarking data collected for the DOM project were willing to share their knowledge with us and eager to participate in the interviews. It should be noted, however, that despite

their openness, not all interviewees or conference and seminar speakers had enough time to cover everything they wanted in detail. Time proved to be the main limitation in the interviews, but in addition to the set of questions provided, the personal interests of the people interviewed also guided the discussion. Because the nature of the labs is to share failures as well as successes, perhaps to an exceptionally great degree, we did not have to dig deep or choose our words carefully to uncover possible failures or areas for development, which might have been the case with more sensitive topics. This does not mean that trust did not play an important role in the interviews; it was necessary for further building and maintaining trust in general. The interviewees did not bring up any actual failures, but they did mention some areas for development. They also discussed some things others should avoid or pay special attention to. We will return to these tips later in this report.

2.2 Summary of the benchmarking data

The benchmarking data collected in the DOM project consist of interviews with lab members, lectures and conference materials freely available online, a few articles, and observations made about lab websites. The key labs are presented in the following table.

Table 1: Labs included in the benchmarking data

Lab	Home	Type of data
Biblioteca Virtual Miguel de Cervantes Lab	Spain, Alicante	DOM interview with one interviewee, and seminar or conference presentations
Bibliothèque nationale de France	France, Paris	DOM interview with two interviewees
Bibliothèque nationale du Luxembourg	Luxembourg	DOM interview with three interviewees
BL LAB	The United Kingdom, The British Library	Seminar or conference presentations
KB data lab Sweden	Sweden, Kungliga bibliotek	DOM interview
KB Denmark	Denmark, Kongelige Bibliotek	DOM interview, and seminar or conference presentations
KB Lab Netherlands	The Netherlands, Koninklijke Bibliotheek	Seminar or conference presentations
Royal Library of Belgium's Digital Research Lab	Belgium, Royal Library	DOM interview with two interviewees
SBB LAB	Germany, Staatsbibliothek zu Berlin	DOM interview with one interviewee
ÖNB LAB	Austria, Österreichische Nationalbibliothek	Seminar or conference presentations

A vital part in gaining a preliminary understanding of what labs are was the above-mentioned book *Open a GLAM Lab* by the GLAM lab community; we even managed to speak to a few of the book's authors during the process. GLAM refers to a much wider community than just libraries: it is an acronym for galleries, libraries, archives and museums. Because many of the authors represent library labs, it was only natural for us to look for inspiration in the book. Our understanding of labs was also influenced by the 2018 datalab report by Pelle Snickars, which has been used to create the model for the Swedish lab. Both the datalab report and the GLAM book offer a more detailed discussion of labs than this report does.

Our benchmarking data also include lab websites, various events associated with digital humanities and other resources, including historian Mila Oiva's dataset *Facilitating Digital Humanities Research 2017* available at the Finnish Social Science Data Archive. The interview questions used in Oiva's dataset caught my attention when we were creating our own interview questions and gave inspiration for some of them. Oiva's dataset does not focus on labs, but instead charts the facilitation of research in three digital humanities centres, thus mainly serving as a reinforcement to the data that we collected.

We analysed the benchmarking data using traditional qualitative content analysis, looking at the data from the context of our own lab project. Because our lab is currently in the planning stage, the analysis focuses on the general nature and purpose of the labs. However, many of our themes follow those of the GLAM book, which offers a host of advice on setting up a GLAM lab. The themes of our report are 1) building a lab, 2) lab offerings, 3) staff structure, 4) facilities, 5) networking and communication, and 6) strategic areas of development. The facts and interpretations are based on the discussions and available material in spring and summer 2020. The motives and stories behind building labs.

3 The motives and stories behind building labs

At the moment, there seems to be a very high demand for labs: several labs have been established in Europe in the past couple of years. One of the main reasons for setting up and developing labs is to promote the use of libraries' digital collections and to create new tools for using the collections. Labs promote the use of digital collections in collaboration with researchers, but other users can also reap the benefits of the work put into exploring what the collections contain and the tools developed to use the data. However, researchers play a special role in this work because the law prioritises research over other uses.

The creation of labs was spurred by meetings in London in 2018 and in Copenhagen in 2019. The British Library Lab has played a central role in the European lab network: it has inspired the foundation of many other labs and actively encouraged people to share their experiences, successes and failures. The lab team has also encouraged people to join the lab network and share ideas with each other. Many labs have indeed learned from one another both in terms of their websites and staff. For example, the Belgian lab's representatives said that they have used the data lab website by the National Library of Luxembourg as a model for building their own website, which was still a work in progress at the time of the interview. In a similar fashion, our project also looks for inspiration in what the other labs offer, what their websites are like and how they present their operating model and the tools required to process digital collections and data.

In general, the lab websites that we explored were easy to use, but some labs only offer information in their local language, making it harder to learn about their work. For example, the French lab's website is in French only, which can make it hard for some to understand the website contents. Some labs also require users to register before they can access all features. On the Austrian lab's website, for example, only registered users can access certain tools. Without publicly available material, we would not have been able to include all these labs in this report. Thanks to the openness that is central to lab work, we have been able to create and maintain collaboration networks necessary for international research purposes.

Many lab teams participate in international conferences that bring together lab people from all over the world to learn about how other labs work. Many of the interviewees mentioned how important the conferences in London and Copenhagen were for establishing their own labs and for networking. Although many conferences discuss digital humanities and libraries more generally, they may offer a programme dedicated to labs. The labs have also organised several international events for exchanging information

targeted at people interested in the field. We will discuss these in more detail under the section ‘Networking and communication’.

An additional incentive for founding a lab has been funding. For example, the French lab received funding from a multiannual research project for developing its operations, and in Luxembourg, the development of operations was expedited by some long-awaited funding secured for digitising their entire collection. Funding has enabled the development and systematic reorganisation of existing functions, but it has rarely been enough to run all the desired operations.

4 The offering and operations of labs

All library labs seem to have certain things in common: digital data, multidisciplinary collaboration with researchers, and openness to innovations. Although labs strive to make their resources as open as possible, national legislation, copyright issues and limited digitising resources impose some restrictions on their openness. In fact, all our interviewees had been asked why certain resources cannot be made public. Countries differ by as much as decades in terms of when they made their collections open, and they also apply different licensing practices, causing confusion especially among researchers who frequently ask why certain data are not available even for national research use. In Luxembourg, for example, only data in the public domain are included in datasets. Copyright matters affect the amount of data available at the labs and partly determine what collections will be digitised to allow as many users as possible to openly access them. For these reasons, all labs try to bring together digital data and researchers.

All the labs develop tools for using their data, and often also for accessing their largest digital collections, newspaper archives. Because newspaper issues have accumulated regularly over a long time, in some cases over centuries, and libraries have centred their digitisation efforts primarily on them, newspapers are the type of material most widely available. To use them in research, researchers often need various downloadable datasets and APIs. Many labs have also developed a wide variety of other tools for processing data and visualising results. For example, the KB Lab in the Netherlands, which has a strong research focus, lists almost 20 tools on its website, more than 10 of which are for analysing newspapers. Based on information available on the lab websites, the KB Lab offers more tools than any other lab, which not only speaks of the resources available in the Netherlands, but presumably also of the historical significance of the newspaper and of the wider freedom of the press. However, that is hardly likely to directly explain the large scale of the newspaper digitising efforts and their use in research.

Many labs divide their website into the following three sections: tools, datasets and APIs. The British Library Lab is an exception to the rule, instead only listing its tools in GitHub. In general, the lab websites offer a brief description of each tool, dataset and API, often with a related image. From this general description, users can then move to a dedicated page, which offers more specific information about the tool, dataset or API and allows users to download the content. Some labs also allow users to filter the content using a category or keyword related to a tool or dataset. On the Netherlands lab website, users can filter content based on category, file format or the tool’s or dataset’s copyright licence. Some labs are only starting out and do not yet have a website or have a website that is still a work in progress.

4.1 Datasets

Based on our benchmarking data, many labs have produced their datasets and tools in collaboration with researchers and for research purposes. The labs have created a wide variety of datasets, most of which focus

on collections that include newspaper page images and OCR text, but they also include datasets related to books, brochures, maps, music and archived websites. The number of datasets offered varies greatly between labs, which in part reflects the fact that the labs are in different stages of development. The British Library Lab offers more than 150 datasets with a very wide variety of digitised collections, whereas the Luxembourg lab only has datasets related to historical newspapers available for downloading on its website.

The data included in the datasets employ various file formats depending on the type of content. Newspaper page images are typically in the JPEG or TIFF formats, and newspaper OCR text is typically in the ALTO XML format or in a regular text file. Many labs have developed various tools for processing the material included in the datasets, with the aim of making the material easier to use. For example, the British Library Lab and the Austrian lab have Jupyter Notebooks that instruct users how to process the data in Python.

Some labs also offer different-sized versions of their datasets to allow users to explore the contents more easily. This way, users do not necessarily have to download and process large amounts of data, but can instead conduct preliminary experiments on the data using a smaller dataset. For example, the Luxembourg lab offers several versions of its newspaper dataset, each in a different size and with different content, from which users can choose the one that best suits their purposes. The Luxembourg lab website offers a detailed description of the differences between the datasets and also explains the file formats included in each set.

4.2 Tools

Any tools developed in collaboration with researchers will primarily be made available in the lab for other researchers to use. The tools designed in labs are made with commonly used programming languages, such as Python, Java and JavaScript, which makes it possible for people of various backgrounds to modify and develop them further. To be successful, developers should have previous experience of the relevant programming language.

Tool documentation varies from tool to tool. Some include very detailed descriptions of how to use the tool and also provide comments in the code. In general, we found that the documentation related to using the tools was adequate, and the most significant shortcomings were associated with the commenting of the code, which makes it harder for users to modify and continue to develop the tool. If a tool is aimed at people who have very limited programming experience but nevertheless need to modify the tool, it is important that code comments and other documentation are fit for purpose. Inadequate documentation may sometimes be due to the incompleteness of the tool and potential further development that is still ongoing. For example, the tools in the Netherlands lab have good documentation and the tool code also includes comments, but the extent of code comments differs between tools. In the Denmark lab, users cannot download any tools or tool source codes, so they can only evaluate the tools' benefits on the lab website on which they can use the tools. Some lab websites had broken links and tools that produced an error message, but this was a rare occurrence and may partly be explained by the development work carried out in the labs.

The tools are mainly licensed using a licence that enables their free use and modification. For example, most of the tools provided by the Netherlands lab have a licence that allows their free use and modification (e.g., GPL, LGPL, MIT, Apache 2.0) and only one tool that is copyrighted.

A few of the labs explicitly state on their website that they offer technical support for the lab's tools and datasets and can also offer other types of support for researchers' own projects. Providing such support gives a positive impression of the lab's operations and encourages users to contact the lab team if they experience any problems. It also encourages people in different labs to collaborate with each other and find projects they could work on together. Especially the British Library Lab and the Austrian lab websites offer a good

overview of how their lab team can help researchers who use the lab's tools and datasets in their own projects.

However, not all labs were able to maintain tools developed for analysing their data, instead transferring the tool administration to universities and research groups, for example. We have given this matter a lot of thought at the National Library of Finland as well: what will we do with the tools researchers develop? Can we administer them ourselves? And can we make sure that tools developed by researchers will be available to other researchers using the same resources? There are many different solutions, but in general, the benchmarking data suggest that libraries should be familiar with the tools developed for using their own digital resources, regardless of who administers the tools. At a minimum, this means that staff members who are familiar with the library's digital collections and data must be able to steer researchers in the right direction, whether they are looking to use or modify tools developed by others. However, the interviews gave the impression that it would be best if all tools and refined data created at the lab were eventually made available to other researchers at the lab. In general, digital humanities researchers place great emphasis on methods and tools. In order to develop tools, labs should be able to guarantee their use in the long term, as sustainably as possible.

4.3 APIs

Many labs offer various application programming interfaces or APIs for accessing their data and the related metadata. The data accessible through an API are usually similar to the data available in the lab's datasets, but by using an API, users can specify in more detail what kinds of data they wish to access and in which format.

Labs generally use the OAI-PMH protocol in APIs that offer access to metadata. This protocol offers access to collections that include different types of material and allows users to determine the format in which they want to retrieve the metadata. The metadata is typically represented in the Dublin Core or MARC21 XML formats. Based on the addresses of any files (e.g., page images or ALTO XML) included in the metadata, these files can also be downloaded. Some labs offer Python libraries that make it easier to use the APIs in Python, and some labs also offer Jupyter Notebooks, which also instruct how to use the APIs in Python. For example, the Netherlands lab has a Python library for using the API, for which they also offer clear instructions. The Austrian lab offers a Notebook in which the use of the API, implemented through the OAI-PMH protocol, is explained in Python using the Sickle library. The labs in the Netherlands, Austria, Luxembourg and Denmark all use APIs that employ the OAI-PMH protocol.

A few labs use IIIF APIs for downloading images and image metadata. The IIIF Image API request allows users to define in detail the size, quality, rotation, file format and portion of the image they want retrieved. Users can also request the metadata related to images through the IIIF Presentation API. The metadata request can include information related to the title, date, file format and height and width of the image. The labs in France and Austria, for example, use the IIIF APIs to relay information about their image collections.

Lab websites have different ways of presenting the copyright and licence information of the lab's tools and downloadable datasets. Some labs clearly state the copyright status of a tool or dataset on its download page, but other labs make visitors look for this information. Most labs make copyright information readily available on their website, but in some labs, the copyright information for certain tools is not provided until the user proceeds to GitHub, from where they can download the tool. The Netherlands lab offers a good example of how to clearly present copyright information for tools and datasets.

4.4 Collaboration with researchers

Many labs organise tech-centric hackathons, datathons and transcribe-a-thons. This is the case with the SBB Lab in Germany and Biblioteca Virtual Miguel de Cervantes Lab in Spain, for example.

At the time of the interviews, not all labs were yet able to offer all of these, or not in a broader manner. For example, the Luxembourg lab is working on an open data website that already offers a large amount of open data in the public domain, but does not yet include interactive demos on using the data. Due to this lack of interaction, the Luxembourg lab would not call itself a lab just yet.

Interaction between the lab team and researchers benefits both parties. Research brings in funding for the labs and creates new uses for the labs' collections and data. Labs provide researchers with data and can, when working in close collaboration, also offer researchers their expertise. This allows researchers to distinguish themselves, much like in any other research. The team at the Biblioteca Virtual Miguel de Cervantes Lab in Spain hopes to see scholarly articles on everything that is done at the lab, regardless of whether the original idea has come from the researcher community or from within the library. Writing openly about lab work is a way of promoting awareness of the lab and improving its standing among researchers. This also increases awareness and appreciation of the collections and creates new ways of using them.

The Spanish lab's team also hopes that all work done at the lab will end up featured as part of teaching, because most researchers also have teaching duties. Students have also been frequently involved in various hackathons, datathons and transcribe-a-thons held to promote the use of digital collections. According to the interviews, involving students in these events allows staff to regularly gain a fresh perspective on the digital collections and increases their potential as awareness of them increases.

In our interviews, we asked the lab representatives about their ongoing collaboration projects with researchers, because we wanted to expand our understanding of how such collaboration works by learning about it in detail. The various projects in different labs shared some similarities both in the topics studied and the technology applied. Several interviews mentioned the sampling of articles and the refining of metadata and descriptions. This was aimed at making the data easier to find.

When we conducted the interviews, a few large-scale research projects were under way in Europe. These projects mainly aimed to improve the OCR of (newspaper) collection texts and develop tools for researchers using these collections. Many of the lab representatives whom we interviewed said that their data are included in the *Impresso* and *NewsEye* projects. Representatives of labs that were not project partners in either project expressed their wish to participate in at least one because they considered such collaboration projects vital to the development of the field. New international collaboration projects were planned and some collaboration projects within the network were being launched during the benchmarking project in spring 2020.

Libraries' own smaller-scale collection projects were also international in nature, although their internationality depended somewhat on the language group because the collections of each country are mainly built around the country's national language. Languages both connect and separate us. Small national languages do not come out on top in collaboration projects because there are wider applications for data in French and German than for data in Finnish and Swedish. In some countries, including Finland and Belgium, the handling of the country's own collections must already cater for their built-in multilingualism. Some funding partners may also require that research focuses on a certain geographical region and language group, as is the case in Belgium.

According to lab representatives, there is a need for more resources for undertaking research projects because in many labs, there is demand for novel and exciting collaborations. Some lab representatives fear that the volume will grow too large for their lab to be able to adequately commit to all collaboration projects.

When the number of projects available exceeds the possible level of commitment, labs must choose which projects they want to participate in. To do so, labs use an application process, only selecting the best or most important projects in terms of data use for long-term collaboration.

5 Funding and staff

According to Shen and Varvel (2013, 556), in developing data management services at the Johns Hopkins University, the two main requirements for providing data services are staff and funding. The two are closely interconnected in labs because funding equals employment opportunities. Securing external funding is particularly important for labs, but acquiring strategic funding for core lab activities seems to be almost as important.

In the labs included in our interviews, the staff most commonly included a research coordinator and a technical expert, both of whom were also more or less familiar with the library's digital collections. As a rule, however, the library staff had to be very familiar with the library's digital collections, or the lab had to have a staff member responsible for the digital collections. The interviewees also highlighted the importance of having legal expertise at the lab, an area which did not receive much attention in Snickars's (2018) report. We would have expected legal expertise to gain an even more prominent role in the interviews because each lab has had to consider legal matters, at least from the perspective of copyright. Copyright matters did indeed come up very frequently in the benchmarking data. Especially representatives of the French lab reported that legal matters were a challenge in their work, also in regard to the agreements made in researcher collaboration. The Luxembourg lab was very careful not to give open access to any copyrighted material. We assume that legal advice is available in some form in each lab because lab members are aware of the exact nature of making digital resources available. The SBB Lab in Germany, for example, reported that legal matters had not resulted in any negative consequences for them. Labs seem to take copyright matters very seriously, which is only natural for organisations that operate under national libraries. It is, however, in the interests of many to open collections and data to wider use.

A staff structure similar to the one reflected in the DOM project data has already been presented elsewhere. In his datalab report, Pelle Snickars recommends that a library lab should be staffed with a minimum of three people: a library lab manager to coordinate work with researchers, a data curator and a developer. Snickars views the lab's role as a link between academic research and the library's services and collections as an integral part of the lab's work, a view that is also apparent in our benchmarking data. Lab staff must be familiar with the processes of academic research and be able to apply for external funding together with researchers. Lab staff should therefore include at least one, or preferably more than one, person with a research background who also has a solid understanding of the library's operations and technical developments (Snickars 2018).

In many of the labs included in our project data, the lab staff were very research-oriented and met the standards set out by Snickars. In Belgium, the digital research lab's staff were all at the postdoc stage, and in Sweden, everyone in the datalab's management holds a PhD. At the Biblioteca Virtual Miguel de Cervantes Lab in Spain, the staff were, at the time of the interview, mainly finishing their doctoral dissertations and will have reached the postdoc stage by the time this report is published. In addition to their duties at the lab, many lab members take part in research projects as researchers, which is the case in the Denmark lab, where the work duties of some staff members include participation in research that uses the lab's own datasets.

Labs must be able to meet the needs of digital humanities researchers and produce datasets that suit individual research projects; in some cases, offering datasets that are too generic and do not meet a research need is a waste of resources. Then again, more generic datasets, such as decade-specific newspaper datasets, serve students and those learning the tools developed for their use. These datasets also allow users to experiment with the data and the tools. In some cases, researchers create their own datasets and possibly also offer them for other researchers to use. In other cases, datasets include copyright restrictions and must be requested specifically. This is the case in France and Denmark, where in the latter the use of web archives for research purposes is quite liberal, but also requires the user to commit to carefully defined terms and conditions.

Not very many labs have a policy for creating datasets; instead, the datasets were created when a need arose. Producing a dataset requires both technical expertise and knowledge of the library environment, and it cannot always be confined to a specific research project. Dataset production often also requires programming skills and an understanding of machine learning and data visualisation. This is especially the case in multidisciplinary settings, which labs usually are by nature.

None of the library labs included in our project offered full-time employment to a great number of people: the average headcount was two or three. Because some labs did not have any funding of their own, the headcount ranged from zero to six. Those who participated in lab work but were not employed at the lab full time had external funding or undertook lab work in addition to their regular library or research duties. They mostly undertook lab duties because they were interested in them or their regular duties intersected with those of the lab. Many labs were run by just one person or a small core team and needed the support of various partners. Lab-related projects often included up to dozens of people whose work was funded either by a university or a foundation offering research funding. These people mainly worked outside the lab facilities and represented the lab's stakeholders from an organisational perspective, bringing the total number of people participating in the work of the lab to several dozen, even over a hundred people. These people formed an expert network for sharing views about research projects and the development of the use of collections and data.

6 Networking and communication

Many lab representatives took pride in how they had helped create a digital humanities network and considered their organisation as a central player in at least their country's national network. Considering the size of the lab community and the stories the interviewees shared about collaboration with researchers and other lab people, the network seems to expand beyond national borders and be very international in nature. Digital humanities centres outside national libraries and digital humanities researchers at universities have also played a key role in building this network.

In the international library lab or GLAM lab network, there are about 20 active participants who frequently talk about their lab work in conferences and seminars. The core network is fairly small, and the people involved in it mostly know each other in person or at least by name. However, the lab network sees no problem in further expanding the network and extending it to South America or Africa, for example. Expanding the network and integrating new institutions into the network is in fact considered desirable, and people work actively to grow the network. The network has indeed become very prominent in many online channels.

In our interviews, the multidisciplinary and interdisciplinary nature of the labs and research networks was never considered a problem. It seemed to be self-evident to all lab members that each collaboration partner is an expert in his or her field and that labs need expertise in different fields. Lab members also considered it vital to have diversity and people who represent different disciplines in order for labs to be agile research partners who keep up with the times.

Digital humanities, a discipline at the intersection of the humanities and technology, came up in some way in all the presentations included in our benchmarking data and in all of the labs' operations. A broad definition of digital humanities seemed to suit people well, allowing them to identify with different paradigms while also feeling that they are part of a group. Interestingly enough, when introducing themselves, many lab representatives positioned themselves either as a non-humanist or a non-techie, i.e. through a negation. By using a negation, they expressed more the direction they were leaning in, rather than saying that they did not have certain qualities at all. Many labs were more humanities-oriented, so the representatives of the Biblioteca Virtual Miguel de Cervantes Lab felt their lab stood out the most because their core team represents computer sciences and has digital humanities researchers only as its stakeholders.

When combined, collaboration activities between labs form a densely criss-crossing network for exchanging ideas, sharing knowledge and getting together for formal and informal meetings. According to the interviews, people in the network are happy to get together to spend the evening either with just the core group or with a wider group. Some prefer more formal communications and activities. Lab members seem to communicate more formally with researchers than with members of their own organisation, perhaps because labs mostly favour a low hierarchy and easy communication.

Electronic communication channels were an integral part of communications between different parties, but face-to-face meetings were considered important for networking. Many lab employees met both external and internal members of their network in person. The interviewees also considered it important that researchers can visit the library in person to get to know the collections and discuss possible research topics. Some meetings with researchers were highly informal and took place in a recreational setting, such as a bar or restaurant, but most meetings with researchers were of the more formal type and included workshops, lectures, symposiums and conferences, as is typical in academia. The labs therefore acted as a channel of academic communication per se, and lab members actively sought information about digital humanities events.

The most commonly mentioned communication channels in our benchmarking data were various instant messaging channels and email. Because labs have a culture of communicating and exchanging ideas openly, lab members were very open to using instant messaging channels. Quick communication was considered easier on Slack or Twitter, and the threshold for sending an instant message was lower than for sending an email that should include opening and closing phrases –although email comes almost as close to instant messaging as actual instant messaging channels when email etiquette is broken. Email was preferred for more formal communications.

In our benchmarking work, too, various electronic channels proved themselves useful and important, particularly in the COVID-19 spring. The pandemic kept people from meeting in person, causing many events to be cancelled or offered online instead. Because we could not meet people, we increasingly conducted our benchmarking process online. Although there are many opinions about how successful virtual interaction can be, it is clear that communicating online changes the nature of the interaction. When technology fails, it can become a problem instead of a solution, but when it works, it can bring us closer together and make communication faster and easier.

7 Facilities

Not all labs have physical facilities, or at least not enough facilities for researchers to use in the long term, making these labs more of a virtual meeting place for people. These meeting places are mainly represented by the lab website, where it is up to the visitor to get to know the demos and tutorials about the use of data, which each lab offered openly to all visitors. In such cases, lab teams had to find other facilities and platforms for meeting stakeholders. According to the interviews, these were often physical spaces – or at least the interviewees hoped that they would have physical spaces in the future.

Many interviews with and presentations by lab members highlighted how essential it is to exchange information and have a physical lab space. The Danish define a lab as a physical space, meaning that if a lab does not have physical facilities, it is not a lab. According to the views presented in our benchmarking data, having a physical space promotes creativity as well as collaboration and the exchange of ideas between researchers and library staff. Representatives of the Luxembourg lab also said that, by definition, a lab must include interaction that enables collaboration and makes the processing of data visible.

At the time of the interviews, people were more or less forced to work remotely and rely more on remote connections than before. The key message of the discussion around this topic was that virtual meetings should not become a permanent *modus operandi*. Many organisers chose to hold their events despite the pandemic, or postpone them by six months, which signals that people prefer to meet face to face rather than make events and meetings entirely virtual. Although the interviewees recognised the many benefits of virtual meetings, they said that virtual events cannot fully replace meeting people and establishing contact in person. In the future, however, it is likely that people will weigh up more carefully whether it is necessary to meet in person, replacing some travel with virtual meetings.

Physical facilities played the most concrete role in the labs that offer long-term research residencies. The researcher-in-residence programme at the KB Lab in the Netherlands accepts early-career researchers based on a call for proposals and the recommendations of a multi-member expert committee. In practice, the researchers accepted to the programme are working on their doctoral dissertation and gain access to the library's facilities and collections for six months. To be accepted, researchers must be able to apply the library's technical development interests to their own research questions. The residency programme is mutually beneficial: it allows researchers to grow professionally and the lab to promote the use of its digital data. In addition to these top-level goals, the lab has many micro-level development goals for promoting research and data use.

The residency programme is not an inexpensive solution. In 2019, the lab's residency budget was EUR 50,000, which had to cover the grant for two researchers for six months and possible library expenses. This means that the grant per researcher is not very substantial, but it is nevertheless a considerable benefit considering that most libraries and universities cannot offer similar facilities for researchers to use regularly over several years. The British Library Lab seems to be the only lab that is able to offer researchers more resources than the Netherlands lab. The British Library Lab hosts several competitions each year, looking for the best solutions for using the British Library's collections and data, rewards researchers who have made the best use of the data, and conducts research collaboration projects in a fashion similar to many other labs with less funding.

According to our benchmarking data, practically all labs would like to offer researchers residency programmes or grants if their budget allowed it. The Biblioteca Virtual Miguel de Cervantes Lab in Spain, for example, would like to set up a research grant programme that would promote the use of its data. The DOM project uncovered a similar wish in the responses to the researcher survey about the use of the digital

collections and services at the National Library of Finland. In this survey, researchers also hoped for a residency programme and for more facilities for research collaboration, but acquiring funding for the residency facilities and grants remains uncertain.

8 Strategic tips for our lab

The interviewees encouraged us to be open to collaboration and networking, and to start small and gradually work our way towards larger goals. As a general guideline, they also recommended keeping things simple. Although lab work requires creativity and innovativeness in the long term, the advice was to commit only to a few important things, those worth prioritising, at a time. Based on the information acquired from the interviews, labs should host as many projects as possible, so long as they benefit the lab's work. It is not enough for labs to provide data: labs must interact with researchers.

Our interviewees pointed out that it is worth publishing the lab's successes and achievements on the lab website, allowing the lab to set an example for others and attract new research projects. They also considered it vital to communicate lab work to external parties in an easily understandable way. Then again, lab members should be able to understand scientific research and participate in writing publications. The interviewees also warned us that bringing together library collections and science will not always be easy. For instance, terms used at the library may not be clear to researchers, and terms used by researchers may not be clear to library staff. This gap can be closed by a PhD-level lab member who is an expert in both communities and acts as a coordinator between the library and researchers.

All interviewees agreed that it is not easy to acquire funding and that the labs who manage to get funding for new projects and development work are lucky. Very few labs were in such a secure position that they did not have to seek additional funding. Because all labs had to apply for research project funding, everyone considered external funding desirable and thought that the lab's expertise and role in a research project should be highlighted in funding applications.

Lab members hoped that organisational hierarchies and traditional ways of doing things would not come in the way of their innovative and creative work. The labs' host organisations were sometimes considered to be rigid and almost the polar opposite of the labs, which represent a new way of thinking and seek to promote research and research tools in collaboration with researchers. This work should not be impeded by unnecessary hierarchies; our benchmarking data showed that what really matters is driving forward the things that are mutually beneficial for both researchers and the lab. This was realised particularly in the possibility of taking on duties based on personal interest.

In general, the host organisation management trusted their lab team and did not restrict or supervise the team's work very much. This freedom either derived directly from the lab's role in the host organisation or from its more informal operating model, and it was therefore not dependent on how strategically committed the management was to the lab's work. Innovation and the freedom to organise one's work seem to be prerequisites for lab work, so they were endorsed. It is also typical of labs to work in a network that is not tied to a physical space, which seemed to give lab members even more freedom in their work.

Some labs had to justify their work to the management more than others, although our benchmarking material does not allow us to offer any numerical estimates. According to our interviews, the entire organisation benefits from the lab's achievements. Labs bring their host organisations recognition and positive publicity, especially when research projects bring in money. Thanks to the labs, the use of the host

organisations' digital collections and data has increased and there are new tools for using them, which is in line with the objectives set for the labs.

Although interviewees considered it essential to state and justify the benefits of the lab's work to the management, they also hoped to be able to solidify their lab's position in the host organisation so that they would not keep having to justify its existence. The interviewees also hoped that the management would participate in the lab's work at a strategic level in order to guarantee sustainability and funding.

When building a lab, the lab team should start small and work towards larger objectives. In the beginning, the team should budget time for bureaucratic challenges and creating a research project policy outlining how many and what kinds of projects to accept. In the early stages, it may be tempting to accept most projects that are possible, but it is important to avoid taking on too many projects: the quality of research projects is more important than their quantity.

9 Summary

Because the lab network is fairly small and labs often only employ a few people, we were able to interview many of the key people in the network. The views offered in the interviews backed up what we had already learned from the labs' public materials and the GLAM lab book. The interviews were nevertheless absolutely necessary: without them, just using the other material available, we would not have been able to form a coherent overall picture of lab work. By contacting people in the lab network, we were also able to tell them about the work we have done in the DOM project at the National Library of Finland. Thanks to our active approach, we were able to connect with new people and stay in touch with old acquaintances. When our lab officially opens, we hope to be able to take a more prominent role. Furthermore, we still wish to see after the active benchmarking the development of the labs in future. We have already seen some details changing during the time we have continued our benchmarking work. Therefore, active interaction between the lab network is something what we are still looking for.

It was surprising to learn how similar the labs were. Although the interviewees brought up different things, we could find many features that make a lab, despite the broad definition of the concept. We will strive to make our future lab a part of this network, becoming a research partner that offers digital collections and data for digital humanities research. In the meantime, we continue to plan and design the Louhos (Mining Lab) in the Digital Open Memory project, funded by the European Regional Development Fund, in collaboration with researchers and the National Library staff. After we have finished collecting data from researchers, we will create a user-oriented service model and draw up a proposal for action.

Data

Source material produced and compiled in the project:

Date	Organisation	Material	Further information
12 March 2020	Staatsbibliothek zu Berlin	Interview, around 1 hour	Interview with one person
30 March 2020	Royal Library of Belgium's Digital Research Lab	Interview, around 1 hour	Interview with two people
24 April 2020	Biblioteca Virtual Miguel de Cervantes Lab	Questionnaire answers and interview, around 30 minutes	Interview with two people
8 May 2020	Bibliothèque nationale du Luxembourg	Replies to preliminary emails and interview, around 45 minutes	Interview with three people
11 May 2020	Kungliga bibliotek Sweden	Interview, around 45 minutes	Interview with one person
15 May 2020	Bibliothèque nationale de France	Interview, around 50 minutes	Interview with two people
11 June 2020	Kongelige bibliotek Denmark	Interview, around 1 hour	Interview with two people

Other material, including seminar and conference videos from the following labs:

ÖNB, BL Lab, KB Lab Netherlands, KB Lab Denmark, NL Scotland

Oiva, Mila (Aalto University): Facilitating Digital Humanities Research 2017 [dataset]. Version 1.0 (2019-11-01). Finnish Social Science Data Archive [distributor]. <http://urn.fi/urn:nbn:fi:fsd:T-FSD3362>

References

Holtzman, Y., Puerta, M., Lazarus, H. and Simon, C. (2011), "An examination of best practices and benchmarking in corporate libraries", *Journal of Management Development*, Vol. 30 No. 1, pp. 134–141. <https://doi.org/10.1108/026217111111098433>, cited on 5 February 2020.

Mahey, M., Al-Abdulla, A., Ames, S., Bray, P., Candela, G., Chambers, S., Derven, C., Dobрева-McPherson, M., Gasser, K., Karner, S., Kokegei, K., Laursen, D., Potter, A., Straube, A., Wagner, S-C. and Wilms, L., with forewords by: Al-Emadi, T. A., Broady-Preston, J., Landry, P. and Papaioannou, G. (2019) *Open a GLAMLab. Digital Cultural Heritage Innovation Labs*, Book Sprint, Doha, Qatar, 23–27 September 2019.

Nyhan, J., Terras, M. & Vanhoutte, E. Introduction. Terras, M., Nyhan, J. & Vanhoutte, E. (edit.) *Defining Digital Humanities*. 1–12. London: Routledge.

Shen, Y. & Varvel, V. E. Developing Data Management Services at the Johns Hopkins University. *The Journal of Academic Librarianship*. 39(6). 552–557. <https://doi.org/10.1016/j.acalib.2013.06.002>, cited on 29 June 2020.

Snickars, Pelle (2018) *Datalabb på KB*.

http://www.kb.se/dokument/Bibliotek/utredn_rapporter/2018/1.2.1-2017-752.pdf

Appendix 1: Interview questions



Leverage from
the EU
2014–2020

Our focus in the library lab is mostly on research services but it is possible to focus on something that you in your lab think would be useful for us.

The following questions are part of the semi-structured interview tradition. You do not have to ask or answer every question in order, and some of them can be left without an answer. The questions serve more as a guideline for the interviewer and interviewee. It is also possible, but not necessary, to follow the order of the questions during the interview.

Organisation

- How would you describe your lab? What is your lab?
- What do you do in the lab? What is your role? What is your educational background?
- What kind of data is used in the lab?
- What is the aim and purpose of the lab?
- Why was it established? For whom?
- What are the daily, weekly, and annual practices of your lab?

Collaboration

- Who are you collaborating with?
- What kinds of skills and knowledge do you have in the lab?
- What are the services that you are currently offering to researchers?
- How well do the researchers know and how easily can they find your services?
- How would you describe the atmosphere among the collaboration partners?

Projects

- What projects do you currently have? What are you working on? Can you describe some of the projects? (topic, status, organisation, funding partners)
- Who are the people involved in the projects, and what are their backgrounds? What are the expected results of the projects?
- What are the minimum infrastructure prerequisites for successful projects?

Funding

- Who funds your lab? Where do the resources come from? (e.g. organisations? foundations?)

- How many people are employed by the lab? How many externally funded people are in the lab? (grant, project funding, etc.)

Communication

- How do you communicate in the lab internally and externally? (meetings, social media, etc.)
- Is there a regular time and place for meetings? (e.g. regular seminars, invited speakers, webinars)
- Have you (or other people involved in the lab) encountered gaps in communication?

Lessons learned

- Have you faced any problems between organisation partners?
 - Any problems when you started?
 - What and how did you overcome those?
 - If not problems, was there anything you would have done differently?

Future plans

- What do you do to improve your current services? What has the feedback been like?
- What kind of service do you think you are not yet offering? What are you supposed to offer?