



# Historical Newspaper Content Mining: findings from the impresso project



# SPEAKERS



**Estelle Bunout**  
Research associate, Leibniz Centre for  
Contemporary History Potsdam



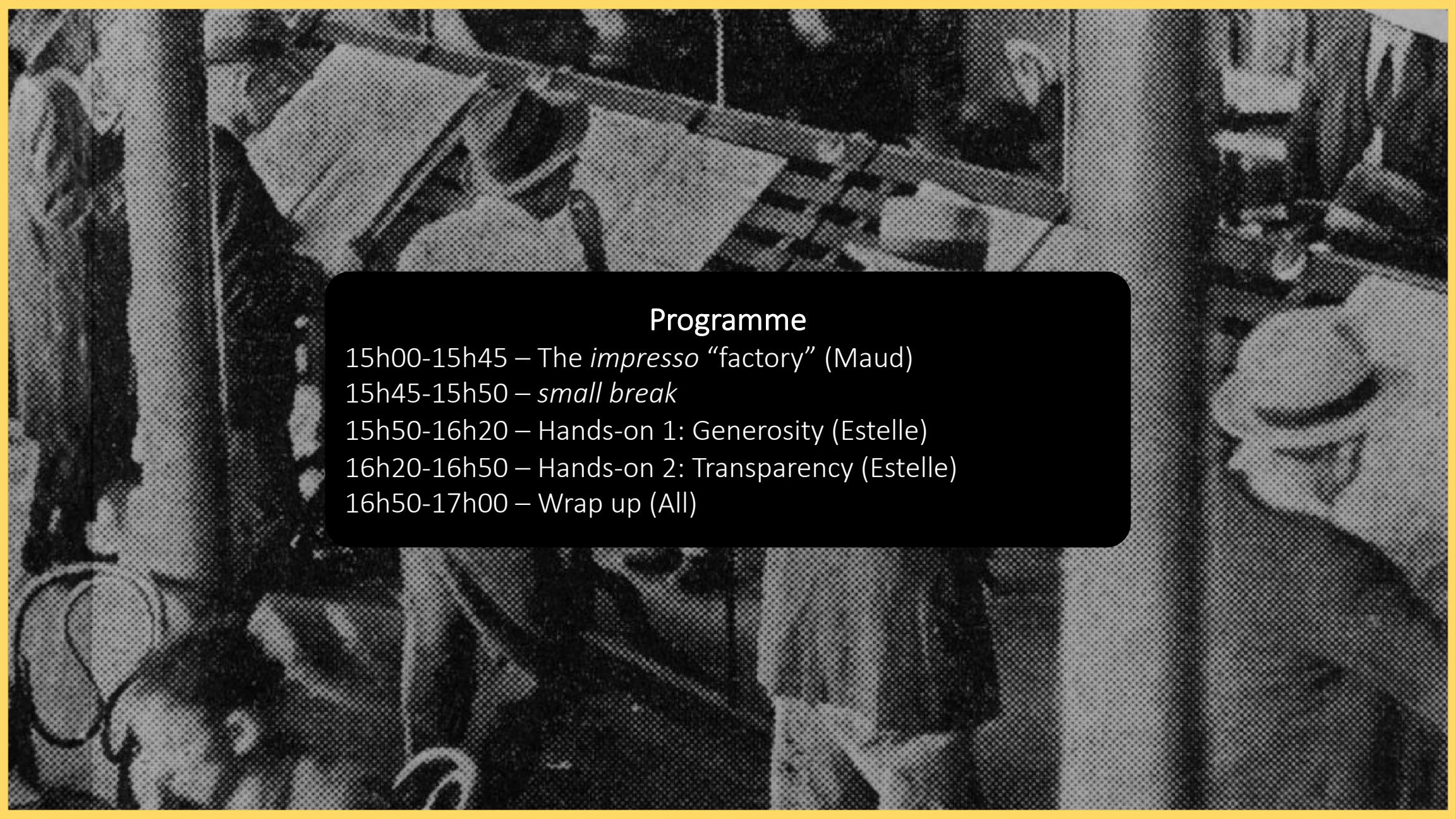
**Maud Ehrmann**  
Research scientist, École Polytechnique Fédérale  
de Lausanne (EPFL)

# NOTES

- **Please turn off your camera and microphone.** You can unmute your microphone when asked.
- **Questions?** Put them in the chat box. We'll put questions to the speakers during the Q&A time and the hands-on sessions.
- **The workshop is being recorded.** All participants will receive a link to the recording shortly.
- **Slides will be uploaded on Zenodo after the workshop**  
All participants will receive a link to the slides shortly.

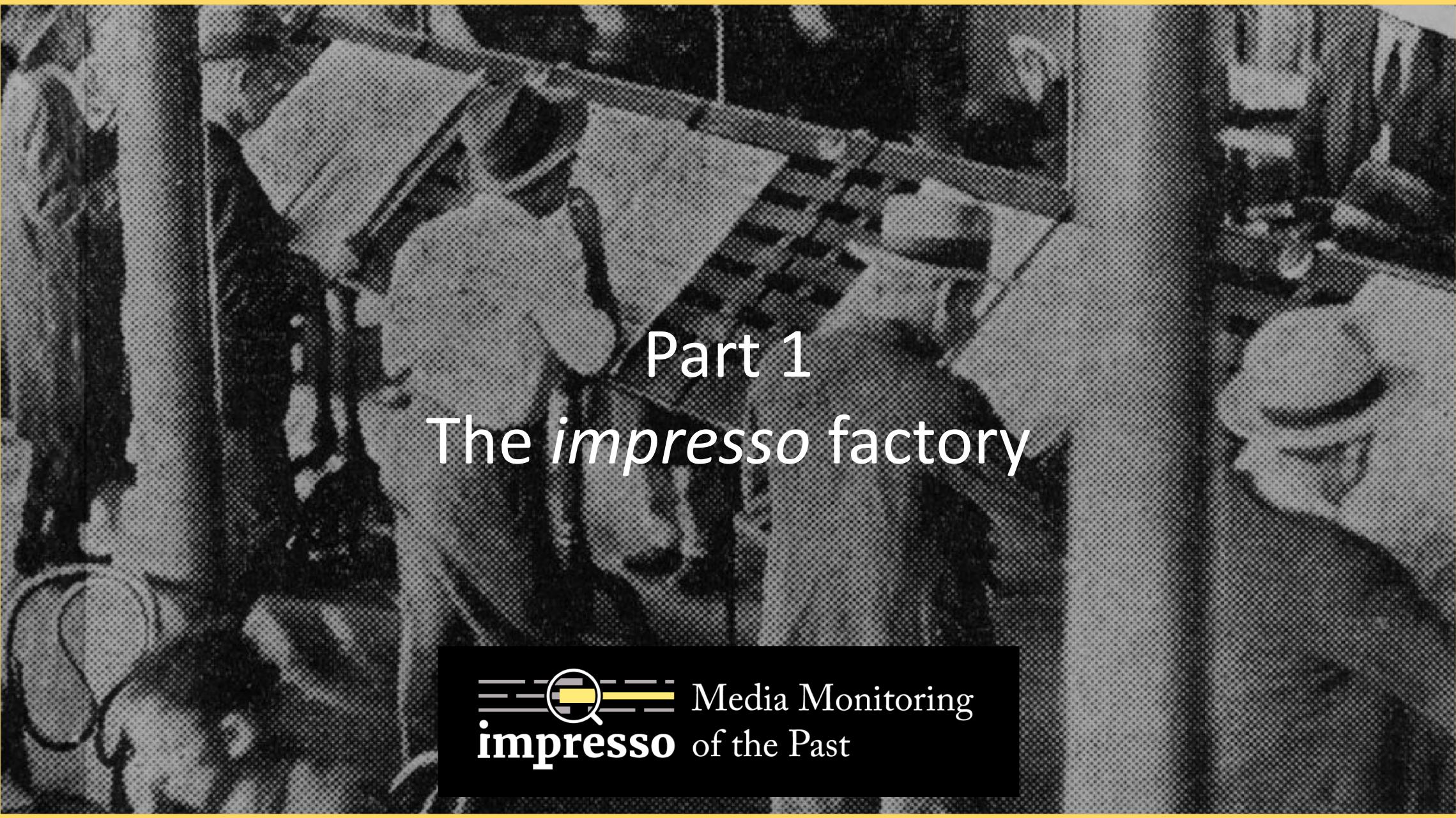


Estelle Bunout, Maud Ehrmann & the *impresso* team  
Workshop AI4LAM - 31.02.2021



## Programme

- 15h00-15h45 – The *impresso* “factory” (Maud)
- 15h45-15h50 – *small break*
- 15h50-16h20 – Hands-on 1: Generosity (Estelle)
- 16h20-16h50 – Hands-on 2: Transparency (Estelle)
- 16h50-17h00 – Wrap up (All)



# Part 1

## The *impresso* factory



Media Monitoring  
**impresso** of the Past

# NEWSPAPERS







British Library National Newspaper Archive - Photo by Christopher Furlong / Getty images

**COLORADO HISTORIC  
NEWSPAPERS COLLECTION**  
Experience Colorado as It Happened

[Help / Register / Log In](#)

[Search](#) [Browse](#) [Topics in History](#) [Help forums](#)

[Support Online](#)

[Search the Collection](#)

[About Our Collection](#)

A service of the Colorado State Library, the Colorado Historic Newspapers Collection (CHNC) currently includes more than 1,000 historical newspaper titles published in Colorado primarily from 1859 to 1923. Due to copyright restrictions, the CHNC does not digitize or make available titles published after 1922, but the CHNC can digitize beyond 1922 if publication date is known.

[Organic support for maintaining and providing access to the CHNC is paid for with money recommended by the Colorado State Library. We continue to add new pages to the CHNC as community funding is needed to pay the costs of digitization.](#)

[Learn more about CHNC](#)

[Browse by...](#)

[County](#)

[Top Journeymen Editors](#)

User name: [Line comedien](#)

1. Judd 102,112  
2. desdemona 260,992  
3. neck 180,433  
4. kate 136,539  
5. RatzenQuack 102,067

[Learn more about correcting text](#)

[Topics in History](#)

The Topics in History series on the Colorado Virtual Library site highlights interesting topics from our past presented through newspaper articles. These topics are some of the historical events that helped shape our state and read the newspaper articles from the time that brought these events to light. The CHNC is a great resource for Colorado.

[Topics in History: Baseball](#)

Topics in History: Baseball  
The Colorado Rockies, a professional major league baseball team, played their first home game in front...

[Topics in History: Nikola Tesla](#)

Nikola Tesla was born on July 10, 1856 in the Austrian Empire in what is...

[Learn more about CHNC](#)

[Visit other newspaper collections](#)

[CHNC Help and user forums](#)

[Topics in History: Past](#)

Listen to readings of selected CHNC articles and hear how people expressed themselves in early Colorado.

[Rocky Mountain National Park - the Birth of a National Park](#)

**Bibliothèque nationale de Luxembourg**  
Kulturmilieus

[Accès](#) [A propos](#) [Contact](#) [Newsletter](#) [Contact](#)

[Chercher un article](#) [Recherche](#)

[Rechercher par date](#) [Rechercher par titre](#)

[Par l'auteur](#)

[Trois journaux du Mouvement écologique en ligne](#)

26.10.2018 Trois journaux supplémentaires, publiés par le Mouvement écologique, sont en ligne à partir d'aujourd'hui sur le site de la BNL. Ils sont intitulés "Le Peuple vert", "Le Génératif" et "Le Kiosque".

[De Kiosque info](#) [De Génératif info](#)

[movement écologique](#)

[A la Une](#)

[Pardate](#)

[Par l'auteur](#)

[movement écologique](#)

[Zukunftsgestaltung zum Land a Riffkin-Prozess:](#)

Nach e wäite Wee:  
grondätzlich Fro si nach net  
ungrosschämt-nin!

**LE TEMPS**

[RECHERCHE](#) [RECHERCHE AVANÇÉE](#) [AIDE](#) [À PROPOS](#)

**Rechercher dans les archives de la Gazette de Lausanne et du Journal de Genève**

Recherche

Ex: Rumus, Confédération helvétique, taillerie, gfb.22.02.1969

Nos partenaires

**EPL**  
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

**Université de Fribourg**  
UNIVERSITÉ DE FRIBOURG  
Centre Universitaire de la Presse et des Médias

**LIBRARY OF CONGRESS**  
The Library of Congress is Chronicing America

**Humanities** **CHRONICLING AMERICA** **HISTORIC AMERICAN NEWSPAPERS**

[Search Pages](#) [Advanced Search](#)

All issues [1799](#) [1803](#) [1807](#) [1811](#) [1815](#) [1819](#) [1823](#) [1827](#) [1831](#) [1835](#) [1839](#) [1843](#) [1847](#) [1851](#) [1855](#) [1859](#) [1863](#) [1867](#) [1871](#) [1875](#) [1879](#) [1883](#) [1887](#) [1891](#) [1895](#) [1899](#) [1903](#) [1907](#) [1911](#) [1915](#) [1919](#) [1923](#) [1927](#) [1931](#) [1935](#) [1939](#) [1943](#) [1947](#) [1951](#) [1955](#) [1959](#) [1963](#) [1967](#) [1971](#) [1975](#) [1979](#) [1983](#) [1987](#) [1991](#) [1995](#) [1999](#) [2003](#) [2007](#) [2011](#) [2015](#) [2019](#)

More Resources

- Historical Digital Newspaper Program
- NWDI Award Recipients
- Newspaper and Current Periodicals Reading Room
- MLC Newsletter & Current Periodicals Reading Room
- Historic Newspapers on Flickr (part of the LC Flickr Commons project)
- Revolvy Words (Help find pictures in historic newspapers!)

100 Years Ago Today: 1.07.1918 (104 issues)

[Austin Daily Statesman](#) (1100p.)  
[Austin, Tex.](#)

[Daily Herald \(1100p.\)](#)  
[Albion, N.Y.](#)

[Daily Mirror \(1100p.\)](#)  
[London, Eng.](#)

[Daily Worker \(1100p.\)](#)  
[New York, N.Y.](#)

**Oesterreichische Nationalbibliothek**

**ANNO Historische Zeitungen und Zeitschriften**

**SPORTSALON** **SCHEIDERL DIARUM CECHE** **NS-FÖRDER**

**Reichspost** **Gierer Volksblatt** **CHARICATUREN** **Volksblatt**

**ANNO – Austria! Newspapers Online**  
Historische österreichische Zeitungen und Zeitschriften online

[Listen](#) [Suchen](#)

[Alphabetische Liste der Zeitungen und Zeitschriften](#) [Suche: Volkstümliche Zeitungen \(1869-1947\)](#)

[Anno wirdzeitliche Beiträge finden Sie auch unter "ANNOhistorisch" Suchfunktion](#)

[15 Jahre ANNO](#)

[Was bietet ZEFTYS?](#)

[ZEFTYS Zeitungsinformationsystem](#)

[Das Zeitungsinformationsystem ZEFTYS bietet den Zugang zu digitalisierten historischen Zeitungen \(Überliegung Image-Digitalisiertes\), zum Teil aber auch \(digitale\), ermöglicht eine bibliographische Recherche zu Zeitungsbeständen in deutschen Bibliotheken und weist weltweit kostenfrei verfügbare Internetressourcen auf. "Zeitung" nach: Bitte beachten Sie, dass die thematische Recherche nicht nur die Bestände der St. Pölten Universitätsbibliothek umfasst, sondern ebenso Einrichtungen aus dem deutschsprachigen Raum Österreich, Südtirol, Schweiz und Liechtenstein nachweist, die nicht im Bestand der St. Pölten Universitätsbibliothek zu Berlin, und](#)

[Was bietet ZEFTYS?](#)

[ZEFTYS weist derzeit insgesamt 276.015 Ausgaben von 183 historischen Zeitungen aus](#)

[In der Amtszeit Preußens](#) [ZEFTYS drei sonst recht seltsame und schwer zugängliche Titel der Bismarckianen Regierungspresse als Faksimile und rechenbararen Volltext. Sie können auch die Kurzadressen: <http://abb.berlin/ampsprese> nutzen \(QR-Code\).](#)

[In der DDR-Presse](#) [zeigt die Relation eines von der Deutschen Forschungsgemeinschaft \(DFG\) finanzierten Projekts drei Zeitungen der DDR-Presse und Volltext erhältlich. Sie können auch die Kurzadressen: <http://abb.berlin/dppresse> nutzen \(QR-Code\).](#)

[Ausgesuchte und intellektuell eindrückliche Internetressourcen](#) [lassen Sie leichter Datenbanken und Information zu Zeitungen im Web finden.](#)

[Die Bibliographische Recherche](#) [lässt Sie Zeitungstitel in der weltweit größten Periodika-Datenbank finden, der Zeitschriftenthesendatenbank \(ZDB\) >, finden.](#)

**Staatsbibliothek zu Berlin** **Prediger Kulturbezirk**

**ANNO** **ZEFTYS** **Start**

**ZEFTYS > Start**

**Start**

[Impressum](#) [Digitale Zeitungstitel](#) [Jahresbericht](#) [Erscheinungsänder](#) [Annoncen](#) [Presse](#) [ODS-Presse](#) [Interessensressourcen](#) [Informationen zu Zeitungen](#) [Über uns](#)

[englisch](#)

[Suchen](#) [ANNO-Suche: Volkstümliche Zeitungen \(1869-1947\)](#)

[Was hat Kreis?](#)

[ZEFTYS weist derzeit insgesamt 276.015 Ausgaben von 183 historischen Zeitungen aus](#)

[In der Amtszeit Preußens](#) [ZEFTYS drei sonst recht seltsame und schwer zugängliche Titel der Bismarckianen Regierungspresse als Faksimile und rechenbararen Volltext. Sie können auch die Kurzadressen: <http://abb.berlin/ampsprese> nutzen \(QR-Code\).](#)

[In der DDR-Presse](#) [zeigt die Relation eines von der Deutschen Forschungsgemeinschaft \(DFG\) finanzierten Projekts drei Zeitungen der DDR-Presse und Volltext erhältlich. Sie können auch die Kurzadressen: <http://abb.berlin/dppresse> nutzen \(QR-Code\).](#)

[Ausgesuchte und intellektuell eindrückliche Internetressourcen](#) [lassen Sie leichter Datenbanken und Information zu Zeitungen im Web finden.](#)

[Die Bibliographische Recherche](#) [lässt Sie Zeitungstitel in der weltweit größten Periodika-Datenbank finden, der Zeitschriftenthesendatenbank \(ZDB\) >, finden.](#)

**POLONA**

Items Search... [Search in content](#)

Sign in

available online [with text recognition](#)

Category [Article](#) [Image](#) [Press and Photo](#) [Book](#) [Manuscript](#) [Card](#)

Author [\(0\)](#) [Author](#) [Title](#) [Subject](#) [Text](#) [Language](#) [Time](#) [Keywords](#) [Genre](#) [Frequency](#) [Form and type](#) [Editor](#) [Publisher](#) [Public place](#) [Subject time](#) [Audience group](#) [Copyright](#) [Source](#) [Project](#)

Industria Materiałowa : polskie i międzynarodowe techniczne redagowane przy współpracy z polskimi i zagranicznymi instytucjami

Mechanika Teoretyczna I Błoszinek, E. 2 (1965), s. 2/3

1957

Liseer Kreisblatt, 1890, Nr. 100 (17 December)

1883-1957

Dziennik Narodowy, B. 5, nr 241 (15 September 1945)

1841-1948

Origine Italica 7 Janvier 1969

**E-NEWSPAPER ARCHIVES.CH** [remplace Presse aktuelle en ligne](#)

Faites une recherche dans la collection

CE QUI EST PASSE LE 7 JANVIER 1969

SUR LA COLLECTION

e-newspaperarchives.ch est le site des journaux suisses et internationaux à la bibliothèque nationale suisse et ses partenaires. Plus d'informations ...

Cette collection de journaux contient 102 titres pour un total de 982.446 éditions, 3.744.540 pages et 20.370.396 articles.

[ETENDUE DE LA COLLECTION](#)

[Carte interactive de la collection](#)

**DIGITAL LIBRARY of GEORGIA**

**GEORGIA HISTORIC NEWSPAPERS**

[HOME](#) [SEARCH](#) [REGIONS](#) [BROWSE](#)

[NEWS](#) [ABOUT](#) [HELP](#) [PARTICIPATE](#)

Search newspaper contents

Recently Added Titles

- 2018-10-24 - The monitor. (1860 - 1875)
- 2018-10-24 - Upon pilot. (1868 - 1869)
- 2018-10-24 - Georgia republican & state newspaper. (1863 - 1865)
- 2018-10-24 - The Georgia evening journal. (1852 - 1853)

Interactive map of Georgia

Click on a region to browse and search the newspaper pages

**The European Library** [Discover](#) [Access our Data](#)

[About](#) [Membership](#) [For Current Partners](#) [Log in](#) [English \(en\)](#) [Advanced search](#)

[Search all European Library content...](#)

[Newspapers Home](#)

The TEL portal has been frozen since 31 December 2015. Over the course of 2017, we will be working to migrate all data and functionalities of European Library content to the new platform. In the meantime, the service may be unresponsive at times and errors may occur; we apologize for any inconvenience. For any problems, please contact us on Twitter and Facebook!

[http://www.europeanlibrary.org/europeannewspapers](#) [http://www.europeanlibrary.org/europeannewspapers](#) [http://www.europeanlibrary.org/europeannewspapers](#)

Search Newspapers

Search within historical newspapers...

Filter by library, date or language

On This Day in History

Show all newspapers titles

Show all newspapers issues

Explore newspapers by source

Explore newspapers by date

Explore newspapers by title

FREQUENTLY ASKED QUESTIONS

**Delfpher**

Doceo also [Delfpher](#) [Delfpher](#)

Ruim 90 miljoen pagina's uit Nederlandse kranten, boeken en tijdschriften

Handeling

Twitter [Newsfeed](#)

**Trove** [Digitized newspapers and more](#)

Search articles for [Keywords](#) [Q](#) [Adv](#)

OR

Browse articles by [Title](#) [Place](#) [Date](#)

Category [Tag](#)

View a lot of all newspaper and gazette titles

Top text correctors

|                   |           |
|-------------------|-----------|
| 1. Jahrfehler     | 6,274,464 |
| 2. neuerwörth     | 5,663,918 |
| 3. neuerwörth     | 3,300,468 |
| 4. MuseumWertheim | 3,171,036 |
| 5. Domfehler      | 2,409,025 |
| 6. neuerwörth     | 2,002,475 |

In this day

Wednesday 28 October 1968

The Rhine Herald (Rhine, Va., Marion, Mo. 1968-1969, 1969-1970)

# Help me

... “expand my query based on common OCR mistakes and spelling variants” + multilingual synonyms



Eegierung  
Begierung  
Regieruug  
Kegierung  
Regiexung

similar words  
+ edit distance

Behorden  
Finanzdirektion  
Staatsrath  
Eegierungen  
Behörde  
Reichregierung  
Staatsrat  
Staatsrathes

similar words

→ lexical processing

# Help me

# editorial

... “find the frequency of word X in  
articles of certain type or thematic”

... “distinguish ‘real’ occurrences from false positives emerging from ads, stock market or meteo pages”

→ segmentation (CV) and classification (NLP)

The image shows the top portion of a historical newspaper page from 'Gazette de Lausanne'. The title 'Gazette de Lausanne' is written in large, bold, black letters at the top center. Below it, in a slightly smaller font, is 'ET JOURNAL SUISSE'. At the very top, the word 'title banner and 'ears'' is written in a large, stylized, handwritten font. On the left side, there is a vertical column of text containing the date 'Vingt-neuvième année' and the address '8, Rue Pépinière, à Lausanne'. On the right side, there is more text including 'LIBERTÉ ET PATRIE', 'N° 375 - Vendredi 26 novembre 1926', and 'ANNONCES: PUBLICITAS'. The overall layout is characteristic of early 20th-century newspaper design.

# culture

# local news

**La réforme des naturalisations**

Il y a quatre ans que le parlement a voté la loi sur les naturalisations. Mais celle-ci n'a pas été appliquée dans toute sa rigueur. Il est donc nécessaire de faire quelques modifications pour assurer l'équité et la sécurité de nos frontières.

La belle œuvre d'un peintre vaudois  
**La chapelle du Collège  
de l'Abbaye de St-Maurice**  
(De notre envoyé spécial)

**Lettre de Berlin**

**Le contrôle militaire de l'Allemagne**

Comment passeront-les des art. 202-208 à l'art. 213 du traité?

apelle la messe pontificale. Avant d'entrer dans la messe pontificale, Mgr M., adresses à son auditoire, forme ensuite des étudiants du collège, une allocution dans laquelle, il renseigne, avec des mots qui font honneur à

mission du Conseil des Etats et après celle de l'Assemblée nationale, il fut nommé à diverses étapes. Un certain naisseur fortuitement à l'étranger, bien que son pareille eussent leur domicile

Dès lors, dès là... le caustique des morts en lèvent le sujet ; et vers le plus expressif des trois qui ferment les vingt-deux strophes de ce poème est arrivé avec chacun des panneaux qui l'illustraient. C'est là, partie nettement séparée de la dé-

sons. Ses chaque station du chemin de fer, il a peint un brancard. Beurie, l'assassin de la jeune tacheuse, est mort sur le fond brûlé des murs. La mort du plaidoir, d'un vénoréaste qui

Paris fut malé, tout spécialement dans la triste affaire du Zapoticon. Mettant ainsi bien en évidence l'absence d'autorité à l'appartement. C'est de moins, maintenant, la ville même qui déclame, la ville même qui déclare que l'ordre succédera à l'anarchie et que l'ordre sera assuré par l'Etat. Ce politico de son prédecesseur, soit placé effectivement malgré toutes les affirmations contraires, au-dessus des derniers rangs commandants les divisions de Reichstag. La presse allemande réplique que l'ordre sera assuré par l'Etat. L'ordre sera assuré par l'Etat. La recherche se tient dans les grottes géantes de Reichstag, sous le regard de son directeur de force : l'engager dans un décret immédiat en utilisant l'assassinat.

Il est donc bon de se pencher sur les deux derniers en résumant et sur l'ensemble de la situation.

frangier. Une des dernières fois qu'il fut constable à feu sous la forme d'une ville alors que le baron de Boulogne et son épouse étaient en visite au château de Fécamp, il fut surpris par l'ordre de l'empereur Louis XIV de faire arrêter l'abbé de Pavaux qui l'avait accusé de complot contre le Roi. Il fut arrêté et emprisonné dans la prison de Rouen où il mourut peu après. Ensuite, il fut transféré à la prison de la Conciergerie à Paris où il mourut également.

Suisse Fr. 2.50  
Etranger „ 4.-

Les aspirants à voler de leurs idées et que leurs préférences visent à la méthode dilatée (pour ne pas dire de la résistance passive) sont d'autant plus nombreux qu'ils se rendent compte que l'effacement de l'ego est le moyen le plus sûr pour atteindre leur but. Cela évidemment, dist il à ses femmes, alors que, pour la seconde fois, il a pu parer une jeune femme à la cause d'égarage. Violante jura à la marche se conçut au contraire. Il fut alors l'assassin de son fils, et, sans l'aider dans la salle communale où Violante lui a propulsé la mort.

gardez les mains entre les siennes. Il y a plus une goutte de sang dans son visage ; son sein se soulève et s'abaisse en une respiration accélérée, ses yeux remplis d'épouvante se fixent dans le vide.

—  
sur François de C. Doutibouze  
—  
la seconde volontiers cette pro-  
position à sa femme, qui répond  
au bout de deux ou trois secondes :  
**N**on, mais je ne signe de vie de  
mort que si l'assassinat est  
dans l'intérêt de la paix et de la  
paix. On t'assure de l'assassinat  
de ma femme, mais je ne signe de vie de  
mort que si l'assassinat est  
dans l'intérêt de la paix et de la  
paix.

— Tu retrouveras nous plus tard, taufs qui se débarrassera de la table. Mais Fini n'est pas partie.

— Qu'est-ce que tu dis, mère ? demanda-t-elle.

C'est une petite créature intelligente, mais un peu trop étroite dans ses idées.

→, distille, de sorte que l'odeur de la pomme de terre éveille les sens de l'adolescent, à la fois curieux et répulsif. Adèle, d'autre part, n'a pas été éduquée au plaisir des sens. Ses mains sont lourdes et fermes comme des maniques, râpeuses et rugueuses, mais elles ont su faire de la magie. C'est pourquoi elle a été nommée ménagère de la maison. Il ne sera donc pas difficile pour elle de préparer un repas délicieux.

Arrive un autre peu à son doigt se rire :  
Elle range bruyamment les verres qui  
entrechoquent, puis elle coupe en deux

# Help me

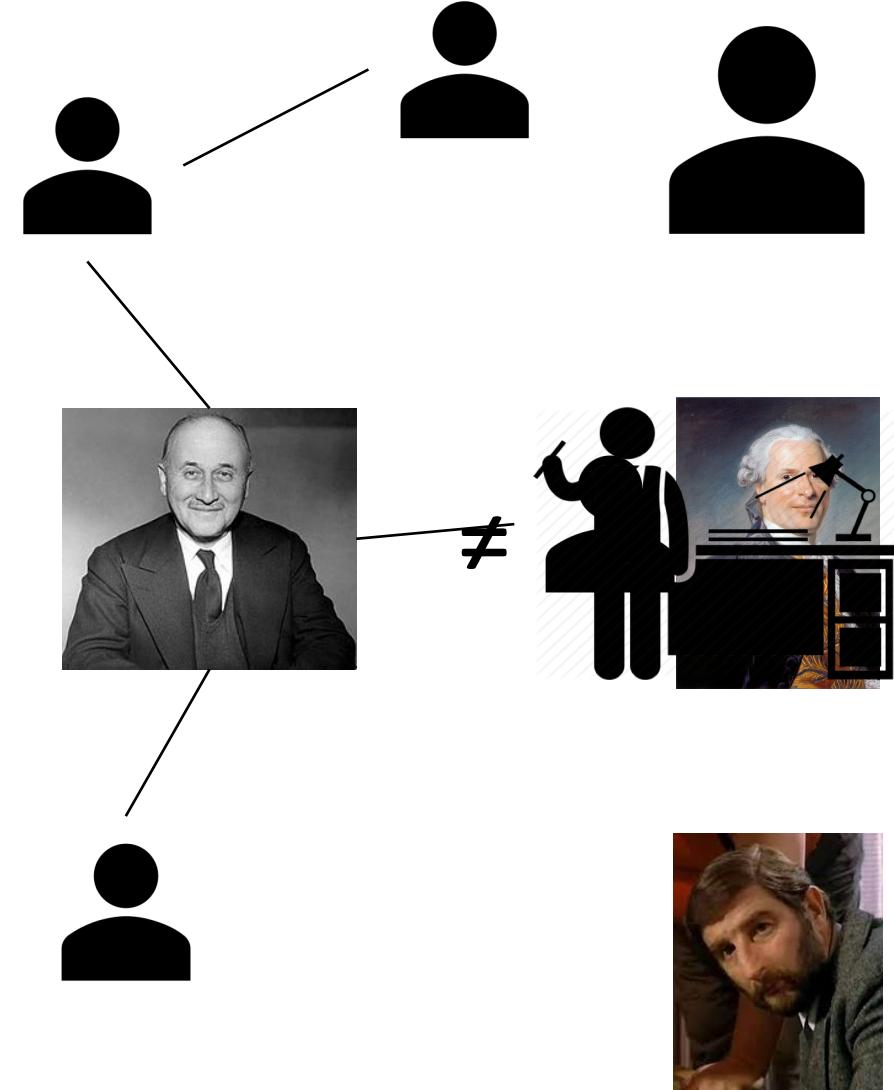
... “ find all articles mentioning Jean Monnet.

And the right one!”

... “ know with whom the name of X is associated.”

... “ find articles mentioning person X as **author** vs. as **main subject** of the article”

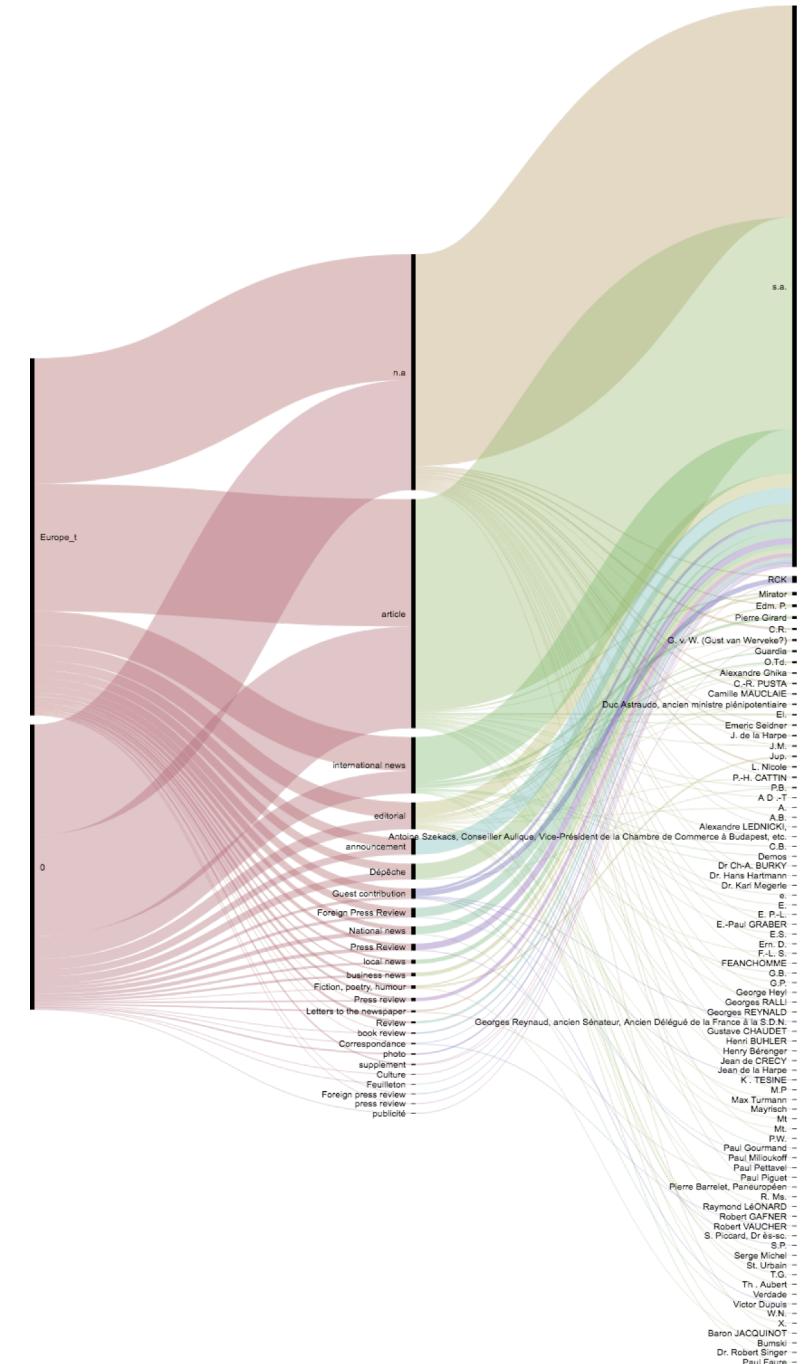
→ named entity processing



# Help me

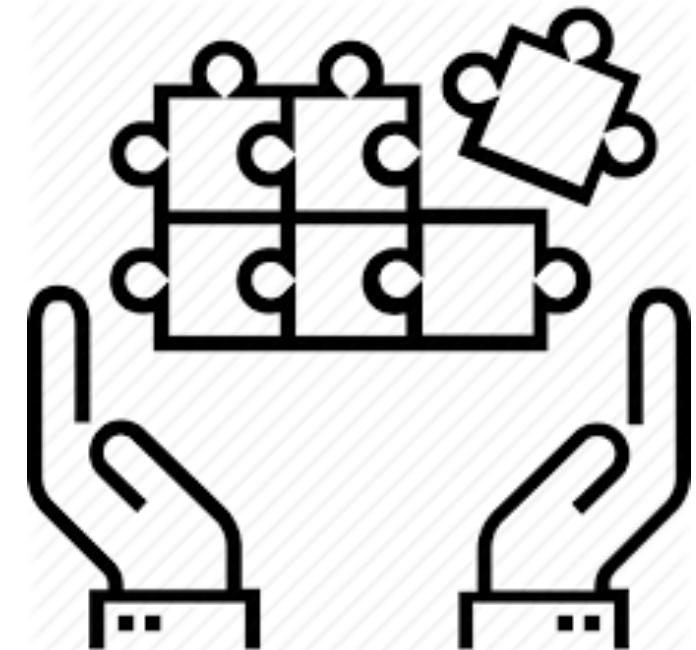
... “ collect and compare sub-collections based on metadata (time, titles, political orientation, etc) but also based on content”

→ interface, metadata



# Help me

- ... “discover similar items I might be interested in”
- ... “appreciate the high-level thematic(s) of the corpus”
- ... “build and interact with my own collection”
- ...



How to enable semantic indexing and  
exploration of large collections of historic  
newspapers?



Media Monitoring  
**impresso** of the Past

**EPFL**

**C<sup>2</sup>DH**



**University of  
Zurich** UZH

**[FNSN**

**Swiss National Science Foundation**



Swiss National Library, SNL



National Library of Luxembourg,  
BnL



State Archives of Valais, AEV.



Swiss Economic Archives, SWA.

**LE TEMPS**

Le Temps

**NZZ**

Neue Zürcher Zeitung, NZZ.

*Unil*  
UNIL | Université de Lausanne

History department, University  
of Lausanne, UNIL.

**infoclio.ch**

infoclio

+ many others institutions



Critical content mining  
of 200 years of  
historical  
newspapers

# The team

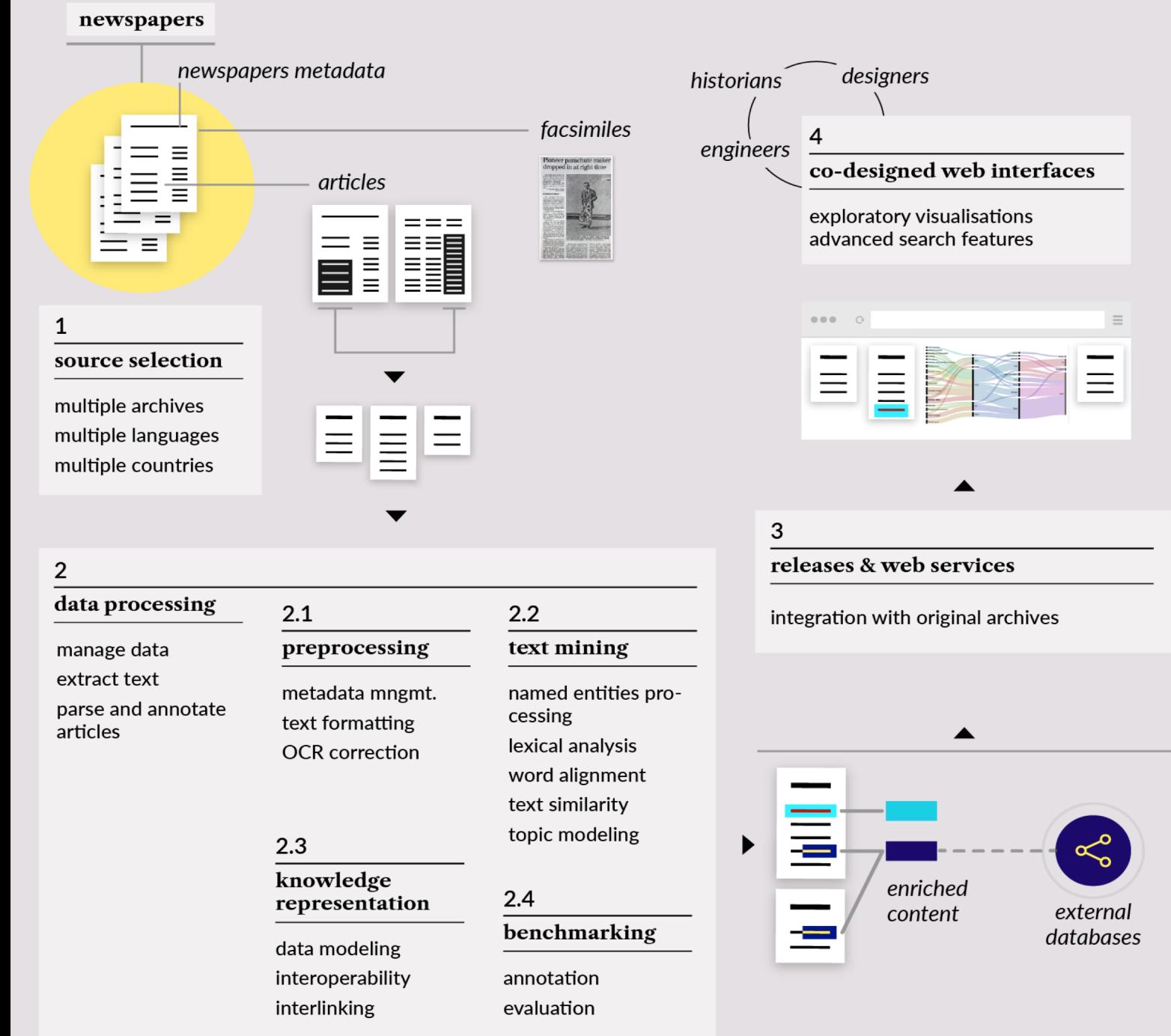
*Estelle Bunout  
Simon Clematide  
Marten Duering  
Maud Ehrmann  
Andreas Fickers  
Daniele Guido  
Frédéric Kaplan  
Peter Makarov  
Matteo Romanello  
Gerold Schneider  
Paul Schroeder  
Benoit Seguin  
Phillip Stroëbel  
Martin Volk  
Thijs van Beek  
Lars Wieneke*



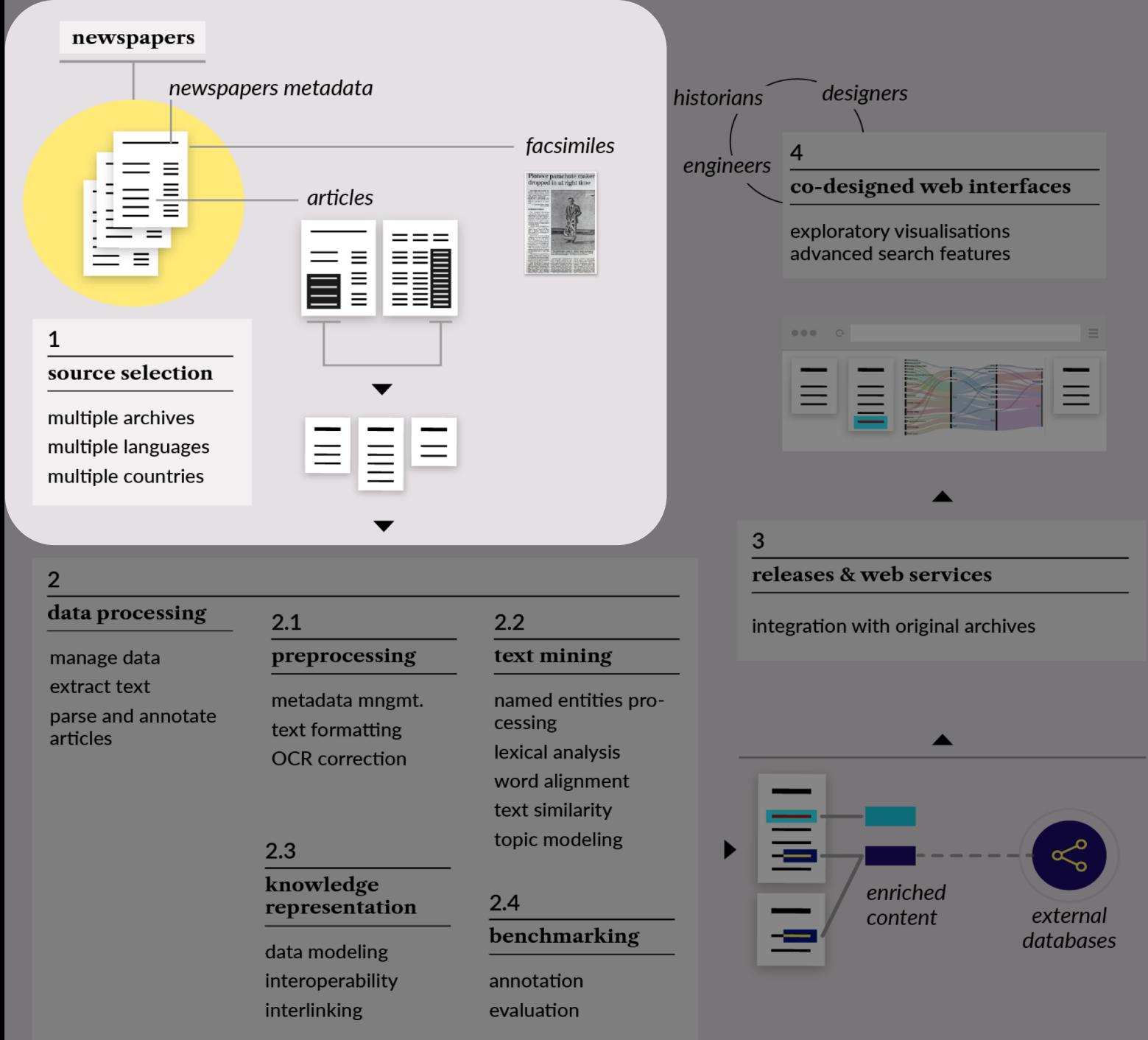
*+ a team of historical advisors  
+ a team of associated historians*



...and our plan:

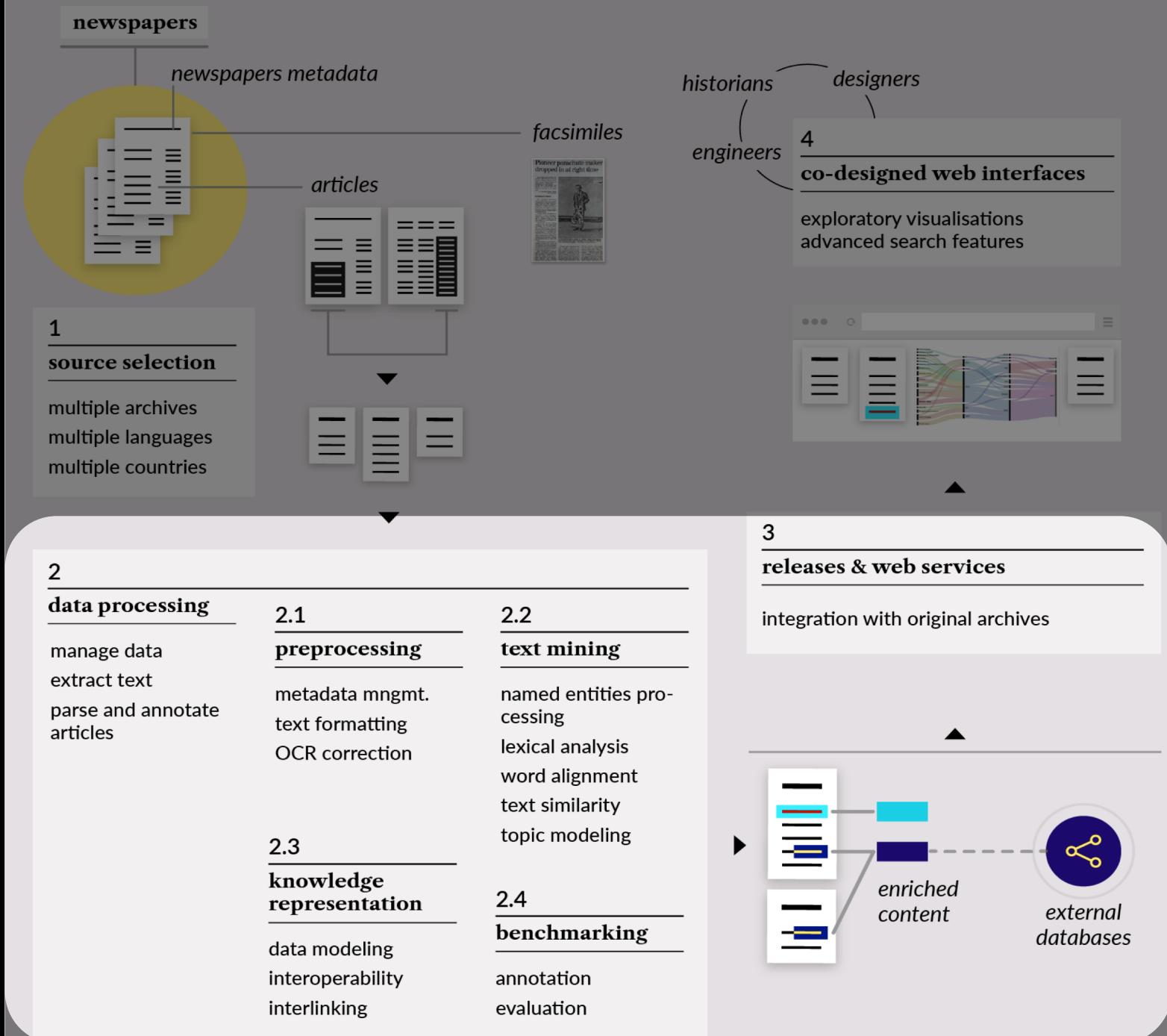


...and our plan:



...and our plan:

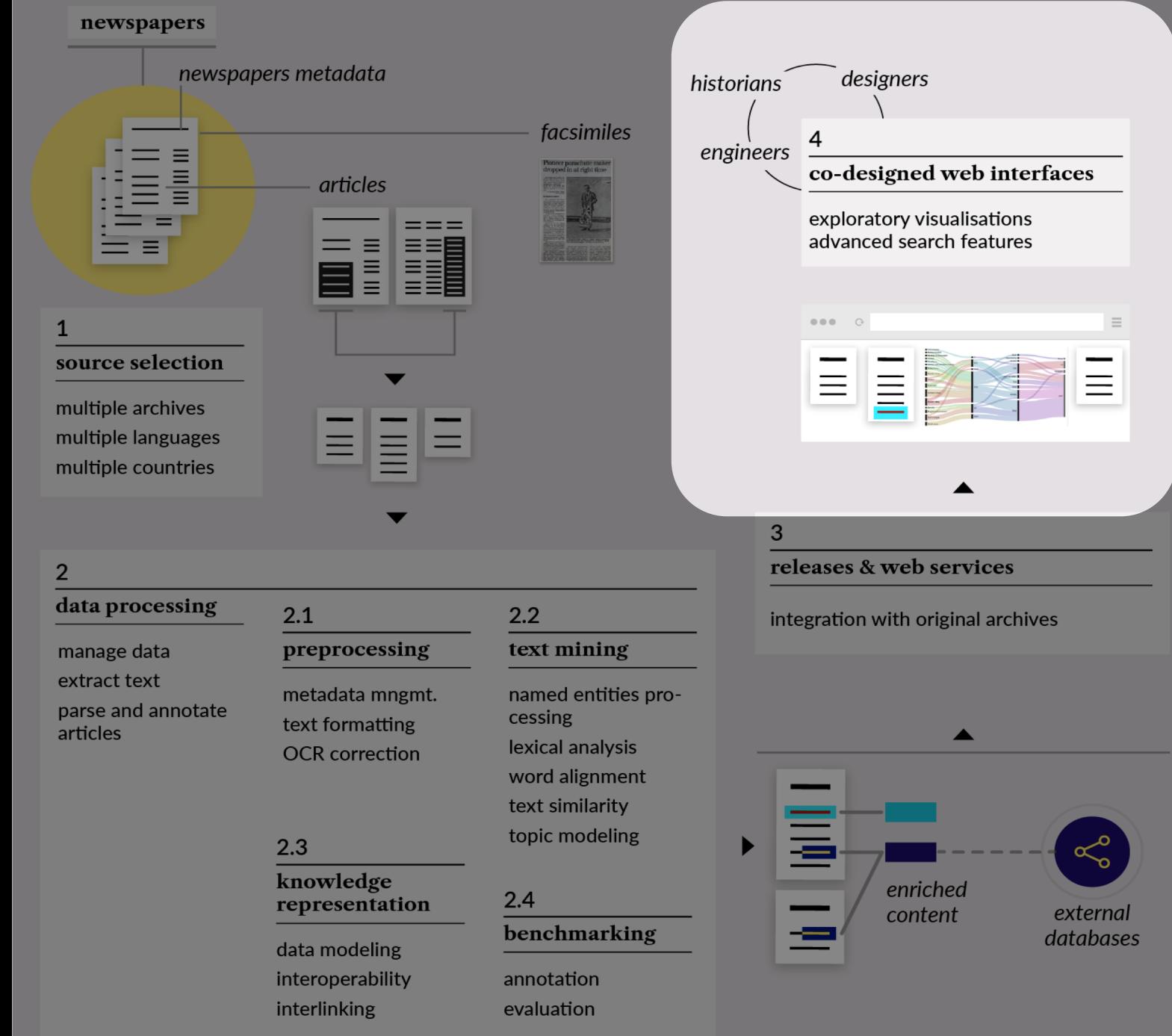
## 1. How to adapt NLP tools to historical texts?



...and our plan:

1. How to adapt NLP tools to historical texts?

2. How to explore complex and vast amounts of data?

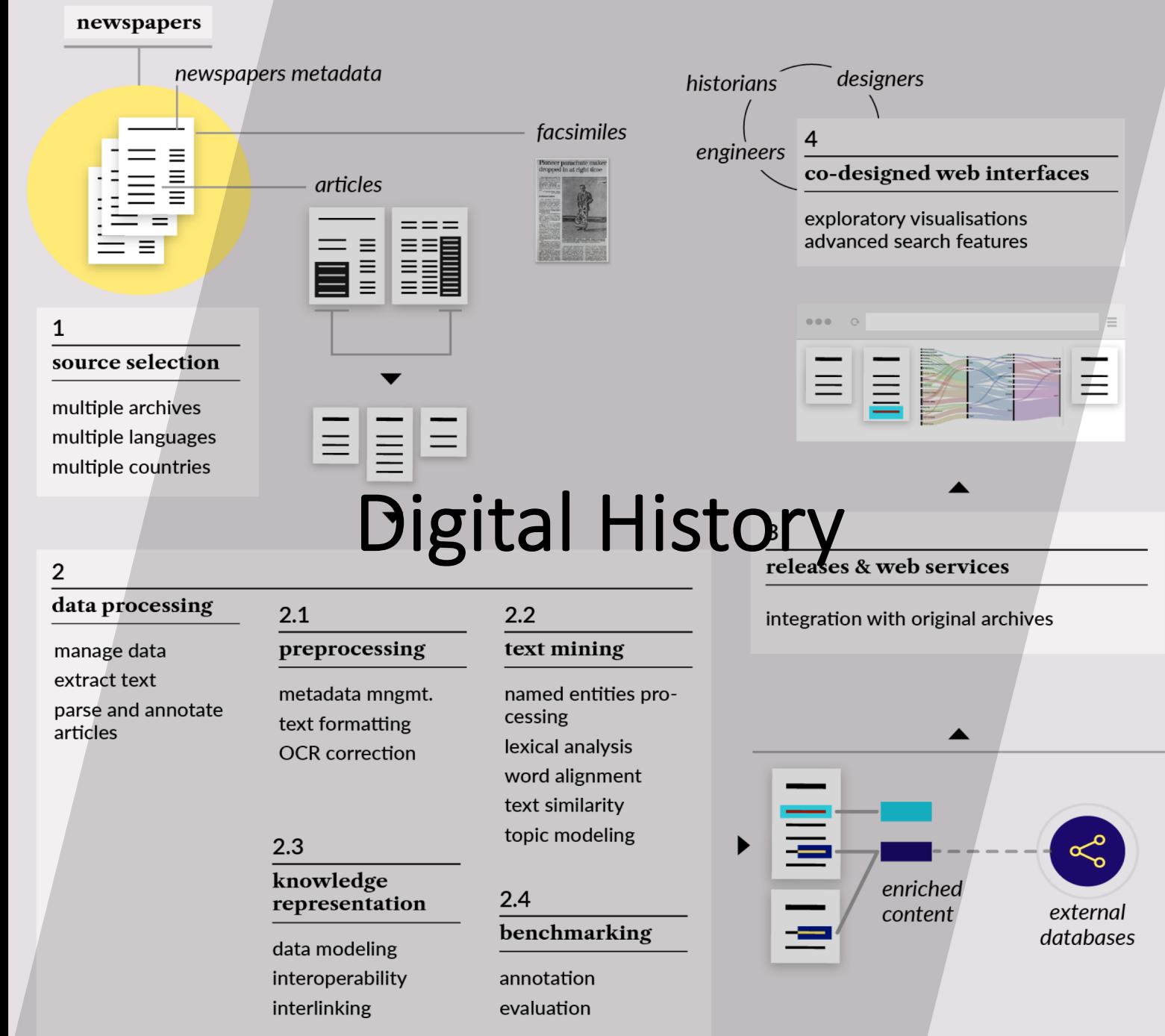


...and our plan:

1. How to adapt NLP tools to historical texts?

2. How to explore complex and vast amounts of data?

3. What is the impact of new tooling on digital scholarship?



# Guiding principles

- Co-design, pull-push
- Experimentation
- Transparency
- Generosity
- Continuum

The screenshot shows the homepage of the Media Monitoring of the Past website. At the top, there's a navigation bar with links for 'Search', 'Newspapers', 'Topics', 'Inspect & Compare', and 'Text reuse'. On the far right are 'FAQ' and 'LOGIN' buttons. The main title 'Media Monitoring of the Past' is prominently displayed in large, white, serif font. Below the title is a search bar with tabs for 'SEARCH ARTICLES', 'SEARCH IMAGES', and 'NGRAMS', and a placeholder 'search for ...'. To the right of the search bar is a button with a magnifying glass icon. A section titled 'IMPRESSO DATA RUNDOWN' provides statistical information: 76 newspapers collected, 600,919 issues, 5,429,656 pages scanned, 47,798,468 content items identified, 3,462,799 images, and 12,493,358,703 words. It also mentions 2 countries of publication and 530,086 named entities disambiguated. Below this, there's a link to the 'blog'. On the left side of the main content area, there's a sidebar with contact information: email (info @ impresso-project [dot] ch), project website ([impresso-project.ch](http://impresso-project.ch)), GitHub ([github: impresso](https://github.com/impresso)), and Twitter ([@impressoproject](https://twitter.com/@impressoproject)). There are also buttons for 'LINES: OFF' and 'DARK MODE: ON'. On the right side, there's a section titled 'Mining 200 years of historical newspapers' with a small icon of a newspaper. Below it, a question 'How can newspapers help understand the past? How to explore them?' is followed by a small icon of a person reading. At the bottom, a note states 'For legal reasons not all content is available in Open Access. To gain full access:' with a link to 'DOWNLOAD NON-DISCLOSURE-AGREEMENT FORM' and a note about returning the signed form to 'info@impresso-project.ch'. In the bottom right corner, there's a small circular icon with a speech bubble and the number '3'.

# keyword suggestion

Media Monitoring  
impresso of its Net

Search... Newspapers Topics Inspect & Compare Text reuse Collections

SEARCH ARTICLES SEARCH IMAGES NGRAMS

GROUP BY ARTICLE ▾

4,081 articles found containing arnhem

arnhem ✓ Contains NOT contains arnhem ADD NEW... FIND SIMILAR...

Enlarge your search! Type one word and obtain a list of surrounding words arnhem French 35

Click on one of the following words to explore your search

arnhem arnhem nimègue zwolle venlo sittard rotterdam tilburg roosendaal eindhoven deventer arnhem tilburg noord groningen doetinchem utrecht veulen moerschoten REMOVE

HIGHLIGHTED TITLES (SEE ALL OPTIONS) check one or more newspaper to filter results

L'Impartial (73 results) ☐ L'Express (63 results) ☐ La Liberté (554 results) ☐ Neue Zürcher Zeitung (540 results) ☐ Journal de Genève (509 results) ☐ Gazette de Lausanne (393 results) ☐ Die Tat (215 results) ☐ Freiburger Nachrichten (199 results) ☐ Le Peuple, La Sentinel (68 results) ☐ L'indépendance luxembourgeoise (53 results) ☐ Luxembourg Wort (53 results) ☐ SMUV-Zeitung (49 results) ☐ Confidérence (40 results) ☐ d'Letzeburger Land (38 results) ☐ Obermosel-Zeitung (36 results) ☐ Escher Tagblatt (12 results) ☐ La tâche syndicale (10 results) ☐ D'Union'n (10 results) ☐ Courrier du Grand-Duché de Luxembourg (7 results) ☐ L'Union'n (7 results) ☐

arnhem nettoyé

Le Peuple, La Sentinelle ⚡ TUESDAY, APRIL 17, 1945 - p. 4 Personal use

Arnhem nettoyé QG allié, 16 avril. On annonce officiellement qu'Arnhem a été entièrement débarrassée de l'ennemi. (Reuter) Arnhem est un important nœud routier et ferroviaire dans une province où la Guerre, dont elle est le chef-lieu, Arnhem, compte 60,000 habitants.

Arnhem nettoyé

Le Peuple, La Sentinelle ⚡ TUESDAY, APRIL 17, 1945 - p. 4 Personal use

Arnhem nettoyé QG allié, 16 avril. On annonce officiellement qu'Arnhem a été entièrement débarrassée de l'ennemi. (Reuter) Arnhem est un important nœud routier et ferroviaire dans une province où la Guerre, dont elle est le chef-lieu, Arnhem, compte 60,000 habitants.

Nouvelle offensive britannique sur Arnhem

Le Peuple, La Sentinelle ⚡ FRIDAY, OCTOBER 6, 1944 - p.8 Personal use

Nouvelle offensive britannique sur Arnhem. Après de la 2 me armée britannique, 5 octobre. (Reuter) Du correspondant spécial Desmond Tighe : Jeudi, le général Dempsey a passé à l'attaque en direction d'Arnhem.

Nouvelle offensive britannique sur Arnhem. Après de la 2 me armée britannique, 5 octobre. (Reuter)

) Du correspondant spécial Desmond Tighe : Jeudi, le général Dempsey a passé à l'attaque en direction d'Arnhem. A trois km. du pont, direction d'Arnhem. A 3 kilomètres du pont, gagée. La nouvelle disant

) Le pont du Lek, près d'Arnhem, était l'un des objectifs principaux des troupes aéroportées. Il est aussi dans ce secteur et d'importantes forces britanniques ont opéré leur jonction au sud d'Arnhem. OBJECTIFS ATTEINTS

QGQ interallié, 6 octobre. (Reuter) L'attaque du général Dempsey, au sud d'Arn d'Arnhem, a atteint

Nouvelle offensive britannique sur Arnhem

Le Peuple, La Sentinelle ⚡ FRIDAY, OCTOBER 6, 1944 - p.8 Personal use

Nouvelle offensive britannique sur Arnhem. Après de la 2 me armée britannique, 5 octobre. (Reuter)

) Du correspondant spécial Desmond Tighe : Jeudi, le général Dempsey a passé à l'attaque en direction d'Arnhem. A trois km. du pont, direction d'Arnhem. A 3 kilomètres du pont, gagée. La nouvelle disant

) Le pont du Lek, près d'Arnhem, était l'un des objectifs principaux des troupes aéroportées. Il est aussi dans ce secteur et d'importantes forces britanniques ont opéré leur jonction au sud d'Arnhem. OBJECTIFS ATTEINTS

QGQ interallié, 6 octobre. (Reuter) L'attaque du général Dempsey, au sud d'Arn d'Arnhem, a atteint

sun LES vous NAVIGABLES pE nom...

La tâche syndicale ⚡ WEDNESDAY, FEBRUARY 8, 1984 - p.4 Personal use

sun LES vous NAVIGABLES pE nom... 23-28 avril' Filia Rheni » Premier jour, lundi: Bâle—Arnhem avec le TEE « Rhelengold » jusqu'à Arnhem. Dîner au wagon-restaurant. Deuxième jour, mardi: Arnhem—Amsterdam

LOCATIONS: Basel ⚡ Dordrecht ⚡

sun LES vous NAVIGABLES pE nom... 23-28 avril' Filia Rheni » Premier jour, lundi: Bâle—Arnhem

With the TEE « Rhelengold » jusqu'à Arnhem. Dîner au wagon-restaurant. Deuxième jour, mardi: Arnhem—Amsterdam dans le port de Rotterdam. Cinquième jour, vendredi: Rotterdam—Arnhem. Nous remontons le Waal, passons à Dordrecht et à Nijmegen, puis sur le Rhin intérieur jusqu'à Arnhem. Sixième jour, samedi: Arnhem—Bâle Retour en train, deuxième classe, jusqu'à Bâle et retour. Train première classe Bâle—Arnhem et retour; y compris

Communication aux porteurs d'obligations de

Gazette de Lausanne ⚡ WEDNESDAY, DECEMBER 3, 1969 - p.2 Personal use

Communication aux porteurs d'obligations de l'emprunt 5 % à 2 x 1967 Algemene Kunststof Unie N. V., Arnhem Nous vous informons

# marginalia

# Named entities

**LA PAIX APPROCHE**

ce la « Paix du Peuple » solennelle

ans les cinq continents, on a  
es les ressources, dépenser des  
iques, galvaniser les énergies,  
tion la science et la technique  
humaines par millions pour org  
e la guerre en vue d'atteindre le  
ctoire.

ura-t-on montrer autant de vol  
ice », de ténacité ; mobiliser éga  
ressources, science, technique ;  
gies ; utiliser bras et cerveaux  
utre but, infiniment plus élevé  
hommes ? Les gouvernements se souviendront-  
es promesses solennelles faites à leurs peui-

**1 A PAIX APPROCHE**  
Sera-ce la « Paix du Peuple » solennellement promise

Jules Humbert-Droz PERSON

1840 1860 1880 1900 1920 1940 1960 1980 2000

apply current search filters (1 filters)

3,811 results in total (from 1834 to 2017)

**ADD AS SEARCH FILTER** **EXCLUDE FROM CURRENT SEARCH**



**Jules Humbert-Droz (Q115791)**  
HUMAN La Chaux-de-Fonds, 1891 - La  
Chaux-de-Fonds, 1971

**W** Swiss communist (1891-1971)  
SOURCE: WIKIDATA ID/Q115791

**MORE...**

cm a su mobiliser  
penser des sommes astro  
er les énergies, mettre à com  
t la technique, sacrifier les  
ns pour organiser et con  
e d'atteindre le but suprême

nt de volonté, d' intel  
é ; mobiliser également riche  
ience, technique ; galvanise  
cerveaux pour atteindre  
plus élevé : le bien-être  
nnements se  
elles faites à leurs peu  
t. Bevin, ministre anglais

du travail, ne proclamait-il, pas, à la conférence  
de la Commission de crise du Bureau international

*How to enable semantic indexing and exploration  
of large collections of historic newspapers?*

5 challenges

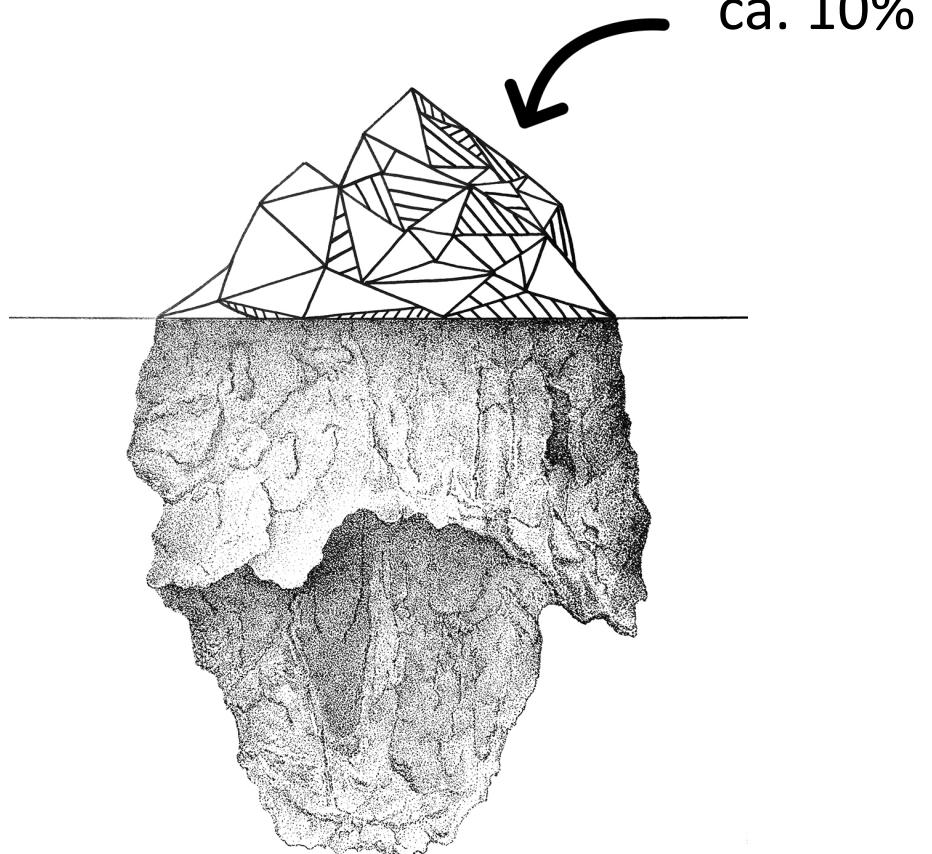
# *How to enable semantic indexing and exploration of large collections of historic newspapers?*

1

Digitized  
newspaper silos

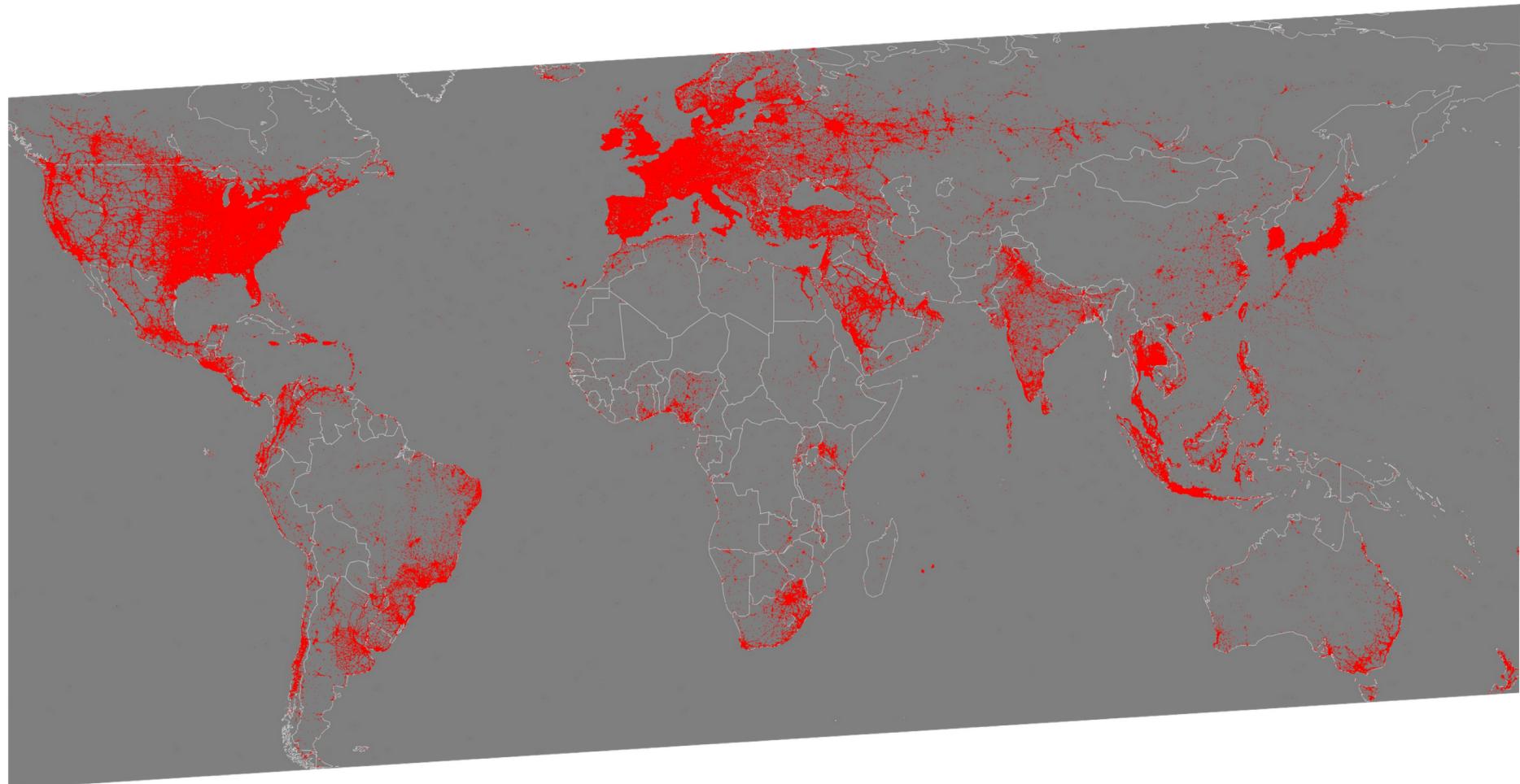
# Shape of the (newspaper) digital landscape

- Partial



# Shape of the (newspaper) digital landscape

- Partial
- Imbalanced



Cf: L. Manovich, *Cultural Data: Possibilities and Limitations of Digitized Archives*, <http://manovich.net/index.php/projects/cultural-data>

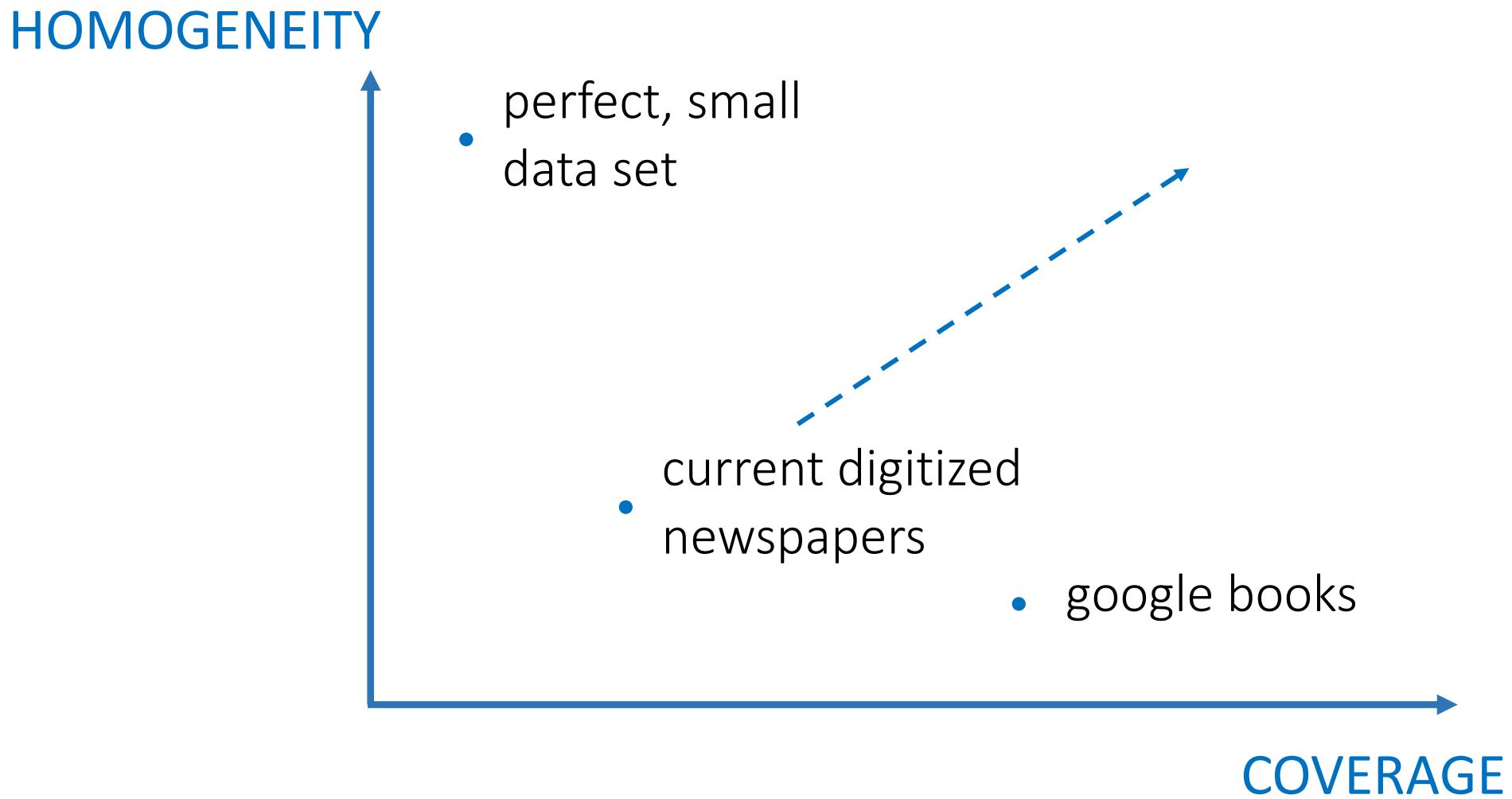
Cf. L. Putnam, *The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast*, *The American Historical Review* (2016)

# Shape of the (newspaper) digital landscape

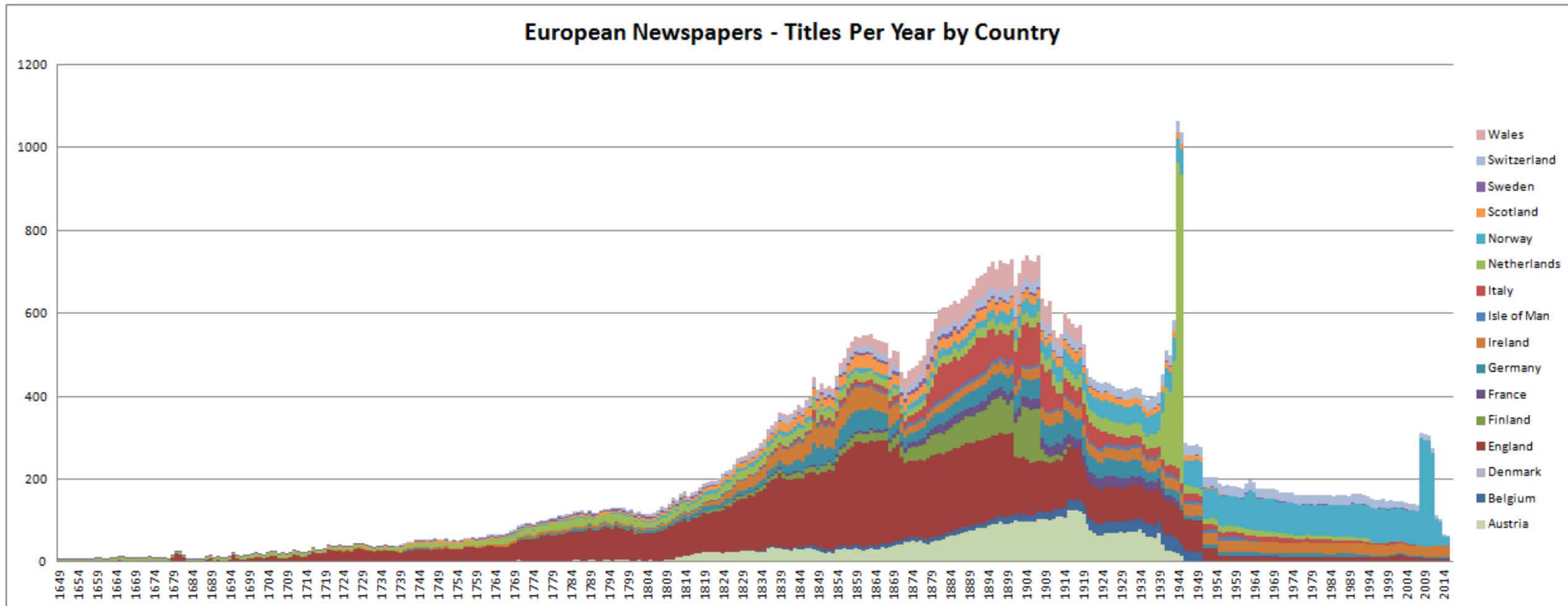
- Partial
- Imbalanced
- Heterogeneous



# Shape of the (newspaper) digital landscape

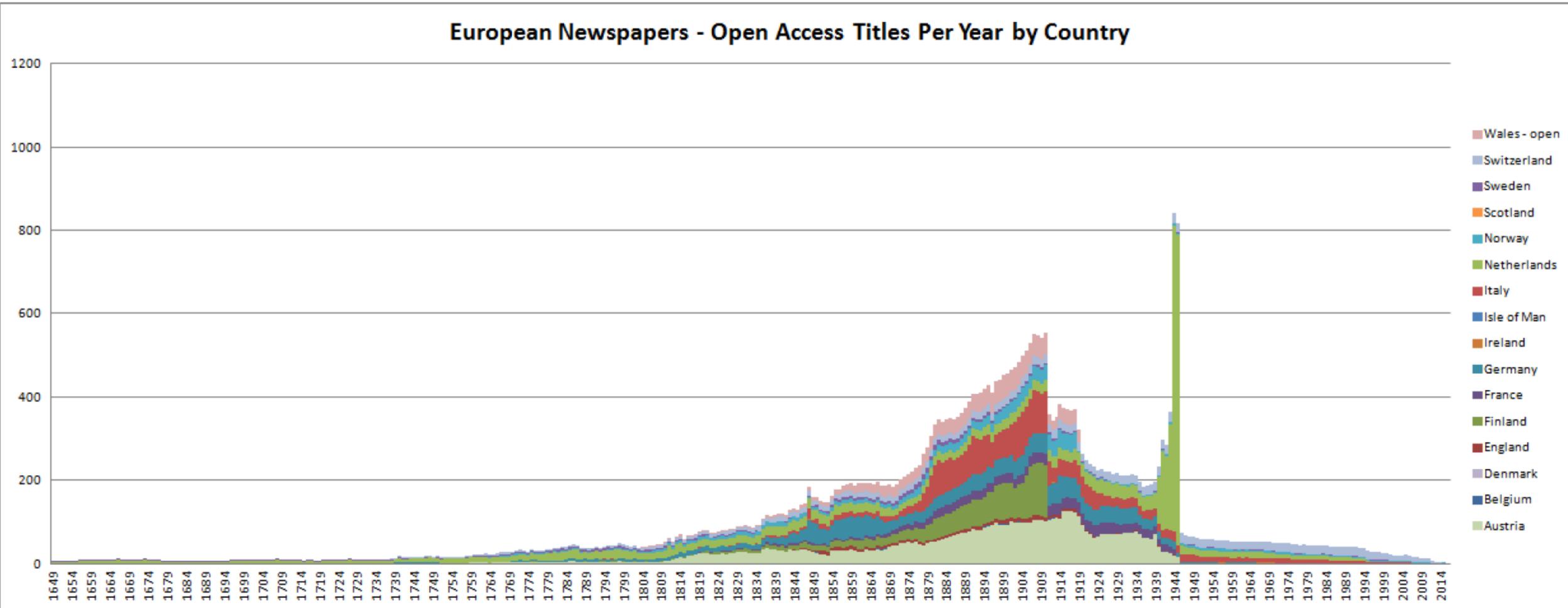


# Access policies



Source: Center for Research Libraries. The State of the Art. A Comparative Analysis of Newspaper Digitization to Date, 2015

# Access policies



# “Data acquisition crusade”

# Tedious and time consuming

- source inventory
  - conventions with right holders
  - physical data acquisition

# Bridging silos?

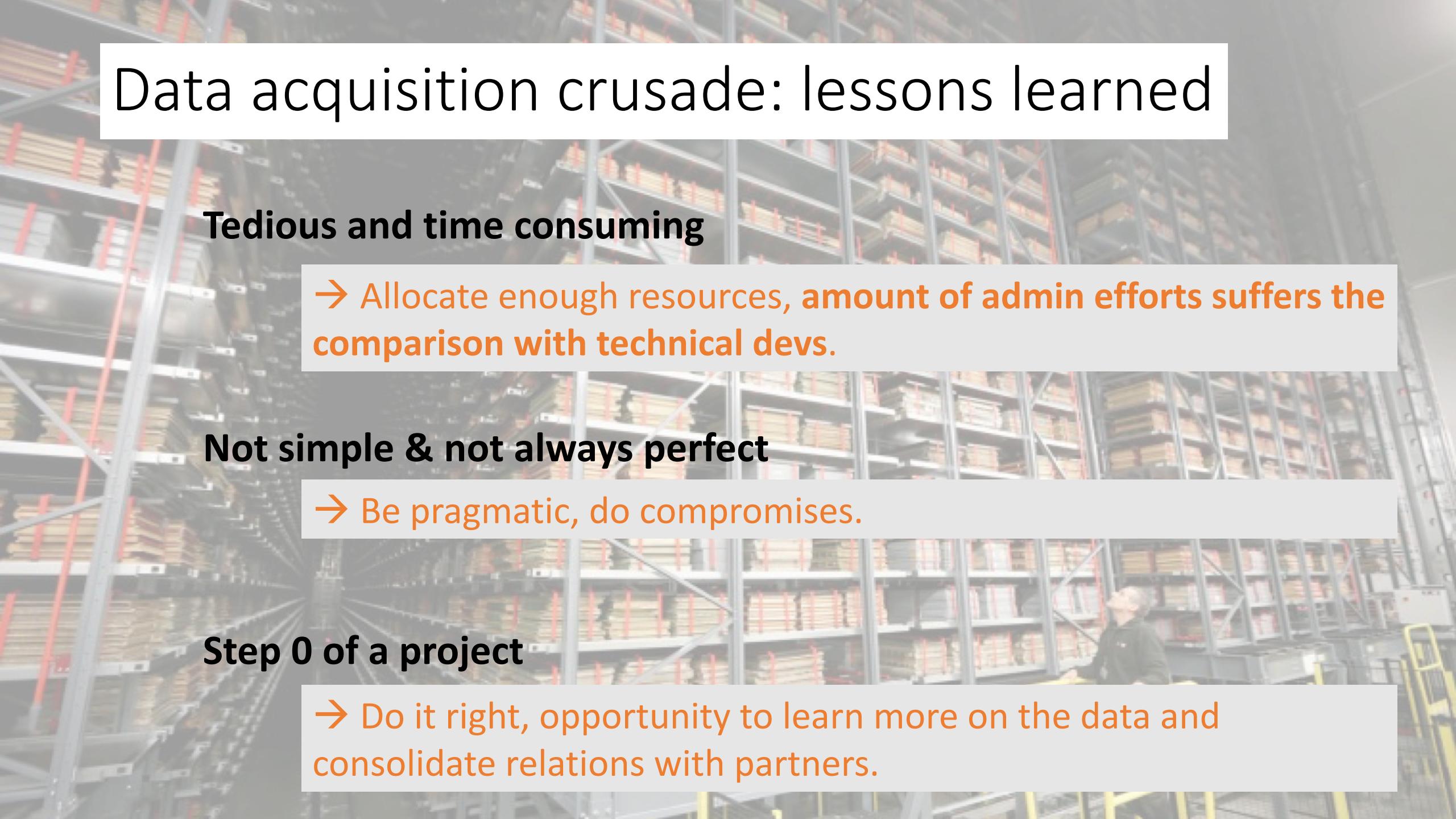
- robotics and politics
- hybrid search strategy
- best argument: demonstrating the potential of digitized newspapers

N E W S  
E  E



**Living with machines**

# Data acquisition crusade: lessons learned

A large warehouse with tall metal shelving units filled with boxes and books. A person is visible in the background, looking up at the shelves.

## Tedious and time consuming

→ Allocate enough resources, amount of admin efforts suffers the comparison with technical devs.

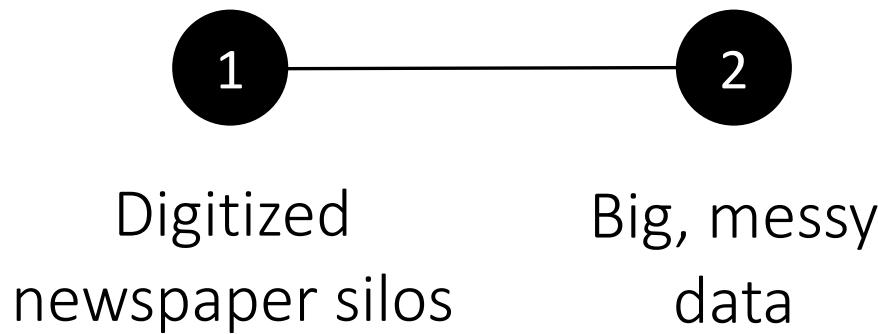
## Not simple & not always perfect

→ Be pragmatic, do compromises.

## Step 0 of a project

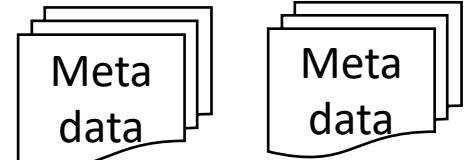
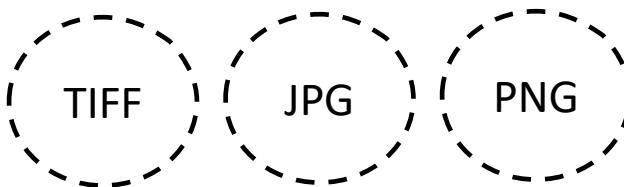
→ Do it right, opportunity to learn more on the data and consolidate relations with partners.

# *How to enable semantic indexing and exploration of large collections of historic newspapers?*



# What happens to digitized items?

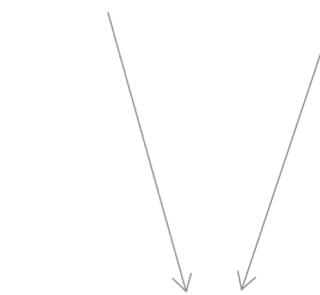
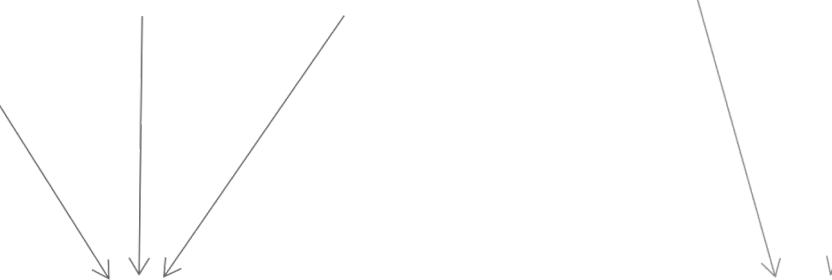
*digital  
archives*



*processable  
data*

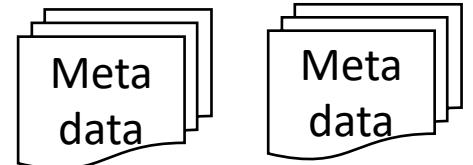
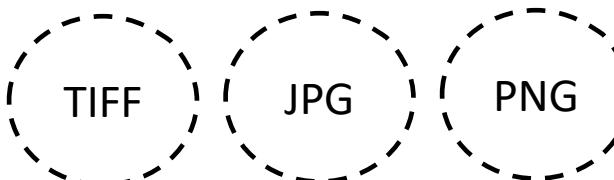
{ JSON }

JPEG2000



# What happens to digitized items?

*digital  
archives*



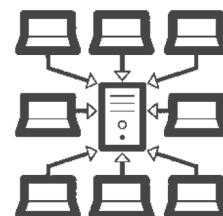
1 year: 25GB

*processable  
data*



{ JSON }

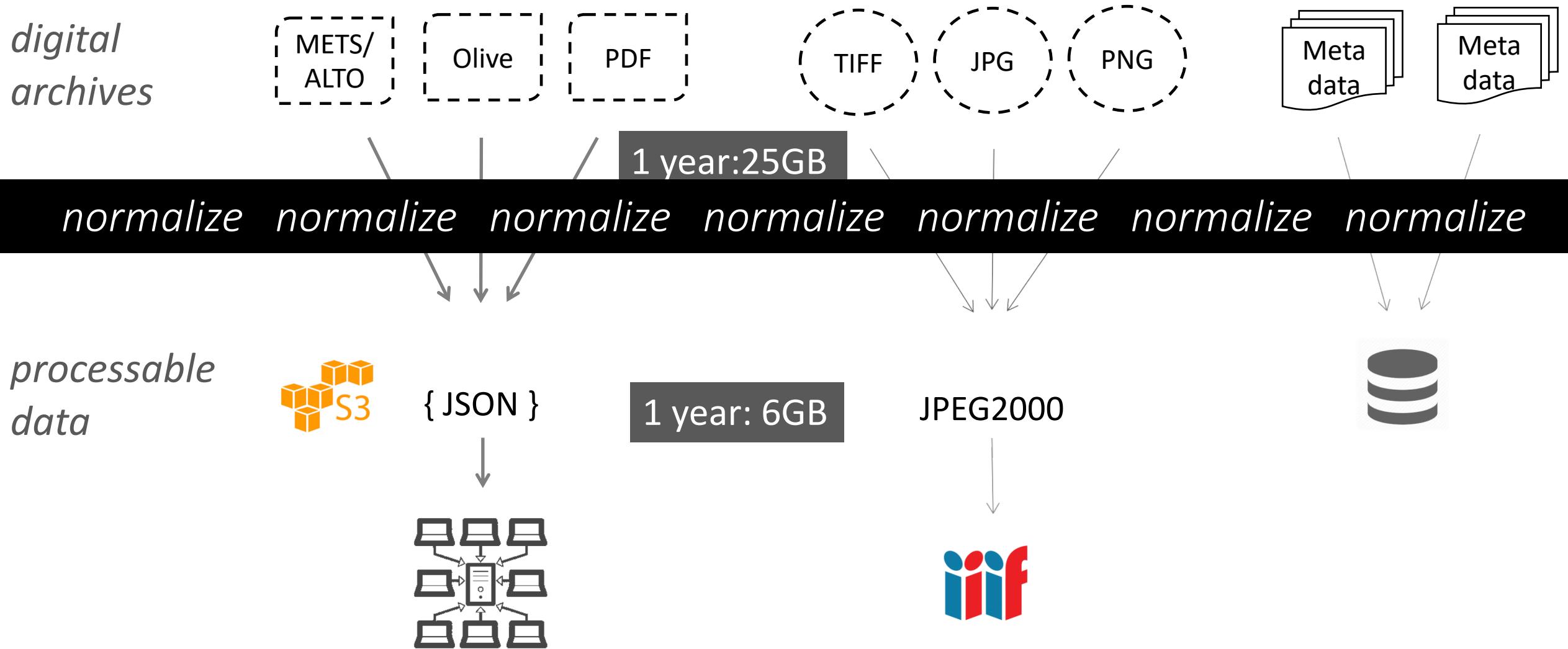
1 year: 6GB



JPEG2000



# What happens to digitized items?



# Lost in translation?

*digital  
archives*



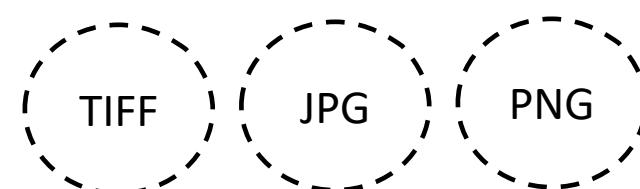
*empty folders*

*corrupted (zip) archives*

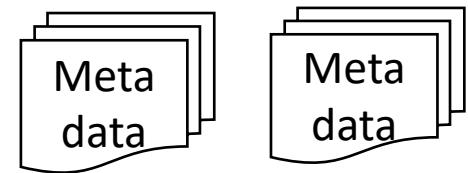
*inconsistencies*

{ JSON }

*processable  
data*



JPEG2000



# Inconsistencies

- mixed up page order
- empty text field (ca. 3,3%)
- duplicated text
- wrong language tag (ca 0.1%)

# Inconsistencies

**Abonnementspreise.**  
Per Einzelblatt 10 Cts. für den kleinen Raum  
für die Schau 20 Cts. für das Ausland 40 Cts.  
Extra 20 Cts. (excl. Innam. Beurkant. Reklamen Art. 1.— der Seite).

N. 122 P Zweihundiebenzigster Jahrgang. Sonntag, 1. Mai 1892.

# Neue Zürcher-Zeitung

**und schweizerisches Handelsblatt.**

Aus der Stadt Freiburg.

Freiburg, 28. April 1892.

Le roi s'amuse. Am letzten Sonntag war hier ein Mitglied des Gemeinderathes an Stelle des verstorbenen Herrn Antonin Boccard zu wählen. Der Auszug der Abstimmung hat dem Bernenchen nach anfänglich beschlossen, sich neutral zu verhalten. Aber er hatte die Rednung, ohne den Wirk und der Färberei, der Partei verlangt unbedingten Gehorsam und mahnt seinen Anhängern zu, für die wichtigsten Dinge und Personen zu stimmen, um sie füllig machen will. Das ist eben mon plaisir. Wer hat er nicht bei der Abstimmung über das Altmühlenmonopol über Nachseine liebenswerte Freunde Theraulax ein Schimpfen geschlagen, der in Bern versprochen hatte, daß der ganze Kanton Freiburg für das Monopol stimmen werde. Herr Python wollte es anders, und es sind andere.

## Die Maiwahlen im Kanton Solothurn.

Im Jahre 1887 gingen die Wahlen in folgender unseligen Finanzkatastrophe unter stürmischer und leidenschaftlicher Agitation vor sich; gegenwärtig verfügt man von einer politischen Auflösung, obgleich uns vom Entscheidungstage nur eine kurze Zeit voraus, verhältnismäßig wenig. Und doch sind die politischen Gegenseite nicht verschwunden und die Opposition hat den Kampf keineswegs aufgegeben. Doch trodern die Aufregungen sich leicht, hat und eine gewisse Ruhe eingesetzt ist, davon wir einerseits der Entscheidendheit, mit der freisinnige Partei stets die Verteilungen überlieferte, beobachtet unterliegt, sowie der Energie u.

politischen Fragen wird sie dogegen mit der liberalen Partei zusammengehen. Die Delegierten-Verfassung der Arbeiterpartei, die am Sonntag in Solothurn stattfand, hat 16 Kandidaten aufgestellt, an ihrer Spitze Staatsammann Füchsel und Fürbress Rehler. In der Stadt Solothurn (10 Vertreter) beanspruchen die Arbeiter drei Vertreter, und diese sind ihnen von den liberalen Parteileitung bereits zugestanden worden; in Lebern (16 Vertreter) 5, in Kriegstetten (14 Vertreter) 2, in Olten (18 Vertreter) 4, in Gossen (10 Vertreter) 2. Die Unterhandlungen mit den Freisinnigen in diesen Bezirken sind noch nicht zum Abschluß gekommen.

Was die Wahlauftaktungen in den einzelnen Bezirken betrifft, so darf als sicher angenommen

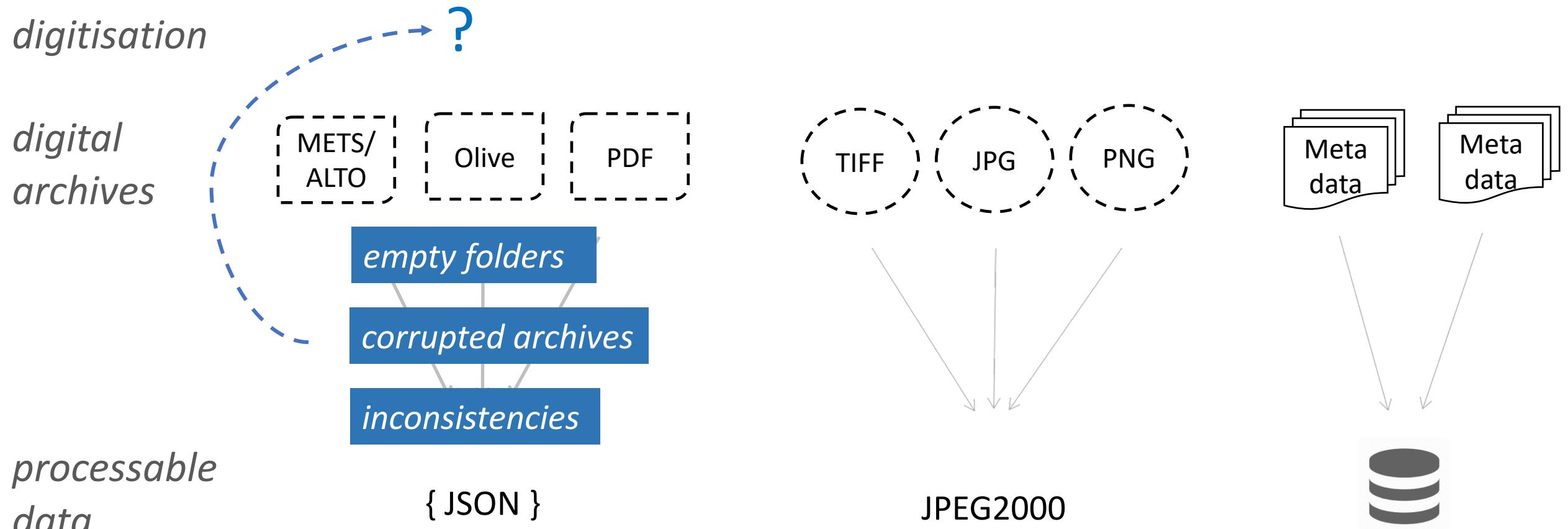
Eulach wird an die Gemeinde Oberwinterthur ein Betrag von 8000 Fr. und an die Gemeinde Gulden 2000 Fr. als Vorleistung ausgerichtet werden. — Die Gemeinde Habsburg erträgt an die Ausgaben für das Strohweizen für 1891 einen außerordentlichen Staatsbeitrag von 500 Fr. — Zum Präsidenten des Regierungsrates wird für die Dauer vom 1. Januar 1892 bis Ende des Amtsdauer gewählt Dr. Regierungsrat z. g. Groß, zum Vizepräsidenten Dr. Regierungsrat Rägeli.

(Witigkeit.) Die Sanitätsdirektion verleiht ein Legat von hundert Franken zu Gunsten der Freiherrnlandsburg Burgbühl als Andenken an eine Verdienste.

Bern.

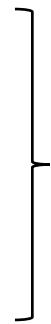
Dr. Otto von Greuz ist zum Lehrer des Deutschen am Gymnasium der Stadt Bern ernannt worden.

# Inputs



# Archives holes

1. publication periodicity



⇒ historical knowledge, often encoded as metadata

2. publication interruption

3. paper copies lost

⇒ preservation history knowledge, rarely encoded as metadata

4. digital copies lost or damaged

⇒ requires digital checks, almost never encoded as metadata

# Actually, holes and imperfections matter

```
# get zip archive and check it
archive = os.path.join(issue_dir.path, "Document.zip")
if not os.path.isfile(archive):
    logger.info(f"No Document.zip in issue {issue_dir.path}. Skipping it.")
    return
else:
    try:
        working_archive = zipfile.ZipFile(archive)
    except zipfile.BadZipfile as e:
        logger.info(f"Corrupted Document.zip in issue {issue_dir.path}. Skipping it.")
    return
```

Count, Store, Compare!

# Lost in translation?

*digital  
archives*



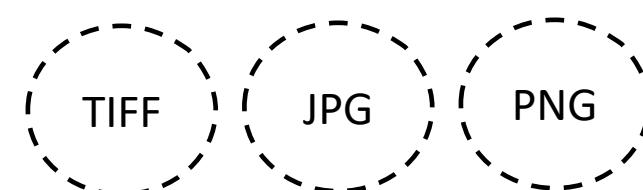
*empty folders*

*corrupted archives*

*inconsistencies*

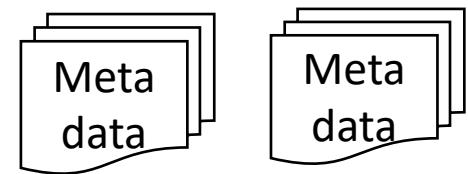
{ JSON }

*processable  
data*



*box coordinates*

JPEG2000



Nightmare

# Lost in translation?

*digital  
archives*



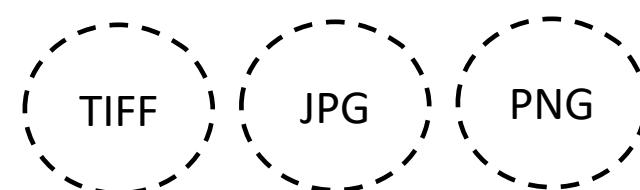
*empty folders*

*corrupted archives*

*inconsistencies*

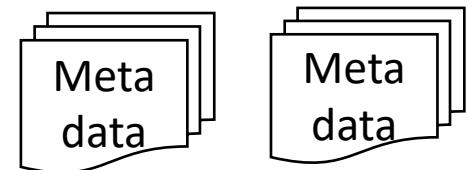
{ JSON }

*processable  
data*



*box coordinates*

JPEG2000

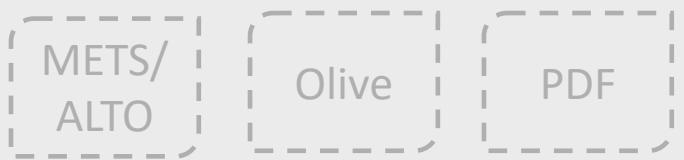


*partial, non  
normalized info*



# Lost in translation?

*digital  
archives*



*empty folders*

*corrupted archives*

*inconsistencies*

*processable  
data*

Harmonization and enrichment needed



*box coordinates*

Quotidien: 17  
Wöchentl.: 10  
Hebdomadaire: 2  
2x par semaine: 4  
3x par semaine: 1  
2x wöchentl.: 6  
Tägl.: 6  
Settimanale: 4  
3x par semaine,: 2  
6x par semaine: 3  
Bi-hebdomadaire: 1  
quotidien,: 1  
trois fois par semaine,: 2  
Quotidien,: 1  
quotidien: 2  
Wöchentl.,: 1  
Mensile: 1  
2x all'anno,: 1  
Quotidiano,: 2  
Quotidiano: 1  
12x all'anno: 1  
3x la settimana: 1  
Halbwöchentl.: 4  
3x wöchentl.: 1  
Ca. alle 1-2 Jahre: 1  
2x monatl.: 1  
2-3x wöchentl.: 1  
6x wöchentl.: 2  
2x wöchentl.: 1  
Annuel: 2  
Ca. 6x jährl.: 1  
Jährl.: 2



*, non  
normalized info*



# Overview of the *impresso* corpus

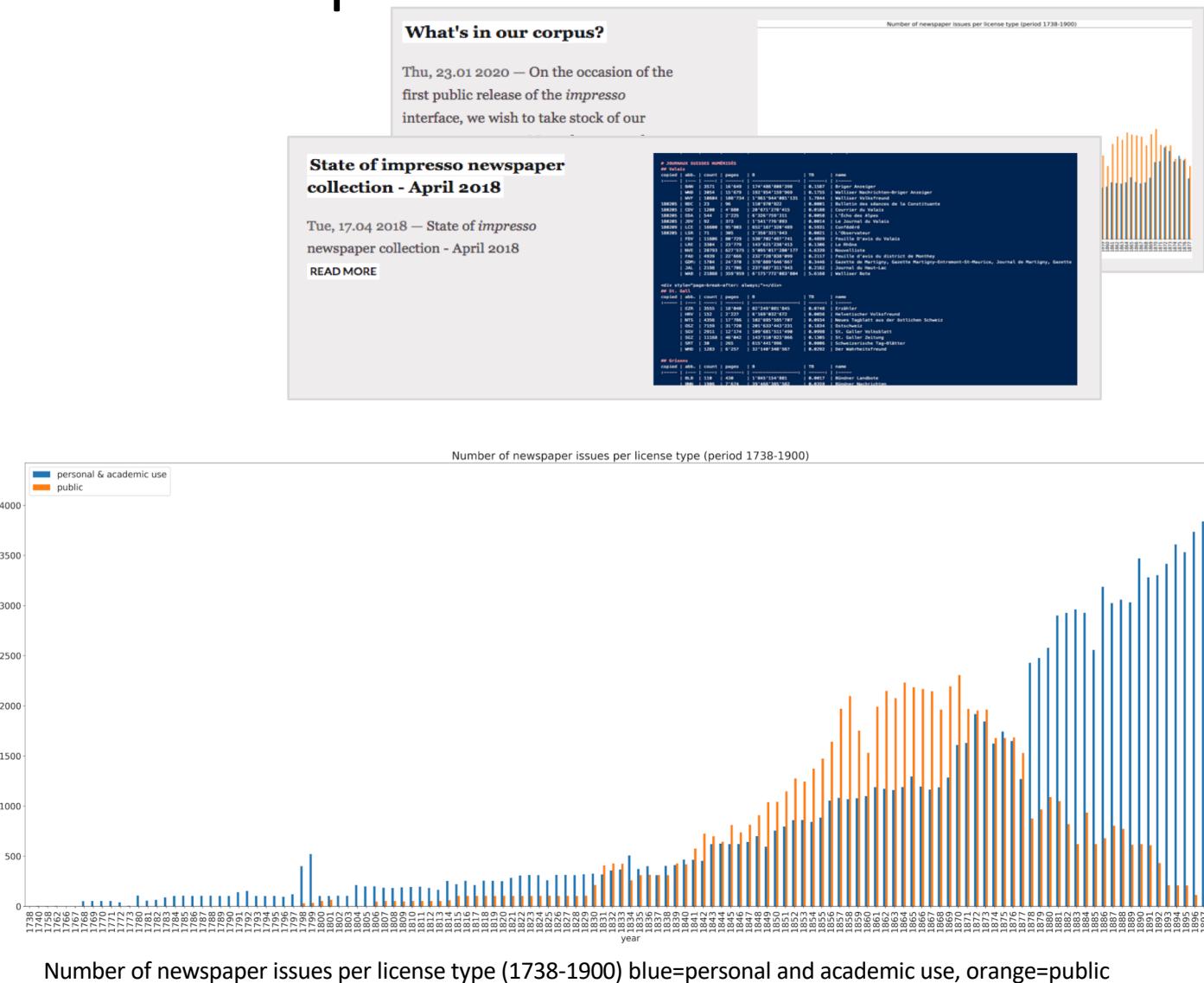
Growing throughout the project:  
**51 (2018), 76 (2019), 93 (2021) titles**

Upcoming release:

- o CH, LU & FR newspapers
- o 58M ‘content items’
- o 15 billion words (token)

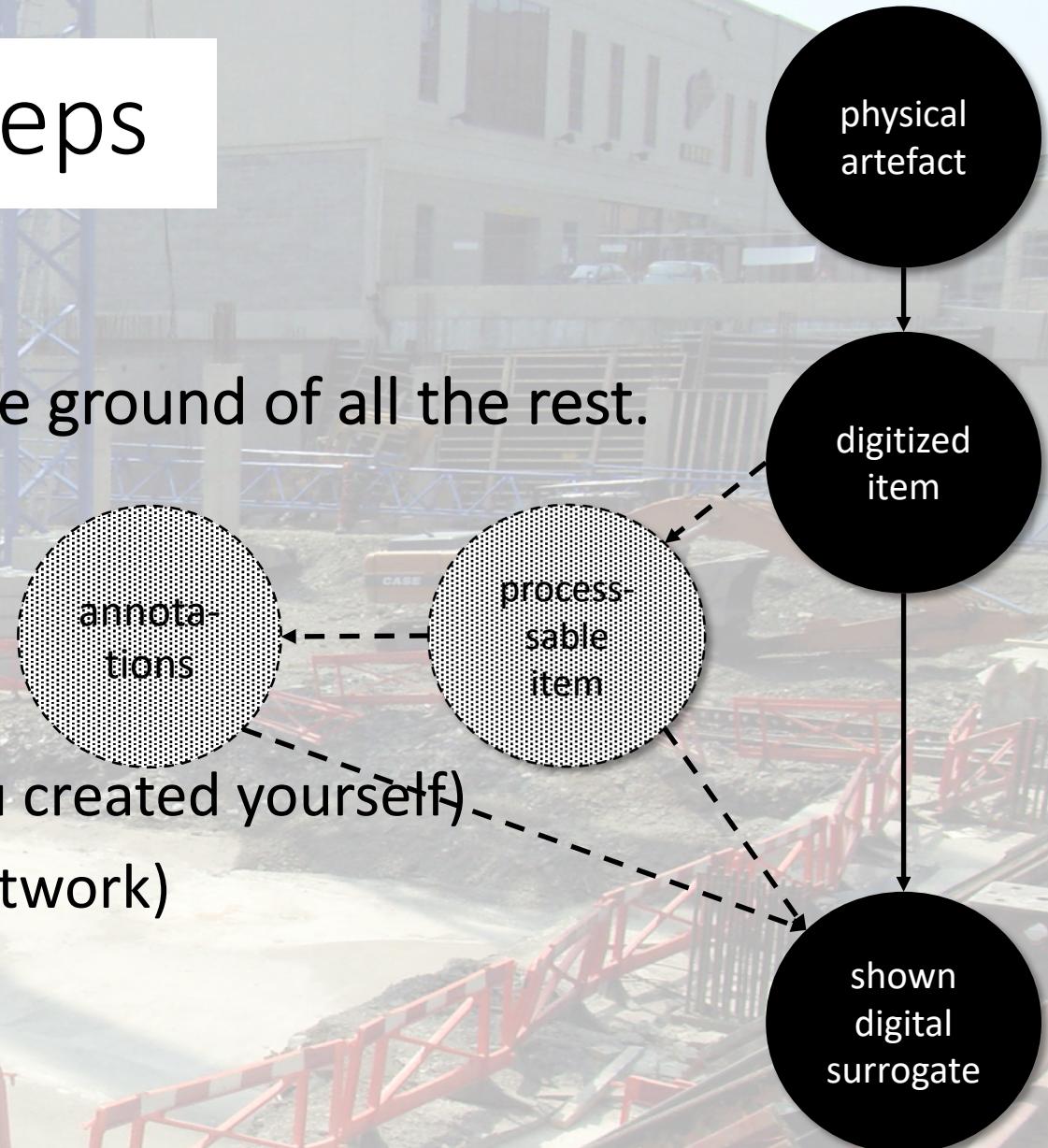
Different rights statements regimens:

- o public domain
- o private and academic use



# A whole world of hidden steps

- Laborious, usually invisible work which lay the ground of all the rest.
- Caring about quality in big data context is not easy:
  - do not trust any data (not even the ones you created yourself)
  - do not trust any infrastructure (S3 pb, IO, network)
  - forget about all your assumptions
  - think big, but triple-check the tiny details



→ We should aim for full tracing of meta-knowledge across actors.

# *How to enable semantic indexing and exploration of large collections of historic newspapers?*





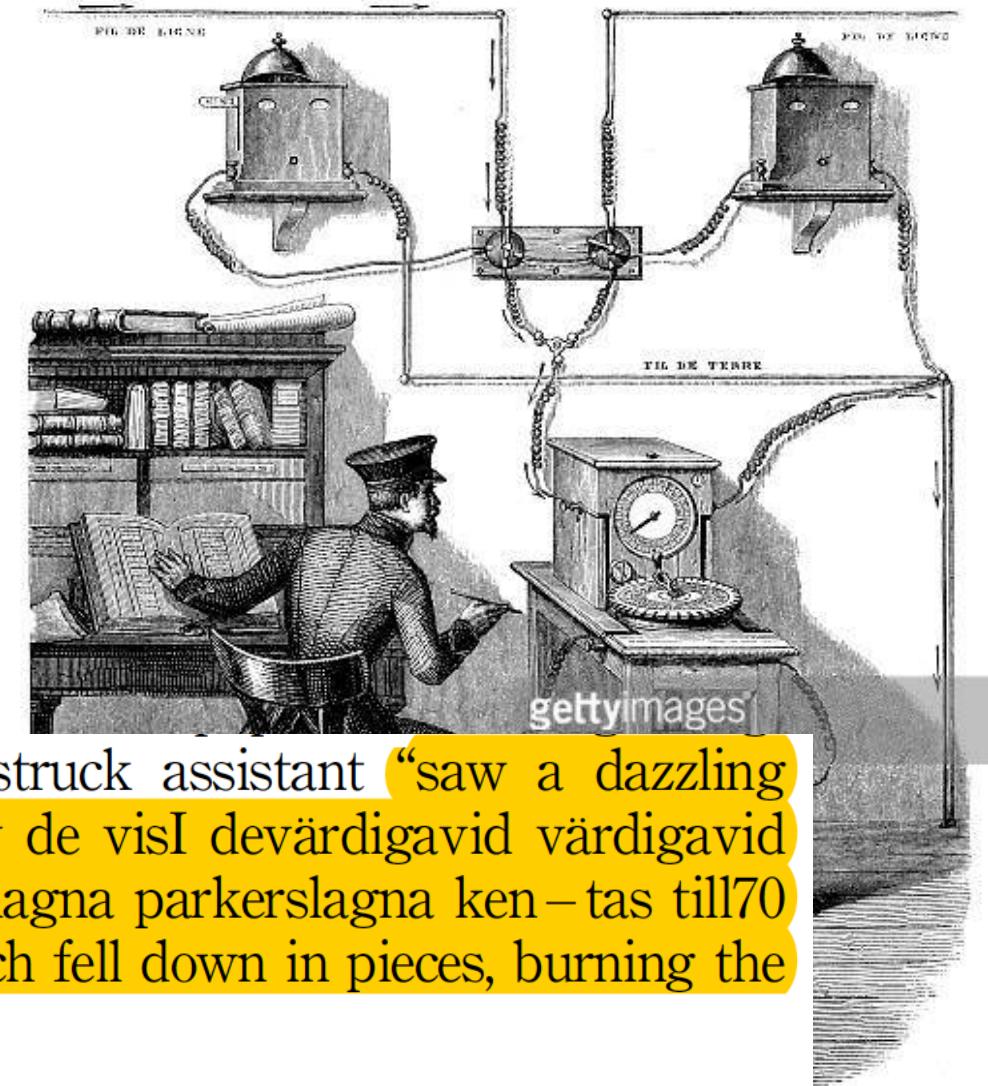
# Invented data?

## Cultural heritage as digital noise: nineteenth century newspapers in the digital archive

2016

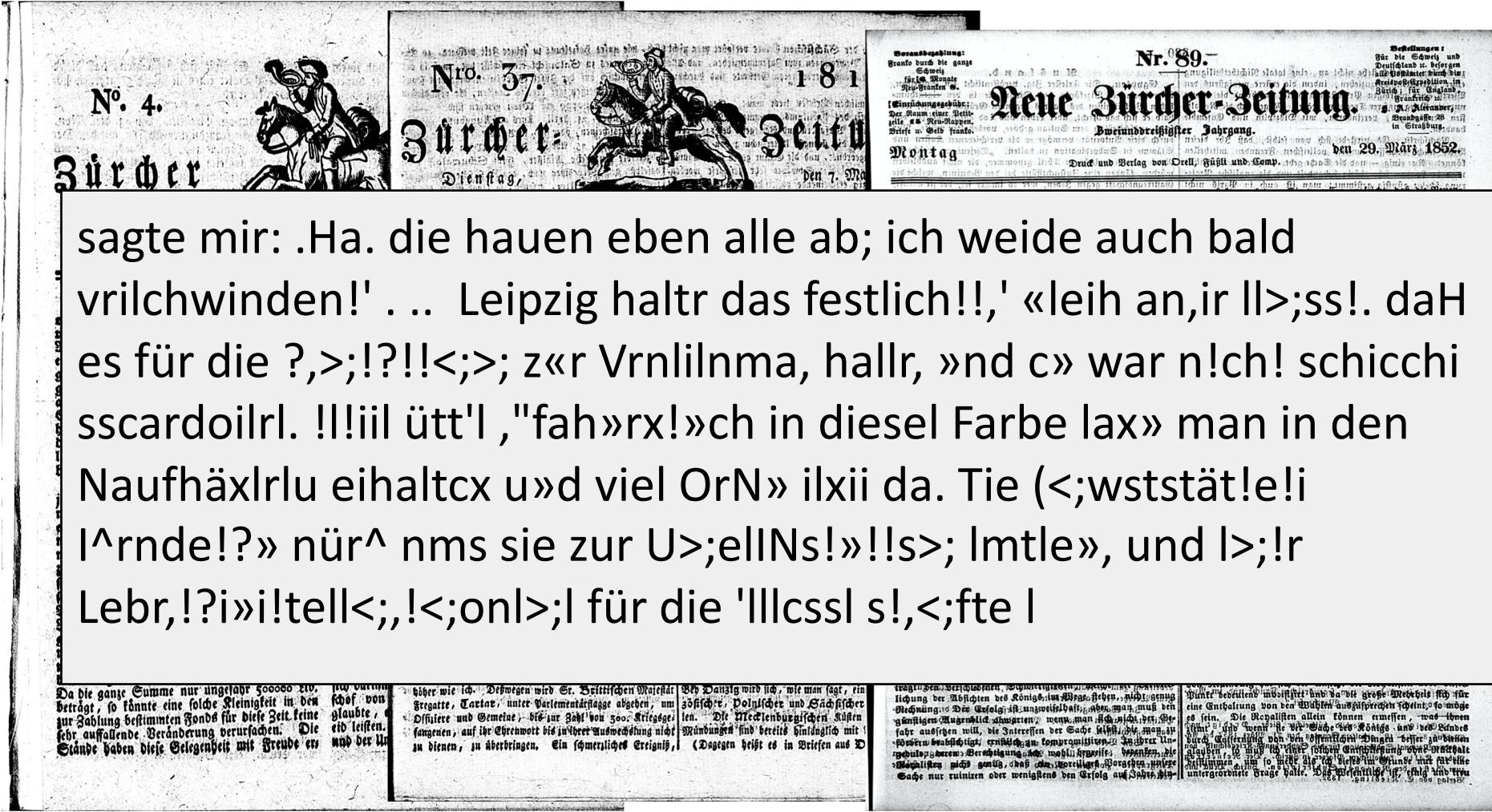
Johan Jarlbrink and Pelle Snickars

telegraph incident, in fact, literally reported that the struck assistant “saw a dazzling light along the wires on the walls conducting electricity de visI devärdigavid värdigavid dejemte fullkommen ihåförvintparkerslagna förvintparkerslagna parkerslagna ken – tas till 70 70 misvärt fruktarsnart tAf eoch sisrans njes ej [...] which fell down in pieces, burning the table and the floor.”



Never printed, but today part of the historical record.

# Quality of Images and Original OCR by NZZ



# Can we do better?

## OCR-2005

N °. 86 . Zürcher Samstag , t  
793 & deN26 .  
Weinmon . MMMM ? urck da  
« franiöische Voll Gouverain  
geworden , mit dem 3 « nnte  
in volitischen Votis » . »  
« ...nntn sen ; «  
... siecht zufrieden damit , im  
Empersts " ; t

**F-Measure: 0.147**

## OCR-2017

Prozeß der Marie Antoinette.  
Nachdem dieselbe am 15. Weinm.  
alten Styls, oder am  
2. des ersten Monat» im 2, en  
Jahre der Republik neuen  
Kaleuderstpls, in den  
Audienz-

**F-Measure: 0.684**

## HTR-2018

Prozeß der Marie Antoinette.  
Nachdem dieselbe am 15. Weinm.  
alten Styls, oder am 23. des ersten  
Monats im 2ten Jahre der Republik  
neuen Kalenderst...  
Audi...  
**F-Measure: ~0.95**  
...lassen hatte, fragte sie der  
Präsident: Wie sie heis- se? „ Ich  
nenne mich, antwortete sie, Marie  
Antoinette von Lotharingen-

# OCR quality assessment

**Goal:** allow users to **estimate** the completeness of their keyword search results

**Principle:** OCR quality  $\cong$  ratio of known words compared to all words

**Measure:**

- **focus** on content words (exit stopwords and punctuation)
- **easy** to grasp by non-technicians

# Example – Quality = $10/13 = 0.77$

Il en est de même de la constitution d'Argovie de 1841 qui n'a pas encore été garantie

|             |   |
|-------------|---|
| Original    | Il en est de mèmede la constitution d'Àrgovie de 1841 qui n'a pas encore été garantie |
| Normalized  | il en est de memede la constitution d argovie de 0000 qui n a pas encore ete garantie |
| Categorized | .. . *** .. MEMEDE .. constitution . argovie .. **** *** . . *** encore *** garantie  |

*Lucerne, Uri, Schwytz, Unterwald, Zug, Fribourg et Appenzell-intérieur.*

|             |   |
|-------------|---|
| Original    | Lucerne, Uri, Schwytz, Unterwald, Zug, Fribourg et Appenzell-intérieur. |
| Normalized  | lucerne uri schwytz unterwald zug fribourg et appenzell inlierieur      |
| Categorized | LUCCRNE uri schwytz unterwald zug fribourg .. appenzell INLERIEUR.      |

. \* = masked character of short or frequent word / RED = unknown word / green = known word

# Named entity processing

**Referential units** which underlie the semantics of texts.

(5Ws:Who, What, Where, When, Why)



**Proper names, definite descriptions** generally of type **Person, Organization, Location**.

Ca. **30%** of content-bearing words in news texts are proper names.

Named entity processing refers to a family of tasks that are part of the larger domain of **Information Extraction**. IE has been defined and formalized in the **90's** with the **Message Understanding Conferences** (MUC) organized by the **US DARPA**. Besides the recognition of proper names of types e.g. **PERS**, **ORG** and **LOC**, named entity processing also implies **disambiguation** and **relation detection**.

**PERSON, ORGANISATION, LOCATION, EVENT, DATE, NLP-TASK**

Good performances when:

- English
- News, generic domain
- Simple typology

# Challenges on historical texts

Suffers from:

- bad OCR

*Constat. iipopje, Buch irest, M'' Lucile*

- language evolution

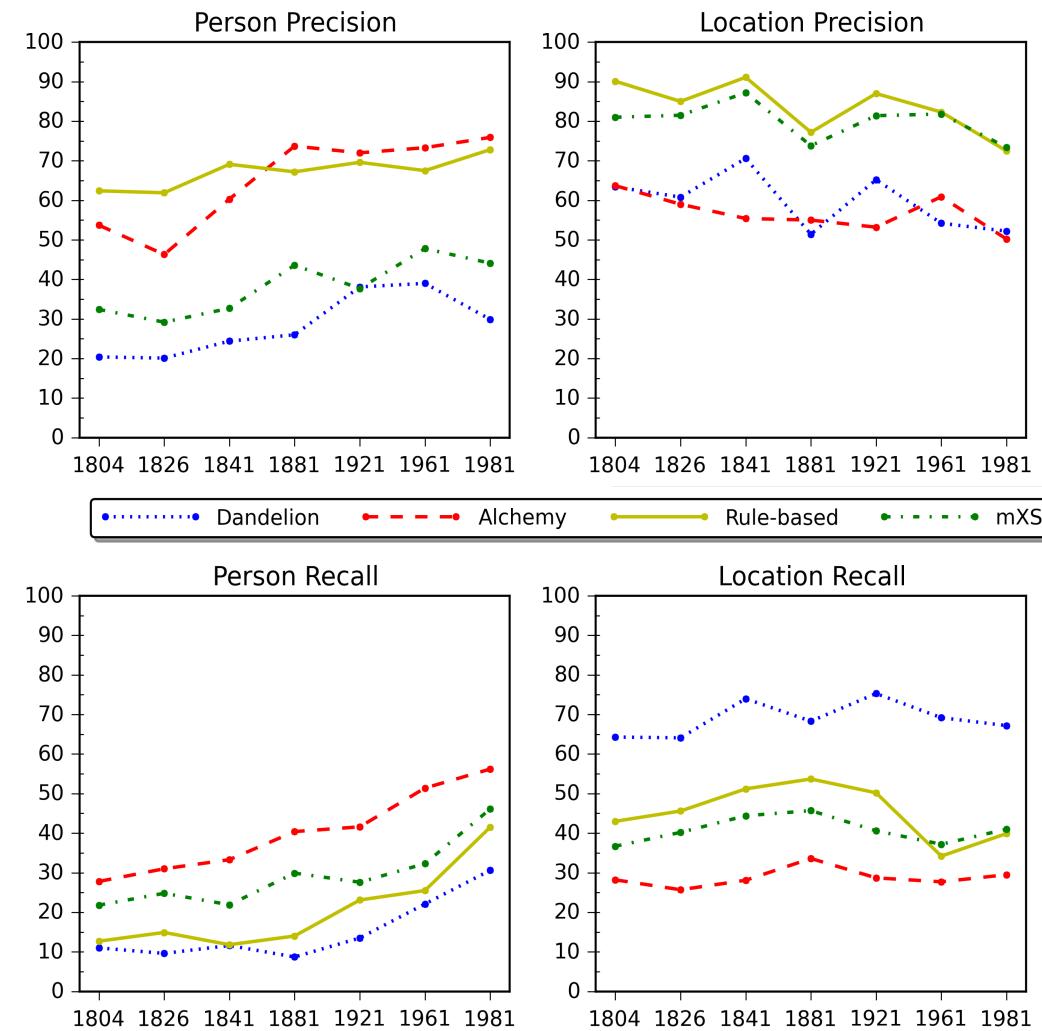
*Härnevi* → *Arnevi*, *Kallmar* → *Kalmar*

- Poor resource coverage

- non VIPs not recognized

- lack of appropriate trigger words

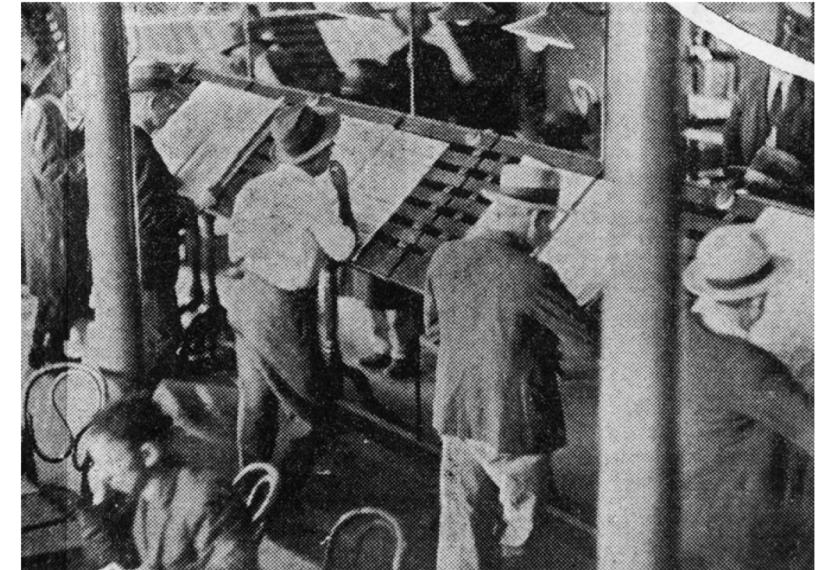
*burgomaster, tailor, munition specialist*



# CLEF-HIPE 2020

- CLEF: Evaluation conference, “Olympic Games of Computational Linguists”
- HIPE (Identifying Historical People, Places and other Entities), “A specific discipline competition”, organized by *impresso*
- Teams try to develop the best system.

Named Entity Processing on Historical Newspapers



## Objectives

- strengthen the **robustness** of approaches;
- enable **performance comparison**;
- foster **efficient semantic indexing** of digitized cultural heritage collections.

# CLEF-HIPE-2020

## Tasks:

- NE recognition and classification
- NE linking



# CLEF-HIPE-2020

## Tasks:

- NE recognition and classification
  - NE linking

Newspapers from CH, LU, US from 1738 to 2019

## **Participation** 13 teams



# CLEF-HIPE-2020

Website: <https://impresso.github.io/CLEF-HIPE-2020/>  
Overviewpaper: <https://infoscience.epfl.ch/record/281054?ln=en>  
Participant papers: <http://ceur-ws.org/Vol-2696/>

## Tasks:

- NE recognition and classification
- NE linking

Newspapers from CH, LU, US from 1738 to 2019

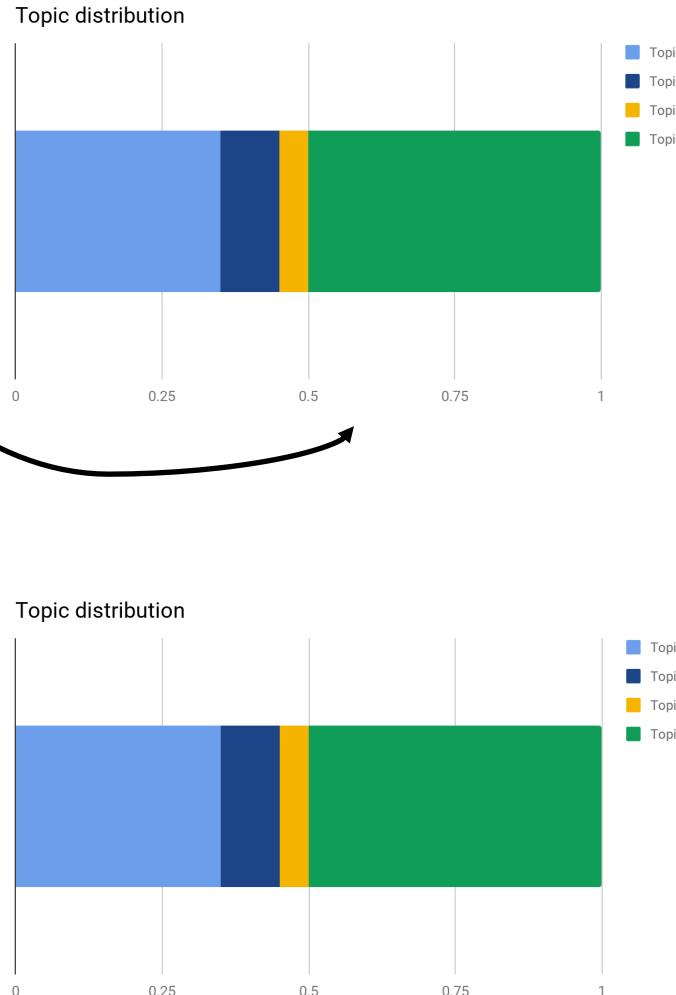
Participation 13 teams

## Main outcomes:

- Contextualized neural embeddings (BERT) are strong
- Careful preprocessing helps to achieve a performance not far behind NER/EL on modern newspapers.



# Topic Modeling



kunst konzert bild theater künstler  
oper musikalisch spielen werk  
künstlerisch orchester aufführung  
ausstellung lied musik vortrag  
stück chor musil programm

deutschen britisch deutsch  
truppe krieg russisch angriff  
flugzeug feindlich schwer front  
armee kampf london japanisch  
havas amerikanisch feind  
alliiert italienisch

zimmer lage vermieten  
verkaufen wohnung garten groß  
auskunft villa hotel schön zürich  
haus pension komfort preis see  
telephon sonnig modern

# OCR errors dominate topics

13/50

drn verde riner drs nnr  
lourde fierer snr znm dri  
vrr wnrde tir zürcher sf  
hauen bri jür eiue

lil eilen lul nil lin ver  
uni llr lullen lun lei van  
vol llo vel lol vll ren  
lesen lit

nicken bitten vnn icl eit  
lill nicht gen liegen clcr  
erden essen gen sel st<sup>+</sup> bereit cke  
liegen nde lei chen seln  
eilen wild ehren slch  
folgen lan schl stellen  
suchen eichen

vita dtt chm dje dew dep  
ftn stm dtm mem hät ftm  
dtr ßch mam  
stch

unc clie icl ncl iic ssc unci  
icr clen vnn clc cll lcl icli  
unss llc nci clic ncni

nicken bitten vnn icl eit  
lill nicht gen liegen clcr  
die klein cit bereit cke  
an immer ckcn

deu leinen tte sep dtt  
sev deutsch teu oct  
dienstag wey ttt ini ftp  
freytag dep beu bep  
neapel ste

auk äis äor sen nnä äio  
aut virä inä aktie van  
veräen obligation versen  
it bitten tur vsr oktober  
äon

werben imb beten beii  
hor finb biefer alt beut  
önnfi bieten bafi hub  
u bcii bcs bitten jii o

dar lir iiiii iit leihen aii  
fiir iri lil tal ifl im  
helfen hui iili iti ifi iif  
uum lullen

bass werben wild leinen  
iahen mehl enn übel wal  
fehlen non tann fühlen  
zul nui zui übn hatt ihi  
wied

# Long story short

A LOT of preprocessing:

- OCR error (post-)correction
- old spelling normalization
- lemmatization

Current models in the interface:

- one per language
- only nouns
- only known lemmas
- elimination of rare and very frequent nouns

# Text reuse

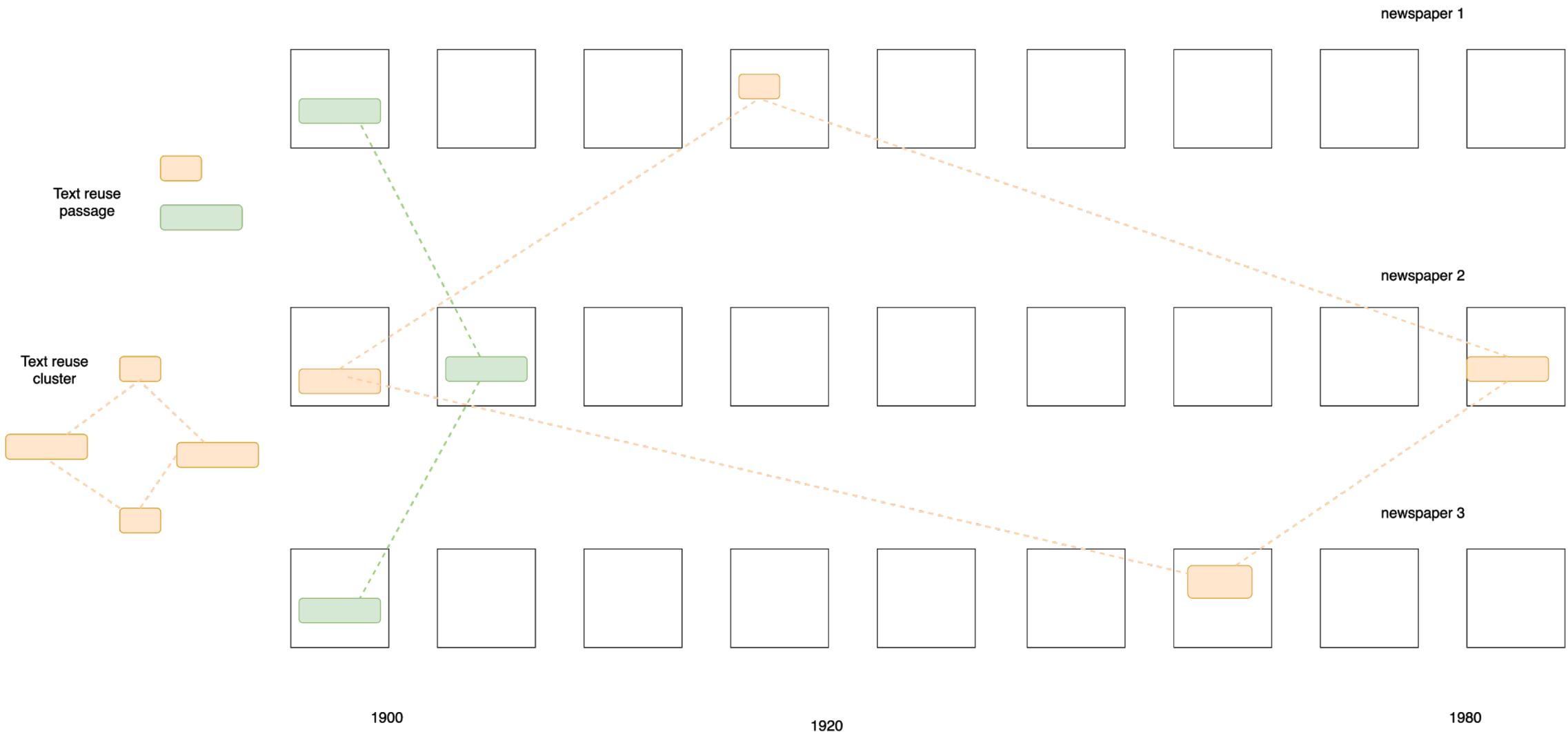
*The meaningful reiteration  
of text, usually beyond the  
simple repetition of  
common language.*

Academic writing → quotation,  
plagiarism

Literature → allusion, paraphrase,  
quotation (intertextuality)

Newspapers → copy/paste  
journalism, republication of news  
articles

# Text reuse passages and clusters



# Passim – output examples

## GDL 03/12/1863

— Mardi il est arrivé un accident, sans suites fâcheuses, au bateau à vapeur parti d'Ouchy pour Genève à 2 heures 20 minutes. Le bateau était arrêté pour le débarquement et l'embarquement devant Nyon, lorsque tout à coup on entendit une détonation, un nuage de fumée (ou de vapeur ?) sortit de la machine et le bâtiment subit une violente secousse. Nous ne savons pas d'une manière précise en quoi consiste la rupture qu'il y a eu ; on parle d'une clavette cassée. Quoi qu'il en soit, le bateau était mis dans l'impossibilité de continuer sa route ; heureusement les voyageurs qui devaient aller plus loin ont pu prendre le train qui passe à Nyon à 5 heures 3 minutes et sont arrivés à bon port, sans autre mal qu'un moment de frayeur. On

## JDG 05/12/1863

— Mardi il est arrivé un accident, sans suites fâcheuses, au Guillaume-Tilt, parti d'Ouchy pour Genève à 2 heures 20 minutes. Le bateau était arrêté Pour le débarquement devant Nyou, lorsque tout à coup on entendit une détonation, un nuage de fumée (ou de vapeur ?) sortit de la hiadiine, et le hû liment subit une violente secousse. Nous ne savons pas d'une manière précise en quoi consiste la rup'" re qu'il y a eu ; on parle d'une clavette cassée. Quoi qu'il en soit, le bateau était mis dans l'impostJuilite de continuer sa route ; h'ureuiipnient les v &lt; y 'geursqui devaient aller plus l &lt;&gt; iri ont pu piendre le train qui passe à Nyon à 5 heures 3 minutes, et sont arrivés à bon poil, sans aulre mal qu'un moment de frayeur. Le



## Noisy, historical texts: recap'

**Challenging.** NLP's legacy is orthogonal to what archives/libraries offer today. Huge time and domain adaptation problem.

**Benchmarking.** But *to what extent?* Organizing a shared task to benchmark current approaches is a huge effort, but it helps advance research on this topic.

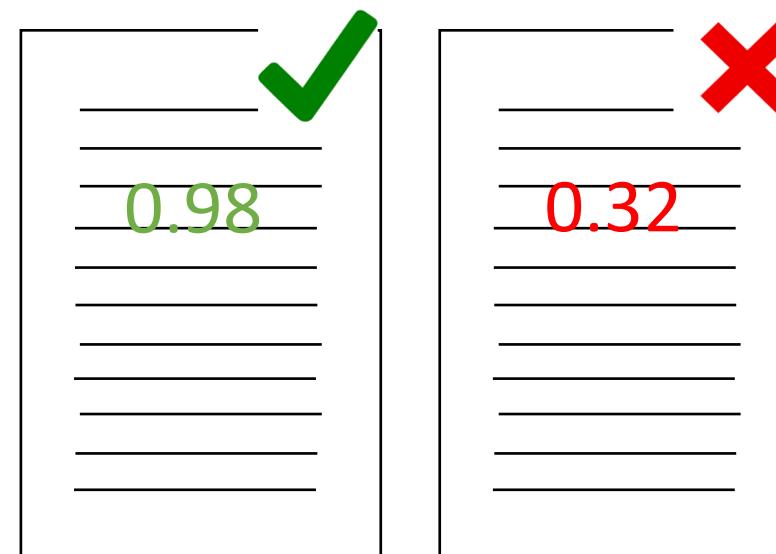
**Critical text mining.** The application of NLP methods and tools ought to be *functional* to the research questions and needs of historians.

Lexical  
processing

*language identification*  
*historical spelling*  
*normalization*  
*lemmatization*  
*word embeddings*

## OCR quality assessment

Lexical  
processing



## Lexical processing

## OCR quality assessment

## Named entity processing

Named entity processing refers to a family of tasks that are part of the larger domain of Information Extraction. IE has been defined and formalized in the 90's with the Message Understanding Conferences (MUC) organized by the US DARPA. Besides the recognition of proper names of types e.g. PERS, ORG and LOC, named entity processing also implies disambiguation and relation detection.

PERSON, ORGANISATION, LOCATION, EVENT, DATE, NLP-TASK

## Lexical processing

## OCR quality assessment



## Named entity processing

## Topic modeling

kunst konzert bild theater  
künstler oper musikalisch  
spielen werk künstlerisch

deutschen britisch  
deutsch truppe krieg  
russisch angriff kampf  
london

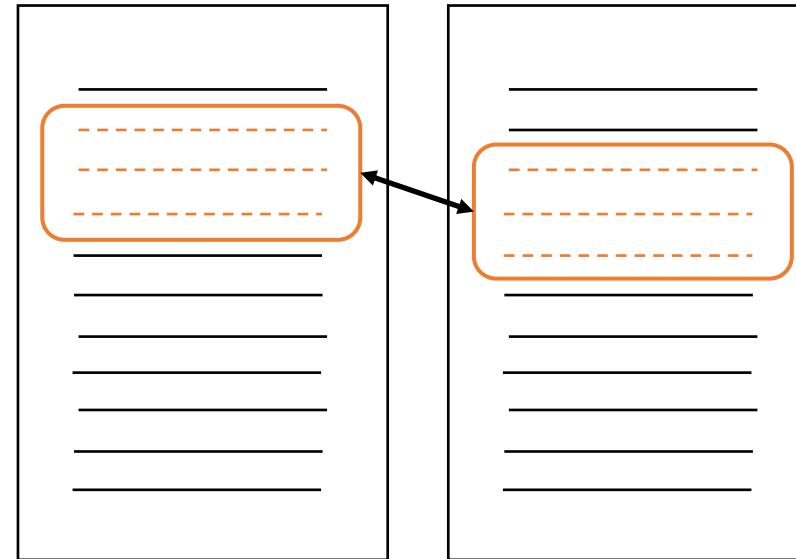
zimmer lage vermieten  
verkaufen wohnung  
garten groß auskunft

Lexical  
processing

OCR quality  
assessment

Named entity  
processing

Topic  
modeling



Text  
re-use

## Lexical processing

## OCR quality assessment

## Named entity processing

## Topic modeling

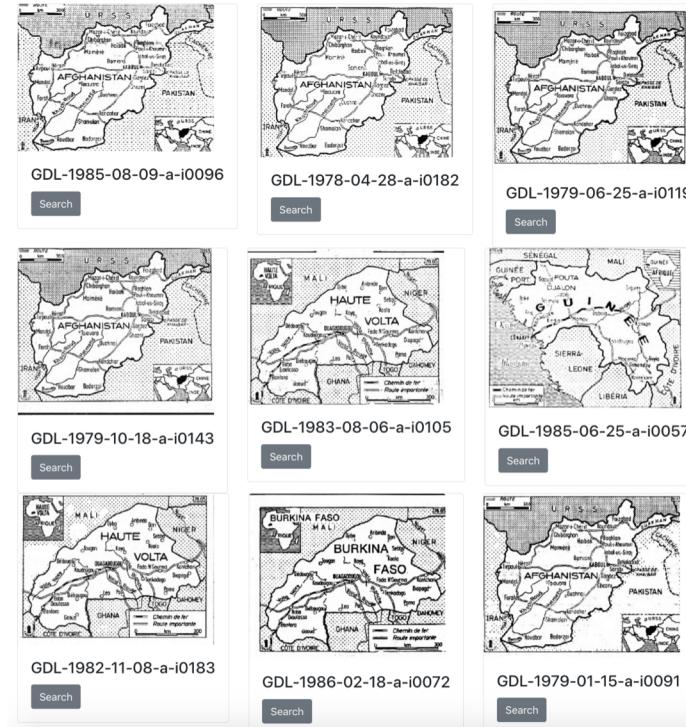


Image  
similarity

Text  
re-use

Lexical  
processing

OCR quality  
assessment

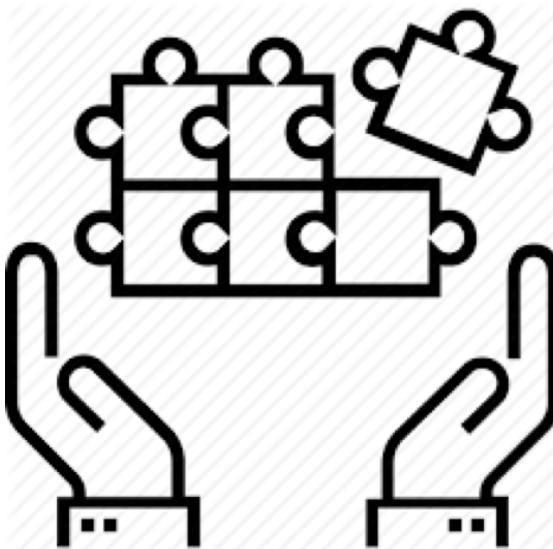
Recommender  
system

Named entity  
processing

Topic  
modeling

Text  
re-use

Image  
similarity



Visit the impresso YT  
channel to learn more  
on each component

Named entity  
processing

OCR quality  
assessment

Topic  
modeling

Lexical  
processing

78 newspapers  
47,876,994  
articles  
200M entities

Text  
re-use

....

Recommender  
system

Image  
similarity

Named entity  
processing

OCR quality  
assessment

Topic  
modeling

Lexical  
processing

search for ...

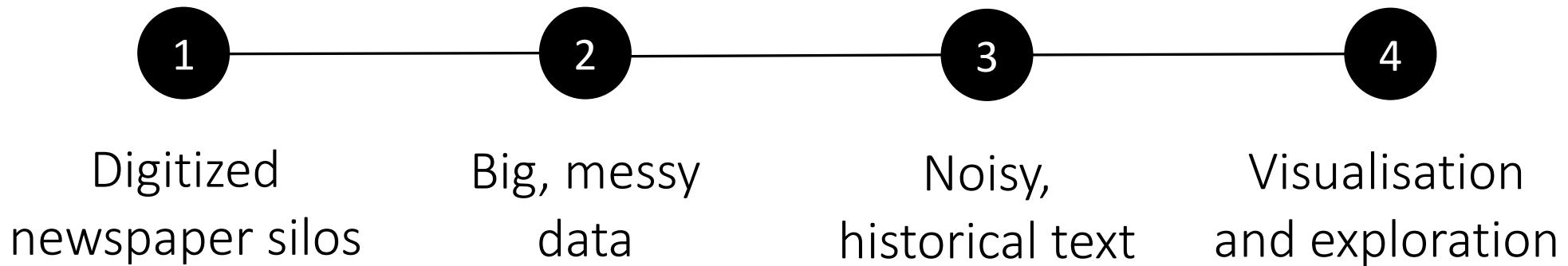


Text  
re-use

Recommender  
system

Image  
similarity

# *How to enable semantic indexing and exploration of large collections of historic newspapers?*



# Survey of existing interfaces

## Families of features

1. Metadata
2. Browsing
3. Search options
4. Result display
5. Result sorting
6. Filtering
7. Viewer
8. Content enrichment
9. User interaction
10. Digitization information
11. Connectivity
12. APIs

24 interfaces  
144 features

|   | California Digital Newspaper Collection | Colorado Historical Newspaper Collection | Chronicling | E-luxembourg                           | The European Library (TEL) | L'express | Georgia Historic Newspapers | Libraria - Ukrainian online periodicals archive | Retronews | StaBi | Swiss Press Online (SPOL) | Tessmann | Le Temps archives | Trove | Number of interfaces | Percentage |     |
|---|---|--|-------------|--|----------------------------|-----------|-----------------------------|---|-----------|-------|---------------------------|----------|-------------------|-------|----------------------|------------|-----|
| 011   | 1738-present                            | 1786-1986                                | 1917-1945   | 1631-1945                              | 1617-1946                  | u         | 1794-2006                   | 1798-1998                                       | 1803-2011 |       |                           |          |                   |       |                      |            |     |
| 28 ca. 2000                                     | 2                                       | 88                                       | 312         | 151                                    | 193                        | 51        | 51                          | 51  | 1294      |       |                           |          |                   |       |                      |            |     |
| u   | u                                       | u  | 35,723      | u                                      | 281990                     | 288,718   | u                           | u   | u         | u     | u                         | u        | u                 | u     | 1                    | 6%         |     |
| u   | u                                       | u  | 15M         |  |                            |           |                             |   |           |       |                           |          |                   |       | 3                    | 18%        |     |
| difficult to kr 1,5 millions                    |   |  | 234,593     | 3M for 50+ titles (ca. 80M t 2,930,362 |                            |           |                             |   |           |       |                           |          |                   |       |                      |            |     |
| y   | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 9                    | 53%        |     |
| n   | n                                       | y  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 8                    | 47%        |     |
| External links                                  | n                                       | y  | y           | n                                      | y                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | y     | 1                    | 6%         |     |
| Description of newspaper                        | n                                       | y  | y           | n                                      | n                          | y         | n                           | y   | y         | y     | y                         | y        | y                 | n     | 4                    | 24%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 8                    | 47%        |     |
| <b>On the search options</b>                    |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Keyword search                                  | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 15                   | 88%        |     |
| Query autocomplete                              | n                                       | n  | n           | n                                      | n                          | y         | n                           | n   | n         | y     | n                         | y        | n                 | n     | 3                    | 18%        |     |
| Search by named entities                        | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Search operators                                | n                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 12                   | 71%        |     |
| Limit the date range                            | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 8                    | 47%        |     |
| Search within newspaper segments / zin          | y                                       | n  | y           | y                                      | y                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 6                    | 35%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>On the result display</b>                    |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Display in distribution over time               | y                                       | n  | n           | n                                      | y                          | y         | y                           | n   | y         | n     | n                         | y        | y                 | n     | 7                    | 41%        |     |
| Display in distribution by newspaper cor        | n                                       | n  | n           | y                                      | y                          | n         | y                           | n   | y         | n     | y                         | n        | y                 | n     | 5                    | 29%        |     |
| Display in distribution by place names i        | y                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | y        | n                 | n     | 1                    | 6%         |     |
| Display of snippet preview                      | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 13                   | 76%        |     |
| Highlight of the searched keywords in f         | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 13                   | 76%        |     |
| Highlight of the searched keywords in C         | y                                       | y  | n           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 9                    | 53%        |     |
| Ngrams  | n                                       | y  | n           | y                                      | n                          | y         | n                           | y   | n         | y     | n                         | y        | n                 | y     | 4                    | 24%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Sort the results</b>                         |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| By relevance                                    | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | n         | y     | y                         | y        | y                 | y     | 13                   | 76%        |     |
| By date   | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 13                   | 76%        |     |
| By newspaper title                              | y                                       | y  | n           | y                                      | y                          | y         | y                           | n   | n         | n     | n                         | n        | n                 | n     | 4                    | 24%        |     |
| By article title                                | n                                       | y  | n           | y                                      | y                          | y         | y                           | n   | n         | n     | n                         | n        | n                 | n     | 2                    | 12%        |     |
| By content type (ad, article, illustration)     | n                                       | y  | y           | y                                      | y                          | y         | y                           | n   | n         | n     | n                         | n        | n                 | n     | 3                    | 18%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Filters (for search and display)</b>         |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Filter by newspaper titles                      | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 14                   | 82%        |     |
| Filter by publishing frequency                  | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Filter by political, religious, ... orientation | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | y         | n     | n                         | n        | n                 | n     | 1                    | 6%         |     |
| Filter by content types (i.e. article, ad, i    | y                                       | n  | y           | y                                      | y                          | n         | n                           | y   | n         | y     | n                         | n        | n                 | n     | 6                    | 35%        |     |
| Filter by events                                | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 1                    | 6%         |     |
| Filter by events                                | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 2                    | 12%        |     |
| Filter by persons                               | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 2                    | 12%        |     |
| Filter by organisations                         | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 2                    | 12%        |     |
| Filter by places mentioned in the text          | y                                       | n  | n           | n                                      | n                          | y         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 3                    | 18%        |     |
| Filter by time period                           | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | n                         | n        | n                 | n     | 12                   | 71%        |     |
| Filter by topics                                | y                                       | y  | y           | n                                      | n                          | y         | n                           | n   | n         | y     | n                         | n        | n                 | n     | 2                    | 12%        |     |
| Filter by geographic region (from newsj         | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | n         | y     | n                         | n        | n                 | n     | 6                    | 35%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Viewer</b>                                   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Facsimile displayed                             | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 15                   | 88%        |     |
| Optional OCRed text display                     | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 11                   | 65%        |     |
| Show full width/height (full page)              | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 13                   | 76%        |     |
| Interactive mini-map of page                    | n                                       | n  | n           | n                                      | y                          | n         | n                           | n   | n         | u     | n                         | n        | n                 | y     | 3                    | 18%        |     |
| Search in viewed page                           | n                                       | n  | n           | n                                      | y                          | n         | y                           | n   | n         | n     | y                         | n        | n                 | n     | 3                    | 18%        |     |
| Option to continue to next page                 | y                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 15                   | 88%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Personal account and user interactions</b>   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Save articles to favourites                     | n                                       | y  | y           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 6                    | 35%        |     |
| Save queries to favourites                      | n                                       | y  | n           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 3                    | 18%        |     |
| Tag articles                                    | n                                       | y  | n           | y                                      | y                          | y         | y                           | y   | y         | y     | y                         | y        | y                 | y     | 3                    | 18%        |     |
| Keep track of viewed materials                  | n                                       | y  | n           | u                                      | n                          | n         | n                           | n   | n         | n     | y                         | n        | n                 | n     | 1                    | 6%         |     |
| Article recommendations                         | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Permalinks                                      | n                                       | u  | y           | u                                      | y                          | y         | y                           | y   | y         | u     | y                         | y        | y                 | y     | 10                   | 59%        |     |
| Export citation                                 | n                                       | u  | n           | u                                      | y                          | n         | n                           | y   | n         | n     | u                         | n        | n                 | n     | 2                    | 12%        |     |
| Option to correct OCR                           | n                                       | y  | n           | y                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Option to correct the OLR, correct or ad        | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Screenshot tool                                 | n                                       | n  | y           | n                                      | y                          | y         | n                           | n   | y         | n     | n                         | n        | n                 | n     | 4                    | 24%        |     |
| Bulk downloads                                  | n                                       | u  | n           | u                                      | y                          | y         | y                           | y   | b         | n     | y                         | y        | n                 | y     | 7                    | 41%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Documentation of the digitisation</b>        |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| OLR at article level                            | n                                       | n  | n           | n                                      | n                          | y         | n                           | u   | n         | y     | y                         | n        | n                 | y     | 5                    | 29%        |     |
| OCR / OLR confidence scores                     | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Documentation of biases and shortcomm           | n                                       | n  | n           | n                                      | u                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Search result sorted by relevance score         | n                                       | n  | n           | n                                      | u                          | n         | n                           | n   | n         | n     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| NER   | n                                       | n  | n           | n                                      | u                          | u         | n                           | n   | n         | n     | y                         | n        | n                 | y     | 2                    | 12%        |     |
| Post-OCR corrections                            | n                                       | y  | u           | y                                      | u                          | u         | n                           | n   | n         | u     | u                         | n        | n                 | y     | 2                    | 12%        |     |
| Digitisation date at title level                | n                                       | n  | u           | n                                      | u                          | y         | y                           | n   | u         | n     | u                         | n        | n                 | n     | 0                    | 0%         |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Connectivity</b>                             |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| Third party identifiers (e.g. VIAF)             | n                                       | n  | n           | n                                      | n                          | n         | n                           | n   | n         | u     | n                         | n        | n                 | n     | 0                    | 0%         |     |
| Links to other repositories, semantic web       | n                                       | y  | n           | n                                      | n                          | n         | y                           | n   | n         | u     | n                         | n        | n                 | n     | 2                    | 12%        |     |
|   |   |  |             |  |                            |           |                             |   |           |       |                           |          |                   |       | 0                    | 0%         |     |
| <b>Overall results of each interface</b>        | 26                                      | 32                                       | 29          | 26                                     | 33                         | 30        | 21                          | 20  | 21        | 24    | 11                        | 34       | 16                | 24    | 25                   | 16         | 35  |
| Pourcentage                                     | 39%                                     | 48%                                      | 44%         | 39%                                    | 50%                        | 45%       | 32%                         | 30%   | 32%       | 36%   | 17%                       | 52%      | 24%               | 36%   | 38%                  | 24%        | 53% |

# Six assessment criteria for digital scholarship

## Source criticism (or *What am I looking at?*)

information on the corpus, on digitization, on OCR acquisition, etc.

## Content search (or *How do I engage with the contents?*)

search on various type of (enriched) material

## Content filtering (or *How do I select?*)

facets, result sorting, filtering

## Generosity (or *How do I discover?*)

corpus presentation, result display modes, recommendation techniques

## User interaction (or *How do I work?*)

personal work space functionalities

## Connectivity (or *How do I go beyond?*)

interlinking at the level of metadata and contents

# Main findings

## 4 generations of interfaces

*1st generation:* **giving access** - publishing keyword-searchable scans

*2nd generation:* **user interactions** - user collections, pdf download, store and share content, edit content

*3rd generation:* **semantic enrichment** - disambiguated named entities and crowd OCR correction

*4th generation:* **transparency**, recommendation

## Features are sparsely covered

on average, interfaces cover 38% of the considered features

### Best scores:

- keyword search
- viewer
- newspaper metadata

### Lowest scores:

- enrichment
- information on collection and on digitization
- query autocomplete
- filter by semantic annotations
- personal account and user interaction
- connectivity

# Open questions

**Audience** - How can we reconcile interfaces made for scholars vs. the general public? Should there be dedicated interfaces for each groups?

**Complexity** - Should all features and enrichments be visible and accessible?

**Tool integration** - Within or outside the interface ? Risk of complexifying the uses vs. an externalization via e.g. download functionalities?

# What is at stake

How best accommodate text analysis research tools and their usage by humanities scholars?

How did we co-design  
the interface



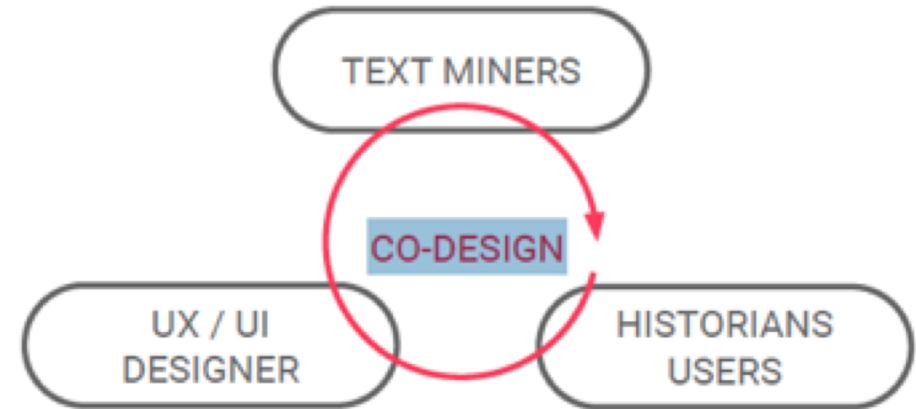
# Co-design and fast prototyping

## Principle

Continuous exchanges to learn from each other; experiment, evaluate

## Implementation

- Workshops
- Associated researchers
- *Impresso* community calls
- One-to-one collaborations



Au quotidien Par époque

**RETRONEWS**  
Le site de presse de la BnF

SE CONNECTER S'ABONNER

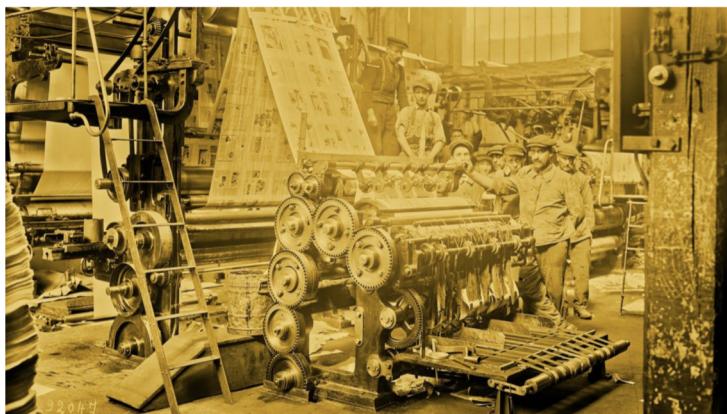
ÉCHO DE PRESSE

# Découvrez la nouvelle version du site RetroNews

---

le 16/04/2018 par RetroNews - modifié le 21/06/2018

ABONNEZ-VOUS



Imprimerie Gellé à Courbeil, Années 1920-1930. Source : Gallica, BnF.

The figure displays a network graph centered around Albert Einstein. The nodes are represented by icons and labels, connected by lines indicating relationships. Key nodes include:

- Mainleus**: A purple icon with a red checkmark.
- Princeton University**: A grey globe icon.
- Institute for Advanced Study**: A grey globe icon.
- Albert Einstein**: A blue icon with a portrait.
- Humboldt University of Berlin**: A grey globe icon.
- Isaac Newton**: A blue icon with a portrait.
- Satyendra Nath Bose**: A blue icon with a portrait.
- Max Planck**: A blue icon with a portrait.
- theori**: A grey icon with a gear.
- equat**: A grey icon with a gear.
- 1915**: A green icon with a clock.
- 1933**: A green icon with a clock.
- 1905**: A green icon with a clock.
- Bern**: A grey icon with a checkmark.

Annotations provide additional context:

- ACT-German-American physicist and founder of the theory of relativity
- ACT-German-American physicist and founder of the theory of relativity

|   | id  | firstname | surname | title        | function | qualifier    | demony       | timespan       | timespan_id |
|---|-----|-----------|---------|--------------|----------|--------------|--------------|----------------|-------------|
| 1 • <b>SELECT * FROM impresso_dev.mentions_ancillary;</b> |     |           |         |              |          |              |              |                |             |
|   | 3   | Claude    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:400-471 | 50-neth-2002   |             |
|   | 4   | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:495-512 | 50-neth-2002   |             |
|   | 5   | Michel    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:535-552 | 50-neth-2002   |             |
|   | 6   | Paul      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:572-589 | 50-neth-2002   |             |
|   | 7   | Claude    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:702-717 | 50-neth-2002   |             |
|   | 8   | Max       | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:839-846 | 50-neth-2002   |             |
|   | 9   | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:886-893 | 50-neth-2002   |             |
|   | 10  | Walter    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:916-925 | 50-neth-2002   |             |
|   | 11  | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:941-950 | 50-neth-2002   |             |
|   | 12  | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:981-990 | 50-neth-2002   |             |
|   | 13  | Fritz     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:126-129 | 50-neth-0      |             |
|   | 14  | Freddy    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:130-133 | 50-neth-0      |             |
|   | 15  | Col       | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:134-137 | 50-neth-0      |             |
|   | 16  | Meinrad   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:138-141 | 50-neth-0      |             |
|   | 17  | Hans      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:142-145 | 50-neth-0      |             |
|   | 18  | Fritz     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:176-179 | 50-neth-0      |             |
|   | 19  | Fritz     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:178-180 | 50-neth-0      |             |
|   | 20  | Adolf     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:178-180 | 50-neth-0      |             |
|   | 21  | Fred      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:183-187 | 50-neth-0      |             |
|   | 22  | Gottlieb  | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:188-192 | 50-neth-0      |             |
|   | 23  | Walter    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:193-197 | 50-neth-0      |             |
|   | 24  | Walter    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:198-199 | 145-neth-0     |             |
|   | 25  | Walter    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:200-201 | 145-neth-0     |             |
|   | 26  | Michel    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:191-196 | 50-neth-0      |             |
|   | 27  | Hans      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:196-199 | 50-neth-0      |             |
|   | 28  | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:202-205 | 2050-neth-0    |             |
|   | 29  | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:206-209 | 2050-neth-0    |             |
|   | 30  | Meinrad   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:210-213 | 2050-neth-0    |             |
|   | 31  | Adolf     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:214-217 | 2050-neth-0    |             |
|   | 32  | Adolf     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:218-221 | 2123-neth-0    |             |
|   | 33  | Albert    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:222-225 | 2123-neth-0    |             |
|   | 34  | André     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:226-229 | 2123-neth-0    |             |
|   | 35  | M. Rölk   | Le      | priédat      | QDL      | 1964-03-09-a | 0008:230-233 | 2123-neth-0    |             |
|   | 36  | Charles   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:349-349 | 349-50-neth-0  |             |
|   | 37  | Charles   | Le      | Madame       | QDL      | 1964-03-09-a | 0008:350-353 | 349-50-neth-0  |             |
|   | 38  | Paul-P.   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:354-357 | 349-50-neth-0  |             |
|   | 39  | Paul-P.   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:358-361 | 349-50-neth-0  |             |
|   | 40  | Danielle  | Le      | Mademoiselle | QDL      | 1964-03-09-a | 0008:362-365 | 349-50-neth-0  |             |
|   | 41  | Paul      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:366-369 | 349-50-neth-0  |             |
|   | 42  | Magdalene | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:370-373 | 349-50-neth-0  |             |
|   | 43  | Edouard   | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:374-377 | 470-495-neth-0 |             |
|   | 44  | Magdalene | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:378-381 | 470-495-neth-0 |             |
|   | 45  | Amélie    | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:382-385 | 470-495-neth-0 |             |
|   | 46  | Charles   | Le      | Madame       | QDL      | 1964-03-09-a | 0008:386-389 | 470-495-neth-0 |             |
|   | 47  | Magdalene | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:390-393 | 470-495-neth-0 |             |
|   | 48  | Jean      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:394-397 | 470-495-neth-0 |             |
|   | 49  | Jean      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:398-401 | 470-495-neth-0 |             |
|   | 50  | Paul      | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:402-405 | 470-495-neth-0 |             |
|   | 51  | CS        | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:406-409 | 470-495-neth-0 |             |
|   | 52  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:410-413 | 470-495-neth-0 |             |
|   | 53  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:414-417 | 470-495-neth-0 |             |
|   | 54  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:418-421 | 470-495-neth-0 |             |
|   | 55  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:422-425 | 470-495-neth-0 |             |
|   | 56  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:426-429 | 470-495-neth-0 |             |
|   | 57  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:430-433 | 470-495-neth-0 |             |
|   | 58  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:434-437 | 470-495-neth-0 |             |
|   | 59  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:438-441 | 470-495-neth-0 |             |
|   | 60  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:442-445 | 470-495-neth-0 |             |
|   | 61  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:446-449 | 470-495-neth-0 |             |
|   | 62  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:450-453 | 470-495-neth-0 |             |
|   | 63  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:454-457 | 470-495-neth-0 |             |
|   | 64  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:458-461 | 470-495-neth-0 |             |
|   | 65  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:462-465 | 470-495-neth-0 |             |
|   | 66  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:466-469 | 470-495-neth-0 |             |
|   | 67  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:470-473 | 470-495-neth-0 |             |
|   | 68  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:474-477 | 470-495-neth-0 |             |
|   | 69  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:478-481 | 470-495-neth-0 |             |
|   | 70  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:482-485 | 470-495-neth-0 |             |
|   | 71  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:486-489 | 470-495-neth-0 |             |
|   | 72  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:490-493 | 470-495-neth-0 |             |
|   | 73  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:494-497 | 470-495-neth-0 |             |
|   | 74  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:498-501 | 470-495-neth-0 |             |
|   | 75  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:502-505 | 470-495-neth-0 |             |
|   | 76  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:506-509 | 470-495-neth-0 |             |
|   | 77  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:510-513 | 470-495-neth-0 |             |
|   | 78  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:514-517 | 470-495-neth-0 |             |
|   | 79  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:518-521 | 470-495-neth-0 |             |
|   | 80  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:522-525 | 470-495-neth-0 |             |
|   | 81  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:526-529 | 470-495-neth-0 |             |
|   | 82  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:530-533 | 470-495-neth-0 |             |
|   | 83  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:534-537 | 470-495-neth-0 |             |
|   | 84  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:538-541 | 470-495-neth-0 |             |
|   | 85  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:542-545 | 470-495-neth-0 |             |
|   | 86  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:546-549 | 470-495-neth-0 |             |
|   | 87  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:550-553 | 470-495-neth-0 |             |
|   | 88  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:554-557 | 470-495-neth-0 |             |
|   | 89  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:558-561 | 470-495-neth-0 |             |
|   | 90  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:562-565 | 470-495-neth-0 |             |
|   | 91  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:566-569 | 470-495-neth-0 |             |
|   | 92  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:570-573 | 470-495-neth-0 |             |
|   | 93  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:574-577 | 470-495-neth-0 |             |
|   | 94  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:578-581 | 470-495-neth-0 |             |
|   | 95  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:582-585 | 470-495-neth-0 |             |
|   | 96  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:586-589 | 470-495-neth-0 |             |
|   | 97  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:590-593 | 470-495-neth-0 |             |
|   | 98  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:594-597 | 470-495-neth-0 |             |
|   | 99  | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:598-601 | 470-495-neth-0 |             |
|   | 100 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:602-605 | 470-495-neth-0 |             |
|   | 101 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:606-609 | 470-495-neth-0 |             |
|   | 102 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:610-613 | 470-495-neth-0 |             |
|   | 103 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:614-617 | 470-495-neth-0 |             |
|   | 104 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:618-621 | 470-495-neth-0 |             |
|   | 105 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:622-625 | 470-495-neth-0 |             |
|   | 106 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:626-629 | 470-495-neth-0 |             |
|   | 107 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:630-633 | 470-495-neth-0 |             |
|   | 108 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:634-637 | 470-495-neth-0 |             |
|   | 109 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:638-641 | 470-495-neth-0 |             |
|   | 110 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:642-645 | 470-495-neth-0 |             |
|   | 111 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:646-649 | 470-495-neth-0 |             |
|   | 112 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:650-653 | 470-495-neth-0 |             |
|   | 113 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:654-657 | 470-495-neth-0 |             |
|   | 114 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:658-661 | 470-495-neth-0 |             |
|   | 115 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:662-665 | 470-495-neth-0 |             |
|   | 116 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:666-669 | 470-495-neth-0 |             |
|   | 117 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:670-673 | 470-495-neth-0 |             |
|   | 118 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:674-677 | 470-495-neth-0 |             |
|   | 119 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:678-681 | 470-495-neth-0 |             |
|   | 120 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:682-685 | 470-495-neth-0 |             |
|   | 121 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:686-689 | 470-495-neth-0 |             |
|   | 122 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:690-693 | 470-495-neth-0 |             |
|   | 123 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:694-697 | 470-495-neth-0 |             |
|   | 124 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:698-701 | 470-495-neth-0 |             |
|   | 125 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:702-705 | 470-495-neth-0 |             |
|   | 126 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:706-709 | 470-495-neth-0 |             |
|   | 127 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:710-713 | 470-495-neth-0 |             |
|   | 128 | Elvia     | Le      | Monsieur     | QDL      | 1964-03-09-a | 0008:714-717 | 470-495-neth-0 |             |
|   | 129 | Elvia     | Le      | Monsieur     | QDL      | 1964-03      |              |                |             |

# Research scenarios

## Bridging the fields: Research scenarios

Tracking the anti-European posture in the public debate in Switzerland and Luxembourg (1848-1945)

1. Goal: Identify the postures in the debates around the European idea
2. Using people and slogans to collect a broad corpus of articles, across several newspaper titles / cluster them according to manual classification and automatically generated information
3. Quantity and diversity of the collected results / types of incarnation of the European idea.

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

1

## Bridging the fields: Research scenarios

Funding secondary education in nineteenth century Europe: structuration of a public debate

1. Research question : how educational policies were addressed by journalists and which data were deemed necessary to support the given point of view;
2. Help me find: articles dealing with educational policies, their "type" and the media associated to them ; linked named entities to "follow them";
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

4

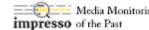


## The coverage of The Battle of Arnhem in European Newspapers (1944-present)

1. Help me find all articles which cover the events at Arnhem: When does coverage start?
2. Help me discover similarities and differences in articles about the battle 1944 to present
3. Help me understand how "popularity" of the battle is changing over time. How could that be measured?

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

2



## Bridging the fields: Research scenarios

How do newspapers reflects the history of computing in Switzerland ?

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

5

## ALIENS IN NEWSPAPERS

What we want to look for:

- Concomitance with an event (astronomical discovery, UFO, release of a movie, publication of a book...)
- "Keyword in context"(-history of representations)
  - Characterization of these non-humans (judgement, anthropomorphism...)
  - Links with semantic fields: war, scare, space, colonization, creatures, intelligence...
  - Mentions of persons (e.g. authors, directors, astronomers...)
- A popular subject? (origin of these newspapers, type of the content-fiction, interviews...size of the articles...)
- Evolutions of representations through time & possible differences between countries

French keywords: extraterrestre, martien, sélénite, alien...

First results on e-newspaperarchives.ch:

- 1289 results for "extraterrestre" from 1896 to 2012
- 26 results for "sélénite" from 1866 to 2010
- 1097 results for "martien" from 1863 to 2009
- 8725 results for "alien" from 1872 to 2014

Difficulties:

- Substantivized adjectives: "extraterrestre" can apply to phenomena and things too
- Homonyms and similar words:
  - "martien" often mistaken for "Martigny" (Swiss city)
  - "alien" often mistaken for "[Woody] alien", "alien", "aliéné" and words with -alien/alien (312 585 results on Retronews !)
  - "alien": English words with numerous irrelevant meanings

3

## Journalistic Cultures during the 19th Century: a Comparison between the NZZ and le JdG

Main focuses (among others): genre change and meta-discourses

The three most important operationalisations (cf. handout)

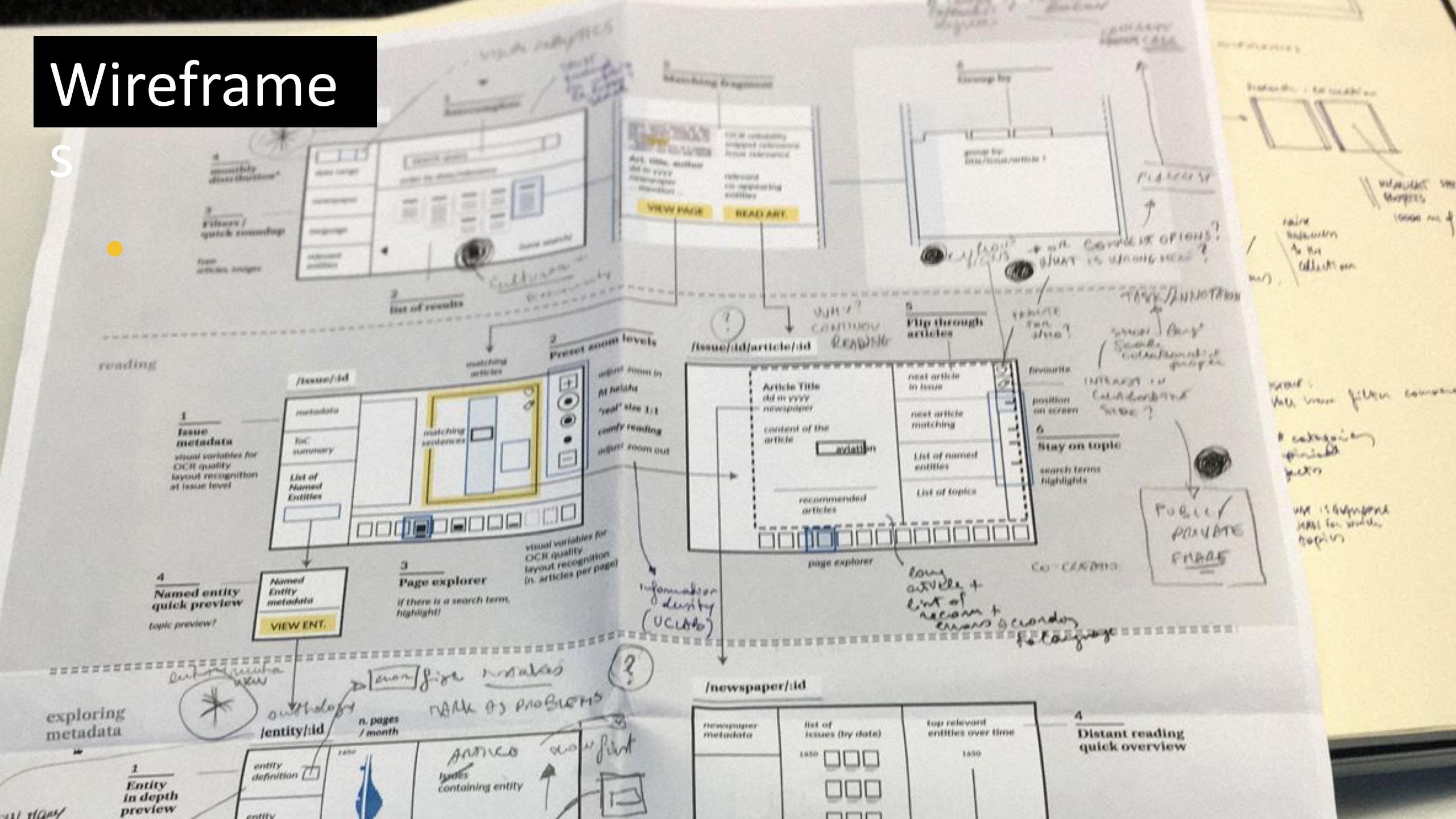
- Help me find the lengths of sections (German: Rubriken, French: rubriques) in both newspapers and the change of these lengths in time (Handout: 1.1.)
- Help me find moments of emergence of sections in newspapers (Handout: 1.1.2.)
- Help me find/collect words and collocations related to sections (Handout: 1.2.1.)

Use-Case for Impresso Project, Tobias von Wachter, 4th and 5th of July 2016

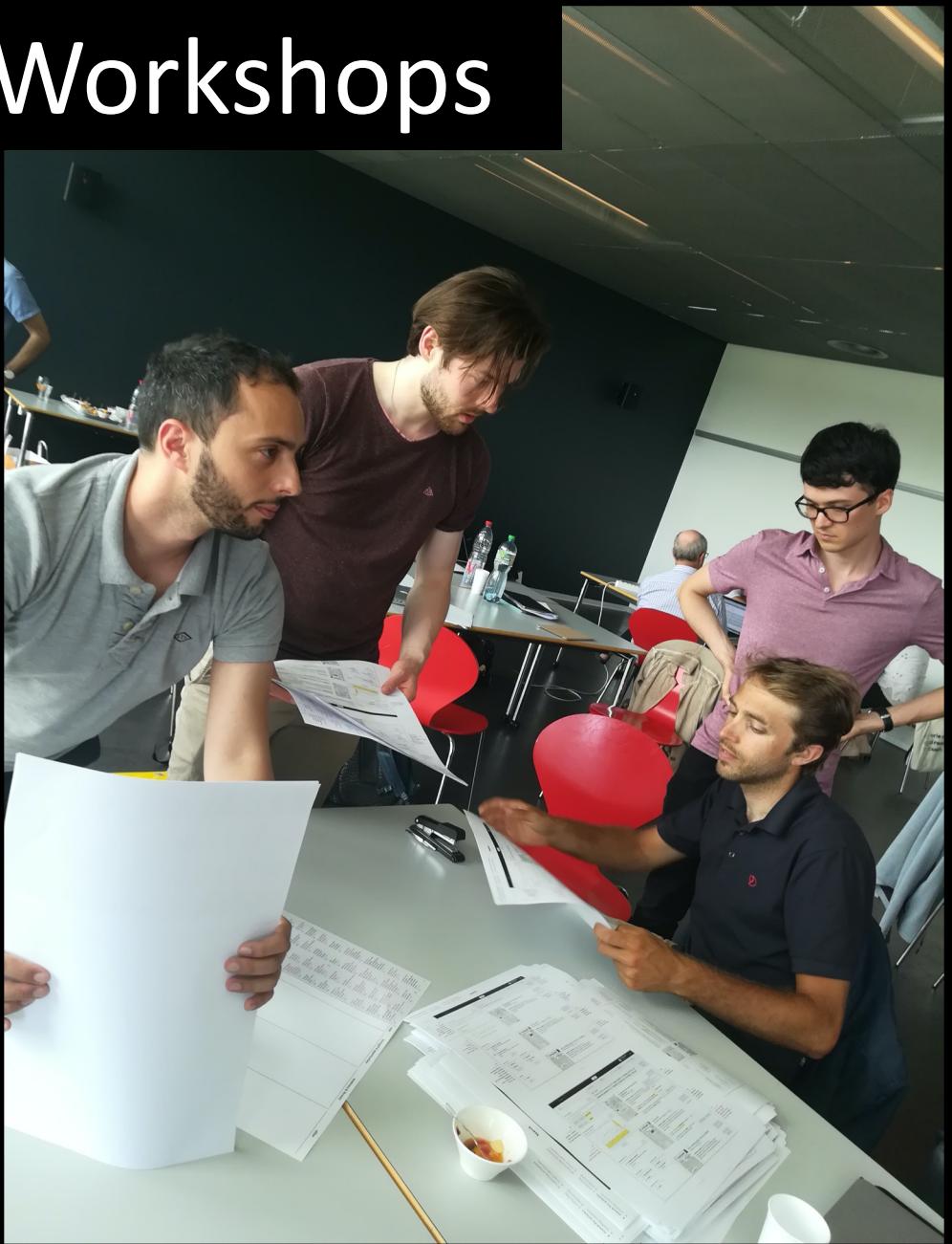
Universität Basel 6

6

# Wireframe



# Workshops



**s** search for articles   **c** collecting  
**R** reading contents   **O** organising

**M** exploring metadata  
**E** visual experiments

code for the SWOT

component description

**S.1**  
search autocomplete:  
based on input,  
suggests words,  
date ranges, named  
entities or  
article categories

red text: uncompleted features

**S.2**  
timeline  
n. results per year.  
This acts as  
"date range" filter

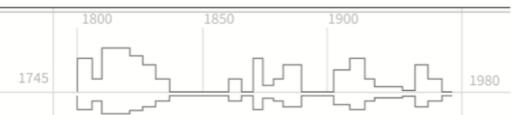
explore ▾

YOUR SEARCH PAST SEARCHES

FULL TEXT SEARCH ... type a text or a date or to start ... X

REFINE ...

PUBLISHED IN (DATE RANGE) TIMELINE OF N. ARTICLES N. ISSUES



select a date range to show articles published

NEWSPAPER TITLES

53625 **La Gazette de Lausanne** quotidien suisse de langue française édité à Lausanne

43864 **Le Journal de Genève** quotidien suisse qui a paru du 1826 au 28 février 1998

MORE ...

ARTICLE TYPE

112 **partisans**

50 **satirique**

MORE ...

TOP NAMED ENTITIES

103 **Napoleon**

50 **Zurich, Suisse** location

MORE ...

LANGUAGE

112 **French**

MORE ...

**S.3**  
metadata filters /  
quick roundup  
how many  
search results  
per newspapers titles,  
languages,  
page content tags,  
format tags

## Search



search

GROUP RESULTS BY ISSUE PAGE ARTICLES SENTENCES

SEARCH SUMMARY  
Find 53625987 articles in our collection.



1 of 112

**Les recherches d'INFORMATION**  
INTERNATIONAL NEWS

**Gazette de Lausanne**  
Les recherches d'une... possibilités JEAN-PIERRE de multiples...



2 of 112

**Les USA et l'AFN**  
INTERNATIONAL NEWS

**Gazette de Lausanne**  
Les USA et l'AFN Un... la terrasse de café o... contre l'assaut...



3 of 112

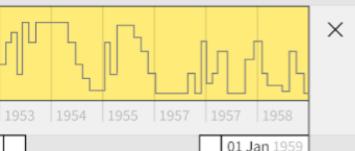
**FRANCO PRÉFÈRE**  
INTERNATIONAL NEWS

**Gazette de Lausanne**  
FRANCO PRÉFÈRE L... PRINTEMPS D'ESPAGNE

| code   | Strength & opportunities | Weaknesses |
|--|--------------------------|------------|
| R.13<br>R.14<br>R.15<br>R.16<br>R.17<br>C.1<br>C.2<br>C.3<br>C.4<br>C.5<br>C.6<br>C.7<br>C.8<br>C.9<br>C.10<br>C.11<br>C.12<br>C.13<br>C.14<br>C.15<br>C.16<br>C.17<br>C.18<br>C.19<br>C.20<br>C.21<br>C.22<br>C.23<br>C.24<br>C.25<br>C.26<br>C.27<br>C.28<br>C.29<br>C.30<br>C.31<br>C.32<br>C.33<br>C.34<br>C.35<br>C.36<br>C.37<br>C.38<br>C.39<br>C.40<br>C.41<br>C.42<br>C.43<br>C.44<br>C.45<br>C.46<br>C.47<br>C.48<br>C.49<br>C.50<br>C.51<br>C.52<br>C.53<br>C.54<br>C.55<br>C.56<br>C.57<br>C.58<br>C.59<br>C.60<br>C.61<br>C.62<br>C.63<br>C.64<br>C.65<br>C.66<br>C.67<br>C.68<br>C.69<br>C.70<br>C.71<br>C.72<br>C.73<br>C.74<br>C.75<br>C.76<br>C.77<br>C.78<br>C.79<br>C.80<br>C.81<br>C.82<br>C.83<br>C.84<br>C.85<br>C.86<br>C.87<br>C.88<br>C.89<br>C.90<br>C.91<br>C.92<br>C.93<br>C.94<br>C.95<br>C.96<br>C.97<br>C.98<br>C.99<br>C.100<br>C.101<br>C.102<br>C.103<br>C.104<br>C.105<br>C.106<br>C.107<br>C.108<br>C.109<br>C.110<br>C.111<br>C.112<br>C.113<br>C.114<br>C.115<br>C.116<br>C.117<br>C.118<br>C.119<br>C.120<br>C.121<br>C.122<br>C.123<br>C.124<br>C.125<br>C.126<br>C.127<br>C.128<br>C.129<br>C.130<br>C.131<br>C.132<br>C.133<br>C.134<br>C.135<br>C.136<br>C.137<br>C.138<br>C.139<br>C.140<br>C.141<br>C.142<br>C.143<br>C.144<br>C.145<br>C.146<br>C.147<br>C.148<br>C.149<br>C.150<br>C.151<br>C.152<br>C.153<br>C.154<br>C.155<br>C.156<br>C.157<br>C.158<br>C.159<br>C.160<br>C.161<br>C.162<br>C.163<br>C.164<br>C.165<br>C.166<br>C.167<br>C.168<br>C.169<br>C.170<br>C.171<br>C.172<br>C.173<br>C.174<br>C.175<br>C.176<br>C.177<br>C.178<br>C.179<br>C.180<br>C.181<br>C.182<br>C.183<br>C.184<br>C.185<br>C.186<br>C.187<br>C.188<br>C.189<br>C.190<br>C.191<br>C.192<br>C.193<br>C.194<br>C.195<br>C.196<br>C.197<br>C.198<br>C.199<br>C.200<br>C.201<br>C.202<br>C.203<br>C.204<br>C.205<br>C.206<br>C.207<br>C.208<br>C.209<br>C.210<br>C.211<br>C.212<br>C.213<br>C.214<br>C.215<br>C.216<br>C.217<br>C.218<br>C.219<br>C.220<br>C.221<br>C.222<br>C.223<br>C.224<br>C.225<br>C.226<br>C.227<br>C.228<br>C.229<br>C.230<br>C.231<br>C.232<br>C.233<br>C.234<br>C.235<br>C.236<br>C.237<br>C.238<br>C.239<br>C.240<br>C.241<br>C.242<br>C.243<br>C.244<br>C.245<br>C.246<br>C.247<br>C.248<br>C.249<br>C.250<br>C.251<br>C.252<br>C.253<br>C.254<br>C.255<br>C.256<br>C.257<br>C.258<br>C.259<br>C.260<br>C.261<br>C.262<br>C.263<br>C.264<br>C.265<br>C.266<br>C.267<br>C.268<br>C.269<br>C.270<br>C.271<br>C.272<br>C.273<br>C.274<br>C.275<br>C.276<br>C.277<br>C.278<br>C.279<br>C.280<br>C.281<br>C.282<br>C.283<br>C.284<br>C.285<br>C.286<br>C.287<br>C.288<br>C.289<br>C.290<br>C.291<br>C.292<br>C.293<br>C.294<br>C.295<br>C.296<br>C.297<br>C.298<br>C.299<br>C.300<br>C.301<br>C.302<br>C.303<br>C.304<br>C.305<br>C.306<br>C.307<br>C.308<br>C.309<br>C.310<br>C.311<br>C.312<br>C.313<br>C.314<br>C.315<br>C.316<br>C.317<br>C.318<br>C.319<br>C.320<br>C.321<br>C.322<br>C.323<br>C.324<br>C.325<br>C.326<br>C.327<br>C.328<br>C.329<br>C.330<br>C.331<br>C.332<br>C.333<br>C.334<br>C.335<br>C.336<br>C.337<br>C.338<br>C.339<br>C.340<br>C.341<br>C.342<br>C.343<br>C.344<br>C.345<br>C.346<br>C.347<br>C.348<br>C.349<br>C.350<br>C.351<br>C.352<br>C.353<br>C.354<br>C.355<br>C.356<br>C.357<br>C.358<br>C.359<br>C.360<br>C.361<br>C.362<br>C.363<br>C.364<br>C.365<br>C.366<br>C.367<br>C.368<br>C.369<br>C.370<br>C.371<br>C.372<br>C.373<br>C.374<br>C.375<br>C.376<br>C.377<br>C.378<br>C.379<br>C.380<br>C.381<br>C.382<br>C.383<br>C.384<br>C.385<br>C.386<br>C.387<br>C.388<br>C.389<br>C.390<br>C.391<br>C.392<br>C.393<br>C.394<br>C.395<br>C.396<br>C.397<br>C.398<br>C.399<br>C.400<br>C.401<br>C.402<br>C.403<br>C.404<br>C.405<br>C.406<br>C.407<br>C.408<br>C.409<br>C.410<br>C.411<br>C.412<br>C.413<br>C.414<br>C.415<br>C.416<br>C.417<br>C.418<br>C.419<br>C.420<br>C.421<br>C.422<br>C.423<br>C.424<br>C.425<br>C.426<br>C.427<br>C.428<br>C.429<br>C.430<br>C.431<br>C.432<br>C.433<br>C.434<br>C.435<br>C.436<br>C.437<br>C.438<br>C.439<br>C.440<br>C.441<br>C.442<br>C.443<br>C.444<br>C.445<br>C.446<br>C.447<br>C.448<br>C.449<br>C.450<br>C.451<br>C.452<br>C.453<br>C.454<br>C.455<br>C.456<br>C.457<br>C.458<br>C.459<br>C.460<br>C.461<br>C.462<br>C.463<br>C.464<br>C.465<br>C.466<br>C.467<br>C.468<br>C.469<br>C.470<br>C.471<br>C.472<br>C.473<br>C.474<br>C.475<br>C.476<br>C.477<br>C.478<br>C.479<br>C.480<br>C.481<br>C.482<br>C.483<br>C.484<br>C.485<br>C.486<br>C.487<br>C.488<br>C.489<br>C.490<br>C.491<br>C.492<br>C.493<br>C.494<br>C.495<br>C.496<br>C.497<br>C.498<br>C.499<br>C.500<br>C.501<br>C.502<br>C.503<br>C.504<br>C.505<br>C.506<br>C.507<br>C.508<br>C.509<br>C.510<br>C.511<br>C.512<br>C.513<br>C.514<br>C.515<br>C.516<br>C.517<br>C.518<br>C.519<br>C.520<br>C.521<br>C.522<br>C.523<br>C.524<br>C.525<br>C.526<br>C.527<br>C.528<br>C.529<br>C.530<br>C.531<br>C.532<br>C.533<br>C.534<br>C.535<br>C.536<br>C.537<br>C.538<br>C.539<br>C.540<br>C.541<br>C.542<br>C.543<br>C.544<br>C.545<br>C.546<br>C.547<br>C.548<br>C.549<br>C.550<br>C.551<br>C.552<br>C.553<br>C.554<br>C.555<br>C.556<br>C.557<br>C.558<br>C.559<br>C.560<br>C.561<br>C.562<br>C.563<br>C.564<br>C.565<br>C.566<br>C.567<br>C.568<br>C.569<br>C.570<br>C.571<br>C.572<br>C.573<br>C.574<br>C.575<br>C.576<br>C.577<br>C.578<br>C.579<br>C.580<br>C.581<br>C.582<br>C.583<br>C.584<br>C.585<br>C.586<br>C.587<br>C.588<br>C.589<br>C.590<br>C.591<br>C.592<br>C.593<br>C.594<br>C.595<br>C.596<br>C.597<br>C.598<br>C.599<br>C.600<br>C.601<br>C.602<br>C.603<br>C.604<br>C.605<br>C.606<br>C.607<br>C.608<br>C.609<br>C.610<br>C.611<br>C.612<br>C.613<br>C.614<br>C.615<br>C.616<br>C.617<br>C.618<br>C.619<br>C.620<br>C.621<br>C.622<br>C.623<br>C.624<br>C.625<br>C.626<br>C.627<br>C.628<br>C.629<br>C.630<br>C.631<br>C.632<br>C.633<br>C.634<br>C.635<br>C.636<br>C.637<br>C.638<br>C.639<br>C.640<br>C.641<br>C.642<br>C.643<br>C.644<br>C.645<br>C.646<br>C.647<br>C.648<br>C.649<br>C.650<br>C.651<br>C.652<br>C.653<br>C.654<br>C.655<br>C.656<br>C.657<br>C.658<br>C.659<br>C.660<br>C.661<br>C.662<br>C.663<br>C.664<br>C.665<br>C.666<br>C.667<br>C.668<br>C.669<br>C.660<br>C.661<br>C.662<br>C.663<br>C.664<br>C.665<br>C.666<br>C.667<br>C.668<br>C.669<br>C.670<br>C.671<br>C.672<br>C.673<br>C.674<br>C.675<br>C.676<br>C.677<br>C.678<br>C.679<br>C.680<br>C.681<br>C.682<br>C.683<br>C.684<br>C.685<br>C.686<br>C.687<br>C.688<br>C.689<br>C.690<br>C.691<br>C.692<br>C.693<br>C.694<br>C.695<br>C.696<br>C.697<br>C.698<br>C.699<br>C.700<br>C.701<br>C.702<br>C.703<br>C.704<br>C.705<br>C.706<br>C.707<br>C.708<br>C.709<br>C.710<br>C.711<br>C.712<br>C.713<br>C.714<br>C.715<br>C.716<br>C.717<br>C.718<br>C.719<br>C.720<br>C.721<br>C.722<br>C.723<br>C.724<br>C.725<br>C.726<br>C.727<br>C.728<br>C.729<br>C.720<br>C.721<br>C.722<br>C.723<br>C.724<br>C.725<br>C.726<br>C.727<br>C.728<br>C.729<br>C.730<br>C.731<br>C.732<br>C.733<br>C.734<br>C.735<br>C.736<br>C.737<br>C.738<br>C.739<br>C.730<br>C.731<br>C.732<br>C.733<br>C.734<br>C.735<br>C.736<br>C.737<br>C.738<br>C.739<br>C.740<br>C.741<br>C.742<br>C.743<br>C.744<br>C.745<br>C.746<br>C.747<br>C.748<br>C.749<br>C.740<br>C.741<br>C.742<br>C.743<br>C.744<br>C.745<br>C.746<br>C.747<br>C.748<br>C.749<br>C.750<br>C.751<br>C.752<br>C.753<br>C.754<br>C.755<br>C.756<br>C.757<br>C.758<br>C.759<br>C.750<br>C.751<br>C.752<br>C.753<br>C.754<br>C.755<br>C.756<br>C.757<br>C.758<br>C.759<br>C.760<br>C.761<br>C.762<br>C.763<br>C.764<br>C.765<br>C.766<br>C.767<br>C.768<br>C.769<br>C.760<br>C.761<br>C.762<br>C.763<br>C.764<br>C.765<br>C.766<br>C.767<br>C.768<br>C.769<br>C.770<br>C.771<br>C.772<br>C.773<br>C.774<br>C.775<br>C.776<br>C.777<br>C.778<br>C.779<br>C.770<br>C.771<br>C.772<br>C.773<br>C.774<br>C.775<br>C.776<br>C.777<br>C.778<br>C.779<br>C.780<br>C.781<br>C.782<br>C.783<br>C.784<br>C.785<br>C.786<br>C.787<br>C.788<br>C.789<br>C.780<br>C.781<br>C.782<br>C.783<br>C.784<br>C.785<br>C.786<br>C.787<br>C.788<br>C.789<br>C.790<br>C.791<br>C.792<br>C.793<br>C.794<br>C.795<br>C.796<br>C.797<br>C.798<br>C.799<br>C.790<br>C.791<br>C.792<br>C.793<br>C.794<br>C.795<br>C.796<br>C.797<br>C.798<br>C.799<br>C.800<br>C.801<br>C.802<br>C.803<br>C.804<br>C.805<br>C.806<br>C.807<br>C.808<br>C.809<br>C.800<br>C.801<br>C.802<br>C.803<br>C.804<br>C.805<br>C.806<br>C.807<br>C.808<br>C.809<br>C.810<br>C.811<br>C.812<br>C.813<br>C.814<br>C.815<br>C.816<br>C.817<br>C.818<br>C.819<br>C.810<br>C.811<br>C.812<br>C.813<br>C.814<br>C.815<br>C.816<br>C.817<br>C.818<br>C.819<br>C.820<br>C.821<br>C.822<br>C.823<br>C.824<br>C.825<br>C.826<br>C.827<br>C.828<br>C.829<br>C.820<br>C.821<br>C.822<br>C.823<br>C.824<br>C.825<br>C.826<br>C.827<br>C.828<br>C.829<br>C.830<br>C.831<br>C.832<br>C.833<br>C.834<br>C.835<br>C.836<br>C.837<br>C.838<br>C.839<br>C.830<br>C.831<br>C.832<br>C.833<br>C.834<br>C.835<br>C.836<br>C.837<br>C.838<br>C.839<br>C.840<br>C.841<br>C.842<br>C.843<br>C.844<br>C.845<br>C.846<br>C.847<br>C.848<br>C.849<br>C.840<br>C.841<br>C.842<br>C.843<br>C.844<br>C.845<br>C.846<br>C.847<br>C.848<br>C.849<br>C.850<br>C.851<br>C.852<br>C.853<br>C.854<br>C.855<br>C.856<br>C.857<br>C.858<br>C.859<br>C.850<br>C.851<br>C.852<br>C.853<br>C.854<br>C.855<br>C.856<br>C.857<br>C.858<br>C.859<br>C.860<br>C.861<br>C.862<br>C.863<br>C.864<br>C.865<br>C.866<br>C.867<br>C.868<br>C.869<br>C.860<br>C.861<br>C.862<br>C.863<br>C.864<br>C.865<br>C.866<br>C.867<br>C.868<br>C.869<br>C.870<br>C.871<br>C.872<br>C.873<br>C.874<br>C.875<br>C.876<br>C.877<br>C.878<br>C.879<br>C.870<br>C.871<br>C.872<br>C.873<br>C.874<br>C.875<br>C.876<br>C.877<br>C.878<br>C.879<br>C.880<br>C.881<br>C.882<br>C.883<br>C.884<br>C.885<br>C.886<br>C.887<br>C.888<br>C.889<br>C.880<br>C.881<br>C.882<br>C.883<br>C.884<br>C.885<br>C.886<br>C.887<br>C.888<br>C.889<br>C.890<br>C.891<br>C.892<br>C.893<br>C.894<br>C.895<br>C.896<br>C.897<br>C.898<br>C.899<br>C.890<br>C.891<br>C.892<br>C.893<br>C.894<br>C.895<br>C.896<br>C.897<br>C.898<br>C.899<br>C.900<br>C.901<br>C.902<br>C.903<br>C.904<br>C.905<br>C.906<br>C.907<br>C.908<br>C.909<br>C.900<br>C.901<br>C.902<br>C.903<br>C.904<br>C.905<br>C.906<br>C.907<br>C.908<br>C.909<br>C.910<br>C.911<br>C.912<br>C.913<br>C.914<br>C.915<br>C.916<br>C.917<br>C.918<br>C.919<br>C.910<br>C.911<br>C.912<br>C.913<br>C.914<br>C.915<br>C.916<br>C.917<br>C.918<br>C.919<br>C.920<br>C.921<br>C.922<br>C.923<br>C.924<br>C.925<br>C.926<br>C.927<br>C.928<br>C.929<br>C.920<br>C.921<br>C.922<br>C.923<br>C.924<br>C.925<br>C.926<br>C.927<br>C.928<br>C.929<br>C.930<br>C.931<br>C.932<br>C.933<br>C.934<br>C.935<br>C.936<br>C.937<br>C.938<br>C.939<br>C.930<br>C.931<br>C.932<br>C.933<br>C.934<br>C.935<br>C.936<br>C.937<br>C.938<br>C.939<br>C.940<br>C.941<br>C.942<br>C.943<br>C.944<br>C.945<br>C.946<br>C.947<br>C.948<br>C.949<br>C.940<br>C.941<br>C.942<br>C.943<br>C.944<br>C.945<br>C.946<br>C.947<br>C.948<br>C.949<br>C.950<br>C.951<br>C.952<br>C.953<br>C.954<br>C.955<br>C.956<br>C.957<br>C.958<br>C.959<br>C.950<br>C.951<br>C.952<br>C.953<br>C.954<br>C.955<br>C.956<br>C.957<br>C.958<br>C.959<br>C.960<br>C.961<br>C.962<br>C.963<br>C.964<br>C.965<br>C.966<br>C.967<br>C.968<br>C.969<br>C.960<br>C.961<br>C.962<br>C.963<br>C.964<br>C.965<br>C.966<br>C.967<br>C.968<br>C.969<br>C.970<br>C.971<br>C.972<br>C.973<br>C.974<br>C.975<br>C.976<br>C.977<br>C.978<br>C.979<br>C.970<br>C.971<br>C.972<br>C.973<br>C.974<br>C.975<br>C.976<br>C.977<br>C.978<br>C.979<br>C.980<br>C.981<br>C.982<br>C.983<br>C.984<br>C.985<br>C.986<br>C.987<br>C.988<br>C.989<br>C.980<br>C.981<br>C.982<br>C.983<br>C.984<br>C.985<br>C.986<br>C.987<br>C.988<br>C.989<br>C.990<br>C.991<br>C.992<br>C.993<br>C.994<br>C.995<br>C.996<br>C.997<br>C.998<br>C.999<br>C.990<br>C.991<br>C.992<br>C.993<br>C.994<br>C.995<br>C.996<br>C.997<br>C.998<br>C.999<br>C.1000<br>C.1001<br>C.1002<br>C.1003<br>C.1004<br>C.1005<br>C.1006<br>C.1007<br>C.1008<br>C.1009<br>C.1000<br>C.1001<br>C.1002<br>C.1003<br>C.1004<br>C.1005<br>C.1006<br>C.1007<br>C.1008<br>C.1009<br>C.1010<br>C.1011<br>C.1012<br>C.1013<br>C.1014<br>C.1015<br>C.1016<br>C.1017<br>C.1 |                          |            |

# Conception of components

## Discussions



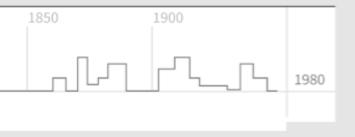
◆ article snippet XL

**LAUSANNE**

**Lausanne** — Les récits au caractère de l'actualité sont rares dans les journaux de Lausanne, mais certains d'entre eux sont tout à fait intéressants. C'est le cas de l'entrevue avec Georges Tissot de l'Institut de psychologie de l'université de Genève, qui a été publiée dans *Le Matin*, le 2 mai. Dans cette interview, Georges Tissot parle de ses recherches sur la psychopathologie de l'adolescence et de son travail de thérapeute. Il explique que les adolescents sont des personnes très sensibles et qu'il est important de leur donner une attention particulière.

EXPORT CITATIONS

DOWNLOAD AS ...



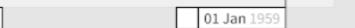
◆ article snippet XL bis

**Gazette de Lausanne**

**Gazette de Lausanne** — Le journal de la ville de Lausanne, fondé en 1850, a été édité par la ville de Lausanne jusqu'en 1900. Ensuite, il a été édité par la ville de Lausanne jusqu'en 1980. Le journal a été édité par la ville de Lausanne jusqu'en 1980.

EXPORT CITATIONS ...

DOWNLOAD AS ...



01 Jan 1959

◆ snippet search timeline

1890 | 310 ISSUES / 259 ARTICLES

SEARCH THIS YEAR

◆ named entity

**Napoleon**

Napoléon Bonaparte (French: [napoleōn bōnapār]; 15 August 1769 – 5 May 1821) was a French statesman and military leader.

1342 ARTICLES

◆ named entity bis

**Pasquale Paoli**

Nationalist Corsican leader.

156 ARTICLES

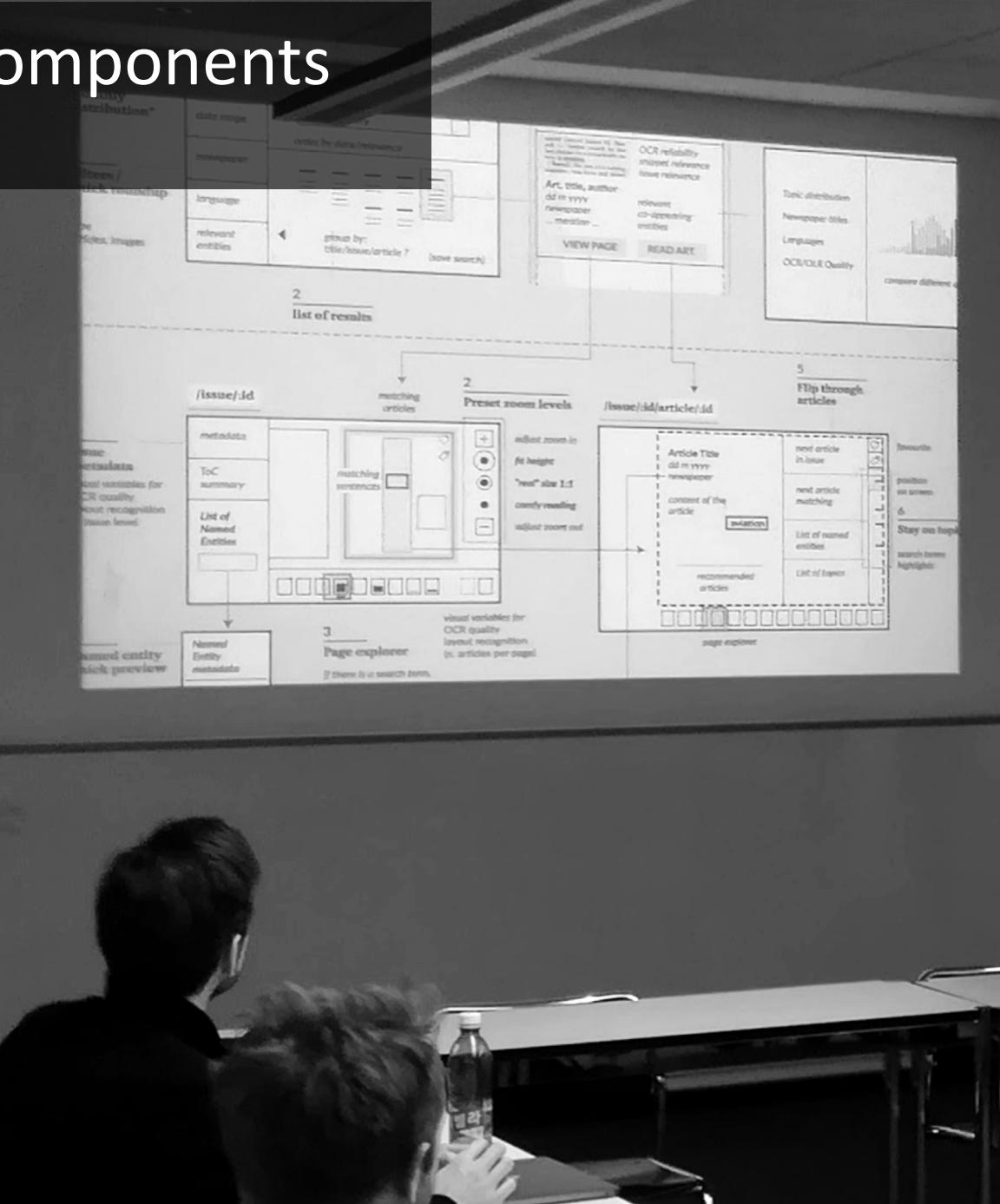
◆ snippet compare

Friday, April 27, 1951

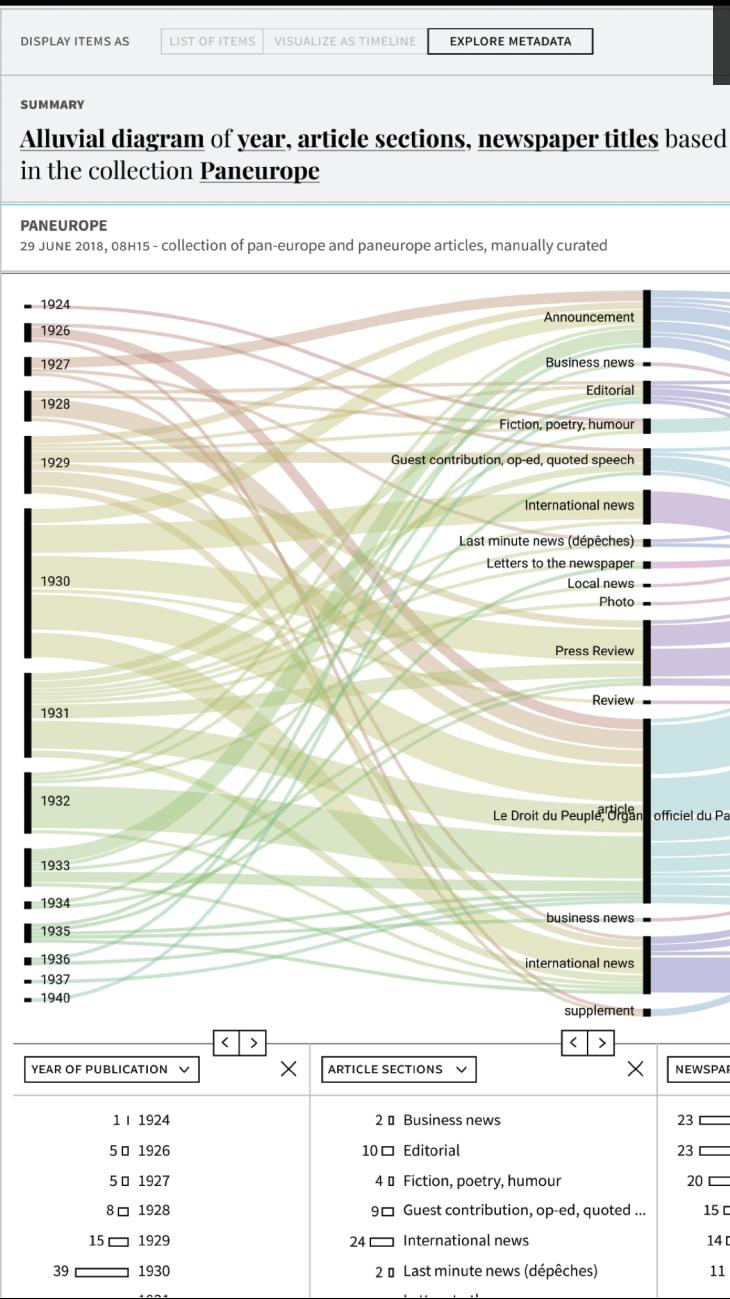
**FRANCO PRÉFÈRE LE TÊTE-A-TÈTE AVEC L'AMÉRIQUE****Gazette de Lausanne**

FRANCO PRÉFÈRE LE TÊTE-A-TÈTE AVEC L'AMÉRIQUE SOUS LE PRINTEMPS D'ESPAGNE ATTENDANT UN TARDIF REPENTIR FRANCO-ANGLAIS ;/ Par notre envoyé spécial Michel CLERC ! ! ,,...

VIEW ARTICLE



personal space / my collections



# Exemple: Inspect & Compare



#design

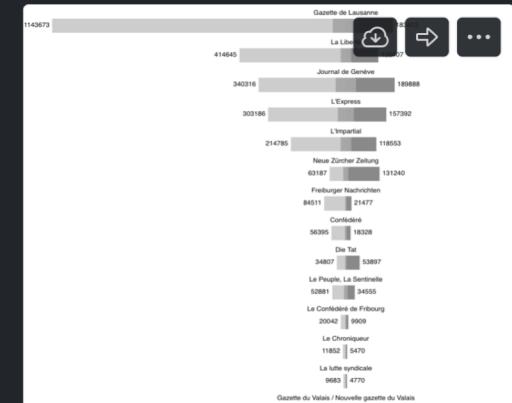
12 | 1 | Add a topic

SMH Zeitung 1971  
Le Monde 2009  
Abo 2015  
St. Galler Zeitung 0  
VFT 2014  
VFT 2014



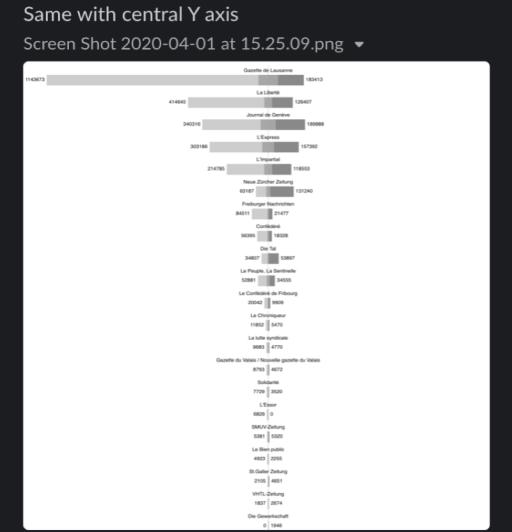
Roman 3:22 PM

What do you think about intersection pattern @daniele, @Pau  
Screen Shot 2020-04-01 at 15.22.09.png ▾



Same with central Y axis

Screen Shot 2020-04-01 at 15.25.09.png ▾



daniele 3:28 PM

fantastic! what if you invert the last two colors? so that the inte

# Exemple: Inspect & Compare

**A** QUERY \* COLLECTION

"arnhem OR arnhem(3 more)"  match · équipe · ligue · club · saise  REFINER ...

search for ...  ADD FILTER...

OPEN IN SEARCH PAGE... (5,586 RESULTS)

**181 results in common**

Distribution of newspapers, named entities and topics for articles which appear both in **(A)** and **(B)**.

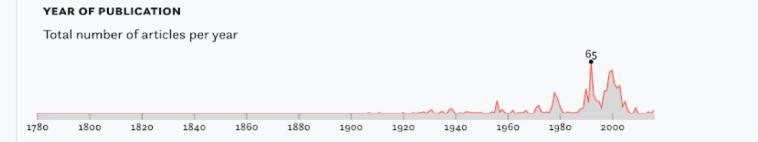
OPEN IN SEARCH PAGE... (181 RESULTS)

**B** QUERY \* COLLECTION

"arnhem OR arnhem(3 more)"  "match"  REFINER ...

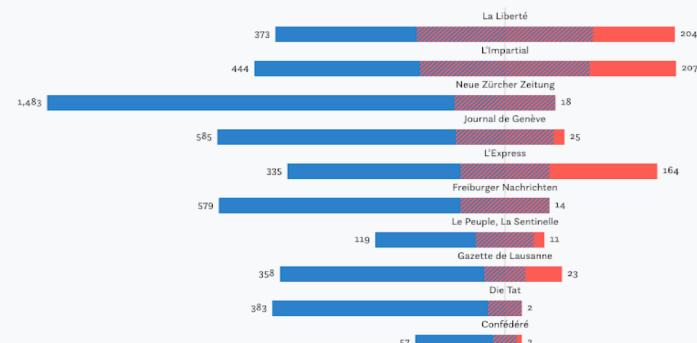
search for ...  ADD FILTER...

OPEN IN SEARCH PAGE... (670 RESULTS)



SCALE: SQUARE ROOT  SORT BY ABSOLUTE INTERSECTION

### NEWSPAPER (10 OPTIONS)



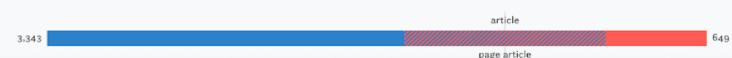
### COUNTRY (1 OPTION)



### LANGUAGE (2 OPTIONS)

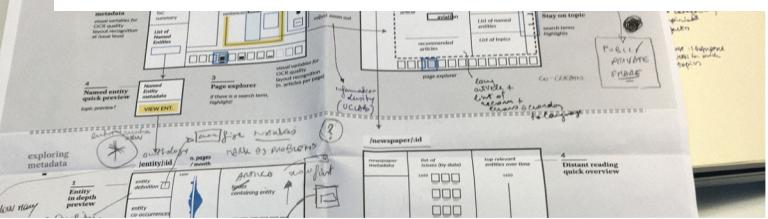


### TYPE (3 OPTIONS)





# Lessons learned: User requirements & tool adoption



With **time, prototypes and practice** original user requirements emerge (chicken-egg problem).

Users need **time and exposure** appropriate new tools.

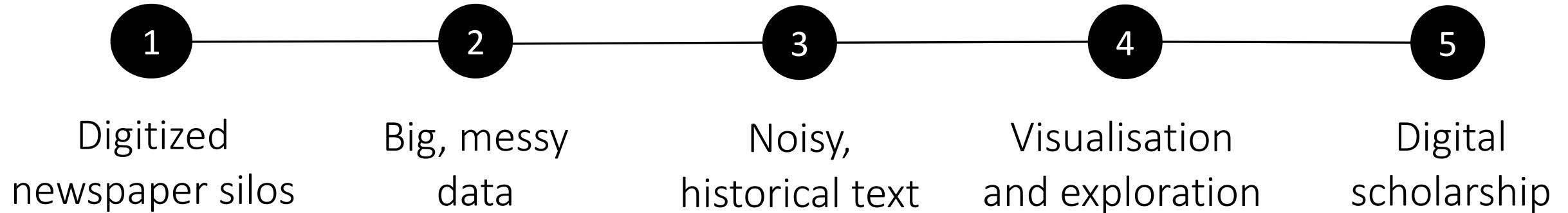
Users are **pragmatic**.

Combinations of **NLP enrichment + IR offer** powerful means to meet the information needs of historians.

**Skill development** is often needed, users susceptible to jumping to conclusions.



# *How to enable semantic indexing and of large collections of historic newspapers?*



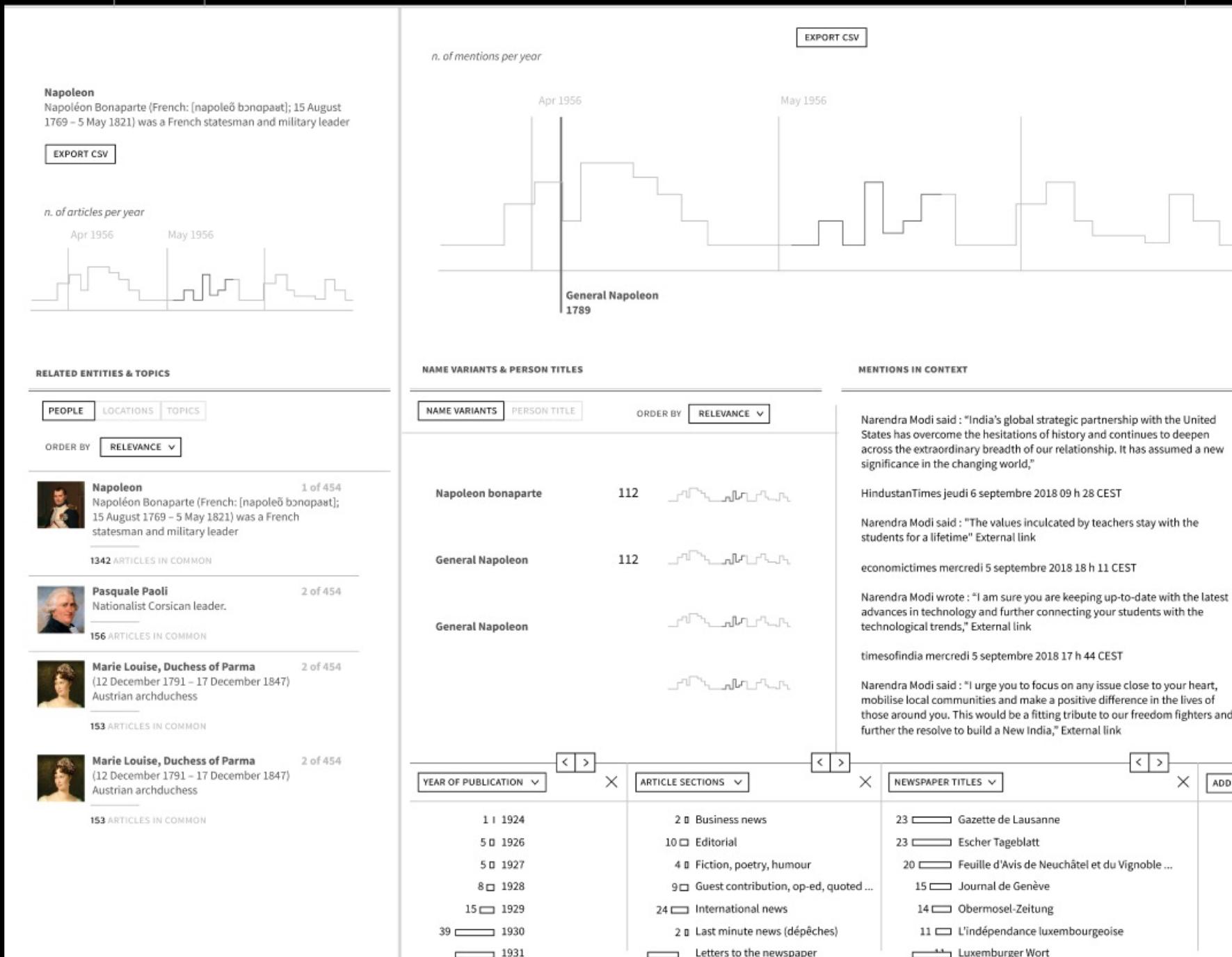
# About digital scholarship

New ways to engage with content also means new ways of using it.

How can historians work with

- automatically extracted, imperfect data?
- algorithmically generated, changing data?





MY COLLECTIONS MY SAVED QUERY

**LIST OF MY COLLECTIONS**

ORDER BY LATEST MODIFIED ▾

**PANEUROPE** 1 of 4  
YESTERDAY, 12H15 - collection of pan-europe and paneurope articles, manually curated  
135 ARTICLES

**UNITED STATES OF EUROPE** 3 of 10  
collection of pan-europe and paneurope articles, manually curated  
8 PAGES 7 ARTICLES

**GUERRE FROIDE & EUROPE** 3 of 10  
29 JUNE 2018, 08H15 - collection of pan-europe  
1 PAGE 2 ARTICLES

**LES CHIMERES** 4 of 4  
collection of pan-europe and paneurope articles, manually curated  
2 ISSUES 8 PAGES 700 ARTICLES

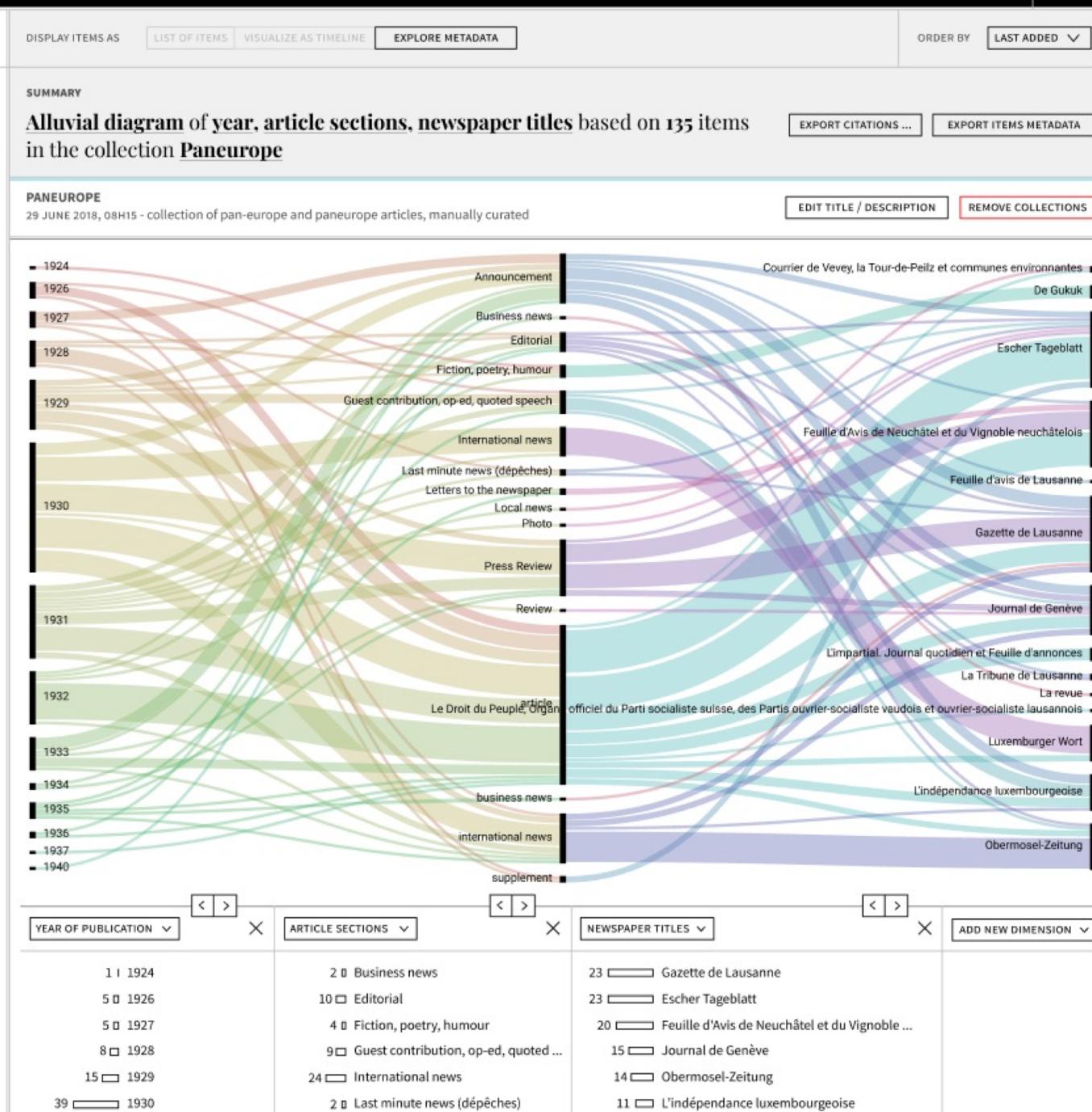
**CONFERENCE REPORTS**  
-  
2 ISSUES 8 PAGES 700 ARTICLES

**ADD NEW EMPTY COLLECTION**

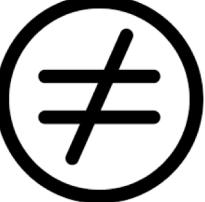
**ALWAYS AVAILABLE COLLECTIONS**

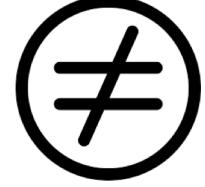
All items in all personal collection contains every article, issue, page you have ever included in one of your collections  
2 ISSUES 8 PAGES 700 ARTICLES

**FAVOURITE ITEMS**  
This list contains  
2 ISSUES 8 PAGES 700 ARTICLES



# Transparency matters

beauty  perfection

imperfection  useless

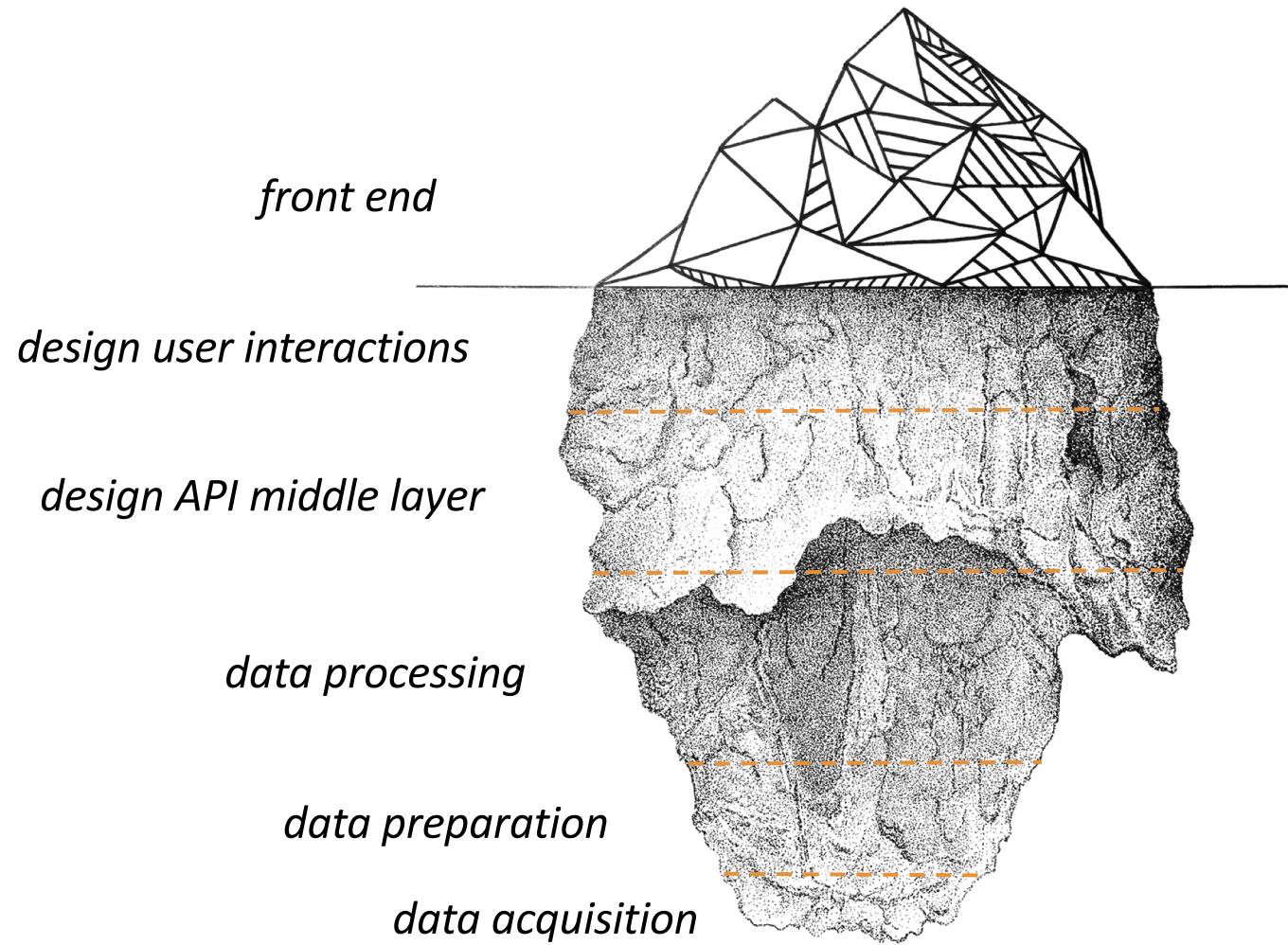
# Transparency matters

We inherit, introduce and propagate biases and variances in our corpus.

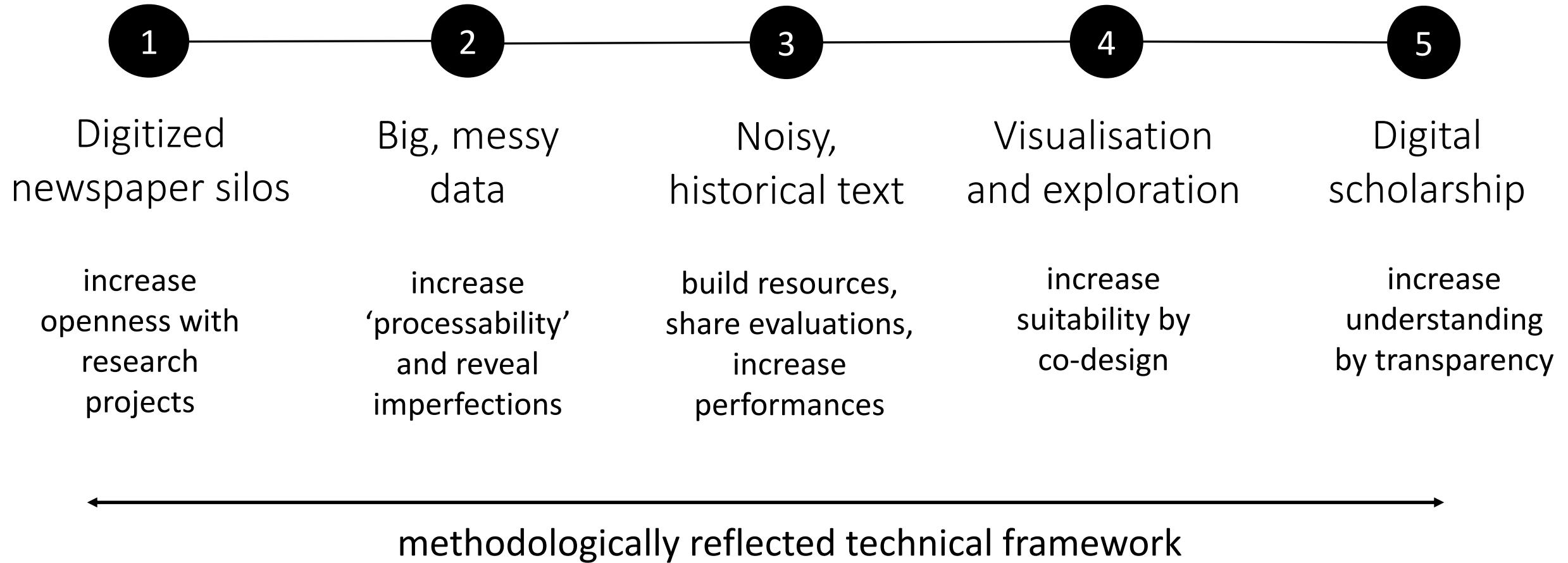
How do they affect content exploration? What exactly do historians need to understand to be able to deal with them?

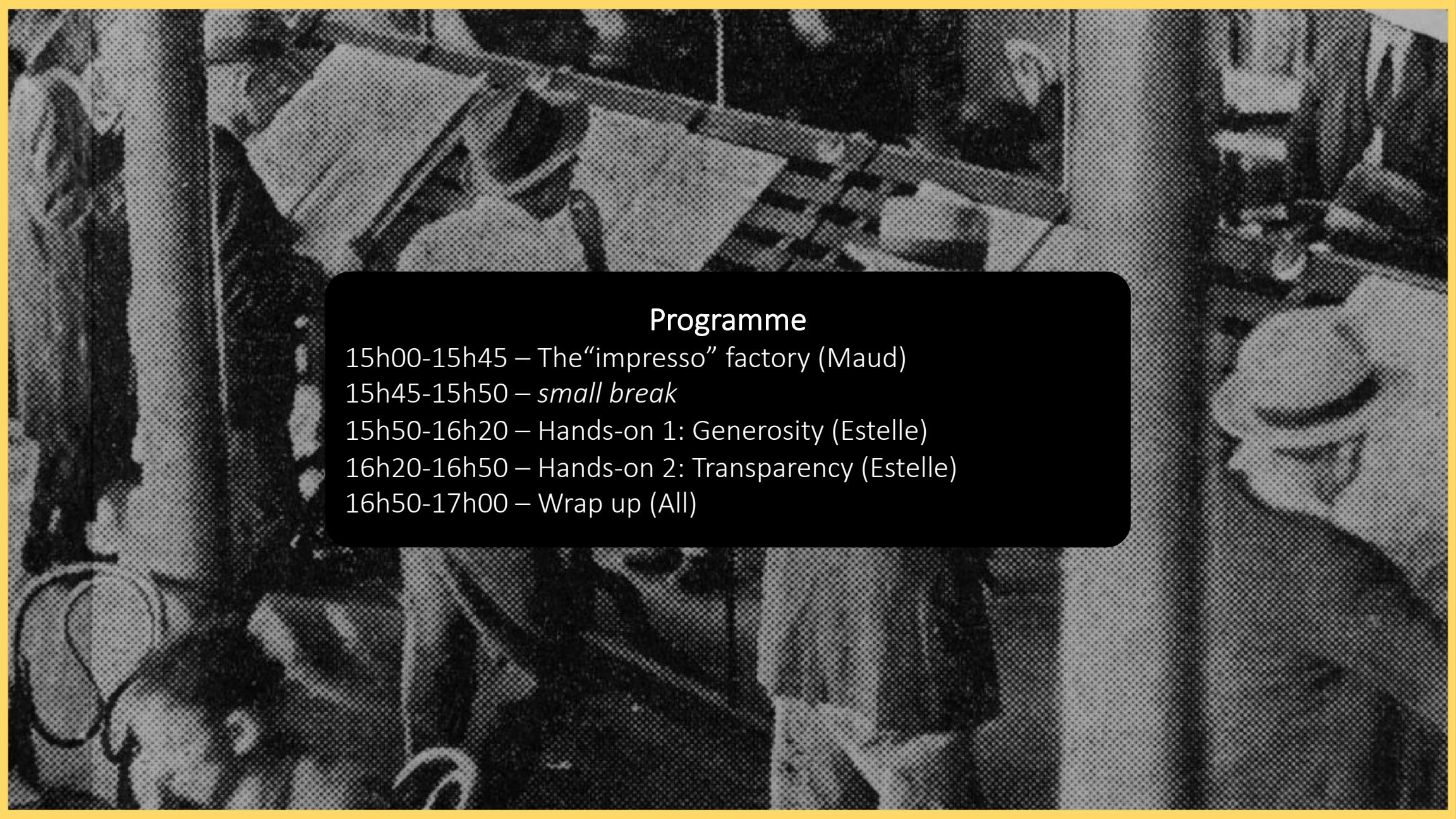
What is the most effective way to communicate these biases and variances?

# What's in the *impresso* app?



Newspapers are important, promising but complex digital assets.





## Programme

- 15h00-15h45 – The “impresso” factory (Maud)
- 15h45-15h50 – *small break*
- 15h50-16h20 – Hands-on 1: Generosity (Estelle)
- 16h20-16h50 – Hands-on 2: Transparency (Estelle)
- 16h50-17h00 – Wrap up (All)



Part 2  
Hands-on session



Media Monitoring  
**impresso** of the Past

# Two guiding principles for the app's co-design



## Generosity

- **exploration**
- **switching scales (drill down, overview)**
- **keep awareness of the context of the query, of an article collection**

## Transparency

- discover novel opportunities in & make **informed** decisions about digitized sources
- provenance and quality to assess the **value** of **imperfect** data

SEARCH ARTICLES

SEARCH IMAGES

NGRAMS

GROUP BY ARTICLE ↘

ORDER BY RELEVANCE ↘

DISPLAY AS LIST ↗ TILES

"robert schuman" ↗



add keyword to search



7,587 articles found containing robert schuman

COMPARE ...

SAVE / EXPORT ↗

## 1. How to discover the newspapers' contents via the impresso app (generosity)

- Guided tour of the app
- Hands-on guided by questions on the use of the app's features
- Collection of your answers

## 2. Taking a step back: understanding the findings (transparency)

- Guided tour of the app
- Hands-on guided by questions on the use of the app's context information
- Collection of your answers

FILTER BY CONTENT LENGTH ↗



FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

check one or more language to filter results

 French (5,491 results) ↗ German (2,052 results) ↗

Après l'échec de M. Antoine Pinay, le ministre de l'Éducation nationale, M. Robert Schuman est chargé d'une mission de conciliation.

Le conflit modérés-SFIO résolu par le MRP ?

PARIS, 19 (A.F.P.). — Le président René Coty a chargé M. Robert Schuman, leader M.R.P., d'une mission de conciliation.

Après avoir été reçu par M. René Coty, M. Robert Schuman a annoncé que le président de la République venait de lui confier une mission spéciale. Il a déclaré qu'il s'agissait d'une visite d'État pour établir une communion entre les deux partis. Il a également déclaré que le résultat de cette mission devrait être connu dans les prochaines semaines.

### M. Robert Schuman est chargé d'une mission de conciliation

L'Express ↗ MONDAY, OCTOBER 21, 1957 – p.1

Personal use

M. Robert Schuman est chargé d'une mission de conciliation. Après l'échec de M. Antoine Pinay, le ministre de l'Éducation nationale, le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A).

LOCATIONS France ↗

PEOPLE Antoine Pinay ↗



SEARCH ARTICLES

SEARCH IMAGES

NGRAMS

GROUP BY

ARTICLE

 "robert schuman" 

Search box

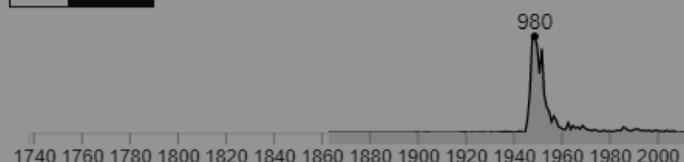
add keyword to search

 Frontpage

FIND SIMILAR WORDS

## PUBLICATION DATE

Number of articles per year

 %  SUM 

ADD NEW DATE FILTER ...

Filters

## FILTER BY CONTENT LENGTH



## FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

check one or more language to filter results

 French (5,491 results)  German (2,053 results) 

DISPLAY AS

LIST

TILES

Human readable summary

7,587 articles found containing robert schuman

COMPARE ...

SAVE / EXPORT

Lycée Robert-Schuman Luxembourg Poste vacant

d'Letzeburger Land FRIDAY, MARCH 18, 1977 – p.14

Personal use (no export)

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose d'engager prochainement un concierge pour les besoins du

LOCATIONS Luxembourg

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose

d'engager prochainement un concierge pour les besoins du Lycée Robert-Schuman à Luxembourg. Préférence

du Lycée Robert-Schuman, boulevard Emmanuel-Servais à Luxembourg, pour le 26 mars 1977 au plus tard

Result list

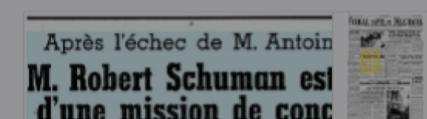
VIEW ADD TO COLLECTION ...

M. Robert Schuman est chargé d'une mission de conciliation

L'Express MONDAY, OCTOBER 21, 1957 – p.1

Personal use

M. Robert Schuman est chargé d'une mission de conciliation Après l'échec de M. Antoine Pinay Le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A.



LOCATIONS France

PEOPLE Antoine Pinay



SEARCH ARTICLES

SEARCH IMAGES

NGRAMS

GROUP BY

ARTICLE ▾

ORDER BY

RELEVANCE ▾

DISPLAY AS

LIST

TILES

"robert schuman" ▾



add keyword to search

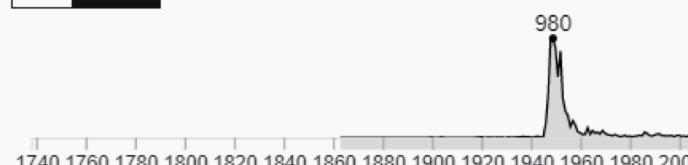


Frontpage

FIND SIMILAR WORDS 

## PUBLICATION DATE

Number of articles per year

% **SUM** 

ADD NEW DATE FILTER ...

FILTER BY CONTENT LENGTH 

645

0

9,900

## FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

check one or more language to filter results

 French (5,491 results)  German (2,053 results) 

GROUP BY ARTICLE ▾

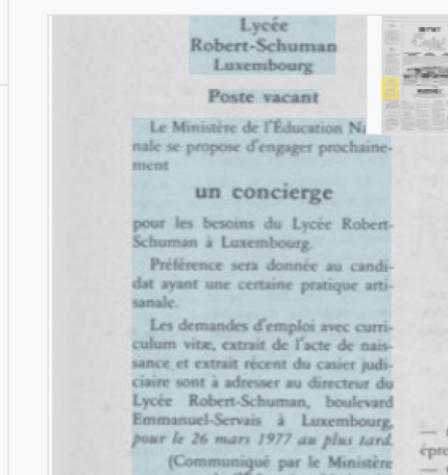
ORDER BY RELEVANCE ▾

DISPLAY AS LIST

TILES

7,587 articles found containing **robert schuman**

COMPARE ...

SAVE / EXPORT [Lycée Robert-Schuman Luxembourg Poste vacant](#)d'Letzeburger Land  FRIDAY, MARCH 18, 1977 – p.14

Personal use (no export)

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose d'engager prochainement un concierge pour les besoins du

LOCATIONS Luxembourg 

Lycée **Robert - Schuman** Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose

d'engager prochainement un concierge pour les besoins du Lycée **Robert - Schuman** à Luxembourg. Préférence

du Lycée **Robert - Schuman**, boulevard Emmanuel-Servais à Luxembourg, pour le 26 mars 1977 au plus tard

VIEW

ADD TO COLLECTION ... [M. Robert Schuman est chargé d'une mission de conciliation](#)L'Express  MONDAY, OCTOBER 21, 1957 – p.1

Personal use

M. Robert Schuman est chargé d'une mission de conciliation Après l'échec de M. Antoine Pinay Le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A.

LOCATIONS France PEOPLE Antoine Pinay 

SEARCH ARTICLES
SEARCH IMAGES
NGRAMS

GROUP BY ARTICLE
ORDER BY RELEVANCE
DISPLAY AS LIST
TILES

"robert schuman"

add keyword to search

Frontpage

PUBLICATION DATE
Number of articles per year

%
SUM

1740 1760 1780 1800 1820 1840 1860 1880 1900 19

FILTER BY CONTENT LENGTH

645

FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

check one or more language to filter results

French (5,491 results)
 German (2,053 results)

7,587 articles found containing robert schuman

Lycée Robert-Schuman Luxembourg Poste vacant
d'Letzeburger Land
FRIDAY, MARCH 18, 1977 – p.14

Personal use (no export)

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose

d'Letzeburger Land — Friday, March 18, 1977

p.14

article FR | p.14



FULLSCREEN: OFF

Lycée Robert-Schuman Luxembourg Poste vacant

LOCATIONS Luxembourg

TOPICS
40.8% FR service · langue · poste · travail · personnel
5.5% FR avril · mai · mars · lieu · juin
4.9% FR scène · théâtre · spectacle · pièce · public

9,900

d'une mission de conc

Le conflit modérés - SFIO résolu par le MRP ?

PARIS, 19 (A.F.P.). — Le président René Coty a chargé M. Robert Schuman, leader M.R.P., d'une mission de conciliation.

Après avoir été reçu par M. René Coty, M. Robert Schuman a annoncé que le président de l'Assemblée devait de lui confier une mission précise : « Les derniers entretiens que vient d'ecrire le président du Conseil, à ce qu'il résulte des révélations de l'Assemblée, sont destinés à faire pression sur le Gouvernement pour qu'il présente au conseil, et entoure de M. Devaux, directeur du budget, Schwartzenbach, directeur des finances extérieures, Blum, haut-commissaire au plan, Blod-Lalieu, député socialiste, et de la plupart des députés de l'opposition, et de M. Rosenstock-Frank, directeur des pris. M. Robert Schuman a aussi en cours les difficultés

Personal use

M. Robert Schuman est chargé d'une mission de conciliation. Après l'échec de M. Antoine Pinay, le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A).

LOCATIONS France

PEOPLE Antoine Pinay

**SEARCH ARTICLES** **SEARCH IMAGES** **NGRAMS**

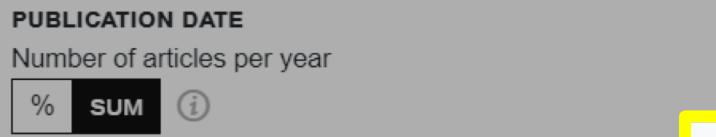
GROUP BY **ARTICLE** ▾

ORDER BY **RELEVANCE** ▾

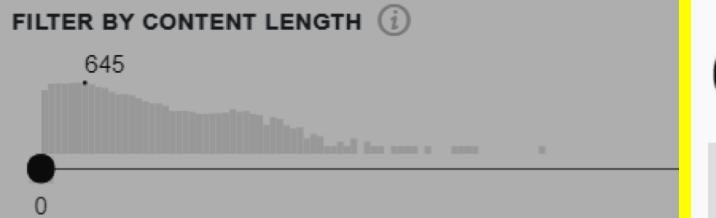
DISPLAY AS **LIST** **TILES**

"robert schuman" X  
add keyword to search  

Frontpage **FIND SIMILAR WORDS**



**ADD NEW DATE FILTER ...**



**FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)**

check one or more language to filter results

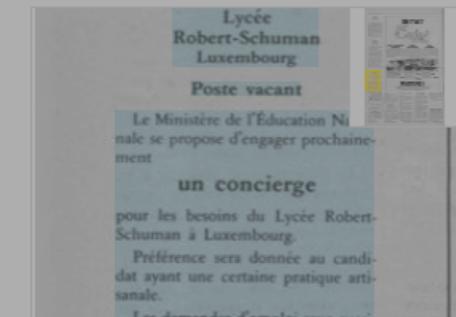
French (5,491 results) ▼

German (2,053 results) ▼

7,587 articles found containing **robert schuman**

**COMPARE ...**

**SAVE / EXPORT** ▼



**Lycée Robert-Schuman Luxembourg Poste vacant**

**d'Letzeburger Land** ▼ FRIDAY, MARCH 18, 1977 – p.14

Personal use (no export)

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose d'engager prochainement un concierge pour les besoins du

**LOCATIONS** Luxembourg ▼

**L'Express** — Monday, October 21, 1957 ◀ p.1 ▶

**article** FR | p.1

**FACSIMILE** 

**TRANSCRIPT**

**FULLSCREEN: OFF**

# M. Robert Schuman est chargé d'une mission de conciliation

**LOCATIONS** France ▼

**PEOPLE** Antoine Pinay ▼

**TOPICS**

35.3% FR président · ministre ·  
gouvernement · général · chef ▼

7.9% FR travail · grève · personnel ·  
chômage · syndicat ▼

17.7% FR conseil · président · conseiller ·  
directeur · chef ▼

4.3% FR conseil · commune · construction ·  
crédit · ville ▼

10.6% FR gouvernement · conférence ·  
accord · question · traité ▼

4.3% FR canton · projet · développement ·  
recherche · région ▼



# Filters : aggregation of library and NLP metadata

Creation of filters based on the metadata:  
help to calibrate the search + information about the collection

## FILTER BY NEWSPAPER TITLES

check one or more newspaper to filter results

- Journal de Genève (421,080 results) ↗
- Gazette de Lausanne (314,103 results) ↗
- L'Express (226,843 results) ↗
- L'Impartial (199,024 results) ↗
- L'indépendance luxembourgeoise (111,928 results) ↗
- Le Peuple, La Sentinelle (74,481 results) ↗
- Confédéré (39,085 results) ↗
- Le Confédéré de Fribourg (20,277 results) ↗
- Courrier du Grand-Duché de Luxembourg (17,436 results) ↗
- Luxembourg (1935) (17,004 results) ↗
- Le Chroniqueur (13,478 results) ↗
- L'Union (13,179 results) ↗
- Luxemburger Wort (8,473 results) ↗
- d'Letzeburger Land (8,226 results) ↗
- L'avenir (5,039 results) ↗
- Courrier du Valais (4,448 results) ↗
- Le Bien public (4,370 results) ↗
- Neue Zürcher Zeitung (3,800 results) ↗
- Le Narrateur fribourgeois (2,665 results) ↗
- L'Essor (2,644 results) ↗

[EXPLORE ALL ...](#)

Metadata

## FILTER BY PERSON

check one or more topics to filter results

- Francis Matthey (5,247 results) ↗
- Louis-Leopold Robert (4,728 results) ↗
- François Mitterrand (4,652 results) ↗
- Yasser Arafat (4,632 results) ↗
- Jacques Chirac (3,998 results) ↗

[EXPLORE ALL ...](#)

## FILTER BY LOCATION

check one or more topics to filter results

- France (355,496 results) ↗
- Paris (281,633 results) ↗
- Lausanne (280,586 results) ↗
- Suisse, Moselle (273,425 results) ↗
- Gare de Cornavin (228,002 results) ↗

[EXPLORE ALL ...](#)

Named entities

## FILTER BY TOPIC

check one or more topics to filter results

- fr ministre · président · gouvernement · affaire · chef (430,702 results) ↗
- fr samedi · dimanche · vendredi · lundi · soir (326,547 results) ↗
- fr président · force · gouvernement · pays · mort (325,304 results) ↗
- fr chambre · gouvernement · ministre · député · ministère (320,971 results) ↗
- fr commission · projet · loi · proposition · conseil (311,543 results) ↗
- fr comité · membre · président · assemblée · association (311,467 results) ↗
- fr loi · droit · initiative · canton · projet (306,266 results) ↗
- fr gouvernement · accord · conférence · question · traité (304,159 results) ↗
- fr parti · gouvernement · socialiste · politique · pouvoir (303,816 results) ↗
- fr guerre · paix · pays · peuple · politique (302,260 results) ↗

[EXPLORE ALL ...](#)

Topics

SEARCH ARTICLES

SEARCH IMAGES

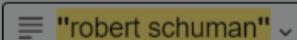
NGRAMS

GROUP BY

ARTICLE ▾

ORDER BY

RELEVANCE ▾

DISPLAY AS   "robert schuman" add keyword to search   Frontpage

PUBLICATION DATE

Number of articles per year

 %

SUM



980

1740 1760 1780

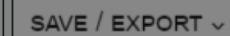
ADD NEW D...

FILTER BY CONTENT LENGTH 

645

9,900

FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

check one or more language to filter results  French (5,491 results)  German (2,052 results) 7,587 articles found containing  robert schuman COMPARE ... SAVE / EXPORT 

## 1. How to discover the newspapers' contents via the [impresso app](#) (generosity)

## 2. Taking a step back: understanding the findings (transparency)

Use case: help me find build a collection on the media coverage of Robert Schuman (1886-1963) in Luxembourg and Switzerland via the [impresso app](#).

### Après l'échec de M. Antoine Pinay, M. Robert Schuman est chargé d'une mission de conciliation

Le conflit modérés-SFIO résolu par le MRP ?

PARIS, 19 (A.F.P.). — Le président René Coty a chargé M. Robert Schuman, leader M.R.P., d'une mission de conciliation.

Après avoir été reçu par M. René Coty, M. Robert Schuman a annoncé que le président de la République venait de créer une commission spéciale. « Les derniers entretiens que vient d'avoir le président de la République, a-t-il dit, ont fait apparaître la nécessité d'une mission de conciliation. Il en résulte toutefois que la conclusion de la crise est domi-

### M. Robert Schuman est chargé d'une mission de conciliation

L'Express  MONDAY, OCTOBER 21, 1957 – p.1

Personal use

M. Robert Schuman est chargé d'une mission de conciliation Après l'échec de M. Antoine Pinay Le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A.

LOCATIONS France PEOPLE Antoine Pinay 



# Try for yourself! Hands-on „generosity“

On the search page, using the search filters and with the confirmation of the human readable summary, answer the following questions and create collections based on each query.

- Q1:** How many articles which contain the keyword “titanic” have been published on the front page?
- Q2:** Which year has the highest number of hits with “titanic” on the front page?
- Q3:** How many articles which contain the keyword “titanic” have been tagged with a topic related to film?
- Q4:** Which year has the highest number of hits with “titanic” and a topic “film”?
- Q5:** Which people have been the most often detected in articles containing the keyword “titanic”?
- Q6:** (Bonus task) On the search page, with the help of the filters, find movies featuring the Titanic sinking which are *not* by James Cameron. (hint: use the same topic but exclude/select other persons).

# Answers to the hands-on session - part 1

*Please find here the answers in forms of links to the impresso app, with the relevant filters activated. These are the answers to the hands-on session of the AI4LAM workshop 'Historical Newspaper Content Mining: findings from the impresso project', held on the 31.03.2021.*

**Q1:** How many articles which contain the keyword “titanic” have been published on the front page? [453](#)

**Q2:** Which year has the highest number of hits with “titanic” on the front page? [1912](#)

**Q3:** How many articles which contain the keyword “titanic” have been tagged with a topic related to film? [2852](#)

**Q4:** Which year has the highest number of hits with “titanic” and a topic “film”? [1998](#)

**Q5:** Which people have been the most often detected in articles containing the keyword “titanic”? [James Cameron](#)

**Q6:** (Bonus task) On the search page, with the help of the filters, find movies featuring the Titanic sinking which are *not* by James Cameron. (hint: use the same topic but exclude/select other persons). [Titanic](#) by Jean Negulesco.

SEARCH ARTICLES

SEARCH IMAGES

NGRAMS

GROUP BY

ARTICLE ↗

ORDER BY

RELEVANCE ↗

DISPLAY AS

LIST ↗

TILES

 "robert schuman" ↗

add keyword to search

 Frontpage

FIND SIMILAR WORDS

PUBLICATION DATE

Number of articles

 %

SUM

ADD NEW DATE FILTER ...

FILTER BY CONTENT LENGTH

645

0

9,900

FILTER BY LANGUAGE OF ARTICLES (4 OPTIONS)

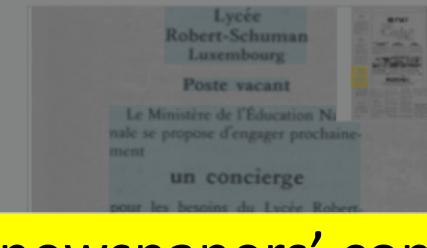
check one or more language to filter results

 French (5,491 results) ↗ German (2,052 results) ↗

7,587 articles found containing robert schuman

COMPARE ...

SAVE / EXPORT ↗

Lycée Robert-Schuman Luxembourg Poste vacant

d'Letzeburger Land ↗ FRIDAY, MARCH 18, 1977 – p.14

Personal use (no export)

Lycée Robert-Schuman Luxembourg Poste vacant Le Ministère de l'Éducation Nationale se propose d'engager prochainement un concierge pour les besoins du

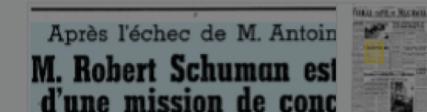
# 1. How to discover the newspapers' contents via the impresso app (generosity)

## 2. Taking a step back: understanding the findings (transparency)

d'engager prochainement un concierge pour les besoins du Lycée **Robert-Schuman** à Luxembourg. Préférencedu Lycée **Robert-Schuman**, boulevard Emmanuel-Servais à Luxembourg, pour le 26 mars 1977 au plus tard

VIEW

ADD TO COLLECTION ... ↗

M. Robert Schuman est chargé d'une mission de conciliation

L'Express ↗ MONDAY, OCTOBER 21, 1957 – p.1

Personal use

M. Robert Schuman est chargé d'une mission de conciliation Après l'échec de M. Antoine Pinay Le conflit modérés-SFIO résolu par le MRP ? PARIS, 19 (A.

LOCATIONS France ↗

PEOPLE Antoine Pinay ↗



SEARCH ARTICLES

SEARCH IMAGES

NGRAMS

GROUP BY

ARTICLE

ORDER BY

RELEVANCE

DISPLAY AS

LIST TILES

 "robert schuman" X

add keyword to search



7,587 articles found containing robert schuman

COMPARE ...

SAVE / EXPORT



1. Search for “[Robert Schuman](#)”: autofill indicates the named entity “Robert Schuman” and “Robert Schumann”

2. Add similar keywords, in FR/[DE](#) to expand the result list

3. Frequencies and filters indicate the distribution of the result list

4. Use of NE “Robert Schumann” to [reduce](#) the result list

5. Use of the topics to narrow down the inspection of the result list

6. Store results in a collection

7. Inspect&Compare the result list to expand and curate it

- [Compare Robert Schuman](#) as named entity and as keyword
- Find out when Robert Schuman was [misspelled](#) as Robert Schumann

Use case: help me find build a collection on the media coverage of Robert Schuman (1886-1963) in Luxembourg and Switzerland via the [impresso](#) app.

# Try for yourself! Hands-on „transparency“

**Q1:** From the newspapers page, can you tell which decade has the most issues per year? the most pages?

**Q2:** Go to the Inspect&Compare component and open in column A the collection "Titanic in movies", and in column B the collection "titanic on front pages". To understand when the overlap peaks and why, please find how many articles appear for the year 1998 in each query (using the mouse over on the frequency graph).

**Q3:** Open the topic page for the topic "film, cinéma, semaine, jean, john", using the arrow at the end of the topic label and choosing "more". What do you notice about its distribution over time? Who are the most frequently mentioned persons?

**Q4:** Back to the Inspect&Compare component. Replace the collection in column B with the query for the person "James Cameron". Who are the most associated persons to both queries (i.e. results in common)?

**Q5:** Switch to the "compare" tab and look at how the collection "titanic in movies" overlaps with the query "James Cameron" as a person. Hover over the graphs and read the overlap proportion. What do you notice when looking at the different facets related to the person James Cameron and the topic "film"?

# Answers to the hands-on session - part 2

**Q1:** From the newspapers page, can you tell which decade has the most issues per year? the most pages?

**A1:** Have a look [here](#) and change the settings of the total frequency graph from "issues" to "pages" and vice-versa.

**Q2:** Go to the Inspect&Compare component and open in column A the collection "Titanic in movies", and in column B the collection "titanic on front pages". To understand when the overlap peaks and why, please find how many articles appear for the year 1998 in each query (using the mouse over on the frequency graph).

**A2:** [in 1998](#), 20 articles containing "titanic" were published on the front page and 1169 containing "titanic" and the topic "film...", resulting in an overlap of 8 articles.

**Q3:** Open the topic page for the topic "film, cinéma, semaine, jean, john", using the arrow at the end of the topic label and choosing "more". What do you notice about its distribution over time? Who are the most frequently mentioned persons?

**A3:** [Gerard Depardieu](#) is the most frequently mentioned person in articles containing this topic.

**Q4:** Back to the Inspect&Compare component. Replace the collection in column B with the query for the person "James Cameron". Who are the most associated persons to both queries (i.e. results in common)?

**A4:** [Leonardo DiCaprio and Kate Winslet](#).

**Q5:** Switch to the "[compare](#)" tab and look at how the collection "titanic in movies" overlaps with the query "James Cameron" as a person. Hover over the graphs and read the overlap proportion. What do you notice when looking at the different facets related to the person James Cameron and the topic "film"? Which claim is false?

**A5:** all articles mentioning James Cameron, that contain the topic "film, cinéma..." are part of "titanic in movies".

# Time to sum up

- The *impresso* app was a real *laboratory* where we experimented a lot.
- NLP enrichments can be combined in surprising ways with IR in order to satisfy the information needs of historians.
- Machine understanding at the service of humans: a milestone towards text mining supporting (improving) historical scholarship

# What's next?

- Infrastructure maintenance and corpus growth.
- Bringing the interface closer to users: upload your texts, download your models, correct annotations, etc.
- HIPE2 under preparation: contact us in case you are interested!

# For more information

- subscribe to the mailing list
- visit the youtube channel
- ask a user account

impresso website:  
<https://impresso-project.ch>

impresso youtube channel:  
<https://www.youtube.com/channel/UCjRGykH-P9m1aA3amjrQuSQ>

impresso application:  
<https://impresso-project.ch/app>

data on Zenodo:  
<https://zenodo.org/communities/impresso/search?page=1&size=20>

code on GitHub:  
<https://github.com/impresso>

# Some publications

More information to come on the new impresso website.

## Highlights

- [Overview of CLEF HIPE 2020: Named Entity Processing on Historical Newspapers](#). CLEF 2020
- [Language Resources for Historical Newspapers: the Impresso Collection](#). LREC 2020
- [Historical Newspaper User Interfaces: A Review](#). IFLA 2019
- [Combining Visual and Textual Features for Semantic Segmentation of Historical Newspapers](#). JDDH 2020.
- [The impresso system architecture in a nutshell](#). Europeana Tech Insights
- [How Much Data Do You Need? About the Creation of a Ground Truth for Black Letter and the Effectiveness of Neural OCR](#)

## Upcoming

- *Digitized newspapers - a new Eldorado for historians ?* Edited volume with De Gruyter
- *Impresso: Historical Newspapers Beyond Keyword Search.* Digital Scholarship in the Humanities.
- *Co-Designing Workflows for the Exploration of Semantically Enriched Historical Newspapers [working title]*



# Media Monitoring of the Past

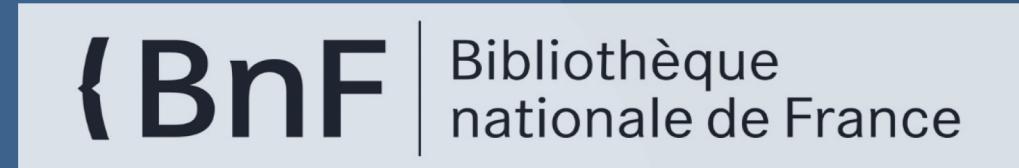
*Estelle Bunout  
Simon Clematide  
Marten Duering  
Maud Ehrmann  
Andreas Fickers  
Daniele Guido  
Frédéric Kaplan  
Roman Kalyakin  
Peter Makarov  
Matteo Romanello  
Gerold Schneider  
Paul Schroeder  
Benoit Seguin  
Phillip Stroëbel  
Martin Volk  
Thijs van Beek  
Lars Wieneke*



Thanks!

impresso-project.ch

twitter.com/impressoproject



Slides by Maud Ehrmann, Estelle Bunout and the *impresso* team

License: [CC BY 4.0](#)



*Credits: These slides are CC BY. Photographs by LIBER, LILLIAD Learning Centre Innovation, Cantonal and University Library of Lausanne. Template by SlidesCarnival.*

