



# Digital Data at KBR

Dr. Julie M. Birkholz & Sally Chambers, presentation for the Belgian Association for Documentation (ABD-BVD)

**KBR** 

  
**GHENT  
UNIVERSITY**



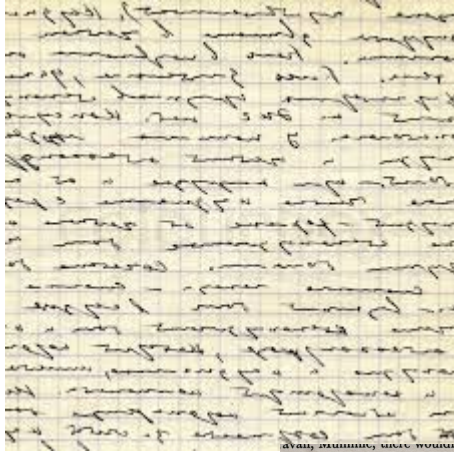
# Introductions

Julie M. Birkholz

- lead of KBR's Digital Research Lab - text & data mining
- Assistant Professor Digital Humanities at Ghent University
- expert in developing digital workflows for text and data mining using cultural heritage materials

Sally Chambers

- research coordinator of KBR's data.kbr.be, Belspo BRAIN project to develop a data platform for sharing datasets of collections
- expert in collections as data for cultural heritage materials
- DARIAH-BE National Coordinator, DARIAH-EU National Coordinator  
Committee Chair



...ms. Oh, do let us go in a caravan."

Mrs. Russell shook her head. "I know it sounds lovely, darling; but we've got to get a caravan? It would cost at least fifty pounds to buy one, and if we had one, Daddy couldn't get away this summer. No, we must make up our minds to do without a holiday this year; but I'll tell you what I'll do: we'll all go to Southend for the day, as we did last year, and have our lunch and tea with us and have a splendid picnic."

"Then we can bathe again," said Bob; "but, oh! I do wish I could have my pony and ride," he added unexpectedly. "You don't know how I long for a pony," he continued, sighing deeply as he remembered the blissful holiday when a friend let him share his little Dartmoor pony and ride occasional rides on it. "Southend is nothing but houses and people," cried Phyllis; "it's not a holiday in this place; and oh! Mummie, I do so long for fields and flowers and fresh air!" she added piteously; and she shook her long brown hair forward to hide the tears in her eyes.

"Never mind, darling, you shall have them one day," answered Mrs. Russell with easy vagueness.

This really was not very comforting, and it was the most fortunate thing that happened at that moment a car stopped at the door.

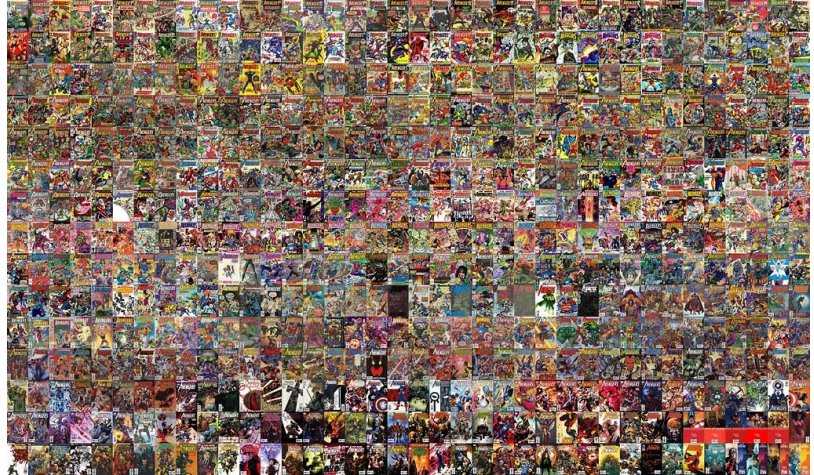
"Uncle Edward!" shouted Bob, rushing from the room. Phyllis brushed her tears so hastily from her eyes that she arrived at the front door almost as he did, and both flung themselves on the tall, kindly-looking man standing beside the car.

"Uncle Edward! Uncle Edward!" they cried. "You've come at last, we've been longing to see you. Oh, how glad we are you're here!"

Now the delightful thing was that their uncle seemed just as pleased to see them as they were to see him, and returned their hugs and greetings with the most cordiality. They were just on the point of dragging him into the house, hanging one on each arm, when he said: "Stop, not so fast. There are some things to fetch in from the car."

So saying he began diving into the back of it and bringing out, not only a suitcase, but various parcels, which he handed out one by one.

"These are the nice chocolates I've brought for your mother," said he, holding



KBR - Royal Library of Belgium is the federal Science/research library of Belgium.

KBR is responsible for *maintaining and making available* a diverse and multilingual collection; preserving, managing and doing research on more than 8 million documents.

- Legal deposit
- Inherently multilingual collection
- Increasing amount of data (on collections)!



@ KBR

KBR maintains lots of materials and data from analogue/source documents, digital objects or versions, born-digital objects, metadata on these objects, these collections as data (e.g. full-text, images), etc...

# What is data?

As librarians, collection managers, cultural heritage experts and researchers, we are constantly confronted with information. Not only how to make sense of it in the human mind! But also easily be able to reference, structure, organize and recall it with the help of computers.

From the moment we have this information in a format that a computer can “read” we have data.

Specifically machine-readable data.



# Data types

- bibliographical data,
- authority data,
- digital data,
- collections as data,
- born-digital and
- e-publication data,
- research data,
- lab data (code).

Each of these are different types of data, that have specific characteristics and are often maintained by diverse groups of experts, physically stored in different places and ways, and thus have different priorities, challenges and opportunities.



# Challenge

The increasing digitization of society and resulting increased digitization of library collections results in KBR needing to develop processes to manage an insurmountable amount of diverse document and data types. And for providing collections available in a machine readable format- or collections as data, demands new processes and practices. This enables us to explore digital humanities research.

# Library Labs: an example of KBR's Labs



Number of research projects working on digital heritage:

- KBR Digital Research Lab
- data.kbr.be
- BE-Social
- BELTRANS

And future:

- CAMILLE
- Data Science Lab
- Literary Lab
- Pop-up Heritage Lab

<https://www.kbr.be/en/research/>

# 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](#) <sup>↗</sup>, 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**

f  in



Tags

---

Open data, Open Science,  
Research

<https://www.kbr.be/en/projects/digital-research-lab/>

# DATA-KBR-BE

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



<https://www.kbr.be/en/projects/data-kbr-be/>

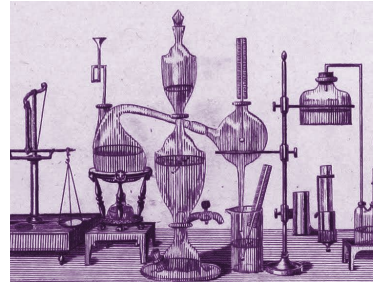




# What are 'library' labs?

"Any library program, physical or digital (or a hybrid) in which innovative approaches to library services, tools, or materials are tested in some structured way before being made part of regular workflows, programs, or mission."

Brooks, M., & Heller, M. (2013) Library labs. *Reference & User Services Quarterly* 52(3), 186-190 (2013).



First library lab was founded in 2013

LIBRARY  
HSILIRB



**Mahendra  
Mahey**

Project Manager,  
British Library Labs



**Adam  
Farquhar**

Head of Digital  
Scholarship



**Eleanor  
Cooper**

Project Officer,  
British Library Labs

## British Library Labs Experiment with our digital collections

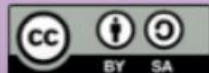
<https://www.bl.uk/projects/british-library-labs>



@BL Labs @britishlibrary @labs@bl.uk @BL DigiSchol

# British Library Labs (BL Labs)

- **Promotes, inspires, and supports the use of the British Library's digital collections and data**
- **Works on projects with researchers, developers, educators, entrepreneurs and artists from around the world**
- **Provides insight into the emerging practice of digital research and helps shape the provision of the British Library's digital services, tools, collections, and data.**
- **Ensures that the intellectual digital heritage we hold is accessible to everyone for research, inspiration and enjoyment**



@BL Labs @britishlibrary @labs@bl.uk @BL DigiSchol

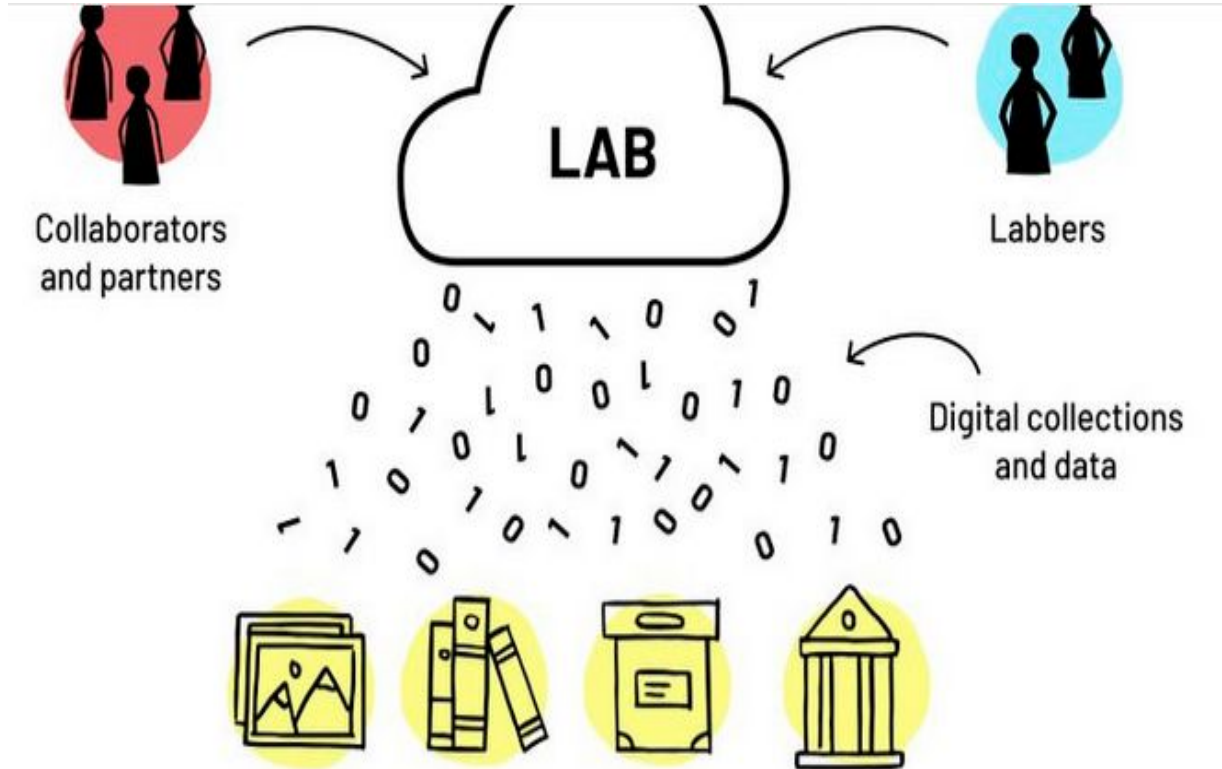
# International GLAM Labs Community



253 people, 60 institutions  
in 30 countries



# What does a Library Lab look like?



<https://glamlabs.io/books/>

# FED-tWIN Research Cooperations

The Digital Research Lab, and other future Labs, are BELSPO - the Belgian Science Policy Office, funded research & cooperation projects

- 10 year projects to bring in research and technical expertise
- In cooperation with Belgian universities
- 50% position @ FSIs & research position @ university
- Started in February 2020, first FED-tWIN project in FSIs

# Launching pad

Digital Research Lab is a launching pad for developing policy plans, research and programs 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.

# What is text and data mining?

- a process for deriving information that can be used or seen as data from machine-readable material to identify and analyze different characteristics or patterns.
- Machine-readable material is any material that a computer can “read” or interpret to store in a semi-structured way. For example an analogue book is only machine readable once it has been scanned, the images have been identified as letters that make up words that thus make up a text.
- This is the process of optical character recognition (OCR) or HTR - the electronic conversion of images of typed, handwritten or printed text into machine-encoded text.
- This machine-encoded text can then be stored on a computer, for example as a PDF, MSWord document, a .txt file, XML, etc... for the user to (re)use and, in the case of research, analyze.

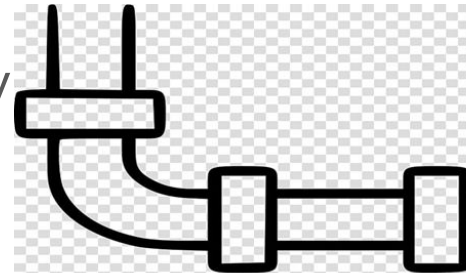
# And data?

- Once this is *machine readable* the information is structured in such a way that the user can identify which parts are needed from a *full-text search*.
- This is afforded by *natural language processing* (NLP) – techniques to computationally process and analyze large amounts of natural language data.
- Once we have those texts which we review as *data*, we then can do further mining, or the analysis of the data implementing different types of analysis *both computational as well as traditional qualitative close reading*; to understand the data, discover new knowledge, and test hypotheses.

Sounds simple but in reality we need to have an overview of the pipeline...

# Pipeline: connecting pieces

- Inherently the work the Digital Research Lab does is connected to all digital heritage projects
  - to develop and ensure sustainable and re-useable research workflows, and facilitate research on the collections.
  - Means also advising and supporting the various Digital Humanities and Digital Heritage projects
  - And working across departments/transversely to achieve goals



# KBR Digital Data Strategy

The KBR Digital Data Strategy outlines a proposal for how to manage the variety of this *data* (e.g. diverse formats needed for different document types). This includes:

- the *volume* (e.g. increased computer storage demands to ensure accurate provenance and high level information on the source documents, digitized versions, and data),
- the *veracity* (e.g. ensuring quality and availability to users in new ways),
- *rights management*,
- and in some cases an increased *volume* of requests and demands of these services (e.g. instead of a researcher consulting one book, they request full-text of one million books).



# KBR Digital Data Strategy + action plan + future agreement with the government 2022

## ***The KBR's Action Plan 2019 - 2021:***

- i) to publish as much information online as possible and facilitate its (re)use;
- ii) to provide the user with easy and sustainable access to the information, both on site and online and
- iii) to provide the user with easy and sustainable access to the information, both on site and online 'ensuring our collections are more accessible and open to the public and humanities researchers'.

# FAIR data

- **Findable**/vindbaarheid



- **Accessible**/toegankelijkheid

- **Interoperable**/interoperabiliteit

- **Reusable**/herbruikbaarheid

# ***BESOCIAL***

set up a sustainable strategy  
for archiving and preserving  
social media in Belgium.

- KBR
- UGent (GhentCDH, MICT, and IDLAB)
- UNamur (CRIDS)
- UCL (CENTAL)



<https://www.kbr.be/en/projects/besocial/>

# BELTRANS - translation flows and their context (1970 - 2020)



Studying intra-Belgian cultural transfers

4 Years project with 3 PhD students working on it

KBR responsible for data management

<https://www.kbr.be/en/projects/beltrans/>

These research projects are all short term projects that require spaces to experiment and for research valorization.

As Labs in the library we have a role in mediating the process with all these actors from the source material to the access to the digital copy and the ability to reuse the digital formats (not just text) for research. But also for wild ideas that need formalizing.

Labs can provide this space to experiment with digital data in digital heritage and digital humanities projects.

# DATA-KBR-BE

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

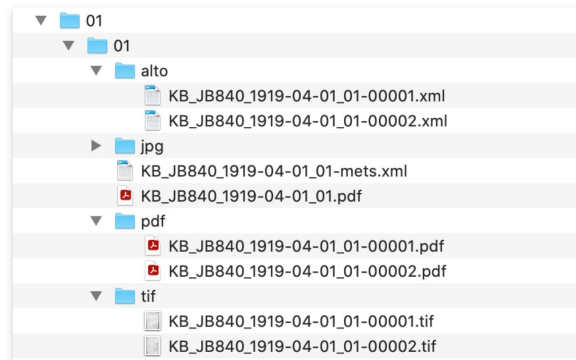
<https://www.kbr.be/en/projects/data-kbr-be/>



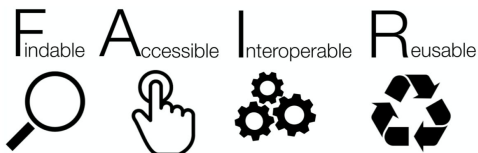
# Digitised Historical Newspapers as Data

## Collections as Data

Providing data-level access to digital collections is a primary challenge for undertaking digital humanities research. In the United States, the flagship initiatives, 'Always Already Computational: Collections as Data' and 'Collections as Data: Part to Whole', define 'Collections as Data' as a "conceptual orientation to collections that renders them as ordered information, stored digitally, so that they are inherently amenable to computation". The initiative was established to document, exchange experience and share knowledge to encourage cultural heritage institutions to implement 'collections as data' in their own institutions. DATA-KBR-BE will kick-start the implementation of 'Collections As Data' in Belgium.



<https://collectionsasdata.github.io>



Providing data-level access to the underlying files of digitised and born-digital cultural heritage resources to facilitate data analysis by means of tools and methods developed in the field of digital humanities

**KBR**

<https://www.kbr.be/en/projects/data-kbr-be/>

# Collections as Data in Belgium

INSPIRATIE



## COLLECTIONS AS DATA

Digitale collecties van cultureel-erfgoedinstellingen, zoals bibliotheken, archieven en musea, worden steeds vaker gebruikt voor digitaal onderzoek in de geesteswetenschappen. Traditionele manieren om toegang te verlenen tot dergelijke collecties, bijvoorbeeld via digitale bibliotheekplatformen zijn niet altijd ideaal voor onderzoekers die datasets rond specifieke onderzoeksvragen willen opbouwen. Het *Collections as Data*-initiatief, afkomstig uit de Verenigde Staten, stimuleert nauwere interdisciplinaire samenwerking tussen cultureel-erfgoedexperten, *digital humanities*-onderzoekers en *data scientists* om samen na te denken over hoe toegang tot digitale collecties verleend kan worden. Zo kan analyse vergemakkelijkt worden met behulp van digitale tools en methoden. Wat is *Collections as Data* precies? Kan het interessant zijn voor cultureel-erfgoedinstellingen in België?

TEKST Sally Chambers (KBR & GhentCDH) en Frédéric Lemmers (KBR)

### COLLECTIONS AS DATA: WAT IS DAT PRECIËS?

*Collections as Data* biedt een nieuwe denkwijze aan over hoe toegang tot digitale (gedigitaliseerde en born-digital) collecties in cultureel-erfgoedinstellingen aangeboden kan worden.

De eerste fase van het initiatief, *Always Already Computational: Collections as Data* (2016-2018), focuste op het uitwisselen van ervaringen en het delen van kennis over potentiële benaderingen om toegang aan te bieden tot het onderliggende data-niveau van digitale collecties.

In de huidige, tweede fase *Collections as Data: Part to Whole* (2019-2021), ligt de focus op de implementatie en het gebruik van *Collections as Data* door een aantal collaboratieve casestudy's. Een belangrijk onderdeel van deze benadering is dat de interdisciplinaire samenwerking tussen cultureel-erfgoedexperten en *digital humanities*-onderzoekers gelijkwaardig moet zijn. Elke *Collections as Data*-casestudy wordt gezamenlijk geleid.

Tot nu toe heeft de implementatie van *Collections as Data* groten-deels plaatsgevonden in de Verenigde Staten, maar het gebeurt langzamerhand ook in Europa. Er zijn al een aantal Europese voorbeelden, onder andere de Data Foundry van de nationale bibliotheek van Schotland, het Open Data Platform van de nationale bibliotheek van Luxemburg en de datasets van het KB Lab van de Koninklijke Bibliotheek van Nederland. Verder is het gebruik van collaboratieve notebooks zoals de GLAM Workbench van Tim Sherratt of de GLAM Notebooks van de labs van de Biblioteca Virtual Miguel de Cervantes, een andere manier om digitale collecties te benaderen op computationele wijze. (*GLAM staat voor galleries, libraries, archives en musea, n.v.d.r.*) Maar, waar staan we eigenlijk met *Collections as Data* in België?

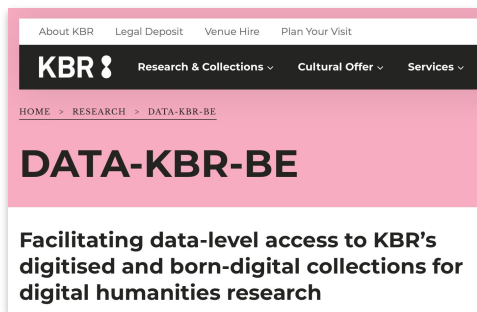
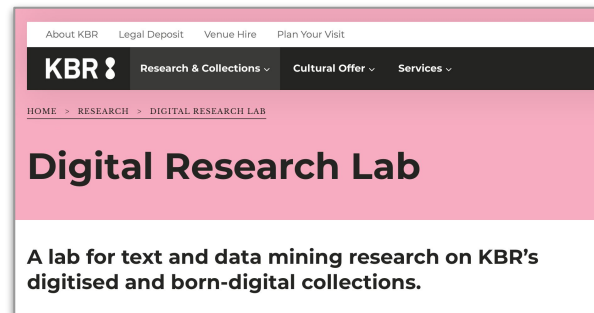
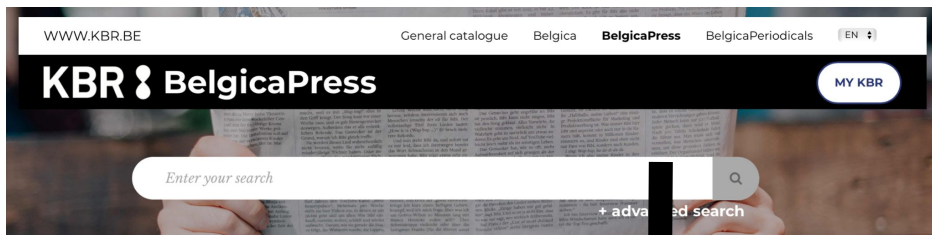


Automatische layout-analyse van de historische kranten van KBR, door IDLab, UGent.

<https://orfeo.kbr.be/handle/internal/7871>

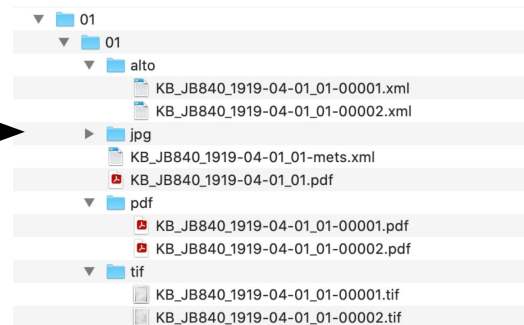
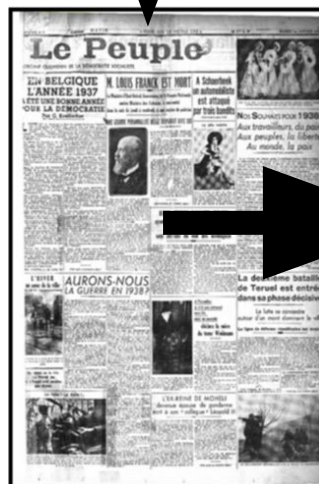


# Digitised Historical Newspapers as Data



<https://www.kbr.be/en/projects/digital-research-lab/>

<https://www.kbr.be/en/projects/data-kbr-be/>



# Newspapers as Data



Bibliothèque nationale du Luxembourg  
Open Data



## STARTER PACK



250MB

of digitised newspapers

- ✓ 5 days of news
- ✓ 5 newspaper issues
  - ✓ 22 pages
- ✓ D'Wäschfra (1868)
- ✓ Public Domain, CC0 (See copyright notice)
  - ✓ Best for getting started & developing

DOWNLOAD (ZIP)

<https://data.bnl.lu/data/historical-newspapers/>

## TEXT ANALYSIS PACK



2GB

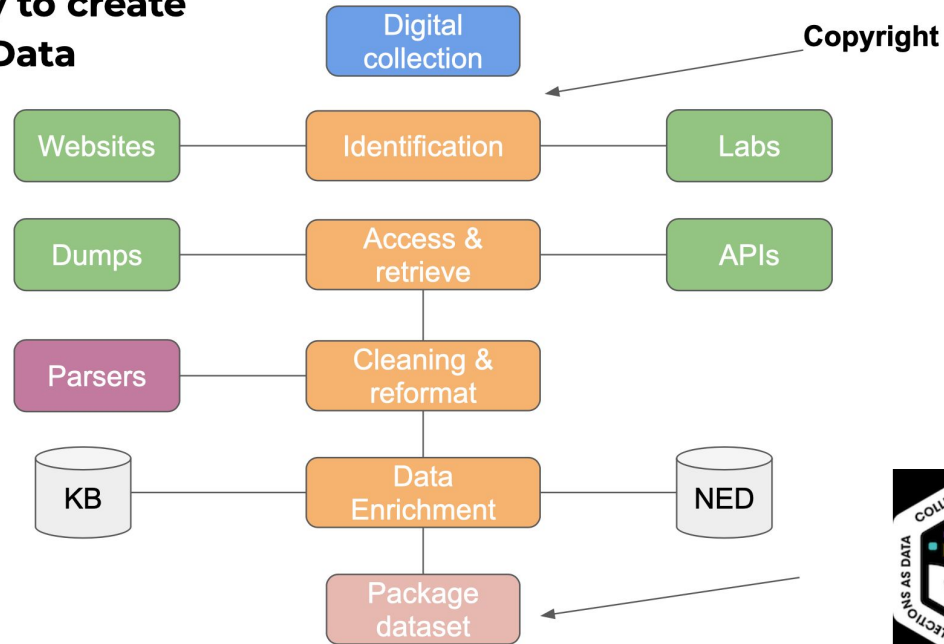
of processed newspapers data

- ✦ 38 years of news (1841-1878)
- ✦ 23663 processed newspaper issues
  - ✦ 97285 processed pages
  - ✦ 510505 extracted articles
- ✦ Public Domain, CC0 (See copyright notice)
- ✦ Best for getting started with text analysis

DOWNLOAD (ZIP)

# Collections as Data Methodology

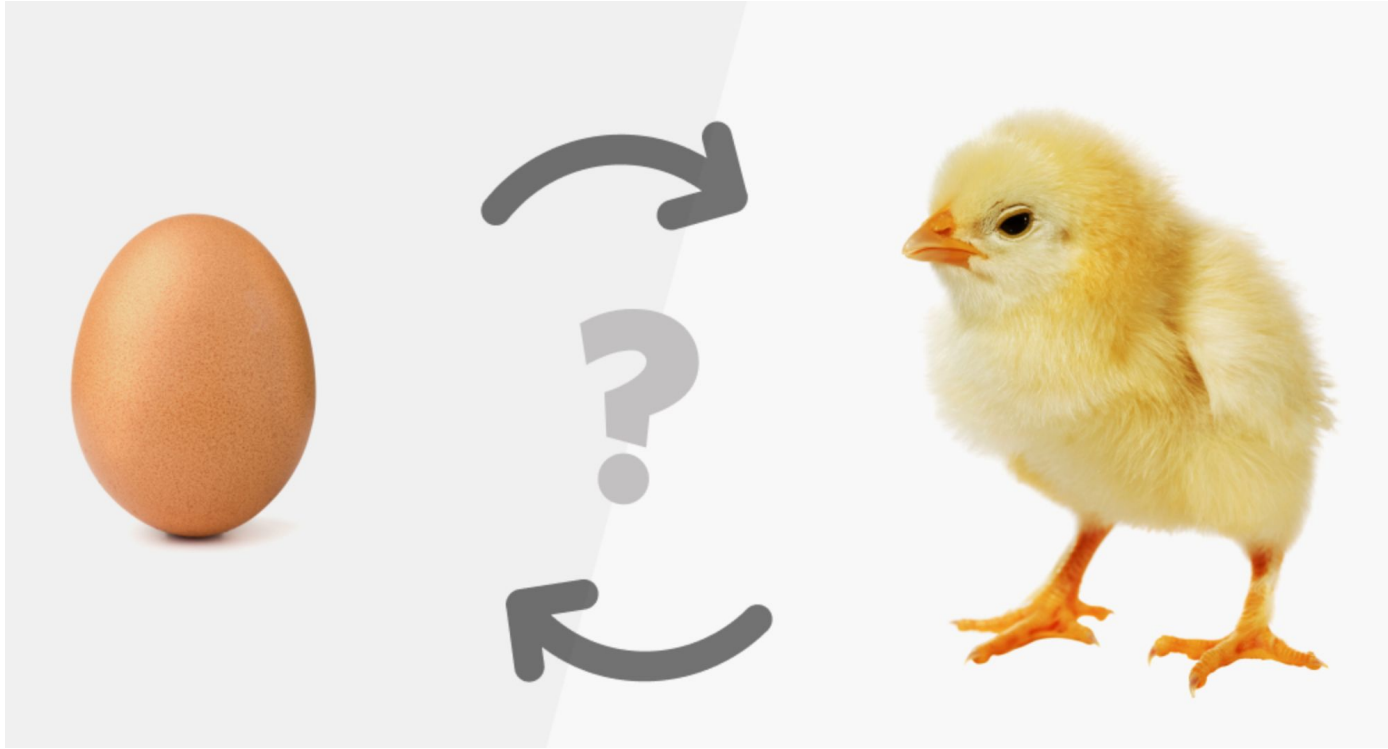
## A methodology to create Collections as Data



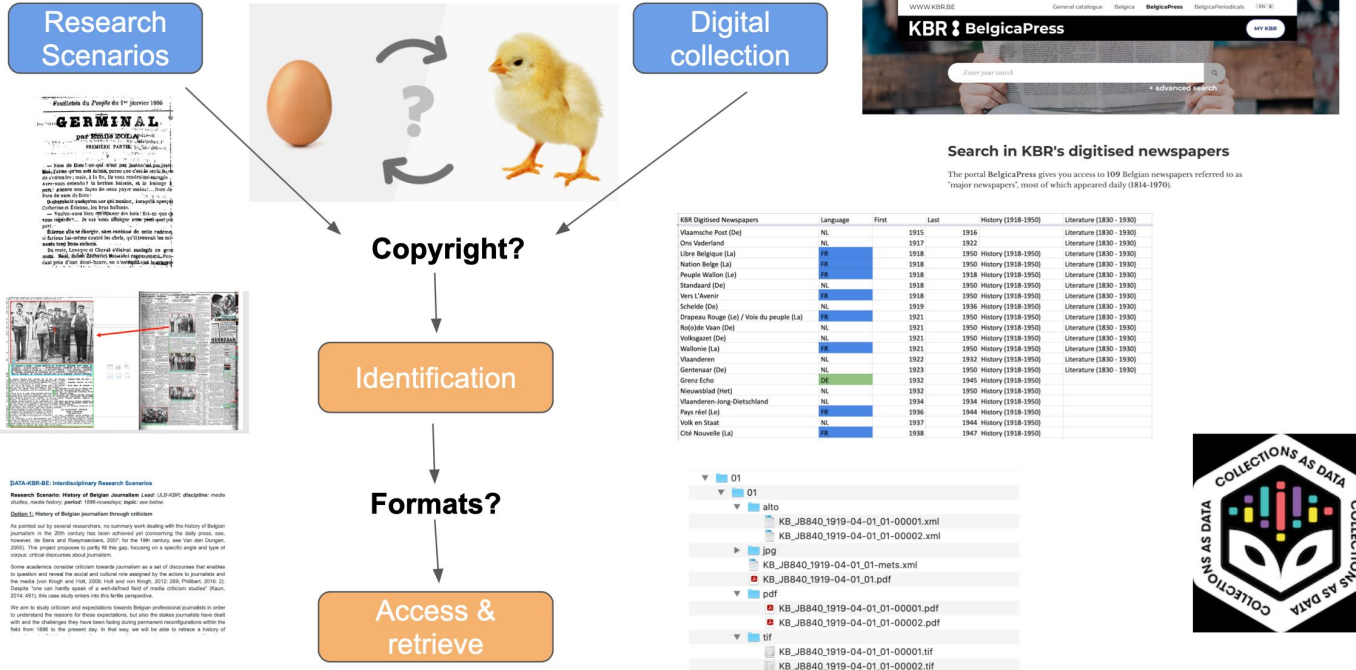
Candela, G., Sáez, M. D., Escobar Esteban, Mp., & Marco-Such, M. (2020). *Reusing digital collections from GLAM institutions*. *Journal of Information Science*:

<https://doi.org/10.1177/0165551520950246> & <http://rua.ua.es/dspace/handle/10045/109460>

# Collections or Corpora?



# Collections as Data @ KBR



**DATA-KBR-BE: Interdisciplinary Research Scenarios**  
**Research Scenario: History of Belgian Journalism** Lead: LEB-ASPC; @kbrpress; media studies; media history; [participatory history](#); [journalism](#); [digital journalism](#)

**Abstract:** History of Belgian Journalism through interdisciplinary approaches

As pointed out by several researchers, no summary work dealing with the history of Belgian journalism in the 20th century has been authored and published in the study period, see, however, de Haene and Depraetere, 2007 for the 19th century, see Van den Driessche 2005). This present proposal is hereby to fill this gap, focusing on a specific angle of the type of impact of digital documentation paradigms.

Some academics consider critical towards journalism as a set of discourses that enables to question and reveal the social and cultural role assigned by the actors in journalism and the media (van Kester and van Kester, 2010; van Kester, 2012; 200; Proffers, 2016; 2). Despite this, the main scope of a well-established field of media criticism (Kester, 2014; 481), the case study enters into this better perspective.

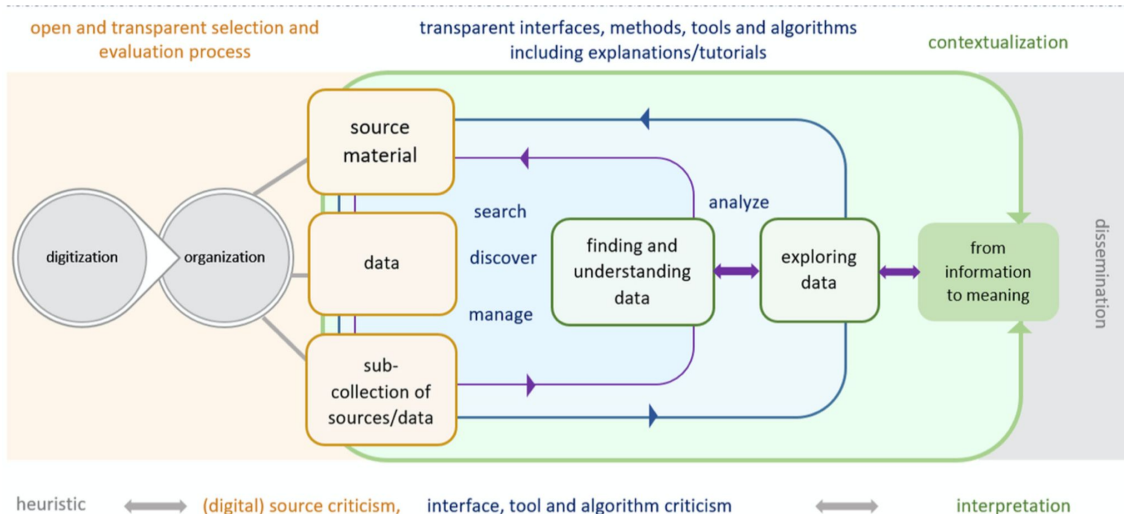
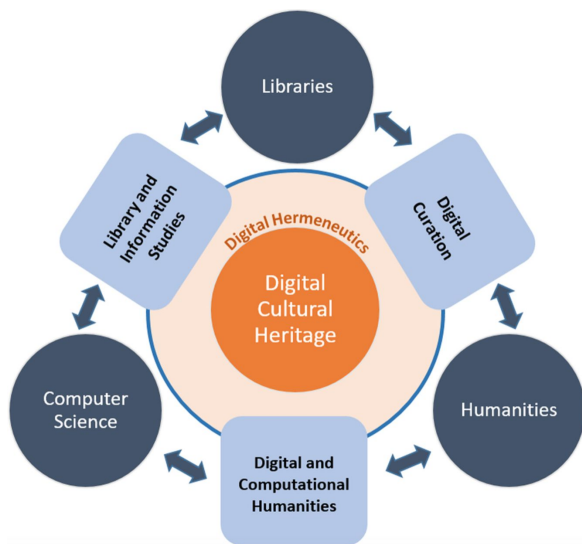
The aim is to study criticism and opposition towards Belgian professional journalists in order to understand the reasons for these expectations, but also the states journalists have dealt with and the challenges they have been facing during prominent moments when the field has been 1880 to the present day. In that way, we will be able to relate a history of

KBR Digitised Newspapers	Language	First	Last	History (1918-1950)	Literature (1830-1930)
Vlaamische Post (De)	NL	1915	1916		Literature (1830-1930)
Deu Vaderland	NL	1917	1922		Literature (1830-1930)
Libre Belgique (La)	FR	1918	1950	History (1918-1950)	Literature (1830-1930)
Nation Belge (La)	FR	1918	1950	History (1918-1950)	Literature (1830-1930)
Peuple Walon (La)	FR	1918	1950	History (1918-1950)	Literature (1830-1930)
Standard (De)	NL	1918	1950	History (1918-1950)	Literature (1830-1930)
Nieuws L'Avenir	FR	1918	1950	History (1918-1950)	Literature (1830-1930)
Scheide (De)	NL	1919	1916	History (1918-1950)	Literature (1830-1930)
Drapeau Rouge (La) / Voix du peuple (La)	FR	1921	1950	History (1918-1950)	Literature (1830-1930)
Revoluie Vaen (De)	NL	1921	1950	History (1918-1950)	Literature (1830-1930)
Volkskrant (De)	NL	1921	1950	History (1918-1950)	Literature (1830-1930)
Wolven (La)	FR	1921	1950	History (1918-1950)	Literature (1830-1930)
Vlaanderen	NL	1922	1932	History (1918-1950)	Literature (1830-1930)
Gemmetar (De)	NL	1923	1950	History (1918-1950)	Literature (1830-1930)
Grenz Echo	DE	1932	1945	History (1918-1950)	Literature (1830-1930)
Nieuwsblad (Het)	NL	1932	1950	History (1918-1950)	Literature (1830-1930)
Vlaanderen Jong Deutschland	NL	1934	1934	History (1918-1950)	Literature (1830-1930)
Pays réel (La)	FR	1936	1944	History (1918-1950)	Literature (1830-1930)
Voik en Staat	NL	1937	1944	History (1918-1950)	Literature (1830-1930)
Ono Nouvelle (La)	FR	1938	1947	History (1918-1950)	Literature (1830-1930)



Candela, G., Sáez, M. D., Escobar Esteban, Mp., & Marco-Such, M. (2020). *Reusing digital collections from GLAM institutions*. Journal of Information Science: <https://doi.org/10.1177/0165551520950246> & <http://rua.ua.es/dspace/handle/10045/109460>

# Corpus Building: an interdisciplinary digital hermeneutics workflow



# Interdisciplinary Research Scenarios

- **Collective Action Belgium** led by [GhentCDH](#), focuses on social history in the Interbellum and World War Two period and aims to trace the dynamics of contention, strikes, demonstrations and other forms of collective action in Belgium as reported in Belgian newspapers;
- **The feuilleton in Belgium**, led by [ACDC](#), focuses on literary studies in the period 1830–1930 and aims to map the publication of literature in Belgian newspapers across the first century of the Belgian nation state;
- **History of Belgian Journalism**, led by [ULB](#) and KBR, focuses on media history from 1886 until now and aims to trace the history of Belgian journalism through the lens of critical discourses about journalism as in Belgian newspapers.



# From collections to corpora

Year	Date	Title	Cause	Number of strikers	Description
1893	12–18 April	<a href="#">Belgian general strike of 1893</a>	Franchise reform	200,000	Successfully led to the establishment of <a href="#">universal male suffrage with plural votes</a> . <sup>[6]</sup> Thirteen strikers were killed and socialist leaders were briefly arrested. <sup>[5]</sup>
1902	10–20 April	<a href="#">Belgian general strike of 1902</a>	Franchise reform and an end to plural vote	350,000	Failed to achieve the abolition of the plural vote as Catholics and Liberals united to oppose constitutional reform. The Belgian Workers' Party had been reluctant to support the strike and it soon descended into violence in Brussels and parts of Wallonia. 12 workers and one policeman were killed. Union membership dropped sharply in its aftermath. <sup>[5]</sup>
1913	14–24 April	<a href="#">Belgian general strike of 1913</a>	Franchise reform	400,000	Carefully planned to avoid the same problems as 1902, the strike gained the promise of electoral reform but its proposals were postponed by the outbreak of <a href="#">World War I</a> and the <a href="#">subsequent German occupation</a> . The policy was finally adopted in 1919. <sup>[3][6]</sup>
1932	7 July–9 September	<a href="#">Belgian general strike of 1932</a>	Pay, working hours and unemployment insurance		Began after a spontaneous strike by <a href="#">coal miners</a> in the <a href="#">Borinage</a> and involved <a href="#">Communist agitation</a> following a severe decrease in living standards and real wages during the <a href="#">Great Depression</a> . Two people were killed during the strike. <sup>[7]</sup>
1936	2 June-2 July	<a href="#">Belgian general strike of 1936</a>	Working hours, paid holiday, union reforms	500,000	Broke out at the <a href="#">port of Antwerp</a> and led to the creation of a National Labour Conference. <sup>[7]</sup> Although influenced by the <a href="#">French Popular Front</a> and held against the backdrop of the <a href="#">Spanish Civil War</a> , it was also supported by Catholic trade unions. <sup>[8]</sup>
1950	24 July-3 August	<a href="#">Belgian general strike of 1950</a>	"Royal Question"	700,000	Chiefly active in <a href="#">Wallonia</a> , the strike contributed to the abdication of <a href="#">King Leopold III</a> on 1 August 1950. At least four strikers were killed.

[https://en.wikipedia.org/wiki/General\\_strikes\\_in\\_Belgium](https://en.wikipedia.org/wiki/General_strikes_in_Belgium)

## Which newspapers are digitised?

Title	Journal No.	Digitally availability	1886	1893	1902	1913	1932	1936	1950	Extraction
1 <a href="#">Vooruit: socialistisch dagblad</a>	JB 809	1884-1889; 1901-1902; 1911-1950	Yes	No	Yes	Yes	Yes	Yes	Yes	1913, 1950
2 <a href="#">Het Volk: antisocialistisch dagblad</a>	JB 785	1911-1916; 18 apr. 1918-31 maart 1927; 1931-1950.	No	No	No	Yes	Yes	Yes	Yes	1913, 1950
3 <a href="#">Vaderland</a>	JB 310	16 maart 1910-30 dec. 1913.	No	No	No	Yes	No	No	No	1913, 1950
4 <a href="#">Le Peuple: organe quotidien de la démocratie socialiste</a>	JB837	1885-1907; 1911-1914; 1918-1940; 1944-1950.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	1913, 1950
5 <a href="#">Le Vingtième Siècle</a>	JB729	6 juin 1895-13 mai 1940	Yes	Yes	Yes	Yes	Yes	Yes	No	1913, 1950
6 <a href="#">La Meuse: journal de Liège et de la Province</a>	JB638	1856; 1858-1882; 1884-5 août 1914; 28 nov. 1918-11 mai 1940; 9 sept. 1944-1950.	No	Yes	Yes	Yes	Yes	Yes	Yes	1913, 1950

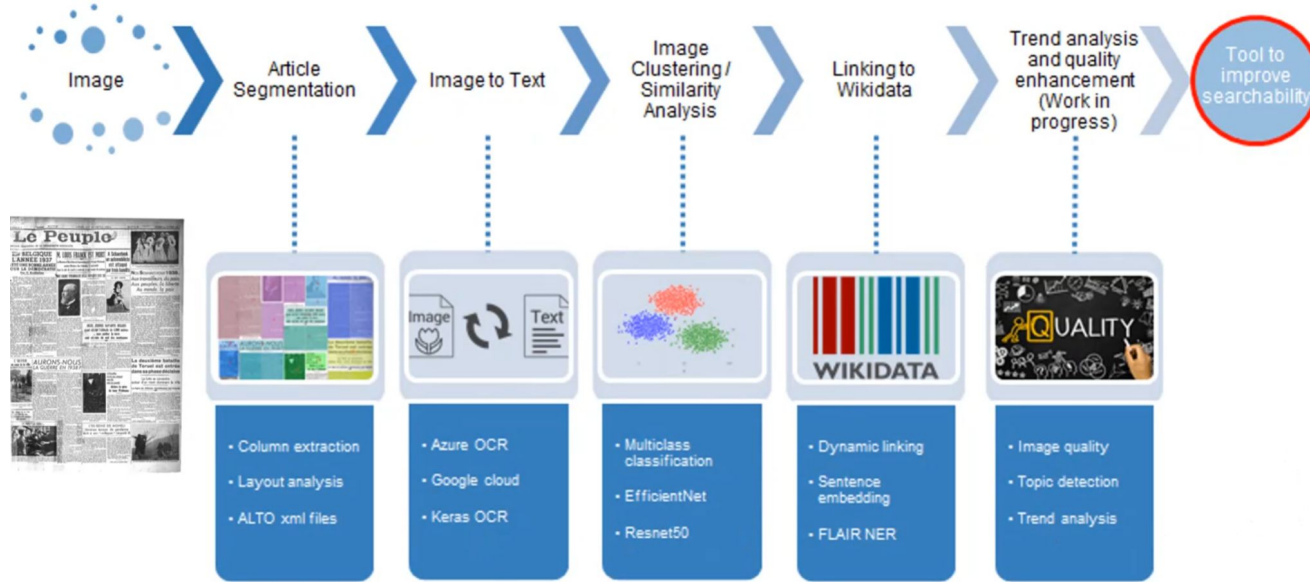
## Importance of historical context

“The newspaper was founded in Ghent in 1884 with the support of the socialist cooperative Vooruit. When the Parti ouvrier belge - Belgische Werkliedenpartij (“Belgian Workers Party”, POB-BWP, 1885) was founded, it was recognised as its official organ for the Flemish part of the country. The Ghent socialist leader Edward Anseele was editor-in-chief, but he also worked as a typographer. Vooruit was published under German censorship during the two world wars. Having reached its peak in the 1950s, it began a slow decline. It was succeeded by the daily *De Morgen* in 1978.”

**Vooruit: socialistisch dagblad**

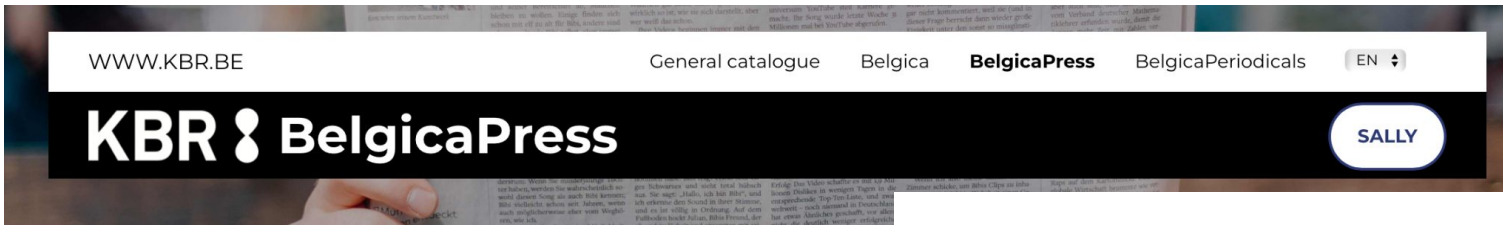


# Interdisciplinary collaboration with data scientists



Dilawar Ali, Kenzo Milleville, Alec Van de Broeck & Steven Verstockt, IDLab, UGent

# German Language Newspapers in Belgium



Your search:  
Title : Grenz Echo  
Language : German

11.865 pages | Sort: **MOST PAGES** | Filter: **YEARS** | **EDITIONS** | **ACCESS**

24-12-1937 - Ed.0 (12)

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8

24-12-1937 - Ed.0 - 1



<b>Titel</b>	<b>Grenz-Echo : Katholisches Volksblatt für die neubelgischen Gebiete Eupen-Malmedy-St. Vith. Organ der Katholischen Partei in den Gebieten Eupen, Malmedy, St. Vith</b>
<b>Bibliografisch adres</b>	Eupen, 1927-
<b>Land</b>	Belgium
<b>Materielle beschrijvi...</b>	f°
<b>Volgende titel</b>	Voortgezet als : <u>Grenz-Echo (und) St. Vither Zeitung</u>
<b>Belgicapress</b>	<a href="#">Consulteren in BelgicaPress</a>

<https://opac.kbr.be/LIBRARY/doc/SYRACUSE/16388591>

# NewspAlper Demonstrator

Dilawar Ali, Kenzo Milleville, Alec Van de Broeck & Steven Verstockt, IDLab, UGent

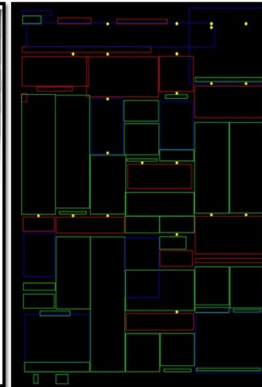
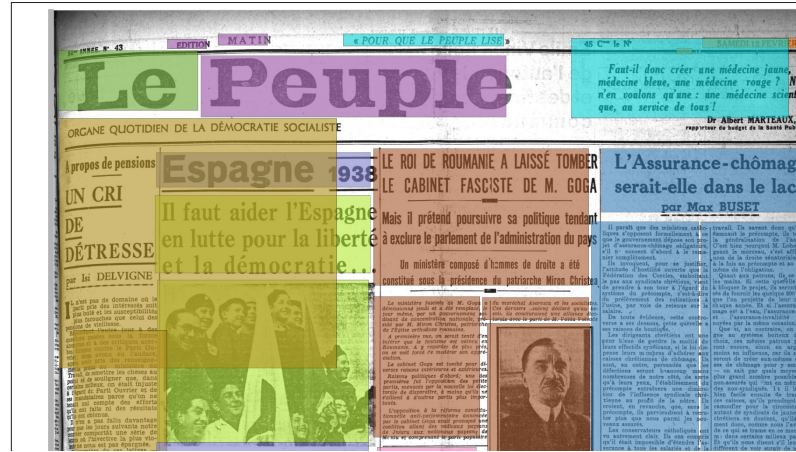


# Newspaper Demonstrator: Document Layout Analysis & Article Separation

## Page Viewer

Click on an article/image to show the recognized text and related articles/images

Le Peuple: Saturday, 12 February 1938, page 1



Document layout analysis using BelgicaPress



# NewspAlper Demonstrator: Named Entity Recognition & Linking

## Page Viewer

Click on an article/image to show the recognized text and related articles/images

Le Peuple: Thursday, 24 February 1938, page 1



## Most similar articles

[Samedi matin, devant de nombreux délégués réunis à la Maison du Peuple de Bruxelles le camarade D.](#)

[TROISIÈME JOURNÉE DE DISCUSSION sur la Politique étrangère au Conseil Général du P.O.B. CORNILLE ME](#)

[DEVANT LE PARLEMENT Dans un discours vivement acclamé Paul-Henri Spaak souligne les mérites de la po](#)

## Article text:

Au Conseil Général du P.O.B. Buset et Paul-Henri Spaak ont introduit un large débat sur LA POLITIQUE EXTÉRIEURE DE LA BEL GIQUE Les délégués ont salué la mémoire d'Edouard Anseele LA DISCUSSION SERA POURSUIVIE MERCREDI PROCHAIN



<https://tw06v072.ugent.be/kbr/>



# Newspaper Demonstrator - Image similarity

## Page Viewer

Click on an article/image to show the recognized text and related articles/images

Le Peuple: Tuesday, 1 March 1938, page 1

DU CITOYEN A. DE BLOCK.

### UN SOLDAT ALLEMAND déserte à bord d'un auto-tank blindé et passe en France

Thionville, 28 février. — Samedi matin, vers 6 heures, les douaniers de Waldwisse, à une vingtaine de kilomètres de Thionville, ne furent pas surpris de voir, venant à travers champs, une auto-mitrailleuse blindée. De toute évidence, l'engin venait d'Allemagne à une vitesse accélérée.

Les autorités frontalières furent immédiatement alertées, mais elles n'eurent pas à intervenir. L'auto blindée s'immobilisa d'elle-même, non sans avoir parcouru plusieurs kilomètres en territoire français. On put alors constater qu'il s'agissait d'une auto blindée « tous terrains », du plus récent modèle, et qu'elle était montée par un seul homme, l'espèce un jeune soldat allemand.

Interrogé immédiatement, ce jeune Allemand déclara qu'il avait fait dans la nuit plus de 100 kilomètres, avant d'arriver en France et qu'il avait

### Staline prépare une nouvelle charrette


Elle comprendra une vingtaine de condamnés dont 7 anciens commissaires du peuple et 4 médecins accusés d'avoir assassiné Gorki!

Boukharine, Rykov, Krestniski, Rakovskii et Iago figurent parmi ces bolcheviks de la première heure menacés de mort

Staline poursuit avec une rigueur inexorable la destruction des leaders bolchevistes de la première heure. Après l'exécution de la vingtaine qui compose la dernière charrette, on pourra compter ceux qui resteront sur les doigts d'une seule main.

Les charges accumulées contre ces derniers accusés présentent le même caractère outrancier que dans les procès précédents. A qui fera-t-on croire que des médecins ont assassiné Gorki, malade des poumons longtemps avant la révolution russe? C'est d'une double-entente sinistre?

Toujours les mêmes questions s'imposent à l'esprit: Que se passe-t-il en U. R. S. S. qui rende ces hécatombes néces-



Most similar images:

Most similar images:



# paix?

# Methods for Data Sharing

The image shows a screenshot of the DATA-KBR web interface. The top navigation bar includes 'Home', 'Login', and 'English (UK)' buttons. The main content area is titled 'Drop-Off Summary' and contains a table of files ready for upload. A file explorer window is overlaid on the right, showing a hierarchical directory structure for the files.

**Drop-Off Summary**

Click on a filename to download that file.  
This drop-off will expire in 28 days.

Filename	Size
JB837-1950-09.zip	2.7 GB
JB837-1950-02.zip	2.2 GB
JB837-1950-06.zip	2.7 GB
JB837-1950-08.zip	2.6 GB
JB837-1950-01.zip	2.2 GB

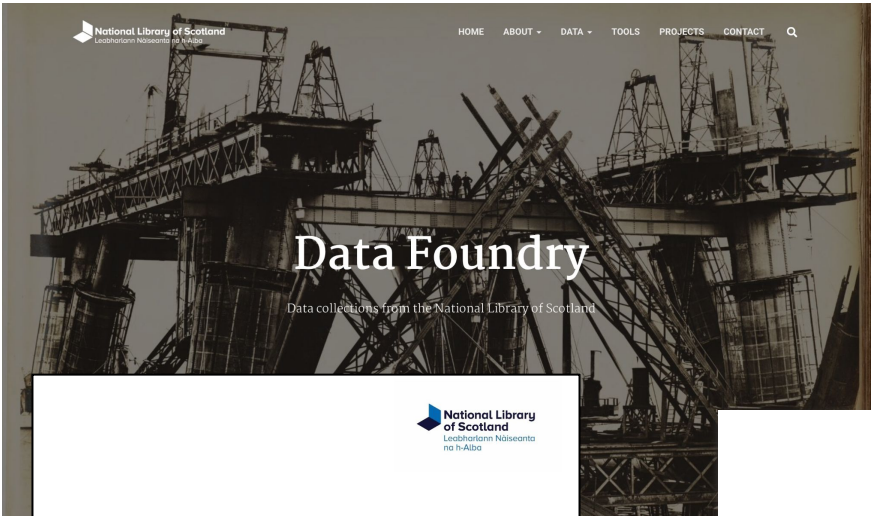
**File Explorer View:**

- 1950
  - 01
    - 03
      - 01
        - alto
          - KB\_JB837\_1950-01-03\_01-00001.xml
          - KB\_JB837\_1950-01-03\_01-00002.xml
          - KB\_JB837\_1950-01-03\_01-00003.xml
          - KB\_JB837\_1950-01-03\_01-00004.xml
          - KB\_JB837\_1950-01-03\_01-00005.xml
          - KB\_JB837\_1950-01-03\_01-00006.xml
        - jpg
          - KB\_JB837\_1950-01-03\_01-00001.jpg
          - KB\_JB837\_1950-01-03\_01-00002.jpg
          - KB\_JB837\_1950-01-03\_01-00003.jpg
          - KB\_JB837\_1950-01-03\_01-00004.jpg
          - KB\_JB837\_1950-01-03\_01-00005.jpg
          - KB\_JB837\_1950-01-03\_01-00006.jpg
        - KB\_JB837\_1950-01-03\_01-mets.xml
        - KB\_JB837\_1950-01-03\_01.pdf
      - pdf
        - KB\_JB837\_1950-01-03\_01-00001.pdf

**1** Royal Library of Belgium us  
Non-Royal Library of Belgi  
address. Start by clicking t  
Royal Library of Belgium us  
clicking "Request a Drop-o

**?** [How secure is DATA-KBR?](#)

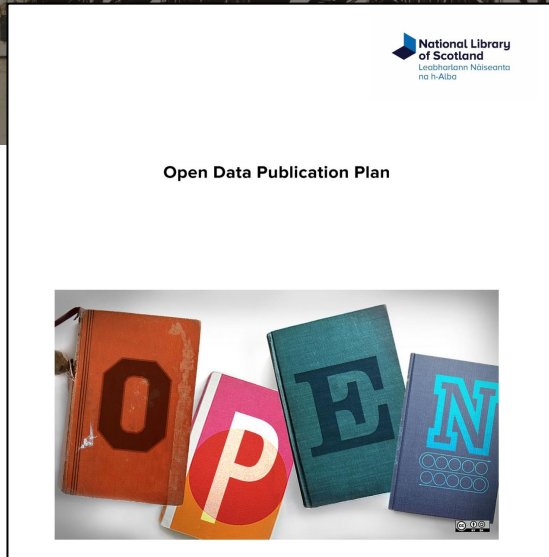
# National Library of Scotland: Data Foundry



Ames, S., & Lewis, S. (2020). Disrupting the library: Digital scholarship and Big Data at the National Library of Scotland. *Big Data & Society*.

<https://doi.org/10.1177/2053951720970576>

<https://data.nls.uk>



## National Library of Scotland data



Open

The National Library of Scotland publishes data openly and in re-useable formats.



Transparent

We take the provenance of our data seriously, and are open about how and why it has been produced.



Practical

We present datasets in a variety of file formats to ensure that they are as accessible as possible.

<https://data.nls.uk/download/national-library-of-scotland-open-data-publication-plan.pdf>



# Historical Data Analysis using Jupyter Notebooks

## NLP Notebooks for Newspaper Collections

*A collection of notebooks for Natural Language Processing*

The following notebooks are aimed particularly at digital humanities scholars who use newspapers as a source. The focus lies on (topic-specific) collection building, a field that is becoming increasingly interesting with better article separation. Very specific research problems are addressed, such as building up collections with ambiguous keywords or working with certain genres. In order to best meet the needs of digital humanities scholars, NLP methods are adapted in new ways, the output is human-readable and the processed newspaper articles can be exported in the form of the original file. In addition, the notebooks allow the users to control the single steps and to choose what is best for their collection. While the NewsEye demonstrator offers the possibility to create datasets quickly and effectively, these notebooks offer possibilities to work on these collections according to specific questions.

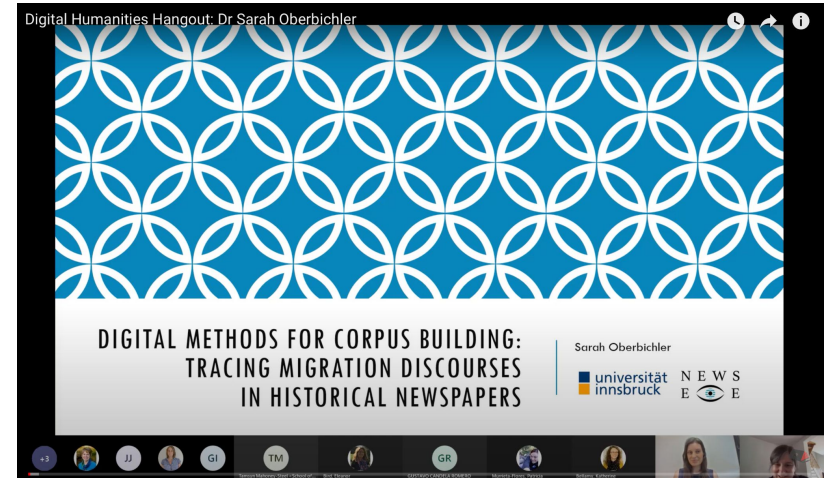
1. [Text classification for topic-specific newspaper collections](#)
2. [Group similar newspaper articles](#)
3. [Discover a newspaper collection with diachronic Ngram clouds](#)
4. [Discourse in Spanish flu coverage](#)
5. [Topic Modeling and Uses of the Term Telegraph in the Context of Journalism](#)

Oberbichler, S. (2020) NLP Notebooks for Newspaper Collections.

<https://github.com/NewsEye/NLP-Notebooks-Newspaper-Collections>

Oberbichler, S. (2021) Digital Methods for Corpus Building: Tracing Migration Discourses in Historical Newspapers, DH

Hangout (Lancaster, Ghent, Lisbon), June 2021: <https://www.youtube.com/watch?v=5w8X0qXP67M>



# Thanks for listening!

**Julie Birkholz & Sally Chambers**

**Julie.Birkholz@kbr.be | Sally.Chambers@kbr.be**

**KBR, Royal Library of Belgium and  
Ghent Centre for Digital Humanities**

