Natural Allies in National Libraries: Open Data Platforms and GLAM Labs

by

Samuel Joseph Ardell

Submitted to the School of Information Management,

Victoria University of Wellington

in partial fulfilment of the requirements for the degree of

Master of Information Studies

October 2021

Abstract

Research Problem

There is an increasing amount of data generated and held by cultural heritage institutions. As national libraries are making this data available through open data platforms, this research explores how features of labs could be utilised to improve the manner in how this data is presented for users, and whether these two apparently distinct approaches could have more to do with each other than might be initially thought.

Methodology

A qualitative Content Analysis approach was used to study the websites of the institutions under study. The population included twenty-two national libraries across three geographic areas: Europe, North America and Australasia.

Results

While some form of open data is offered by the majority of national libraries, only a handful have developed detailed platforms for users, whereas the majority present open data in technical, undeveloped webpages as bare data dumps or through APIs. Labs exist in a number of institutions, and can be distinguished between those that present an external service, and those that are internal only. The external platforms present unique and creatives ways to present data to users, and have features that are suitable to adopt for open data presentation in national libraries.

Implications

Cultural heritage Institutions should consider not just *what* they are making available, but also *how* they are making data available. Institutions should seek ways to get users to engage, via shaping and presenting information in a relatable way. Lessons can be learned from how labs use and present data in innovate ways, in order to engage staff and users. Applying the features of and learnings from the lab experience to how institutions present open data could be a means to supplement the often dry, technical information presented. Exploring this in a specific institution as a case study would be the next step in this process.

Table of Contents

1. Introduction

1.1 Topic and Background

Digital collections are an ongoing and increasingly important facet of cultural heritage (CH) in national institutions. In New Zealand, both the National Library (n.d.a) and National Archives (n.d.) are pursuing large scale digitisation of their collections. The process of digitisation itself produces more digital material in the form of metadata, through the creation of records, enhanced record descriptions, and techniques such as OCR (optical character recognition). If available, this data can be utilised by researcher and users to reveal new insights both within CH and possibly other areas also.

While this means that access to data about and from collections is becoming increasingly possible, there are challenges with how this data is being made available which means access may remain largely theoretical rather than practical. These challenges exist due to the practical and conceptual limitations CH institutes may unknowingly erect when they make data available. Part of the reason is a lack of contemporary experience, as has been noted: "cultural heritage institutions have rarely built digital collections or designed access with the aim to support computational use" (Wittmann et al., 2019, p.50). This itself stems from the fact institutions have largely worked with and in an environment of analog material, and are still largely operating in that paradigm (Zwaard, 2019, p.157).

The idea of open data represents a potential paradigm shift. The Open Knowledge Foundation (n.d.a) defines *open* as where "anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness". The open data movement argues that such an approach benefits wider society (Kitchin, 2014, p.48).

In the CH sector open data can be viewed in terms of the overall picture of access rights, and where the sector is generally on these points. Such an aim is what initiatives such as the Open GLAM Survey aimed to uncover, particularly in the areas of rights and means of access (McCarthy, 2019a). Against this macro view can be contrasted the micro view of data, which might look at a particular institution in deep focus. By looking at open data

through the lens of national libraries, this study leans towards the micro view. However, given the signifiance of national libraries, how they provide open data has some implications which extend far beyond their existence as single institutions.

Aside from open data platforms, a number of national libraries institutions are exploring the use of labs. Labs are dynamic instruments which "take elements of existing core services, knowledge, skill and engagement practices, such as digitisation, collections, exhibitions and communities, and pivot and reimagine their collective relevance to collaborators and audiences" (Mahey et al., 2019, p.36). As such, they have been seen by their proponents as "one of the most significant and disruptive influences on cultural heritage organisations in the age of digital production and transformation" (Mahey et al., 2019, p.150). Labs are one method institutions have utilised to facilitate engagement, experimentation, and explore new possibilities with digital collections and collections data. As such, this study will look at how these labs provide access and context to data in the national libraries under study.

Both open data and labs challenge the usual use-case of born or digitised material, whereby a digital surrogate is studied in isolation from other material and (often) with the use of minimal metadata. The Collections as Data (CAD) movement likewise challenges the traditional conception of collections and data in CH institutions (Padilla, 2018, p.296). CAD aims "to extend the research value of digital collections beyond traditional use and to give researchers more flexible access to [these] collections by facilitating access to the underlying data" (Wittmann et al., 2019, p.49). The ability for researchers or users to utilise these approaches are dependent on heritage institutions recognising the value of such approaches, and accordingly providing both the materials and the tools to do so. This study defines open data as the material, and suggests that labs could be one of the tools (there are no doubt others), or at least provides a model for the way data collections can be presented for users in a way that encourages further engagement.

1.2 Research Objectives

The aim of this study is to identify the current state of open data in CH institutions, and discuss this in light of the experiences and aims of GLAM Labs. Put another way, this study intends to address the following questions:

- How are national Libraries providing access to open data?
- What are the features of labs within national libraries?
- How might open data platforms benefit from the types of services labs offer researchers?

2. Literature Review

The following discussion will overview literature relating to open CH data, CH open data, GLAM Labs, and will touch on the CAD movement. The first consideration is regarding the scope of what is being made available, which will be done in terms of its relation to CH.

2.1 Big Data and Cultural Heritage

2.1.1 Big Data Overview

As a concept, Big Data (BD) can be described as "the idea that computers can gather trillions of pieces of information about billions of different things and find useful patterns in that information" (Hoy, 2014, p.322). While this fits as a general term for BD, there is both no agreed precise definition (Hoy, 2014, p.321), and no exact set of features that describe all BD (Kitchin & McArdle, 2016, p.8).

Despite these conceptual disagreements, there are other aspects that do fine agreement. Discussions of BD often refer to the 'three Vs' of volume, velocity, and variety (Reinhalter & Wittmann, 2014, p.365). Volume refers to the size of data, usually referred to in petabytes or larger; velocity is the great speed at which data can be created, changed, and destroyed; variety refers to the range of data types or formats that exist (Reinhalter & Wittmann, 2014, p.364-366). While recognising the importance of velocity, Kitchin & McArdle (2016, p.8), argue that BD does not necessarily require volume or variety, and

instead can be better defined by its 'exhaustivity' (Kitchin & McArdle, 2016, p.8): that is, its tendency to include all possible data from a dataset, as opposed to 'small' data which is typically a sample or representation of a dataset (Kitchin & McArdle, 2016, p.7).

Besides salient features, discussion of BD usually also touch on some of its drawbacks. These include the potential loss of individual privacy, which can result from the collection and storage of large amounts of data relating to an individual (Hoy, 2014, p.322; Zhan & Widén, 2018, p.133; Reinhalter & Wittmann, 2014, p.368), and the related issue of data ownership (such as whether the individual has ownership over data about themselves held by third parties) (Hoy, 2014, p.322).

2.1.2 Big Data and Cultural Heritage

The relationship between BD and CH is usually viewed in a framework of positive benefits. For instance, Hoy (2014) argues that the library sector is well-situated to work with BD, particularly in the areas of data preservation and the development of datasets. Other potential roles include assistance in organising data, facilitating access to datasets from outside and within the library, and advising on issues such as copyright (Zhan & Widén, 2018, p.136). Another potential role is helping users directly work with BD (Hoy, 2014, p.324), which arguably falls in line with the sector's traditional role of facilitating access to information (Reinhalter & Wittmann, 2014, p.364).

However, limitations or issues are also noted. For instance, getting to the point where staff can fulfill the above roles requires up-skilling (Zhan & Widén, 2018, p.135; Reinhalter & Wittmann, 2014, p.369). There is also a need to become generally more familiar with the tools of BD (Hoy, 2014, p.324).

Specific types of libraries and BD have been discussed. In the context of research libraries, the development of data librarians (Reinhalter & Wittmann, 2014, p.364), and whole data service departments (Reinhalter & Wittmann, 2014, p.368-369) has been noted. Data librarians provide some of the services mentioned above, serving as "resources for

instruction on data discovery, data management plans, data analysis, and visualization tools such as geographic information systems (GIS)" (Reinhalter & Wittmann, 2014, p.369). BD is much less discussed in relation to certain types of libraries, such as public libraries, though there are a few studies such as Zhan & Widén (2018), who found that while there may be some knowledge of BD among staff, there is often little practical experience (Zhan & Widén, 2018, p.137).

2.2 Data and Collections

2.2.1 Digital Collections

Cultural heritage is seeing continual growth in the size and range of digital material. For one, institutions have seen both increasing digitisation and online digital collections (Wittmann et al., 2019, p.49; Windhager et al., 2019, p.2312). Moreover, with the advent of automated analysis, OCR, transcription, and visualisation into digitisation workflows, more data is being produced by this process than ever before (McGillivray et al., 2020, p.8).

To be useable data must also be of good quality. In response to the challenge of creating and maintaining good data, Wilkinson et al. (2016, p.1) argue that all data must have the features of Findability, Accessibility, Interoperability, and Reusability (FAIR). A key feature of this approach is that the focus is on both human and computer processes (Wilkinson et al., 2016, p.3), and applies to both content and the accompanying metadata (Wilkinson et al., 2016, p.4). Sula (2013, p.18) presents a "continuum of information content associated with cultural heritage institutions", which outlines the range of activities associated with human-computer approaches and primary-secondary (or meta) data (Sula, 2013, p.18-21). While FAIR has been seen as a useful framework (e.g. McGillivray et al., 2020, p.17-18; Candela et al., 2020, p.2), it has also been critiqued as not affording enough attention to the specific requirements of cultural heritage material, such as they can be both records as well as data (Barbuti, 2020, p.174-175). As such, the addition of four additional Rs: Readability, Relevance, Reliability, and Resilience, has been suggested (Barbuti 2020, p.175-176).

2.2.2 Computational Analysis

Computational analysis is the employment of computer technology on CH material. Computational analysis involves same techniques as "text mining, data visualization, mapping, image analysis, audio analysis, and network analysis," and can be used to search for patterns and answer new questions across much larger datasets (Padilla et al., 2019, p.19). The potential scope of material is huge, covering digital collections across virtually all formats (Coleman, 2020, p.14). There is likewise great potential for GLAM institutions to benefit from opening current and legacy material to computational analysis (Candela et al., 2020, p.1-2). Researchers are realising this potential, and though still small in number, are increasingly desirous of access to such data (Padilla, 2018, p.297). However, the scope for computational analysis is currently limited by a number of factors, such as the ongoing focus on textual collections (Padilla, 2017, p.8), as well as the fact that the publishing of analysis-ready data is a low priority for GLAM institutions (Candela et al., 2020, p.2).

2.2.3 Open Data

Fundamentally, the ability to perform analysis on data is determined by the ability to access data. According to the Open Knowledge Foundation's *Open Definition* (n.d.a), in order to be open data must be capable of being "freely used, modified, and shared by anyone for any purpose". Open data can be published through institutional websites or through cross-institutional portals (Roued-Cunliffe, 2020, p.40), although the information available through portals is usually more limited in scope (Roued-Cunliffe, 2020, p.54).

The benefits of open data include providing transparency, creating value, and encouraging participation (Open Knowledge Foundation, n.d.c). Other arguments for open data include making organisations transparent and accountable, making citizens more informed, improving organisation efficiency and governance, improving an organisation's profile and reputation, and contributing to wider commercial value (Kitchin, 2014, p.55-56). For the researcher, open data offers a number of benefits besides

easy access and use, as it offers the potential for new ways to understand data through the use of such approaches as data visualisation or the combination of open data with additional datasets (Roued-Cunliffe, 2020, p.76): the latter being seen as a key strength of open data (Roued-Cunliffe, 2020, p.93).

Given its utility, the availability and use of open data is naturally a central concern. The Open Knowledge Foundation (n.d.b) states that data which is "useful, usable and used" has the potential to become truly valuable, by its transformation into open knowledge. The connection between availability and use is seen by some in the CH sector: "[t]he first requirement for heritage to be used and reused is for it to be available" (Roued-Cunliffe, 2020, p.57).

Critiques of open data include that it "lacks sustainability, utility and usability" (Kitchin, 2014, p.61), and that the focus for open data projects is primarily on making data available, and less on actual use (Kitchin, 2014, p.64).

2.3 GLAM

2.3.1 Open GLAM

Open GLAM is a subset of the open data movement, and promotes the adoption of open data among GLAM institutions (OpenGLAM, n.d.b). Open GLAM is supported by a network of organisations in the area of open data, including the Creative Commons organisation and the Wikimedia Foundation (OpenGLAM, n.d.c). The OpenGLAM movement is currently doing research towards a "Declaration on Open Access for Cultural Heritage" (OpenGLAM, n.d.d). The movement also has a set of principles, one of which affirms the commitment to "[o]penness to collaboration and to novel forms of user engagement are essential if cultural heritage institutions are to realise the full potential of the internet for access, innovation and digital scholarship" (OpenGLAM, n.d.a).

Open GLAM has struggled at times to establish its visibility in the GLAM sector. A 2019 survey of the GLAM community's attitude towards Open GLAM's principles found they

were neither widely known nor seen as practically applicable (Scann, 2019). As a result, the declaration aims to provide a more detailed guide "that provides workable definitions, goals, and standards for making digital cultural heritage available, accessible, and reusable" (OpenGLAM, n.d.d). The need to get an overview of the state of open data in the CH sector was the motivation for launching another survey around the same time.

2.3.2 Open GLAM Survey

The Open GLAM Survey was launched in 2019 as a means of crowd-sourcing an overview of the state of CH open data. Specifically, the researchers behind the survey set out to answer the following questions: "How many cultural heritage institutions make their digital collections available for free reuse? How do they do this, and where is open access most prevalent?" (McCarthy, 2019a). The survey is an ongoing crowd-sourcing effort to capture the global state of open data, with a focus "on digital surrogates of objects in the public domain ... [and] covers objects and data that GLAMs make available on their own websites and on external platforms like Wikimedia Commons, Europeana, the German Digital Library and GitHub" (McCarthy, 2019a).

Overall, the authors found the situation restrictive and confusing. One of the reports the researchers provided from the survey results was on institutions' use of rights statements. They found that many institutions did not clearly articulate whether metadata available on their website was in the public domain, though the same data might carry such a rights statement on aggregate portals, and that a third did not have any reference to rights for metadata (McCarthy, 2019b). They also found that GLAMs tend to maintain restrictive control of their digital collections (McCarthy, 2019c). They also note that "[f]urther research and tools to facilitate quantitative analysis and enable new insights would be valuable" (McCarthy, 2019c).

2.3.3 GLAM Labs

GLAM Labs are an international phenomenon, and like the state of open data before the GLAM survey, can be difficult to locate and consider in practical terms. The existence of

the International GLAM Labs Community (n.d.) makes doing so easier, as they record member institutions on their website. The website is also where they make their guidebook 'Open a GLAM Lab' available for download. The book was written in a few days by Mahey et al. (2019) at a gathering of members. The book discusses a range of topics which are pertinent to the existence and utilisation of labs, including "characteristics, aims and objectives, processes and prospects, tools and services, as well as legal, financial and operational issues" (Mahey et al., 2019, p.8-9). Central to their purpose are some aspects which have been mentioned in this discussion, including the issues and opportunities presented by big data (Mahey et al., 2019, p.10) and the provision of collection data to users (Mahey et al., 2019, p.100).

There are a range of GLAM Labs, which operating in different institutions naturally have somewhat different focuses. These include "National and State-based Library Labs" (Mahey et al., 2019, p.37), "University Labs" (Mahey et al., 2019, p.37), "Museum or Gallery Labs" (Mahey et al., 2019, p.38), "Archival Labs" (Mahey et al., 2019, p.39), and "Individual Labbers" (Mahey et al., 2019, p.40). Mentioned specifically in this last group is Tim Sherratt and his GLAM Workbench. The Workbench is a unique resource which offers "a collection of tools, tutorials, examples, and hacks to help you work with data from galleries, libraries, archives, and museums (the GLAM sector)" (Sherratt, n.d.). While the focus is on Australasia (reflecting the creator's career in CH in this area), the tools offer a unique instance of an extra-institutional approach for utilising CH data. The labs community is also involved in the CAD movement (Mahey et al., 2019, p.100).

2.4 Collections as Data
2.4.1 Definition
Collections as data (CAD) is a relativity recent initiative in the area of computational collections. CAD, as defined by arguably its most central figure Thomas Padilla, is an approach which features a "conceptual orientation to collections that renders them as ordered information, stored digitally, that are inherently amenable to computation" (Padilla et al., 2019, p.7). CAD promotes access to the underlying collections data

(Wittmann et al., 2019, p.49) via the means of datasets which are computationally ready to be used by researchers (Candela et al., 2020, p.1).

While CAD can be seen as part of the general growing awareness of and movement towards computational approaches by both researchers and institutions over the last few years, it is also a specific "movement" (Wittmann et al., 2019, p.49) which grew out of a funded project from 2016 to 2018 known as "Always Already Computational: Collections as Data" (Padilla et al., 2019). The project aimed to "document the implications of collections as data work across cultural heritage organization functions, practices, and roles" (Padilla et al., 2019, p.9). As part of their final report they produced a framework (Padilla et al., 2019), which is a key means of understanding the perspective and aims of the movement. A second phase of the project, titled "Collections as Data - Part to Whole" (n.d.), is currently underway, and aims "to foster the development of broadly viable models that support implementation *and* use of collections as data."

2.4.2 CAD Framework & Principles

The CAD framework itself is comprised of a series of reports which were produced during the project before being combined in the *Final Report* (Padilla et al., 2019). The statement of principles, known as the Santa Barbara Statement on Collections as Data, is "designed to help guide practitioners through the practical, theoretical, and ethical dimensions of collections as data work" (Padilla et al., 2019, p.10), and covers 10 principles, which include: encouraging the use of collections (#1); the importance of ethical considerations (#2); the importance of lowering barriers to use (#3); the need to be guided by user requirements (#4); the importance of providing access to documentation (#5); openness as a key conceptual framework (#6); the need to build in interoperability (#7); the desire for transparency (#8); the need to allow the analysis of both digital object data and metadata (#9); that CAD is a journey and not a destination (#10) (Padilla et al., 2019, p.20-21).

3. Research Design

3.1 Strategy

This study takes an inductive, exploratory, qualitative approach. A qualitative approach "look[s] at characteristics, or *qualities*, that cannot be entirely reduced to numerical values. A qualitative researcher typically aims to examine the many nuances and complexities of a particular phenomenon" (Leedy & Ormrod, 2016, p.24). An inductive approach "begins not with a pre-established truth or assumption but instead with an observation", and which "uses specific instances or occurrences to draw conclusions about entire classes of objects or events" (Leedy & Ormrod, 2016, p.37). An exploratory research design involves a researcher using a qualitative approach "to get a general sense of characteristics, phenomena, and/or issues related to the topic of study" (Leedy & Ormrod, 2016, p.331), before moving into specific aspects of study. Finally, a distinguishing feature of qualitative research is that it can be highly iterative between the phases of data collection and analysis (Leedy & Ormrod, 2016, p.260).

3.2 Methodology

3.2.1 Content Analysis

Content Analysis (CA) is the primary research method for this study. Content analysis can be generally defined as "a detailed and systematic examination of the contents of a particular body of material for the purpose of identifying patterns, themes, or biases" (Leedy & Ormond, 2016, p.275). Traditional CA is often described as systematic (Neuendorf, 2017, p.1; Herring, 2010, p.234), and seen through an objective, quantitative lens of analysis (Neuendorf, 2017, p.1) via a coding schema (Herring, 2010, p.234) using *a priori* parameters (Neuendorf, 2017, p.9). The goal of this traditional approach is "to produce *counts* of key categories and measurements of the *amounts* of other variables" (Neuendorf, 2017, p.21). Although a quantitative approach is perhaps more usual with CA, it is possible to utilise CA in a qualitative manner (Leedy & Ormond, 2016, p.275-276). Doing so works for a number of reasons. For one, the strict adherence to an *a priori* design is often not feasible, due to the necessity of having some level of pre-existing familiarity with the content under study (Leedy & Ormond, 2016, p.320). It could be argued this is

especially true for *exploratory* studies. Moreover, Herring (2010, p.234) notes that CA is not only about identifying patterns, but can also be used "for making inferences about intentions and effects". The result is that there is greater allowance for subjective assessments in CA than might be initially considered.

3.2.2 Web Content Analysis

CA has had a mixed experience with studies of web content. On the one hand, CA has a long history of being utilised in web studies (Herring, 2010, p.234). On the other hand, CA has been used in such a way that traditional approaches "are rarely followed strictly" (Herring, 2010, p.236). Much of this has to do with the inherent difficulties of applying CA to the internet and the volume, velocity, and variety of content (Herring, 2010, p.236-237). The speed of change makes approaches such as random sampling "infeasible in many cases" (Herring, 2010, p.237), a sentiment echoed by Kim & Kuljis (2010, p.373).

3.3 Data Collection

Data has been collected by conducting an analysis of the content of the national libraries' websites. The websites were searched for instances of open data, whether they existed on an identifiable data platform, or spread across several pages on a website, as well as the existence and state of labs. Data was recorded in a qualitative manner in a written form, with both observations and general impressions of structure and content, and eventually complied into a coding table (see 4.2.3 and 4.2.5). A small amount of quantitative data was also collected via the websites of both the GLAM Labs Community and CAD movement.

3.4 Data Analysis

The qualitative data was analysed using a Grounded Theory (GT) approach. Also known as Grounded Theory Method (Bryant & Charmaz, 2007), GT is a qualitative approach (Leedy & Ormond, 2016, p.276), which "comprises a systematic, inductive, and comparative approach for conducting inquiry for the purpose of constructing theory" (Bryant & Charmaz, 2007, p.1). GT allows for a great deal of flexibility in how data is collected and analysed: data is analysed from the outset of the study, with important features being formed and reformed by continual analysis (Leedy & Ormond, 2016, p.274),

in a generally iterative manner (Bryant & Charmaz, 2007, p.1). This is how the open data and labs pages were analysed for this study. For the analysis of quantitative data, Python via Jupyter Notebooks was used to combine data and create graphs, which were used to gain an understanding of certain aspects.

3.5 Population

The population of the study was selected using a purposive sampling approach (Leedy & Ormrod, 2016, p.183), and involved selecting countries from the regions of Australasia, Europe, and North America. These areas were selected due to their close cultural similarity, the concentration in these areas of open data platforms, glam labs, and CAD movement participants, and their use of English as either a primary or close secondary language, thus minimising the potential for cultural or language misunderstandings. From these three regions, a total of twenty-two National Libraries were identified comprising the following countries: Australia, Austria, Belgium, Canada, Denmark, England, Finland, France, Germany, Ireland, Italy, Luxembourg, Netherlands, New Zealand, Norway, Portugal, Scotland, Spain, Sweden, Switzerland, United States, and Wales. The specific counties were selected due to the relativity developed state of open data and labs in these countries. Europe was also limited arbitrarily to the western part for practical reasons relating to the scope of the study.

3.6 Limitations

There are a number of general and specific limitations in the study. General limitations include the study's validity. Leedy & Ormrod (2016, p.278) state that in order to be valid, data in qualitative studies must be "reasonably *accurate* with regard to the characteristics and dynamics of the entities or situation being studied." Accuracy is dependent on having a means measurement that is reliable. The majority of data from this study is qualitative, being based on the subjective assessments of the author, and are therefore arguably neither reliable nor repeatable. Specific limitations include the narrow geographic and cultural range of the study, and the narrowness of the chosen type of institution. Finally, as there could be features which were missed by the author, the analysis and results can in no way be seen as definitive, but rather only indicative.

4. Findings

4.1 GLAM Labs/CAD - Overview

Data collected from the members list on the International GLAM Labs Community website (n.d.) can be used to get a general sense of labs' geographic spread and prominence according to institutional type. The data shows that labs are growing in number. Using the Internet Archive's Wayback Machine, the number of registered partners shows steady growth from 44 in November 2019, to 60 in November 2020, and 69 as of October 2021. The findings also show that the location of labs is largely concentrated in one geographic area, with the majority located in Europe, and in particular the United Kingdom and Spain (9 each out of a total 41), followed by Australia (6) and the United States (5):
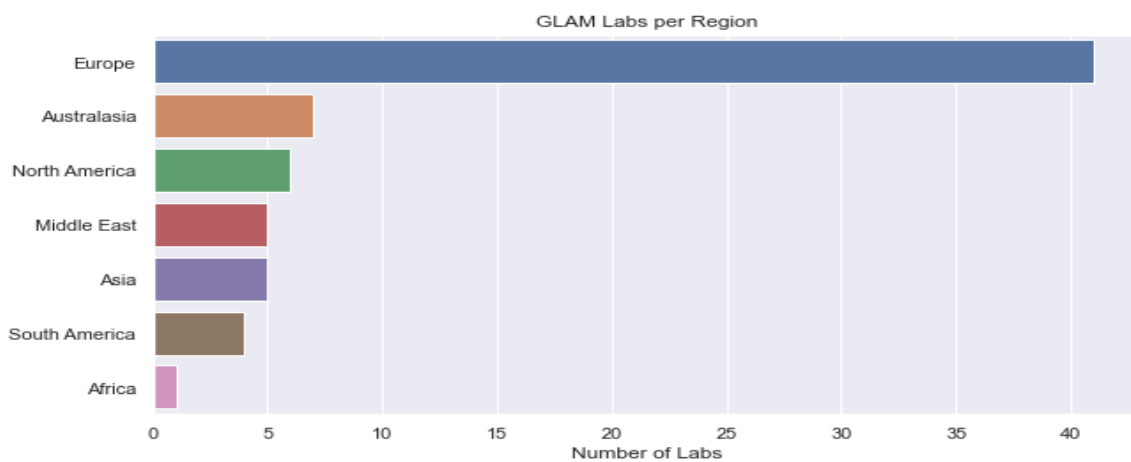
In terms of types of institutions labs are found in, the majority are libraries and universities:
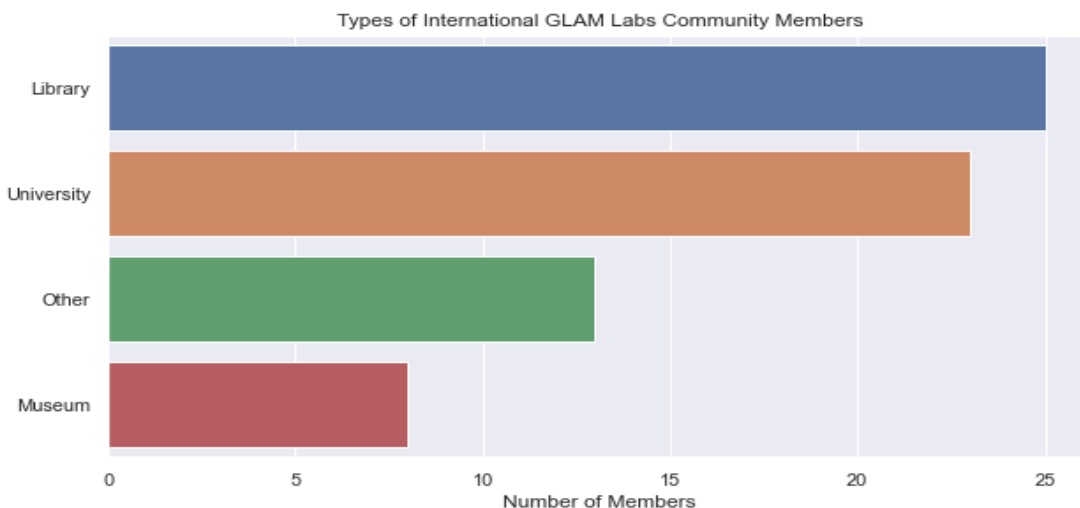
Data collected from the CAD project websites provides some insight into the general state of the movement. The data shows that the initiative is also geographically concentrated, in this case in just two countries:

CAD Initiative Participants by Country

However there is

CAD Initiative Participants by Institute Type
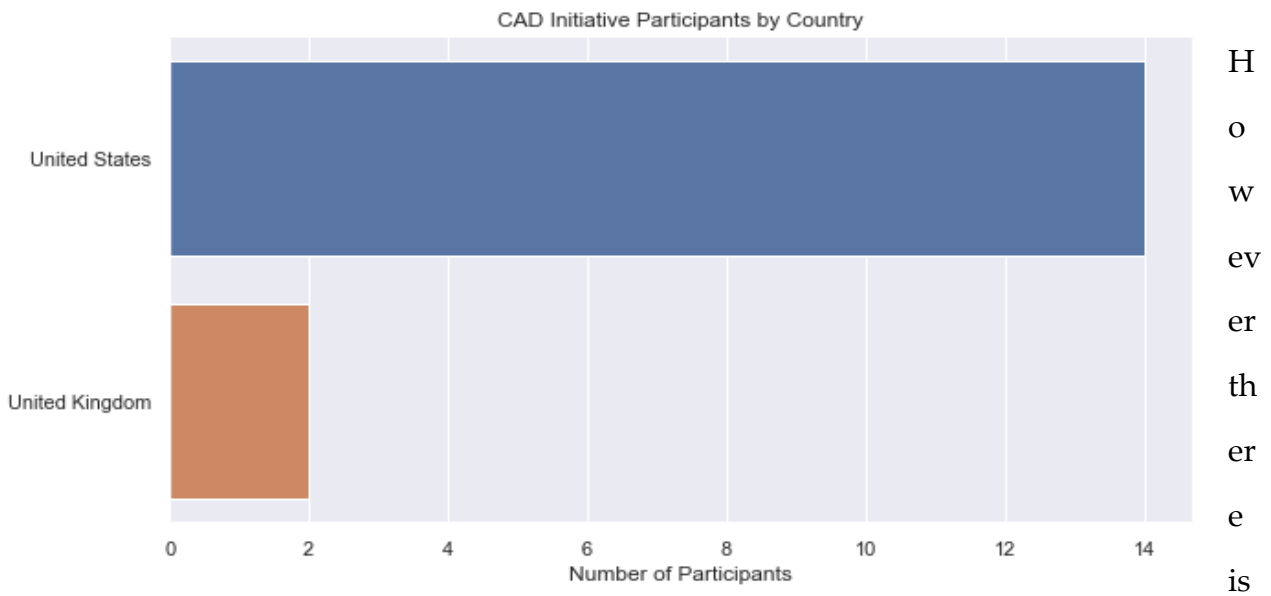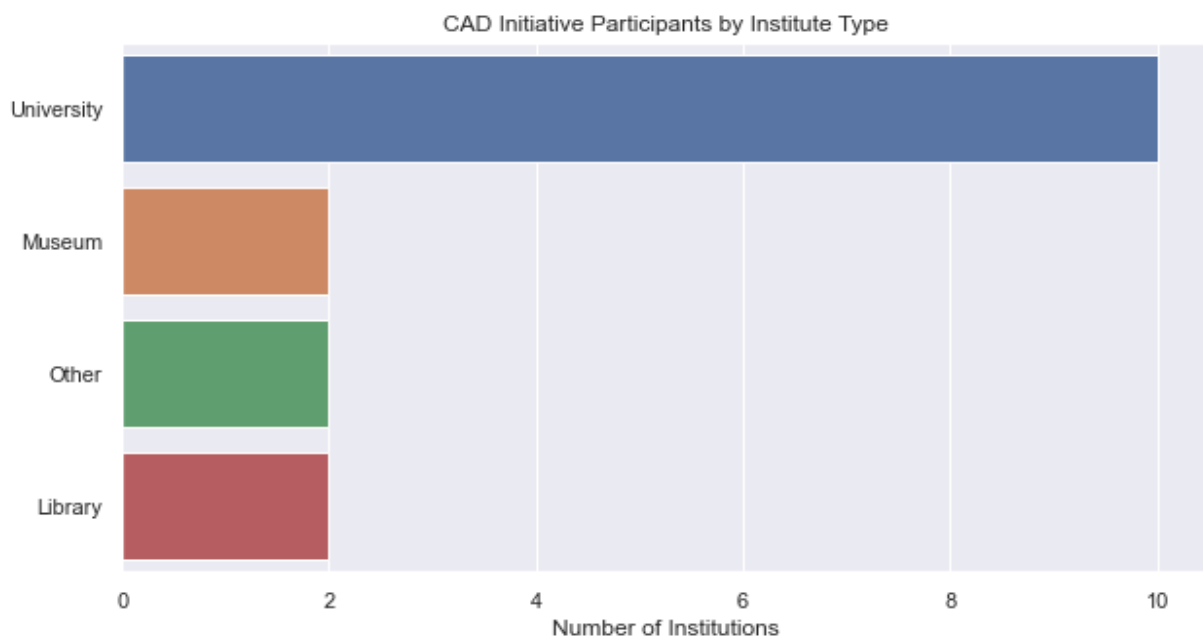
a more even spread among the types of institutions which are members of the initiative:

4.2 NL Open Data and Labs

4.2.1 Overview

The results from the analysis of the NL websites under study indicate that around 17/22 institutions have some kind of open data or linked data page. Of these, 14 are in Europe,

2 are in North American, and 2 are in Australasia. Analysis of the NL websites indicate that a total of 10/22 appear to have some kind of lab service. Of these, 9 are in Europe, 1 is in North America, and 0 is in Australasia. A small number of institutions were found to have no discernible open data or lab features, as in the case of Ireland and Italy.

4.2.2 Coding Table Definitions – Open Data

The NL open data results were compiled into a table (4.2.3), with the following categories emerging during the course of the analysis:

Type (OD/LD/NA). Does the platform/page refer to open data (OD), linked data (LD), or neither (NA).

Platform/Page (PF/PG). Is open data offered via a dedicated platform (PF), or via one or more webpages (PG) as a subsection of another page/section of the website.

Detailed (Y/N). Is the platform/page comprised of multiple pages (Y), or a single page (N) contextual/additional information (Y), or is minimal information provided (N).

Download (Y/N). The platform/page allows for any amount of data downloading.

API (Y/N). The platform/page presents or directs users to an API. This does not mean that the institution does not have an API outside of the platform/page.

Tools/Tutorials (Y/N). The platform/page includes any tools or tutorials for users.

Examples (Y/N). The platform/page presents or links to at least one example of data use or analysis.

4.2.3 Coding Table – Open Data

Country (Institution)	Name	Type	Platform/ Page	Detailed	Download	API	Tools/ Tutorials	Examples	URL	Screenshot
Australia (NL of Australia)	Create something	NA	PG	Y	Y	Y	N	Y	https://trove.nla.gov.au/about/create-something	Appendix 9.1
Austria (NL of Austria)	Linked Open Data	LD	PG	N	Y	N	Y	N	https://labs.onb.ac.at/en/dataset/lod/	Appendix 9.2
Belgium (Royal Library of Belgium)	DATA-KBR-BE	OD	PG	N	N	N	N	Y	https://www.kbr.be/en/projects/data-kbr-be/	Appendix 9.3

Country	Service	Type	PG/PF						URL	Appendix
Canada (Libraries and Archives Canada))	Open Data from LAC	OD	PG	N	N	N	N	N	https://www.bac-lac.gc.ca/eng/discover/open-data/Pages/open-data.aspx	Appendix 9.4
Finland (NL of Finland)	Open Data Service	OD, LD	PF	Y	Y	Y	N	N	https://data.nationallibrary.fi/	Appendix 9.5
France (NL of France)	Bnf Data	LD	PF	Y	N	N	N	N	https://data.bnf.fr/en/	Appendix 9.6
Germany (NL of Germany)	Metadata Services	OD	PG	Y	Y	N	N	N	https://www.dnb.de/EN/Professionell/Metadatendienste/metadatendienste_node.html	Appendix 9.7
	Linked Data Service	LD	PG	N	Y	N	N	N	https://www.dnb.de/EN/Professionell/Metadatendienste/Datenbezug/LDS/lds_node.html	Appendix 9.8
Luxembourg (NL of Luxembourg)	Open Data Platform	OD, LD	PF	Y	Y	Y	Y	Y	https://data.bnl.lu/	Appendix 9.9
Netherlands (NL of the Netherlands)	Data services & APIs	OD, LD	PF	N	Y	Y	N	Y	https://www.kb.nl/en/resources-research-guides/data-services-apis	Appendix 9.10
New Zealand (NL of New Zealand)	Open data	OD	PG	Y	Y	Y	N	Y	https://natlib.govt.nz/about-us/open-data	Appendix 9.11
Portugal (NL of Portugal)	Open Data	OD, LD	PF	Y	Y	Y	N	N	https://opendata.bnportugal.gov.pt/eng_index.htm	Appendix 9.12
Scotland (NL of Scotland)	Data Foundry	OD	PF	Y	Y	N	Y	Y	https://data.nls.uk/	Appendix 9.13
Spain (NL of Spain)	DATOS.BNE.es	LD	PF	Y	N	N	N	N	https://datos.bne.es/inicio.html	Appendix 9.14

Country	Name	Type	Platform/Page						URL	Appendix
Switzerland (NL of Switzerland)	Open Data policy	OD	PG	N	N	N	N	N	https://www.nb.admin.ch/snl/en/home/information-professionals/mediation/opendata.html	Appendix 9.15
United Kingdom (British Library)	Free dataset downloads	OD, LD	PG	Y	Y	N	N	N	https://www.bl.uk/collection-metadata/downloads	Appendix 9.16
	BNB Linked Data Platform	LD	PF	Y	Y	Y	Y	N	https://bnb.data.bl.uk/	Appendix 9.17
United States (Library of Congress)	Selected Datasets	NA	PG	Y	Y	N	N	N	https://www.loc.gov/collections/selected-datasets/about-this-collection/	Appendix 9.18
	ID.LOC.GOV – Linked Data Service	LD	BF	Y	N	Y	N	N	https://www.bibframe.org/	Appendix 9.19
Wales (NL of Wales)	NLW Data	NA	PG	N	Y	N	N	N	https://www.library.wales/collections/activities/research/nlw-data	Appendix 9.20

4.2.4 Coding Table Definitions – Labs

The NL lab results were compiled into a table (4.2.5), with the categories largely carried over from the previous section, albeit with some minor changes:

Type (Ext/Int). Does the lab have an external interface for users (Ext), or are its services provided internally only (Int)

Platform/Page (PF/PG). Is the lab offered via a dedicated platform (PF), or via one or more webpages (PG) as a subsection of another page/section of the website

Detailed (Y/N). Is the lab comprised of multiple pages (Y), or a single page (N) contextual/additional information (Y), or is minimal information provided (N)

Download (Y/N). The lab allows for any amount of data downloading

API (Y/N). The lab presents or directs users to an API. This does not mean that the institution does not have an API outside of the lab

Tools/Tutorials (Y/N). The lab includes any tools or tutorials for users

Examples (Y/N). The lab presents or links to at least one example of data use or analysis.

4.2.5 Coding Table – Labs

Country (Institution)	Name	Type	Platform/ Page	Detailed	Download	API	Tools/ Tutorials	Examples	URL	Screenshot
Austria (NL of Austria)	ONB Labs	Ext	PF	Y	Y	N	Y	Y	https://labs.onb.ac.at/en/	Appendix 9.21
Belgium (Royal Library of Belgium)	Digital Research Lab	Int	PG	N	N	N	N	N	https://www.kbr.be/en/projects/digital-research-lab/	Appendix 9.22
Denmark (Royal Library of Denmark)	KB Labs	Ext	PF	Y	N	N	Y	Y	https://labs.kb.dk/	Appendix 9.23
Germany (NL of Germany)	DNBLAB: Access to Data Sets and Digital Objects	Int	PG	N	Y	N	Y	N	https://www.dnb.de/EN/Professionell/Services/WissenschaftundForschung/DNBLab/dnblab.html	Appendix 9.24
Netherlands (NL of the Netherlands)	KB Lab	Ext	PF	Y	N	N	Y	Y	https://lab.kb.nl/	Appendix 9.25
Norway (NL of Norway)	Digital Humanities Laboratory	Int	PG	N	N	Y	Y	N	https://www.nb.no/en/forskning/digital-humanities-laboratory/	Appendix 9.26
Spain (NL of Spain)	BNElab	Ext	PF	Y	N	N	Y	Y	https://bnelab.bne.es/en/	Appendix 9.27
Sweden (NL of Sweden)	KBLab	Int	PG	N	N	N	N	N	https://www.kb.se/in-english/research-collaboration/kblab.html	Appendix 9.28
United Kingdom (British Library)	British Library Labs	Ext	PG	Y	Y	N	Y	Y	https://data.bl.uk/?_ga=2.132338445.283969606.1636598558-9182090.1636598558	Appendix 9.29
United States (Library of Congress)	LABS	Ext	PF	Y	Y	Y	Y	Y	https://labs.loc.gov/	Appendix 9.30

4.3 NL Open Data

4.3.1 Extent of Platforms

The national libraries' websites are the most common means of providing access to their open data. However, the complexity of their open data offerings varies hugely, from dedicated and highly detailed platforms to single pages with minimal information. Instances of the former include the National Library of Luxembourg's (n.d.) Open Data Platform, and the NL of Scotland's (n.d.b) Data Foundry. Other sites are less developed but have decent offerings. Among these are the NL of Finland's (n.d.) Open Data Service, and the NL of New Zealand's (n.d.b) open data page, and the Library of Congress' (n.d.e) datasets page. Some institutions present open and linked data together, for example the British Library's (n.d.c) "Collection Metadata" page. The NL of France (n.d.) provides a platform for linked data only. A few institutions only offer a single page, such as the NL of the Netherlands (n.d.a), and the NL of Wales (n.d.).

4.3.2 External Platforms

Rather than providing data via their own websites, some institutions direct users to external platforms, which are usually government open data portals. Library and Archives Canada (n.d) is one example. A similar case is the NL of Switzerland (n.d), whose page "Open Data policy" simply states "[a] selection of our data can be downloaded from the portal opendata.swiss." The NL of Portugal (n.d) has an open data portal, which notes that users can use the government data portal or the European Data Portal.

4.3.3 Bulk Downloads

Several institutions make open data available for download directly through their websites. One is the already mentioned British Library's metadata section. The NL of New Zealand's open data page presents a range of datasets comprised of metadata from the catalogues and authority names. The NL of Luxembourg and the NL of Scotland both present metadata as well as collection content available for download, though the collection data only related to a small number of collections. Newspapers appear to be

the most extensive datasets available, as seen in the collections of the NL of Luxembourg, and the recent pilot project run by the NL of New Zealand. Both institutions provide small bulk download packages with which users can get familiar with the larger datasets. In addition to the Library of Congress's datasets page (n.d.a), it also has downloads available through on its newspaper collections, Chronicling America (Library of Congress, n.d.e).

4.3.4 API Access

APIs access is a feature of several sites. Some institutions present APIs with data downloads, or APIs alone. Examples of APIs and downloads being presented together include the NL of Finland, the NL of Luxembourg, and the NL of New Zealand. The Library of Congress refers to APIs on their page "LC for Robots" (n.d.d), though this page is located within the labs section. Although it may seem that APIs are less common than bulk downloads, this is not necessarily the case. While mainly outside the scope of this study, in addition to their own collections many national libraries operate or contribute to a larger national CH portal which may also provide API access. For some institutions, the API is presented as the main way to access data, such as the NL of Australia (n.d.) which only offers a small sample of data for download as a means of previewing data from the Trove API.

4.3.5 Tools for Researchers

A small number of institutions have gone beyond providing basic access by developing or considering more specific tools. Examples include creating tailored datasets for users, creating guides for users, and directing users to tools. The British Library (n.d.b) has created what it calls "Researcher Format (CSV) datasets", which are "themed datasets [that] are available in comma-separated value (CSV) format for you to analyse the data using utilities or applications such as Open Refine." The Library of Congress (n.d.c) provides researchers with a guide specifically for the library's datasets. The NL of Luxembourg (n.d.) offers a number of services. For one, it provides data in different formats to assist users with different skill levels, as well as pointing to tools such as

Jupyter Notebooks, map tools, and external programs (which require coding) for analysis and visualisation. The NL of Scotland's Data Foundry (n.d.a) likewise has a page for tools, and references Jupyter Notebooks and a few other tools.

4.3.6 Developing Data Platforms

Finally, it appears that a small number of institutions are developing platforms. The NL of Belgium (n.d.a) is currently undertaking a data project which it describes as "a collaboration between cultural heritage experts, digital humanities researchers and data scientists ... to facilitate data-level access to KBR's Collections." Similarly, the NL of Wales (n.d.) is pursuing a data project which is an "initiative that provides access to digital collections held by NLW using methods other than the traditional catalogue or library websites." While the exact scope and outcome of these projects are uncertain as of the writing, these developments nonetheless show that the creation of data specific platforms is of interest to institutions.

4.4 NL Labs

4.4.1 Extent of Labs

Like open data offerings, labs vary greatly in both their size and their services. The most developed labs are those which have their own dedicated lab area on the institution's website. Examples include the labs at the NL of Austria (n.d.c), the Royal Library of Denmark (n.d.b), the NL of the Netherlands (n.d.b). While these labs are interactive, some are detailed but do not offer such level of interaction. Examples include the NL of Spain (n.d.c), the Library of Congress (n.d.f), the British Library (n.d.a), and the NL of Germany (n.d.). Some labs have little interactivity and little of a public face. The labs at the NL of Norway (n.d.), the NL of Sweden (n.d.), and Royal Library of Belgium (n.d.b) are such examples.

4.4.2 Services - Data Access

A range of services are provided across NL labs. One of the most common is the offering of collections data in one form or another. The lab at the NL of Austria (n.d.c)

offers a wide range of datasets, including newspapers from 1568-1879, historic postcards, linked open data, web-archive, bibliographic metadata, pamphlets from 1848, music manuscripts and the papyrus collection. The same lab also has 'collections' (National Library of Austria, n.d.b) made up of datasets, and which allow users or staff to create custom data collections that they can either keep private or share. The lab at the British Library (n.d.b) has its own 'Labs datasets', which include metadata and text created through OCR, and were created as the result of collaboration with staff or researchers. The lab at the NL of Germany (n.d.) offers access to "(theme-based) data sets", such as "Free online publications" and "German NL: World War I collection 1914–1918". The lab at the NL of the Netherlands (n.d.b) provides access to a range of datasets which include images, text from newspapers and books, and data from the web archive. The British Library lab (n.d.b) and the NL of Spain lab (n.d.c) state that they are developing more offerings of user-friendly file formats such as CSV. On the other hand, the lab at the NL of Denmark (n.d.) only presents datasets that it has created, and does not allow users to download.

4.4.3 Services - Tools for Researchers

Besides offering data through their websites, some labs also provide tools to analyse data. The labs at the NL of Austria, the NL of the Netherlands, and the Royal Library of Denmark are most evident in this respect. The tools they offer come in many forms. Tools include those that visualise changes in language over time (Royal Library of Denmark, n.d.a). Other tools show how to create interactive graphs with Jupyter Notebooks and search the collections with SPARQL (National Library of Austria, n.d.g). The lab at the NL of the Netherlands (n.d.d) has 29 listed tools, and lets users search and filter tools based on categories such as the data content, type of tool, file format, and copyright. The tools also cover the main forms of analysis, with 13 relating to textual analysis, and 11 relating to data visualization (National Library of the Netherlands, n.d.d).

4.4.4 Services - Assistance for Researchers

Labs also offer assistance to researchers. Some labs do so by offering direct assistance with research projects. The NL of Austria (n.d.a) lab trailed a free service called "Bring your Project" for 1 ½ years until August 2020, which offered data, tools, and specialist staff time. While this particular service does not appear to have continued, the lab does offer researchers assistance with individual projects, for instance with the analysis and visualisation of collections data (National Library of Austria, n.d.d). Often help is provided via an application process. The lab at the NL of Germany (n.d.) has a yearly 'DH-Call', which researchers can apply to if they want assistance with "automated analyses of protected full texts, images or AV media". The NL of Sweden (n.d.) lab lets researchers place an application for assistance. Other ways of helping include providing tutorials or teaching. Using Jupyter Notebooks is a common way of providing tutorials: the NL of Germany (n.d.) uses notebooks to teach "analysing and working with our open-access items" including text and data mining and the NL of the Netherlands (n.d.d) lab provides "walk-throughs for preparing and visualising the data."

4.4.5 Services – Innovation for Institutions

Finally, labs also appear to provide a service back to the institution. One way they do so is by demonstrating innovative uses of the librarys' collections. For example, the NL of Austria lab acts as a showcase for projects created with lab's tools and data, such as their contribution to a current exhibition (National Library of Austria, n.d.f). The NL of Spain lab (n.d.b) likewise hosts digital experiments with collection data, such as their "Interactive Books" project. The Royal Library of Denmark (n.d.b) lab presents applications they developed "to visualize, engage or showcase the different materials or collections that we have available, to inspire and deepen the knowledge of what collections we actually have, and hopefully expand the use of these." The NL of Netherlands (n.d.c) lab explored data from the Dutch digital archive Delpher, in order to explore and visualise the growth of terms relating to pandemics. The NL of Spain lab (n.d.a). created "Community BNE", a crowd-sourcing platform which offers the opportunity for users to get involved and contribute to collections. The United States lab also developed a crowd-sourcing platform "By the People" (n.d.b), which is a "web-based

crowd-sourcing application where anyone with an internet connection can transcribe documents from Library of Congress digitized collections."

5. Discussion

5.1 Introduction

The following discussion will cover some of the salient features of the open data and lab offerings presented by the national libraries under study, consider how labs could be used to assist the delivery service of open data, and conclude with some specific directions libraries could take with their CH open data. Open data may be more difficult to access than is intended, due to the focus on bare bulk downloads or API access, which together serve to favour certain types of user. Part of this difficult may stem, on the institution's side, from an under-appreciation of the difference between government and CH open data. Initiatives live the CAD movement offer some possible considerations for institutions, and indeed seem to be already having an impact. Labs present perhaps the greatest opportunity to utilise open data in new ways, due to their role as places which encourage innovation, their experience working with diverse CH data, including development tools, and helping users. Along with CAD they may also present an opportunity to address deeper issues in CH open data and engage with different communities. Finally, they could be used as a medium for shaping and presenting data users for users in an engaging manner.

5.2 Open Data Environment

5.2.1 Accessing Data

The aims of the open data movement provide a lens with which to consider the current state of national libraries' open data. According to Kitchin (2014, p.48), "[t]he open data movement seeks … both opening up data for wider reuse, but also providing easy-to-use research tools that negate the need for specialist analytic skills." The current state of open data currently fulfills a portion of that description. On the one hand, bespoke open data platforms are being developed, as seen by the platforms of the national libraries of Luxembourg and Scotland. These provide more opportunities for wider reuse, by making

access easier and use clearer. However, the majority of institutions are not anywhere near this point, with most providing bare pages with overly technical information, and few examples or discussion of use. The offerings of downloadable data are wide, but are generally presented in a somewhat confusing manner, sometimes across multiple sections of a websites.

5.2.2 APIs

APIs are a common feature of national libraries' open data offerings. The Open GLAM survey results (McCarthy, 2019a) showed that many institutions provide some sort of API access. This is especially true of national libraries, whether that is through their own website or through a wider portal – indeed, the way they were noted in this study perhaps underrepresented how common they are. In many ways, the prominence of APIs is a positive result. By some measurements, providing data through an API is considered to be the greatest degree of openness possible (Roued-Cunliffe, 2020, p.19). Of course how well depends on the institution. In truth, there are some institutions that do this very well. One example is the NL of Australia: Roued-Cunliffe (2020, p.114) says that this library's API, offered via its portal Trove, "is probably the best-documented and most user-friendly example of a heritage data API that we have to date." Considering that the same institution offers little in the way of direct download capabilities, and that the API serves as the API for 10 other Australian institutions, it makes sense that the institution has focused on this resource. And while the state of APIs in CH is worthy of praise, though it does come at a cost.

5.2.3 APIs – Some Considerations

The cost of an over-reliance on APIs and bare downloads comes in many forms. One issue is that the complexity of these services may turn researchers and users away. APIs are technical and require a degree of user familiarity with code and other concepts. Moreover, it can be unclear what an institution offers, either by API or download. Roued-Cunliffe (2020, p.4-5) points to the Library of Congress, noting that "while these digital collections

and their access points are shared on the web page 'LC for Robots', it is not quite clear what data is available through the different APIs and bulk data downloads."

The result is that libraries are perhaps inadvertently favouring a specific type of user. Many institutions have separate pages for these technical users or 'developers,' for example Europeana and DPLA (Roued-Cunliffe, 2020, p.53), and DigitalNZ and Trove. While anyone who has used an API with Python or another programming language could say using an API is somewhat straightforward, it is debatable whether this would be the case for many researchers or potential users. And while data downloads may offer more practical access, in many cases these downloads are not presented in a much less technical manner. Moreover, even if users could access the data, whether they would know how to manipulate the data enough to produce some utility from the data is another question.

5.2.5 CH and Government Open Data

The issue with how downloadable data is presented on national library sites possibly reflects another issue, which is the tension between the aims of open government and CH institutions. While both aim to make data available, there is a difference in the type of data. For one, CH data is usually less prominent among open data offerings, due to the fact that often has "marginal economic value" (Kitchin, 2014, p.56). This means that there are from the outset less potential users of CH open data. Another issue may be that, whereas open government data is often 'tidy', so can be presented in easy to use spreadsheet or CSV formats, CH data is often inherently unstructured or 'messy' data. As such, the fields on which these two data types are presented are not equal: one requires a lot more work (and with potentially less 'economic' reward). A compudong factor is the wider environment, which is determined by each country's unique history and experience of open data (Kitchin, 2014, p.54). Adopting some of the approaches of CAD could mitigate some of these issues for CH open data.

5.2.6 CAD and CH Open Data

National libraries' open data initiatives could benefit from utilising CAD approaches in a number of ways. The first and primary benefit would be in the way CAD can address issues inherent in the data itself. Zeigler (2020, p.27) points out that "[t]he collecting practices of cultural institutions have long been marred by the racial bias of the archivists and curators who build collections. The decisions made about what is collected are colored by the opinions of those doing the collecting." Through digitisation, such a situation "might result in a non-representative digitised collection" (Mahey et al., 2019, p.103). In effect, putting open data 'out there' without context, while well-intentioned, could in fact create harm. For instance, flatly presenting data may mean an apparent 'natural' predominance in viewpoints of one group over another (Zeigler, 2020, p.30). Approaches such as CAD emphasise the importance of providing data with context (Zeigler, 2020, p.39). Indeed, whether influenced by CAD or not, this is partly being seen in the more developed open data platforms, such as the NL of Luxembourg and Scotland. These institutions, and various other NLs, are members of GLAM Labs, CAD, or both.

5.3 Open Data and Labs

5.3.1 Innovation Spaces

Labs can assist open data initiatives by providing services that fit within the framework of open data. One of the primary aims of the open data movement is "opening up data for wider reuse" (Kitchin, 2014, p.48). However, Roued-Cunliffe (2020, p.26) argues that due to legal concerns institutions are often conservative with how they make open CH data available, and so the sector "needs innovation and experimentation, and this is difficult to sustain in an environment of fear." This is one area where labs could play a role, as experimental practice is one of the key features of labs. This is seen in the websites that offer a variety of tools and services. They are also the place where staff can test new tools and means of how data is provided to users or communities. Moreover, a lab provides a safe space to try out new ideas, as "[l]abs can rapidly prototype and test emerging technology and processes at a much smaller scale and cost" (Mahey et al., 2019, p.42).

5.3.2 Working with Data

Labs can be a partner for open data initiatives due to their experience of working with a range of data. This is useful since, as Roued-Cunliffe (2020, p.40) explains, CH data comes in many different forms depending on the type of collection and institution: "galleries, as the material here is almost exclusively visual and published as images with metadata. Some libraries make the content of books and newspapers available as digitised images and full text. Archives often have lists of people who can be searched for – for example in census records, passenger lists, etc. …. Museums often have collections of objects that can be published as images in a similar way to the art published by galleries. Some museums have material that is geographically located and can be published on a map." Labs typically work with across teams and with a great variety of collection types: maps, photos, manuscripts, objects, music, books, programs, and more (Mahey et al., 2019, p.34). Because of their experience with data, and their ability to experiment with data, labs can provide the venue where different uses and possibilities with data can be explored.

5.3.3 Developing Tools

Another way that labs can assist open data initiatives is by helping to provide the means by which data can be understood. Referring back to Kitchin (2014, p.48), one of the aims of the open data movement is "providing easy-to-use research tools". However, the same author (Kitchin, 2014, p.64) found that this usually not the case for open data sites, as "[m]any sites also lack appropriate tools and contextual materials to support data analysis." Most libraries did not provide any tools, and while the national libraries of Scotland and Luxembourg did refer to tools, these were more general rather than specific tools for the data provided. Labs could help fill this gap. One of the primary purposes of labs is to "facilitate [the] adoption of innovative and modern tools and methods for content delivery and user engagement" (Mahey et al., 2019, p.42). Comparing the tools offered on the open data sites with the tools that labs provided, the labs tools are much more institutional and collection specific. This means that labs can develop specific tools to ease users into exploring and using the open data.

5.3.4 Helping Users

A greater role for labs in the provision of tools in order to assist users would fit with another part of Kitchin's (2014, p.48) definition of the aims of open data, that being to provide tools which "negate the need for specialist analytic skills." Arguably, the current state of open data in NLs does not negate this need, either via the use of bare data downloads or technical APIs. This issue was identified in the wider open data environment in 2014, when Kitchin (2014, p.64) found that many sites "operate more as data holdings or data dumps, lacking the qualities expected in a well-organised and run data infrastructure, such as clean, high-quality, validated and interoperable data that comply with data standards and have appropriate metadata and full record sets (associated documentation)." Labs can provide a way of avoiding this situation. As seen by the labs in this study, documentation and context is a part of their services. Moreover, a central aspect of labs is looking at collections from the user perspective (Mahey et al., 2019, p.42). By adding contextualising information through documentation and examples, labs could supplement open data by offering a more user-focused perspective.

5.3.5 Engaging Communities

Labs can also assist open data initiatives by providing the means for the creation and improvement of existing data. Labs can do this by testing approaches which facilitate user engagement with collections. Examples of labs doing this are seen in the form of tools developed which gather user feedback. One demonstrated example of this has been their role in crowdsourcing projects. These projects "frequently exist in parallel to Labs, offering the possibility to collaborate and reintegrate the user-generated data back to the organisation through the Lab. Various forms of user generated data exist and crowdsourcing initiatives are not the only source" (Mahey et al., 2019, p.110). The crowdsourcing platforms from the labs at the NL of Spain and the Library of Congress, which came from a lab project, illustrate the success which labs can have in providing the means for users to contribute back to an institution. This in turn can expand the quality of the open data provided by the institution.

5.3.6 Addressing Imbalances

Labs make open data more engaging for specific communities by providing a means through which cultural and other imbalances in collections can be addressed. Imbalances can manifest not only in physical holdings, but also "[w]here Indigenous cultures have been colonised, the result of digitisation has been the continued colonisation of their cultural heritage by these organisations" (Mahey et al., 2019, p.43-44). On the contrary to this, Mahey et al. (2019, p.43-44) argue that "[l]abs work in the space of decolonisation and …. can redefine how institutions work with communities represented in and by their collections, to seek diverse collaborations to reimagine how their stories are told, how cultural heritage ownership is conceived and to create new pathways for mutual understanding." With open data, a lab can both be a vehicle for identifying issues within data, while also presenting data specific to a group or community in a manner that is meaningful for that community.

5.3.7 Curating Data

Having a role in shaping data raises the question of how far data could be curated for users to facilitate use and understanding. Adopting a more active role in how data is curated makes some sense. For one, this is already seen in the labs which have created 'researcher' datasets as CSV files. Secondly, it would counteract the assumptions which are sometimes made about open data itself, as Kitchin (2014, p.62) notes: "[i]mplicit in most discussions on open data is the notion that the data are neutral and objective in nature and that everyone has the potential to access and use such data … However, these are not the case." Thirdly, 'messy' data is an obstacle to use, and indeed is a substantial issue in open heritage data (Roued-Cunliffe, 2020, p.117). On the other hand, curating a dataset "requires quite and effort and is not always possible", but "does provide users with a clean and easy-to-use collection" (Mahey et al., 2019, p.106). Despite the effort involved, other institutions are utilising curation and outreach more than previously, such as Europeana (Roued-Cunliffe, 2020, p.70). Even if it is only overseeing the creation of 'tidy' data for the benefit of users, data curation could be one of the services that a lab could provide alongside or within an open data platform.

5.3.8 Storytelling

Further along the curation path, labs could be used to supplement open data offerings by providing an active framework of understanding for users. One way this could be done is by adopting aspects of storytelling with data. While this might appear strange, it should be pointed out that storytelling is not in fact that unusual in CH. Roued-Cunliffe (2020, p.60) notes that "[o]ne of the most important and common ways in which heritage is used is through story-telling and outreach. This is often within the confines of more or less official structures like museums, published books, TV or film." Some labs are creating individual pieces of a story already, either by producing specific datasets, tools for specific tasks, or by presenting code or other contextual information showing how a dataset was created. Presenting these various aspects together in a cohesion narrative could demonstrate to users how open data was created, shaped, and could be utilised. And while there is no doubt a danger in being overly-prescriptive in these efforts, there are arguably equal dangers for CH sector in not doing so. The danger is being under utilised. Roued-Cunliffe (2020, p.70) notes that Europeana, one of the largest and best resourced platforms in CH, is "surprisingly under-used in the larger context of things. In my experience, not many people outside of heritage professionals know about Europeana, and even among heritage professionals only a few use it."

5.3.9 Storytelling with Visualisations

One of the ways which labs can assist data storytelling is through the use of data visualisations. Visualisations can assist both general users and researchers. For researchers, visualisations "throw a new lens on collections to help generate new research questions" (Mahey et al., 2019, p.43). Beyond this group of users, visualisations "make heritage datasets more widely available for a wider audience" (Roued-Cunliffe, 2020, p.72). As seen, some labs already employ visualisations, either through the tools they offer, or through providing such services to researchers. Some labs use visualisations quite frequently: data visualisations are the most common type of tool at the NL of Denmark's lab. As labs, and the staff who operate in labs, gain experience with data analysis and visualisations, these skills could readily transfer to the open data made available through a

NL's platform. The result would be a range of data visualisations presented alongside open data, which users could use as a means of exploring, extracting, or understanding the data that is offered.

6. Conclusion

The open data space is complex and changing. In the cultural heritage sector, national libraries are assuming a role within this data environment, whether driven by the necessity to contribute to government open data initiatives, or by their own desire to make their data more widely available.

The great strength of open data is its scope and variety. In national libraries, open data comes in a variety of forms, whether through the institutions itself, through aggregate data portals, or through government portals. This open data can be accessed in different ways, from file downloads either concentrated on a single page or spread across numerous parts of a website, to an API via an institution or portal, which can access vast quantities of data. In all, the efforts so far indicate the focus has been on infrastructure, on the *how* of making data available.

However, there seems to have been less consideration of a key corollary of how, which is *why*. How and why should not be considered in isolation, but rather should be considered together. By focusing on how of open data, and not on the how and why together, there is a risk of building a service that is under used. It may be helpful to consider the history of endeavors where the effort goes into one aspect alone, as Roued-Cunliffe (2020, p.58) warns: "[l]essons learnt in the 21st century show that we cannot afford an attitude of 'if you build it they will come'."

Such as attitude has perhaps manifested itself in the current state of open data in national libraries. Specific issues include: many National Library data platforms lack easy to manipulate datasets, but are instead comprised of a variety technical file formats; data

downloads are spread across a plethora of linked and un-linked data platforms, and are difficult to locate and get a sense of what is and is not available; while providing a single access point, APIs require knowledge of programming or knowledge of how to manipulate and clean data; documentation for APIs and downloads vary, as done the number of examples for users, from decent to bare bones. The result may be that many researchers are put off from even trying.

There are emerging exceptions, as seen in the open data platforms that have recent appeared in some libraries. The National Library of Scotland's Data Foundry, and the National Library of Luxembourg's Open Data platform are two examples of an approach which looks more closely about both how and why. Coincidently or not, one is a member of CAD, and the other is a GLAM Labs member. However, while both do present a more friendly and clear interface, they are fundamentally portals of data access, rather than understanding. Having the platform provides the crucial foundations, but the building on the top of these foundations can be assisted by labs, or by adopting the features that labs have shown themselves to be strong in: engagement, creativity, and understanding.

National Libraries are the ideal institutions to host such combined initiatives. Practically, these institutions have access to the raw data, either through their own large collections, or centre points for aggregate portals, and they have the size and infrastructure needed, both financially and in terms of personnel. Labs can either build the tools for researchers, or they can at least present models which researchers can use to build out their own tools. If the tools and models are retained, they can be used for future projects and for future users.

The need to change focus has been noted in the past: "open data need to be rethought as a service process …. considering the needs and expectation of end-users" (Kitchin, 2014, p.52). This need is especially true for cultural heritage institutions. Open data should not be considered as a second- rate resource, made available by institutions in a technical sense, but in reality in a way which requires the users to do most of the work. Rather, open data should be a service, part of the many services that cultural heritage institutions, and

national libraries in particular, provide. It is a service which labs and open data platforms can provide together, and in doing so represent the best parts of cultural heritage: a mix of technology and people, working with the remnants of the past for the benefit of the present, and always with an eye towards the future.

7. References

Ames, S., & Lewis, S. (2020). Disrupting the library: Digital scholarship and Big Data at the National Library of Scotland. *Big Data & Society*, 7(2), 1-7. https://doi.org/10.1177/2053951720970576

Archives NZ. (n.d.). *Archives 2057 strategy*. Retrieved May 3, 2021, from https://archives.govt.nz/publications/archives-2057-strategy

Barbuti, N. (2020). Creating Digital Cultural Heritage with Open Data: From FAIR to FAIR5 Principles. In M. Ceci, S. Ferilli, & A. Poggi (Eds.), *Digital Libraries: The Era of Big Data and Data Science* (p. 173–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-39905-4_17

British Library. (n.d.a). *British Library Labs.* Retrieved September 25, 2021 from https://www.bl.uk/projects/british-library-labs

British Library. (n.d.b). *British Library Labs Datasets.* Retrieved September 25, 2021 from https://data.bl.uk/bl_labs_datasets/

British Library. (n.d.c). *Collection Metadata.* Retrieved September 25, 2021 from https://www.bl.uk/collection-metadata/

British Library. (n.d.d). *Downloads.* Retrieved September 25, 2021 from https://www.bl.uk/collection-metadata/downloads

Bryant, A., Casper, M., Charmaz, K., ... & Wittgenstein, L. (2007). Introduction: grounded theory research: methods and practices. In *The SAGE handbook of grounded theory* (pp. 1-28). SAGE Publications Ltd, https://www.doi.org/10.4135/9781848607941

Candela, G., Sáez, M. D., Escobar Esteban, Mp., & Marco-Such, M. (2020). Reusing digital collections from GLAM institutions. *Journal of Information Science*, 1-17. https://doi.org/10.1177/0165551520950246

Coleman, C. N. (2020). Managing Bias When Library Collections Become Data. *International Journal of Librarianship*, *5*(1), 8–19. https://doi.org/10.23974/ijol.2020.vol5.1.162

Collections as Data - Part to Whole. (n.d.). Retrieved September 30, 2021, from https://collectionsasdata.github.io/part2whole/

Cordell, S. A., & Gomis, M. (2017). So what are you going to do with that?: The promises and pitfalls of massive data sets. *College & Undergraduate Libraries*, *24*(2–4), 350–363. https://doi.org/10.1080/10691316.2017.1338979

Dalbello, M. (2011). A genealogy of digital humanities. *Journal of Documentation*, *67*(3), p.480–506. http://dx.doi.org.helicon.vuw.ac.nz/10.1108/00220411111124550

De Sa Pereira, M. (2019). Mixed Methodological Digital Humanities. In Gold M. & Klein L. (Eds.), *Debates in the Digital Humanities 2019* (pp. 405-412). Minneapolis; London: University of Minnesota Press. https://doi:10.5749/j.ctvg251hk.37

Herring, S. C. (2010). Web Content Analysis: Expanding the Paradigm. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp.233–249). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_14

Hoy, M. B. (2014). Big Data: An Introduction for Librarians. *Medical Reference Services Quarterly*, p.33(3), 320–326. https://doi.org/10.1080/02763869.2014.925709

Hubbard, M. (2015). *What are Digital Scholarship and Digital Humanities?* LMU Library News. https://librarynews.lmu.edu/2015/09/what-are-digital-scholarship-and-digital-humanities/

Hughes, L., Constantopoulos, P., Dallas, C. (2016). Digital Methods in the Humanities: Understanding and Describing their Use across the Disciplines in Schreibman, S., Siemens, R., & Unsworth, J. (Eds.), *A new companion to digital humanities* (pp.150-170). John Wiley & Sons, Ltd.

International GLAM Labs Community. (n.d.). *List Members*. Retrieved September 25, 2021
from https://glamlabs.io/list-members/

Kim, I., & Kuljis, J. (2010). Applying Content Analysis to Web-based Content. *Journal of
Computing and Information Technology*, 18(4), 369-375.
https://doi.org/10.2498/cit.1001924

Kirschenbaum, M. (2012). What is Digital Humanities and What's it doing in English
Departments? In M. K. Gold (Ed.), *Debates in the Digital Humanities* (NED-New
edition, pp.3–11). University of Minnesota Press.
http://www.jstor.org/stable/10.5749/j.ctttv8hq.4

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their
Consequences*. SAGE Publications Ltd. https://www-doi-
org.helicon.vuw.ac.nz/10.4135/9781473909472

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the
ontological characteristics of 26 datasets. *Big Data & Society*, 3(1), 1-10.
https://doi.org/10.1177/2053951716631130

Kovarova-Simecek, M., Wagner, S. C., & Karner, S. (2019). *ONB Labs – An Open Digital Hub
of Cultural Heritage.* Retrieved September 28, 2021, from
https://www.semanticscholar.org/paper/ONB-Labs-An-Open-Digital-Hub-of-
Cultural-Heritage-Kovarova-Simecek-
Wagner/2db97516533251bedda34943fa6a35a0a163e85c

Library and Archives Canada. (n.d.). *Open Data*. Retrieved September 25, 2021 from
https://www.bac-lac.gc.ca/eng/discover/open-data/Pages/open-data.aspx

Library of Congress. (n.d.a). *Chronicling America.* Retrieved September 25, 2021 from
https://chroniclingamerica.loc.gov/

Library of Congress. (n.d.b). *By the People: LC Crowdsourcing.* Retrieved September 25, 2021
from https://labs.loc.gov/work/experiments/crowd/

Library of Congress. (n.d.c). *Datasets.* Retrieved September 25, 2021 from

    https://guides.loc.gov/datasets

Library of Congress. (n.d.d). *LC for Robots.* Retrieved September 25, 2021 from

    https://labs.loc.gov/lc-for-robots/

Library of Congress. (n.d.e). *Selected Datasets.* Retrieved September 25, 2021 from

    https://www.loc.gov/collections/selected-datasets/

Library of Congress. (n.d.f). Retrieved September 25, 2021 from https://labs.loc.gov/

Leedy, P.D. & Ormrod, J.E. (2016). *Practical research: planning and design* (11th ed).

    Edinburgh Gate, England: Pearson Education.

Mahey et al. (2019). *Open a Glam Lab*. International GLAM Labs Community. Retrieved

    September 26, 2021, from https://glam-

    labs.s3.amazonaws.com/media/dd/documents/Open_a_GLAM_Lab-10-

    screen.9c4c9c7.pdf

McCarthy, D. (2019a). Uncovering the global picture of Open GLAM. *Open GLAM*.

    https://medium.com/open-glam/uncovering-the-global-picture-of-open-glam-

    af364aadeeee

McCarthy, D. (2019b). Licensing policy and practice in Open GLAM. *Open GLAM*.

    https://medium.com/open-glam/licensing-policy-and-practice-in-open-glam-

    49c867b49de8

McCarthy, D. (2019c). Open Access Scope in Open GLAM. *Open GLAM*.

    https://medium.com/open-glam/open-access-scope-in-open-glam-70461bec2bca

McGillivray, B., Alex, B., Ames, S., Armstrong, G., Beavan, D., & Ciula, A., … Willcox, P.

    (2020). *The challenges and prospects of the intersection of humanities and data science: A*

    white paper from The Alan Turing Institute. 1-30. https://doi.org/10.17863/CAM.56765

National Library of Australia. (n.d.). *Bulk Download*. Retrieved September 25, 2021 from

    https://trove.nla.gov.au/about/create-something/bulk-download

National Library of Austria. (n.d.a). *Bring Your Project*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/topic/bring-your-project/

National Library of Austria. (n.d.b). *Collections*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/datasets/

National Library of Austria. (n.d.c). *Datasets*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/datasets/

National Library of Austria. (n.d.d). *Drama 1600-1815*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/topic/drama-1600-1815/

National Library of Austria. (n.d.e). *Historic Newspapers*. Retrieved September 25, 2021
from https://labs.onb.ac.at/en/topic/historic-newspapers/

National Library of Austria. (n.d.f). *Pasetti Digital*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/topic/pasetti-digital/

National Library of Austria. (n.d.g). *Tools*. Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/tools/

National Library of Austria. (n.d.h). Retrieved September 25, 2021 from
https://labs.onb.ac.at/en/

National Library of Finland. (n.d.). Retrieved September 25, 2021 from
https://data.nationallibrary.fi/

National Library of France. (n.d.). Retrieved September 25, 2021 from
https://data.bnf.fr/en/

National Library of Germany. (n.d.). Retrieved September 25, 2021 from
https://www.dnb.de/EN/Professionell/Services/WissenschaftundForschung/DNBLa
b/dnblab.html

National Library of Luxembourg. (n.d.). Retrieved September 25, 2021 from
https://data.bnl.lu/

National Library of New Zealand. (n.d.a). *Digitisation Plan 2019–2022*. Retrieved May 3, 2021, from https://natlib.govt.nz/about-us/strategy-and-policy/digitisation-plans/digitisation-plan-2019-2022

National Library of New Zealand. (n.d.b). *Open Data* Retrieved September 25, 2021 from https://natlib.govt.nz/about-us/open-data

National Library of the Netherlands. (n.d.a). *Data Services & APIs*. Retrieved September 25, 2021 from https://www.kb.nl/en/resources-research-guides/data-services-apis

National Library of the Netherlands. (n.d.b). *Datasets*. Retrieved September 25, 2021 from https://lab.kb.nl/datasets

National Library of the Netherlands. (n.d.c). *Delpher Demo Pandenucs Through Time*. Retrieved September 25, 2021 from https://lab.kb.nl/tool/delpher-demo-pandemics-through-time

National Library of the Netherlands. (n.d.d). *Tools*. Retrieved September 25, 2021 from https://lab.kb.nl/tools

National Library of Norway. (n.d.). *Digital Humanities Laboratory*. Retrieved September 25, 2021 from www.nb.no/en/forskning/digital-humanities-laboratory/

National Library of the Netherlands. (n.d.b). Retrieved September 25, 2021 from https://lab.kb.nl/

National Library of Portugal. (n.d.). *NLW Data*. Retrieved September 25, 2021 from https://opendata.bnportugal.gov.pt/eng_index.htm

National Library of Scotland. (n.d.a). *Tools*. Retrieved September 25, 2021 from https://data.nls.uk/tools/

National Library of Scotland. (n.d.b). Retrieved September 25, 2021 from https://data.nls.uk/

National Library of Spain. (n.d.a). *Comunidad*. Retrieved September 25, 2021 from https://comunidad.bne.es/

National Library of Spain. (n.d.b). *Interactive Books*. Retrieved September 25, 2021 from https://bnelab.bne.es/en/project/interactive-books/

National Library of Spain. (n.d.c). Retrieved September 25, 2021 from https://bnelab.bne.es/en/

National Library of Sweden. (n.d.). *KBLab*. Retrieved September 25, 2021 from https://www.kb.se/in-english/research-collaboration/kblab.html

National Library of Switzerland. (n.d.). *Open Data*. Retrieved September 25, 2021 from https://www.nb.admin.ch/snl/en/home/information-professionals/mediation/opendata.html

National Library of Wales. (n.d.). *NLW Data*. Retrieved September 25, 2021 from https://www.library.wales/collections/activities/research/nlw-data

Neuendorf, K. (2017). Defining content analysis. In *The content analysis guidebook* (pp.1-35). SAGE Publications Ltd. https://www-doi-org.helicon.vuw.ac.nz/10.4135/9781071802878

Open Knowledge Foundation (n.d.a). O*pen Definition.* Retrieved October 1, 2021, from https://opendefinition.org/

Open Knowledge Foundation (n.d.b). *What is open?* Retrieved October 1, 2021, from https://okfn.org/opendata/

Open Knowledge Foundation (n.d.c). *Why open data?* Retrieved October 1, 2021, from
https://okfn.org/opendata/why-open-data/

OpenGLAM (n.d.a). *OpenGLAM Principles*. Retrieved September 28, 2021, from
https://openglam.org/principles/

OpenGLAM (n.d.b). *What*. Retrieved September 28, 2021, from https://openglam.org/what/

OpenGLAM (n.d.c). *Why*. Retrieved September 28, 2021, from https://openglam.org/why/

OpenGLAM (n.d.d). *Towards a Declaration on Open Access for Cultural Heritage*. Retrieved
September 28, 2021, from https://openglam.pubpub.org/

Padilla, T. (2018). Collections as data: Implications for enclosure. *College & Research
Libraries News*, *79*(6), 296–300. https://doi.org/10.5860/crln.79.6.296

Padilla, T. (2017). *On a Collections as Data Imperative*. 1-11
https://escholarship.org/uc/item/9881c8sv

Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. (2019). *Final Report
- Always Already Computational: Collections as Data*.
https://doi.org/10.5281/zenodo.3152935

Pauwels, L. (2012). A Multimodal Framework for Analyzing Websites as Cultural
Expressions. *Journal of Computer-Mediated Communication*, 17(3), 247–265.
https://doi.org/10.1111/j.1083-6101.2012.01572.x

Poole, A. H., & Garwood, D. A. (2018). "Natural allies": Librarians, archivists, and big data
in international digital humanities project work. *Journal of Documentation*, *74*(4), 804–
826. https://doi.org/10.1108/JD-10-2017-0137

Reinhalter, L., & Wittmann, R. J. (2014). The Library: Big Data's Boomtown. *Serials
Librarian*, *67*(4), 363–372. https://doi.org/10.1080/0361526X.2014.915605

Roued-Cunliffe, H. (2020). *Open heritage data: an introduction to research, publishing and
programming with open data in the heritage sector*. London, UK: Facet Publishing.

Royal Library of Belgium. (n.d.a). *Data-KBR-BE*. Retrieved September 25, 2021 from
    https://www.kbr.be/en/projects/data-kbr-be/

Royal Library of Belgium. (n.d.b). *Digital Research Lab*. Retrieved September 25, 2021 from
    https://www.kbr.be/en/projects/digital-research-lab/

Royal Library of Denmark. (n.d.a). *Smurf*. Retrieved September 25, 2021 from
    http://labs.statsbiblioteket.dk/smurf/

Royal Library of Denmark. (n.d.b). Retrieved September 25, 2021 from https://labs.kb.dk/

Scann (2019). OpenGLAM Principles: Ways forward to Open Access for cultural heritage.
    https://medium.com/open-glam/openglam-principles-ways-forward-to-open-
    access-for-cultural-heritage-81ebb3559b4

Sherratt, T. (n.d.). *GLAM Workbench*. Retrieved September 28, 2021, from https://glam-
    workbench.net/

Schnapp, J., *A Short Guide to the Digital_Humanities* (2012), in Burdick, Anne, et al.
    Digital_Humanities (pp.121-136). MIT Press. *ProQuest Ebook Central*,
    https://ebookcentral-proquest-
    com.helicon.vuw.ac.nz/lib/vuw/detail.action?docID=3339545

Sula, C. A. (2013). Digital Humanities and Libraries: A Conceptual Model. *Journal of
    Library Administration*, *53*(1), 10–26. https://doi.org/10.1080/01930826.2013.756680

Vinopal, J., & McCormick, M. (2013). Supporting Digital Scholarship in Research Libraries:
    Scalability and Sustainability. *Journal of Library Administration*, *53*(1), 27–42.
    https://doi.org/10.1080/01930826.2013.756689

Wilkinson, M. D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., & Baak, A. …
    Mons, B. (2016). The FAIR Guiding Principles for scientific data management and
    stewardship. *Scientific Data*, 3(160018). 1-9 https://doi.org/10.1038/sdata.2016.18

Windhager, F., Federico, P., Schreder, G., Glinka, K., Dork, M., Miksch, S., & Mayr, E.
    (2019). Visualization of Cultural Heritage Collection Data: State of the Art and

Future Challenges. *IEEE Transactions on Visualization and Computer Graphics*, *25*(6), 2311–2330. Scopus. https://doi.org/10.1109/TVCG.2018.2830759

Wittmann, R., Neatrour, A., Cummings, R., & Myntti, J. (2019). From Digital Library to Open Datasets: Embracing a "Collections as Data" Framework. *Information Technology and Libraries*, *38*(4), 49–62. https://doi.org/10.6017/ITAL.V38I4.11101

Zhan, M., & Widén, G. (2018). Public libraries: Roles in Big Data. *The Electronic Library*, *36*(1), 133–145. https://doi.org/10.1108/EL-06-2016-0134

Ziegler, S. L. (2020). Open Data in Cultural Heritage Institutions: Can We Be Better Than Data Brokers? *Digital Humanities Quarterly*, 14(2). http://www.digitalhumanities.org/dhq/vol/14/2/000462/000462.html#p11

Zwaard, K. (2019). Building Institutional and National Capacity for Collections as Data. In Padilla, T., Allen, L., Frost, H., Potvin, S., Russey Roke, E., & Varner, S. *Final Report - Always Already Computational: Collections as Data* (pp.156-158). https://doi.org/10.5281/zenodo.3152935

8. Bibliography

Bingham, N. J., & Byrne, H. (2021). Archival strategies for contemporary collecting in a world of big data: Challenges and opportunities with curating the UK web archive. *Big Data & Society*, (8)1, 1-6. https://doi.org/10.1177/2053951721990409

Bonacchi, C., & Krzyzanska, M. (2019). Digital heritage research re-theorised: Ontologies and epistemologies in a world of big data. *International Journal of Heritage Studies*, *25*(12), 1235–1247. https://doi.org/10.1080/13527258.2019.1578989

Brennan, C. (2018). Digital humanities, digital methods, digital history, and digital outputs: History writing and the digital revolution. *History Compass*, *16*(10), 1-12. https://doi.org/10.1111/hic3.12492

Cresce, R. D., & King, J. (2017). Developing collaborative best practices for digital humanities data collection: A case study. *College & Undergraduate Libraries*, 24(2–4), 226–237. https://doi.org/10.1080/10691316.2017.1326330

Fitzpatrick, K. (2012). The Humanities, Done Digitally. In M. K. Gold (Ed.), *Debates in the Digital Humanities* (NED-New edition, pp.12–15). University of Minnesota Press. http://www.jstor.org/stable/10.5749/j.ctttv8hq.5

Piñeiro-Naval, V., Nuevo, L. A. S., & Lazcano, F. de J. M. (2014). A study of web content analysis on cultural heritage. *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*, 659–665. https://doi.org/10.1145/2669711.2669970

Piñeiro-Naval, V., & Serra, P. (2019). How Do Destinations Frame Cultural Heritage? Content Analysis of Portugal's Municipal Websites. *Sustainability*, 11(4), 947. https://doi.org/10.3390/su11040947

Posner, M. (2016). What's Next: The Radical, Unrealized Potential of Digital Humanities. In Gold M. & Klein L. (Eds.), *Debates in the Digital Humanities 2016* (pp.32-41). Minneapolis; London: University of Minnesota Press. https://www.jstor.org/stable/10.5749/j.ctt1cn6thb.6

Poremski, M. D. (2017). Evaluating the landscape of digital humanities librarianship. *College & Undergraduate Libraries*, 24(2–4), 140–154. https://doi.org/10.1080/10691316.2017.1325721

Stevens, G. (2017). New Metadata Recipes for Old Cookbooks: Creating and Analyzing a Digital Collection Using the HathiTrust Research Center Portal. *The Code4Lib Journal*, 37. https://journal.code4lib.org/articles/12548

9. Appendix

9.1

9.2

9.3

KBR

Research & Collections ˅   Cultural Offer ˅   Services ˅   SUPPORT US

ACCUEIL > RESEARCH > DATA-KBR-BE

DATA-KBR-BE

Facilitating data-level access to KBR's digitised and born-digital collections for digital humanities research

About the project

The DATA-KBR-BE project is a 24 month project (2020-2022) financed by the Belgian Science Policy Office (Belspo) as part of the Belgian Research Action through Interdisciplinary Networks, BRAIN 2.0 programme. It is an interdisciplinary collaboration between cultural heritage experts, digital humanities researchers and data scientists. The aim of the DATA-KBR-BE project is to facilitate data-level access to KBR's Collections for Open Science.

DATA-KBR-BE will optimise KBR's existing ICT infrastructure to stimulate sustainable data-level access to KBR's digitised collections for digital humanities research. For this project, research teams at the universities of Ghent (GhentCDH and IDLab) and Antwerp (ACDC) will work closely together with the digitisation, collections and ICT experts at KBR to co-design three interdisciplinary research scenarios that will extract relevant thematic datasets from KBR's digitised historical newspaper collection, BelgicaPress for reuse and analysis in the field of digital humanities.

The key outputs of this project will include:

Table of Contents

About the project
Collections as Data
Data-level access to collections
Digital Humanities Research
Harnessing the expertise of Data Science
Laying the foundations
International collaboration
Partners
Contact

Tags

Digital humanities, Newspapers, Open

9.4

Library and Archives Canada   Canada

Search BAC-LAC.gc.ca   Search

Discover the Collection ˅   Search the Collection ˅   Services for the Public ˅   Services and programs ˅

Home → Discover the Collection → Open Data from LAC

Renewing our web presence   ✕

Open Data from LAC
Related Links
Databases

Open Data from LAC

In March 2011, Library and Archives Canada (LAC) joined the Government of Canada (GC) initiative on open data, which provides all Canadians with access to GC datasets and associated metadata in the public domain. For the launch of this initiative, LAC released two datasets: **Soldiers of the First World War—CEF** and **Maps, Plans and Charts of Canada**, and subsequently developed a long-term plan for releasing more public domain datasets.

- Open Data Portal
- Canadian Subject Headings in SKOS/RDF
- Government of Canada Core Subject Thesaurus in SKOS/RDF format

Older Government Data

Share this page   👍 Like

Date modified: 2020-09-09

Terms and Conditions | Transparency

About us
Our mandate

News
News releases

Contact us
Address

Stay connected
Social media

9.5

National Library of Finland
Open Data and Linked Data Service

Search works, persons, organizations and subjects:

[ ] Go

Open Data Service

Access open data sources maintained by the National Library of Finland

Data Catalog

Our data catalog showcases the open data sets and APIs of the National Library.

Data sets

Open data sets available for reuse.

APIs

Interfaces that provide access to our data.

Tweets by @NatLibFiData

NLF Open Data
@NatLibFiData
The #Fennica #LinkedData service
data.nationallibrary.fi is up and running again!
It was shut down over the holidays due to
server issues. Apologies for the
inconvenience.

Jan 18, 2020

NLF Open Data Retweeted

Osma Suominen
@OsmaSuominen
Finally the IFLA #LRM, #BIBFRAME and

Linked Data

Linked Bibliographic Data

The national bibliography Fennica is available as
RDF Linked Open Data.

**Browse our Linked Data
Collections**

- Fennica

Access

- SPARQL endpoint
- Linked Data Fragments server
- Downloadable data dumps

Documentation

- Fennica RDF data model documentation
- What is included in Fennica (in Finnish)
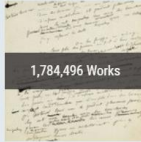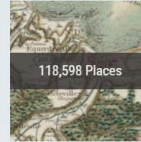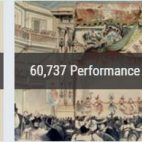
9.6

{BnF Data

Se repérer sur data ? | Que demander à data ? | Le projet data | Semantic web and data model | Sparql | Français | English | Deutsch

Reference information about authors, works and topics.

[search data.bnf.fr]

2,376,496 Authors   1,784,496 Works   193,580 Topics   118,598 Places   2,709 Dates   60,737 Performance   354,587 Journals

Informations legales | footer.contact | version 2.12.4

9.7

DEUTSCHE NATIONAL BIBLIOTHEK
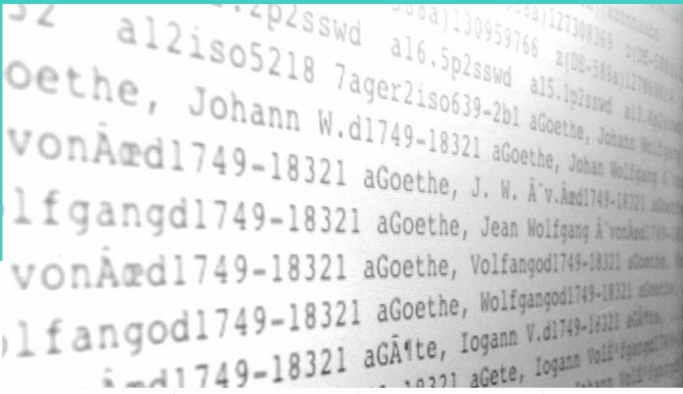
☰ MENU   CORONA MEASURES   DNB FOR USERS   DNB PROFESSIONAL   👤   🔍

Home › DNB Professional › Metadata Services

METADATA SERVICES

- ﹀ Which metadata is available?
- ﹀ How can I obtain data?
- ﹀ Which export formats are available?
- ﹀ Overview and costs
- ﹀ Stay informed!

All bibliographic data from the German National Library and the authority data from the Integrated Authority File is available free of charge for general use in according with our **business modell** under the terms of **Creative Commons Zero (CC0 1.0)** ↗ .

9.8

DEUTSCHE NATIONAL BIBLIOTHEK
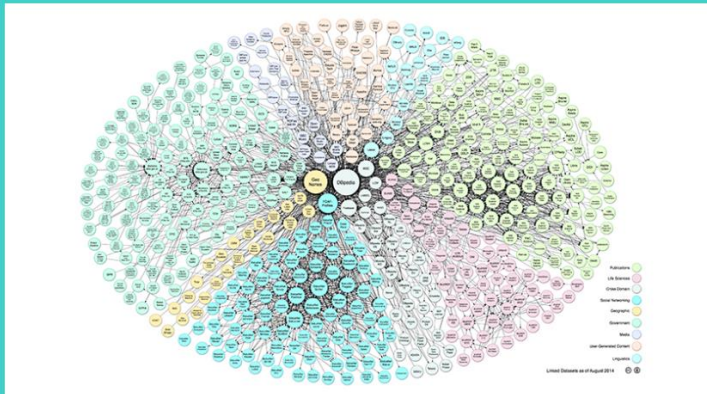
☰ MENU   CORONA MEASURES   DNB FOR USERS   DNB PROFESSIONAL   👤   🔍

Home › DNB Professional › Metadata Services › Linked Data Service

LINKED DATA SERVICE

- ﹀ Overview
- ﹀ Integrated Authority File (GND)
- ﹀ Bibliographic data
- ﹀ Test data
- ﹀ Subscription Terms and Terms of Use
- ﹀ Further development and service information
- ﹀ Frequently asked questions (FAQ)
- ﹀ Documentation
- ﹀ Download
- ﹀ Contact

9.9
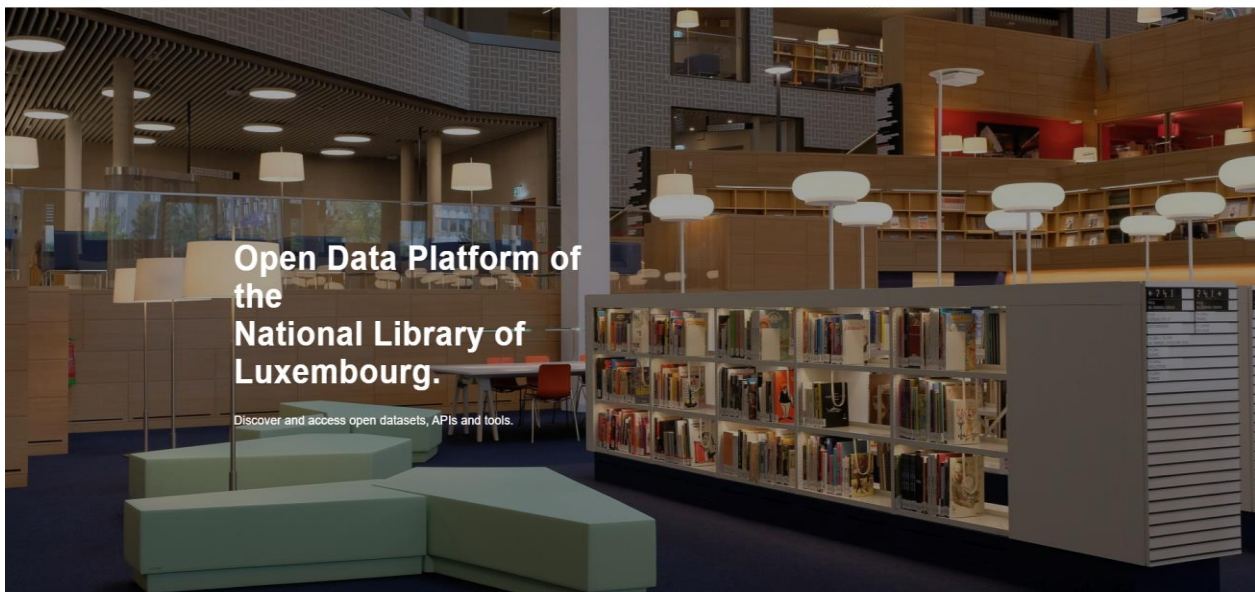
9.10
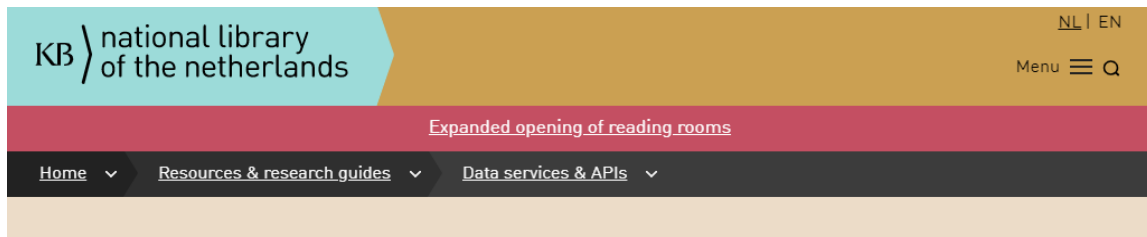
Data services & APIs

Digital illustrations, metadata and texts from the KB-collections are available for re-use. The data can, depending of the license, be used for research, web applications and other services. The data are available in linked data, as download, through an API (Application Programming Interface) using SRU or OAI-PMH or via Wikipedia Commons.

Open data sets
The following data sets may be used freely:

> The Delpher open newspaper archive , consisting of texts (OCR, ALTO, XML) of Dutch public domain newspapers published between 1618 until 1876 (as a download)
> Dutch Parliamentary Papers (Staten Generaal Digitaal), in cooperation with the House of Representatives (via OAI-PMH and SRU)
> Early Dutch Books Online (EDBO or DPO) in cooperation with the university libraries of Amsterdam (UvA) and Leiden
> Medieval Illuminated Manuscripts in cooperation with Museum Meermanno | Huis van het Boek
> Watermarks in Incunabula printed in the Low Countries
> Short Title Catalogue Netherlands (STCN; as linked data)
> Title information of the Digital Library for Dutch Literature (Digitale
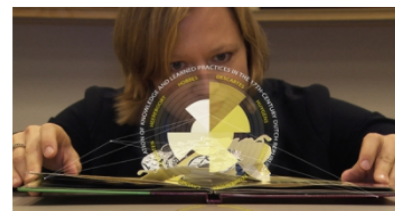
Digital humanities

Explore our services and programmes for digital humanities research

Go to the overview

9.11

9.12

9.13

9.14

The bibliographic data portal of the National Library of Spain

Datos.bne.es proposes to the user a new way of approaching the collections and resources of the National Library of Spain.

It is a data publishing project such as Linked Open Data, based on Web technologies and standards.

See help for search instructions and tips.

Buscar una persona, grupo, entidad, el título de una obra o un tema. Ej.: Leonardo da Vinci, ONU, Grapes of Wrath, Aztecas

9.15

9.16

9.17

9.18

9.19

9.19

9.20

9.21

9.22

Digital Research Lab

A lab for text and data mining research on KBR's digitised and born-digital collections.

The KBR Digital Research Lab serves to facilitate text and data mining research on KBR's diverse, multilingual digitised and born-digital collections. This includes supporting the digital access of textual sources, stimulating the (re)use and research of these digital sources, data and metadata of these collections.

Through a unique long-term cooperation with the Ghent Centre for Digital Humanities ⬈, Ghent University, the Lab seeks to serve as a site of **research and experimentation** for providing advice and support for realising these digital projects, specifically the use of computational tools such as **text and data mining**, and **digital humanities research** on KBR's collections. Thus it serves as a site for collaboration between KBR's departments, as well as academia, cultural heritage institutions and the larger public to develop digital pipelines, advise on digital tools and methods, and implement artificial intelligence on the digitised and born-digital textual collections to look at the collections new ways.

Tags

Open data, Open Science, Research

Contact person

Prof. dr. Julie M. Birkholz, Assistant Professor of Digital Humanities at the Ghent CDH at Ghent University, Ghent,

9.23

9.23

9.24

9.25

9.26

9.27

Welcome to BNElab

The project by the National Library of Spain for digital reuse

We want knowledge to become much more.

We want to explore and discover new uses for the BNE data and digital collections. Providing tools, products and services to be used for research, teaching and learning... but also for gastronomy, tourism, design and science... and thus re-create and generate something new. Generating more knowledge, and making our cultural heritage more prominent and useful in society.

A driving force for personal, economic and social growth.
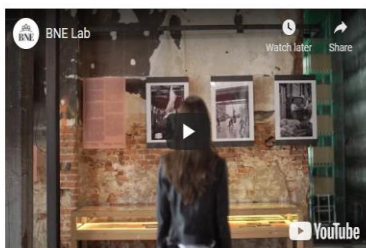
9.28

You are here: | Go to:
The library | Collaboration and development

Open today: 11–15   Find quickly   Languages   search

Home page • In English • Research collaboration • KBLab

KBLab

KBLab enables researchers to engage in large-scale analysis of KB's collections. We also develop language models and work with artificial intelligence.

This information is also available in Swedish.

KBLab is a national research infrastructure for digital humanities and social science. Through the lab we provide access to KB's collections in structured and quantative form. This makes it possible for researchers both to seek new answers and pose new questions in their research.

The lab is also a forum for play and innovation. We design and make models to analyse the library's vast collections of text and audio-visual materials in new ways.

Contact

E-mail: kblabb@kb.se
Visiting address: Karlavägen 100, Stockholm.

National Library of Sweden

9.29

9.30