

5^e JAAR

Donderdag, 4 Mei 1899

N^o 18

Dimanche 7 Mai 1899

40^e année — N^o 19

18 Mei

Zaterdag

1899

Pressing Matters. Digitising, Unlocking and Researching Belgium's Historical Press Collections

Brecht Deseure

WEEKBLAD VOOR DUIVENMAATSCHAPPIEN EN LIEFHEBBERS
VERSCHEIJNENDE ELKEN ZONDAG.

ABONNEMENTSPRIJS:

BUREAU EN REDACTIE:

De redigeer aanteekeningen van prijksporten en uitgedrukte brieven
aan de redactie.
De redigeer aanteekeningen van prijksporten en uitgedrukte brieven
aan de redactie.

1^{re} ANNEE.

Dimanche 21 mai 1899.

N^o 8.

Les Petites Affiches du Luxembourg

Journal spécialement consacré aux annonces

PARAISANT TOUTS LES DIMANCHES

Distribué gratuitement à tous les établissements publics, hôtels et cafés de la ville et de la province

ANTVERPIA

ET

SCALDIS PHILATÉLIQUE

RÉUNIS

paraissant le 1^{er} de chaque mois.

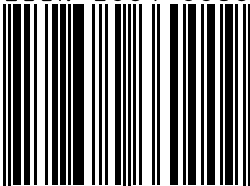
Les Carnets du Lafu

Laboratoire des pratiques et des identités journalistiques

N. 6 - 2026



ISSN 2684-6608



9 772684 660009

Les Carnets du LaPIJ

Laboratoire des pratiques et
des identités journalistiques

N. 6 - 2026

Pressing Matters. Digitising, Unlocking and
Researching Belgium's Historical Press Collections

Brecht Deseure

Comité éditorial

Juliette DE MAEYER (Université de Montréal)

Carlo GUBITOSA (Université libre de Bruxelles)

Manon LIBERT (Université de Mons)

Lia SEIXAS (Université Federal de Bahia)

Florian TIXIER (IJBA, Bordeaux Montaigne et ULB)

Alexia VIDALENCHE (ULB et Université Paul Valéry Montpellier 3)

Éditeurs responsables : Manon LIBERT, David DOMINGO, Florence LE CAM

Σαράη



Lancée en juin 2020, la collection Les Carnets du LaPIJ est consacrée à l'analyse des identités professionnelles et des transformations du journalisme. Sa vocation est de diffuser les travaux de recherche menés au sein du Laboratoire des pratiques et des identités journalistiques (LaPIJ). Les Carnets du LaPIJ sont disponibles en libre accès.

Numéros précédents :

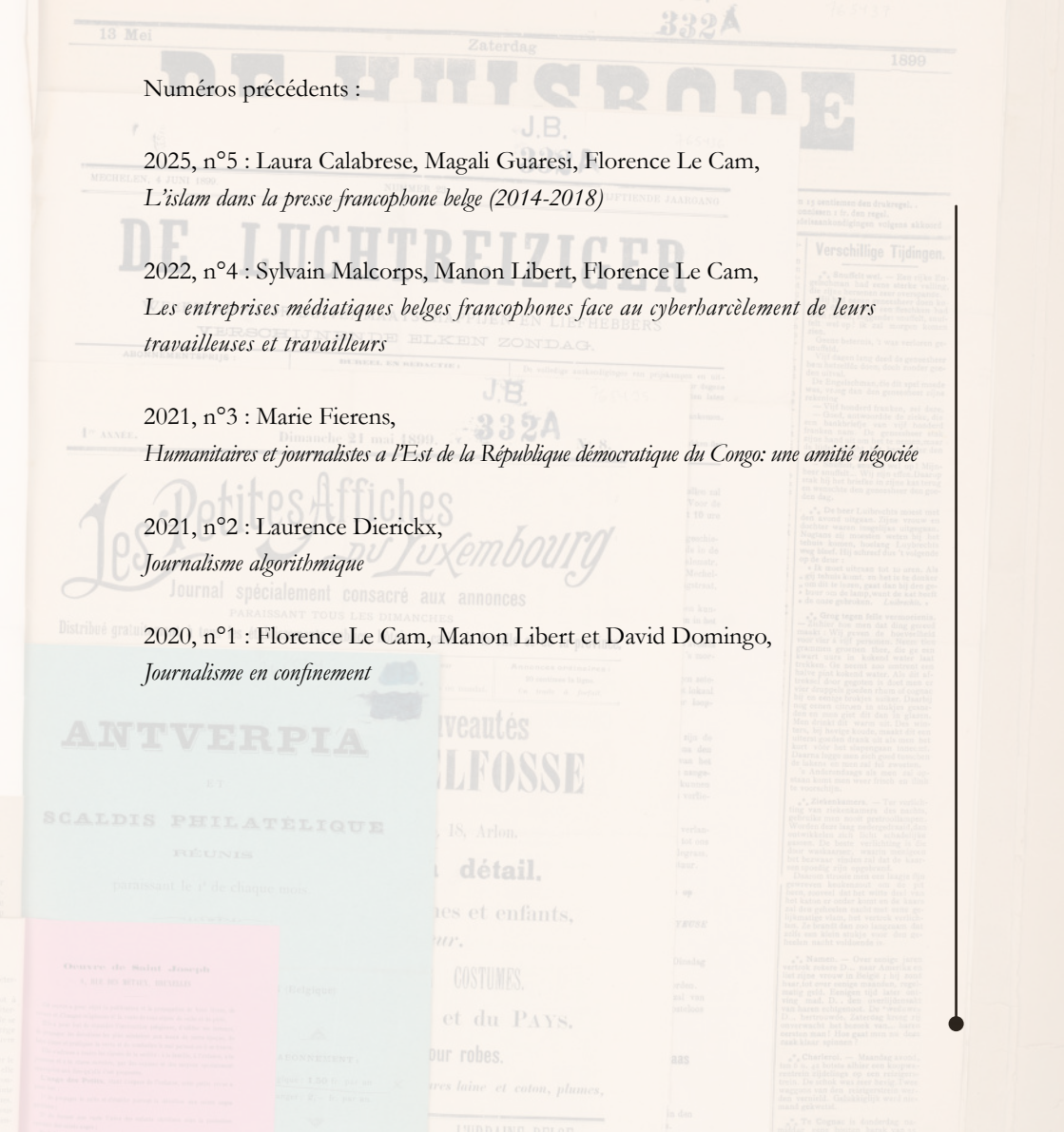
2025, n°5 : Laura Calabrese, Magali Guaresì, Florence Le Cam,
L'islam dans la presse francophone belge (2014-2018)

2022, n°4 : Sylvain Malcorps, Manon Libert, Florence Le Cam,
Les entreprises médiatiques belges francophones face au cyberharcèlement de leurs travailleuses et travailleurs

2021, n°3 : Marie Fierens,
Humanitaires et journalistes à l'Est de la République démocratique du Congo: une amitié négociée

2021, n°2 : Laurence Dierickx,
Journalisme algorithmique

2020, n°1 : Florence Le Cam, Manon Libert et David Domingo,
Journalisme en confinement



Remerciements

This volume is dedicated to the memory of Marc D’Hoore (1963–2023), curator of newspapers at KBR, expert in the history of the Belgian press, and a valued colleague and friend.



Summary

The landscape of historical print media is rapidly evolving. Historical newspapers and periodicals are increasingly being made digitally accessible — on the one hand to safeguard their long-term preservation, as this heritage is inherently fragile, and on the other to meet the growing demand for data to support the development of automated research methods in the digital humanities, a field that has seen remarkable progress in recent years.

Belgium, too, holds rich collections of historical print media that are progressively being digitised. However, the specific Belgian context presents both heritage professionals and researchers with a range of challenges. Historical newspaper and periodical collections are dispersed across numerous large and small institutions, and digitisation initiatives are undertaken at various institutional levels, often without central coordination.

To foster dialogue among the different stakeholders, the research centre *CAMille*, *Centre d'archives sur les médias et l'information* (KBR-ULB), organised the conference *Belgian Print Media in the Age of Digitisation: Current Challenges of Research into the History of Media and Journalism in Belgium* in 2023. This volume of the *Carnets du LaPIJ* was conceived as a continuation of the discussions initiated at the conference, offering an overview of the current challenges and developments in the field. It brings together contributions from curators, heritage professionals, digital humanities scholars, and other academic researchers.

In their contributions, the authors reflect on both the opportunities and the challenges presented by the digitisation of printed media, addressing issues such as collection management, metadata, data quality, OCR, copyright restrictions, and the integration of digital humanities methods. This volume is therefore essential reading for scholars of media history, digital humanities, and heritage studies, as well as for professionals engaged in the stewardship of historical press collections.

Résumé

Le paysage de la presse imprimée historique évolue rapidement. Les journaux et périodiques anciens sont de plus en plus rendus accessibles sous forme numérique — d'une part pour garantir leur conservation à long terme, ce patrimoine étant intrinsèquement fragile, et d'autre part pour répondre à la demande croissante de données nécessaires au développement de méthodes de recherche automatisées dans le domaine des humanités numériques, un secteur qui a connu des avancées remarquables ces dernières années.

La Belgique possède elle aussi de riches collections de presse imprimée historique, qui font progressivement l'objet de projets de numérisation. Cependant, le contexte institutionnel spécifique de la Belgique confronte tant les professionnels du patrimoine que les chercheurs à un ensemble de défis. Les collections de journaux et de périodiques historiques sont dispersées entre de nombreuses institutions, grandes et petites, et les initiatives de numérisation sont menées à différents niveaux institutionnels, souvent sans coordination centrale.

Afin de favoriser le dialogue entre les différents acteurs concernés, le centre de recherche CAMille, Centre d'archives sur les médias et l'information (KBR-ULB) a organisé en 2023 la conférence *Belgian Print Media in the Age of Digitisation: Current Challenges of Research into the History of Media and Journalism in Belgium*. Ce volume des *Carnets du LaPIJ* a été conçu comme une prolongation des discussions amorcées lors de cette rencontre, offrant un aperçu des défis actuels et des développements récents dans ce domaine. Il rassemble des contributions de conservateurs, de professionnels du patrimoine, de spécialistes des humanités numériques ainsi que d'autres chercheurs académiques.

Dans leurs contributions, les auteurs examinent à la fois les opportunités et les défis liés à la numérisation de la presse imprimée, abordant des questions telles que la gestion des collections, les métadonnées, la qualité des données, la reconnaissance optique de caractères (OCR), les restrictions liées au droit d'auteur et l'intégration des méthodes des humanités numériques. Ce volume constitue ainsi une lecture essentielle pour les chercheurs en histoire des médias, en humanités numériques et en études patrimoniales, ainsi que pour les professionnels impliqués dans la préservation des collections de presse historique.

Table of Contents

Brecht Deseure, Pressing Matters. Introduction **12**

Part 1. State of the Art

1 Kaspar Beelen, Reflections on a Workshop: Research Challenges and Opportunities for Belgian Print Media in the Age of Digitisation **25**

2 Christoph De Spiegeleer, Digital Access to Historical Newspapers in Belgium and Abroad. From Endangered Heritage to a Gold Mine for Machine Learning Technology **37**

Part 2. Digitisation: Institutions, Collections, Workflows

3 Brecht Deseure, Frédéric Lemmers & Sally Chambers, Belgian Press Heritage Online. Digitisation of Printed Media Collections at KBR **54**

4 Sophia Rochmes, *Nieuwe Tijdingen*: Towards a Newspaper Preservation Programme for Flanders **82**

5 Anthony Leroy & Ingrid Mayeur, Transforming ULB Libraries Digitisation Service: Going Beyond Scanning Our Own Collections to Also Become a Data Provider for Our Researchers **101**

6 Roxanne Wyns & Katrien Weyns, From Paper to 'Paper'. Facilitating Structural Digitisation and Reuse of Print Media Preserved in a Heritage Institution **121**

7 Joris Colla, News from Around the Church. The Participatory Digitisation and Contextual Disclosure of Local Parish Journals **138**

8 Jacques Gillen, Newspapers in the Mundaneum's Collection **151**

9 Nick Coorevits, *Historische Kranten: 20 Years of Digitising and Unlocking Newspapers in the Westhoek Region* **165**

10 Alexander Soetaert & Hilde Van Parys, *The Digitisation of Historical Newspapers in Bruges Public Library: Experiences and Perspectives* **183**

Part 3. Research Projects

11 Morgane Ott & Sébastien Hermans, *Constructing a Digital Corpus of Interwar Belgian Magazines to Investigate the Representations of Fine Arts. An Artpresse Case Study* **197**

12 Erwin Dejasse, *Studying Comics in the Belgian Press through Digitised Corpora: Challenges and Opportunities* **229**

13 Florence Le Cam, Brecht Deseure & Max De Wilde, *CAMille. A Research Centre on the History of Belgian Journalism* **241**

14 Sébastien de Valeriola, *News Nuggets: Mining Imperfect Historical Press Data* **260**

Bibliography **269**

Pressing Matters. Introduction

Brecht Deseure

The value of the press as a historical source is beyond doubt. Humanities scholars (including historians, linguists, sociologists and political scientists) have been using historical newspapers and periodicals for decades to gain insights into various aspects of past societies. Newspapers and magazines touch upon virtually all facets of society, including, of course, the history of the press itself. They serve not only as sources of factual information but also as invaluable materials for studying mentalities and discourse.¹ However, the way these sources are consulted has changed dramatically in a short period of time. Since the beginning of the century, libraries and other heritage institutions have invested heavily in making their historical press collections digitally available. The manual, analogue, and time-intensive consultation of paper originals or microfilm copies has in many cases been replaced by automated searches of digitised periodicals in an online viewer or at the data level.²

This wave of digitisation was primarily driven by conservation concerns: newspaper and magazine paper is often fragile and inherently has a limited lifespan. While microfilm was once the default medium for libraries seeking to preserve their historic press collections, new digital techniques have proven to be a more attractive alternative, for many reasons. One result of this change is that institutions have evolved from being repositories of physical objects to being providers of digital collections and data. Increasingly, the challenge is not only to safeguard collections but also to meet the changing expectations of the public.³

1. KRIEG, Alice. Analyser le discours de presse, *Communication*. 2000, vol. 20, n° 1. Available online: <http://journals.openedition.org/communication/6432>. DOI: <https://doi.org/10.4000/communication.6432>; MOIRAND, Sophie. *Les discours de la presse quotidienne. Observer, analyser, comprendre*. Paris (France): Presse universitaires de France, 2017.

2. BINGHAM, Adrian. The Digitization of Newspaper Archives: Opportunities and Challenges for Historians. *Twentieth Century British History*. 2010, vol. 21, n° 2. DOI: <https://doi.org/10.1093/tcbh/hwq007>.

3. PRÉMEL, Benjamin. La numérisation de la presse rétrospective à la BnF. *Le Temps des médias*. 2007, vol. 8, n° 1, 267-278. DOI: <https://doi.org/10.3917/tm.008.0267>; LALLICH-BOIDIN, Geneviève & BRUN-PICARD, Céline. Collections de presse, numérisation et patrimoine des bibliothèques. *Culture & musées*. 2013, vol. 21, n° 1, pp. 89-110.

The altered consultation modalities have, of course, had a profound impact on research. The status of historical press sources and the information extracted from them has been thoroughly transformed.⁴ While the material aspect of historical press sources is largely obscured, their textual content is gaining significantly in visibility. Digitised press sources offer enormous gains to the researcher in terms of time, volume, and accessibility. Millions of pages of historical text are now just a click away for both the general public and specialised researchers. Thanks to the massive scale of available data (historical press collections are extremely large, both physically and textually), these collections have become an ideal playground for the development of digital text — and increasingly, image — analysis techniques. This constantly evolving field enables us to interrogate historical press sources in previously unimagined ways.

However, the digitisation and accessibility of historical press sources, as well as the research conducted with them, present significant challenges for heritage institutions, the public, and researchers alike. The needs and priorities of these three groups are not always aligned, and supply and demand do not always match. Issues such as source selection, available resources, data access, reliability of metadata, copyright constraints, OCR quality, and institutional cooperation are just a few of the many challenges. Furthermore, the progress of digitisation and the accessibility of historical press sources are strongly dependent on the (inter)national context, and individual developments are liable to unfold differently in different countries and parts of the world.

Successful networks and large-scale research projects dedicated to historical print media have been established in several European countries, including Impresso (Switzerland and Luxembourg),⁵ NewsEye (Austria, Finland, France

4. EHRMANN, Maud, BUNOUT, Estelle & CLAVERT, Frédéric. Introduction. In: BUNOUT, Estelle, EHRMANN, Maud & CLAVERT, Frédéric (eds.). *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. Berlin (Germany): De Gruyter Oldenbourg, 2023, p. 4. DOI: <https://doi.org/10.1515/9783110729214>. FICKERS, Andreas. Towards a New Digital Historicism? Doing History in the Age of Abundance. *VIEW Journal of European Television History and Culture*. 2012, vol. 1, n° 1, pp. 19–26; KOOLEN, Marijn, VAN GORP, Jasmijn & VAN OSSENBRUGGEN, Jacco. Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice. *Digital Scholarship in the Humanities*. 2019, vol. 34, n° 2, pp. 368–385, DOI: <https://doi.org/10.1093/llc/fqy048>.

5. <https://www.impresso-project.ch/>.

and Germany),⁶ and Numapresse (France).⁷ This volume focuses on the situation in Belgium. In 2020, the research centre CAMille was launched at the Université libre de Bruxelles and the Royal Library of Belgium (KBR).⁸ In addition to serving as a centre of expertise for the history of Belgian journalism, CAMille also functions as a laboratory for the development of automated research techniques applied to digitised newspapers and periodicals (see the contributions by Florence Le Cam, Brecht Deseure, and Max De Wilde, as well as Sébastien de Valeriola in this issue). As a FED-tWIN project, supported by both the university and the national library, it serves as a bridge between the research community and collection management.

From the outset of the project, it became clear that the digitisation, accessibility, and study of historical press collections in Belgium involve a wide range of actors operating at different institutional levels. The Belgian state structure presents specific challenges that hinder a centralised approach: culture and heritage are not federal competences but fall under the authority of the language communities, with the exception of the Federal Scientific Institutions (one of which is KBR). Within the communities themselves, there exists a considerable diversity of actors, including local, supra-local and regional heritage institutions, as well as museums and universities.

It seemed beneficial to initiate a dialogue between the numerous stakeholders involved in these operations. To this end, CAMille took the initiative to organise a conference to assess the state of affairs in Belgium. The conference *Belgian Print Media in the Age of Digitisation. Current Challenges of Research into the History of Media and Journalism in Belgium* took place on 4 and 5 September 2023.⁹ Presentations by Belgian heritage professionals and researchers were inter-

6. <https://www.newseye.eu/>.

7. <http://www.numapresse.org/>.

8. <https://www.kbr.be/en/projects/camille/>.

9. The conference was organised by the research centre CAMille, KBR, the department of literary studies at Ghent University and the Ghent Centre for Digital Humanities, with support of the *Laboratoire des Pratiques et Identités Journalistiques* (LaPIJ), the *Centre de Recherche en Information et Communication* (RESIC), the Research Foundation Flanders (FWO) and the *Fonds de la recherche scientifique* (FNRS). The organising committee consisted of Julie M. Birkholz, Brecht Deseure, Florence Le Cam and Marianne Van Remoortele.

spersed with keynotes by international specialists who, through their pioneering research, are setting the tone in this field. This volume of the Carnets du LaPIJ was conceived as a continuation of the discussions initiated at the conference, offering an overview of the current challenges and developments in the field.¹⁰

Part I. State of the Art aims to identify some key threads and common themes. In his essay, **Kaspar Beelen** (School of Advanced Study, University of London) shares a personal recollection of the conference *Belgian Print Media in the Age of Digitisation*, for which he provided the concluding remarks. He highlights various challenges and opportunities related to working with digitised sources at scale and touches upon the fragmentary nature of the conservation and digitisation of historical press collections in Belgium. Among the possible ways forward, he emphasises the need for a comprehensive understanding of the newspaper landscape via reliable data work and advocates a *collections as data* approach to enhance accessibility and facilitate computational analysis of the historical press.

Next, **Christoph De Spiegeleer** (Liberas) outlines the state of the art of newspaper digitisation in Belgium. He discusses both the promises and challenges of researching Belgian digitised newspapers, seeking to determine whether they truly represent the digital Eldorado they are sometimes claimed to be.¹¹ In Belgium, he concludes, truly integrated (inter)national accessibility of digitised newspaper heritage for different user groups is still a thing of the future. Among the main challenges facing the field, he identifies: 1) fragmentation of digitisation initiatives in Belgium, caused by institutional realities; 2) the differing expectations of institutions and academic researchers regarding digitisation and the data provided; and 3) the coherence and quality of datasets. He also highlights key trends in the management of digitised collection objects, including the *collections as data* movement, the rise of artificial intelligence, and the potential offered by expertise developed in several pioneering research projects. These themes and challenges reappear throughout the rest of the volume.

10. <https://lapij.ulb.ac.be/rubrique/publications/carnets-lapij/>.

11. See: BUNOUT, Estelle, EHRMANN, Maud & CLAVERT, Frédéric (eds.). *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. Berlin (Germany): De Gruyter Oldenbourg, 2023. DOI: <https://doi.org/10.1515/9783110729214>.

De Spiegeleer's contribution reveals that a Belgian Gallica,¹² Delpher,¹³ or Chronicling America¹⁴ remains a distant dream. In Belgium, digitisation initiatives are not centrally coordinated but are instead divided among numerous players at national, regional, and local levels. These actors often take valuable initiatives and make significant progress within their respective domains.

In Part 2. Digitisation: Institutions, Collections, Workflows, Challenges, the state of newspaper digitisation in Belgium is examined from the perspective of the institutions involved. These institutions range from large national players to universities, archives of ideological movements (*landelijke culturele archieven*), public libraries, and heritage cells. While this volume cannot represent all stakeholders, the contributions below provide a representative overview of the different levels, actors, and initiatives, as well as the often similar challenges they face.

The largest player in Belgium is KBR. Not only does it hold the most extensive collection of historical Belgian newspapers and periodicals, but through its digital platforms BelgicaPress¹⁵ and BelgicaPeriodicals,¹⁶ it is also the only institution providing central platforms for digitised press collections from across Belgium. **Brecht Deseure** (KBR-ULB), **Frédéric Lemmers** (KBR), and **Sally Chambers** (British Library) elaborate on this role. At the beginning of the century, several federal rescue plans for fragile paper heritage were undertaken, positioning KBR as a pioneer in the digitisation of historical newspapers and periodicals in Belgium. KBR digitises using both its own resources and through external funding, placing a strong emphasis on collaboration with other institutions. As a result, significant thematic and historically relevant collections of periodical publications from the early seventeenth century to the present have been made available online. The Digit department conducts in-house digitisation using its own scanners (sometimes working from microfilm) and coordinates larger projects executed by exter-

12. <https://gallica.bnf.fr/>.

13. <https://www.delpher.nl/>.

14. <https://chroniclingamerica.loc.gov/>.

15. <https://www.belgicapress.be>.

16. <https://www.belgicaperiodicals.be/>.

nal service providers. Influenced by developments in other national libraries, KBR is currently developing a platform to make digitised and born-digital collections available as open data and curated datasets. Nonetheless, digitisation faces practical challenges and constraints, including financial resources, OCR quality, and copyright restrictions.

As culture in federalised Belgium is a community matter, responsibility for heritage digitisation lies at different institutional levels. Consequently, a single, central newspaper platform for the entire country is difficult to achieve. Important initiatives have also been taken at the Flemish level. **Sophia Rochmes** (Flanders Heritage Libraries) discusses the initiatives of Flanders Heritage Libraries¹⁷ and *meemoo*, the Flemish institute for archives,¹⁸ to develop a coordinated digitisation programme for historical newspapers in Flemish heritage collections. More than 100 organisations across Flanders, ranging from large institutions such as university libraries and national cultural archives to smaller organisations such as local libraries, city and municipal archives, and local historical societies hold collections of historical newspapers. A crucial tool for tracking which titles are held where and identifying what has been digitised is Abraham, an online catalogue of Belgian newspapers preserved in Flemish libraries and other collecting institutions, comprising over 12,500 titles.¹⁹

Flanders Heritage Libraries and meemoo launched the project *Nieuwe Tijdingen*²⁰ with the aim of moving away from stand-alone projects and instead developing a centrally coordinated programme for the digitisation of Flemish newspaper collections, complementing national-level initiatives by KBR. The proposed programme, building on *News of the Great War*,²¹ *Abraham 2020*,²² and

17. <https://vlaamse-erfgoedbibliotheken.be>. Since February 2026 [BibliotheceairErfgoed.be](https://bibliotheceairerfgoed.be).

18. <https://meemoo.be/>.

19. <http://www.krantencatalogus.be>.

20. <https://www.bibliotheceairerfgoed.be/projecten/nieuwe-tijdingen>.

21. <https://nieuwsvandegrooteoorlog.hetarchief.be/en>.

22. <https://krantencatalogus.be>.

Het Archief,²³ is structured around three essential and interconnected pillars: digitisation, digital archiving, and online access. Collection managers have indicated that 3.34 million pages from 549 titles are in urgent need of digitisation, a collection which will be prioritised in the proposed newspaper programme. Since August 2025, newspaper collections from over 20 Flemish organisations are offered via hetarchief.be.

Digitisation in Francophone Belgium is following a different trajectory compared to the national and Flemish levels. The *Fédération Wallonie-Bruxelles* has developed a platform for accessing digitised heritage,²⁴ yet no specific plan exists for centrally coordinated digitisation of historical press collections. In Flanders, a notable institution is the Hendrik Conscience Heritage Library,²⁵ which functions both as a city library and as a nationally recognised Flemish heritage institution. It has already digitised significant series of historical newspapers and periodicals, including well-known titles such as *Nieuwe Tijdinghen* and *Gazette van Antwerpen*.

A key player in both the digitisation of and research into historical newspaper collections are the universities. Significant historical press collections are hosted by universities such as Ghent, Leuven, and Liège. Groundbreaking research on the automated analysis of digitised newspapers is conducted at institutions including the Quadihum centre at ULB (see the contribution by Sébastien de Valeriola in this issue)²⁶ and the Ghent Centre for Digital Humanities.²⁷

The ULB Libraries are also highly active in this field, as demonstrated in the contribution by **Anthony Leroy** (ULB) and **Ingrid Mayeur** (ULiège). Whereas the ULB Libraries' Digitisation Service initially focused on preserving and providing access to its own collections, it has since expanded to include large-scale projects such as the digitisation of Belgian press and literary journals. Today, the service aligns with the *collections as data* paradigm, offering machine-readable

23. <https://hetarchief.be/>.

24. <http://www.numeriques.be/>.

25. <https://consciencebibliotheek.be/>.

26. <https://quadihum.ulb.be/>.

27. <https://www.ghentcdh.ugent.be/>.

digital corpora for researchers. Among the service's priorities are data accessibility, interoperability, and long-term digital sustainability. The authors outline the internal workflow, management principles, and optimised processes that have been developed in response to the evolving needs of researchers. In doing so, the ULB Libraries fully exploit their position at the intersection of preservation and research.

Archival institutions also play a crucial role in preserving and providing access to historical press collections. A distinctive feature of the Belgian context is the existence of a series of archives of ideological movements (*landelijke culturele archieven*), reflecting the societal pillarisation that has profoundly shaped the country's history. Among these archives are the aforementioned Liberas,²⁸ the centre for the history of the liberal movement and the freedom ideal in Belgium, which has built an extensive collection of digitised newspapers and periodicals, and KADOC, the Interfaculty Documentation and Research Centre on Religion, Culture, and Society at KU Leuven.²⁹ **Roxanne Wyns** (KU Leuven Libraries) and **Katrien Weyns** (formerly KADOC-KU Leuven, now VAI) present in their contribution the project *Journals in the Picture*,³⁰ through which this archival institution investigated the automation of the selection process within its in-house digitisation workflow. This process was optimised through the enrichment of bibliographic descriptions. Based on these experiences, the authors reflect on the use of linked (open) data in the heritage sector. A second project, *iCANDID*,³¹ launched by KU Leuven Libraries' digital provider Libis, yielded significant insights into the collection and normalisation of, as well as access to, big data in accordance with FAIR (Findable, Accessible, Interoperable, Reusable) principles.³² While these methodologies hold great promise, they also present challenges: while a semantic web approach could considerably reduce

28. <https://www.liberas.eu/en/>.

29. <https://kadoc.kuleuven.be/>.

30. https://kadoc.kuleuven.be/8_projecten/2022/tijdschriften_picture/2022_03_tijdschriften-in-the-picture/.

31. <https://icandid.t.libis.be/>.

32. WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, Ijsbrand Jan and others. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data*. 2026, vol. 15, n° 3. DOI: <https://doi.org/10.1038/sdata.2016.18>.

the workload for data normalisation, researchers and heritage professionals do not always possess the necessary skills to work with semantic data.

Joris Colla (KADOC-KU Leuven) presents a case study on the innovative digitisation and preservation trajectory of a specific type of periodical at KADOC. The parish journal *Kerk en Leven* and its predecessors rank among the most widely read periodicals in Flanders and constitute an important historical source. The project *News from Around the Church*³³ aims to digitise and provide access to the collections of local parish journals preserved throughout Flanders and Brussels. Over a million pages of the parish journal have been scanned and ingested into KU Leuven's Integral Archiving System. Moreover, these digital collections are systematically linked to contextual information clusters on the historical development of the parish landscape in ODIS, a bilingual (Dutch-English) contextual online database on the history of civil society.³⁴ As a result, users are offered various possibilities for the contextual exploration of KADOC's digital collection of parish journals through related information clusters.

An exceptional position in the realm of Belgium's press heritage is occupied by the Mundaneum, a private archival institution based in Mons. It found its genesis in the vision of pioneering documentalists Paul Otlet and Henri La Fontaine, who sought to create a comprehensive repository of global knowledge. In his contribution, **Jacques Gillen** (Mundaneum) presents these press collections and discusses their digitisation. The Mundaneum³⁵ provides a home for the collection of the International Press Museum, which was established in 1907 with the aim of providing an overview of all periodical publications worldwide. It contains specimens from over 100,000 newspaper titles from across the globe, dating from the 18th century to the mid-20th century. Although this collection has not yet been subjected to a large-scale digitisation campaign, specific sub-collections have been digitised and made publicly accessible, partly thanks to financial support from the Wallonia-Brussels Federation.

A remarkable development in the field of digitisation is the progress made by institutions at the local and regional levels. The website krantencollectie.be,

33. https://kadoc.kuleuven.be/8_projecten/2021/2021_06_parchiebladen.

34. <https://www.odis.be/>.

35. <https://mundaneum.org/>.

for instance, hosts historical newspapers from the cities of Leuven and Halle,³⁶ whereas *Kempens Erfgoed* collects press sources from the Campine region.³⁷ A true success story is the digitisation of local newspapers in the West Flemish region of Westhoek. This project began twenty years ago as a digitisation initiative for historical newspapers in the city of Ypres. Over time, collaborations were established with heritage services in surrounding towns and villages, including Poperinge, Diksmuide, and Veurne, transforming the website into a heritage database where more than seventy different newspaper titles from the Westhoek region can be consulted digitally.³⁸ Nearly all locally preserved historical newspapers from the mid-19th to the late 20th century are now available online. **Nick Coorevits** (CO₇) of the intermunicipal partnership *CultuurOverleg Zeven*,³⁹ which coordinates the project, outlines its history and highlights future challenges. The collaboration between city archives, local heritage services (*erfgoedcellen*), and regional partnerships, have been essential to the project's success. However, the scale of the project now risks surpassing the capacity of local institutions, prompting consideration of initiatives at the Flemish level in which meemoo, the Flemish institute for archives, is expected to play a crucial role.

That local policy choices can lead to successful digitisation campaigns is illustrated by various city archives and libraries, including the city archives of Aalst⁴⁰ and Sint-Truiden.⁴¹ A particularly successful example is that of the Bruges Public Library,⁴² which, since 2014, has digitised and made available online approximately 700,000 pages from local newspapers published between 1792 and the early 2000s on the *Erfgoed Brugge* platform.⁴³ **Alexander Soetaert** and **Hilde Van Parys** (both Bruges Public Library) elaborate on the history

36. <https://krantencollectie.be/>.

37. <https://www.kempenserfgoed.be/>.

38. <https://historischekranten.be/>.

39. <https://www.co7.be/>.

40. <https://aalst.courant.nu/>.

41. <https://www.udesk-sast.eu/ucat/sast/>.

42. <https://brugge.bibliotheek.be/erfgoed>.

43. <https://erfgoedbrugge.be/>.

and challenges of the project. As the Bruges newspapers are offered on the same website as the image bank containing heritage images, specific challenges occur concerning metadata and filtering. As with all press digitisation projects, copyright restrictions must be taken into account, particularly concerning recent publications. Furthermore, the scale of the data that will be collected and its storage poses challenges that extend beyond the local level.

As highlighted in several of the aforementioned contributions, many institutions face the challenge of aligning their digitised newspaper collections with the needs of both the general public and specialised researchers. These two target audiences often have diverging requirements when it comes to data formats and technical standards. While the average visitor may be happy with simply consulting digitised files in a viewer, researchers frequently require access to raw data.

Part 3: Research Projects showcases several innovative, ongoing research projects that develop and apply diverse methodologies to process large corpora of digitised historical press materials in a (partly) automated manner.

The Artpresse project (KBR – ULiège – KU Leuven),⁴⁴ which focuses on the rich collection of interwar illustrated periodicals within the KBR collections, is an excellent illustration of the opportunities and challenges associated with working with large-scale digitised periodicals. As a manually created dataset, the Artpresse catalogue is positioned between digital archives on the one hand and the analytical promises of *collections as data* on the other. **Morgane Ott** (KBR/ULiège) and **Sébastien Hermans** (KBR/KU Leuven) emphasise the importance of thorough cataloguing and historical preparatory work in compiling an extensive digitised corpus of mass media magazines. Their analysis of the editorial networks and the contexts in which these magazines were produced has led to a comprehensive study of Belgian magazine culture. Over 750,000 pages from approximately 130 illustrated periodicals have been digitised through a research-driven, small-scale mass digitisation workflow, in close interaction with the intermedial study of the fine arts through the lens of family magazines. Drawing on data collected in an internally illustrated and indexed catalogue, the authors also examine how fine arts were featured in selected family magazines during the early 1930s in relation to Return to Order discourses.

44. <https://www.kbr.be/en/projects/artpresse/>.

Another relatively unknown and under-researched Belgian press corpus is the focus of **Erwin Dejasse** (ULB-KBR), who explores automated methods for detecting and analysing comics in digitised youth and comic magazines (as opposed to comic albums). By applying computer vision techniques, a workflow for the automated extraction of comics has been developed, leading to the discovery of extensive corpora that were previously largely unknown. Advances in stylometry are also expected through the application of digital humanities methodologies, including the recognition of stylistic influences between different illustrators and the attribution of comic authorship to one or more artists.

Digitised press collections are, of course, also an invaluable source for research into the history of media and journalism. The research centre *CAMille*, *Centre d'archives sur les médias et l'information*, a joint initiative of ULB and KBR, was established to advance the study of Belgian journalism through the development of digital research methodologies. **Florence Le Cam** (ULB), **Brecht Deseure** (ULB-KBR), and **Max De Wilde** (ULB-Université de Genève) discuss the origins and objectives of the project, with particular attention to its research methods. The project has developed its own platform featuring a selection of digitised historical newspapers and other press-related documents, serving as a testing ground for research methodologies. The careers and media presence of Belgian journalists are examined through the development of automatic signature detection and the automated reconstruction of media careers. The prosopographical aspect of the project is pursued via the creation of a database of Belgian journalists, developed within the RDF data framework of SPARQLb. A significant sub-study focuses on media discourse concerning female journalists, analysed with a combination of quantitative and qualitative methodologies.

The development of automated research methods within CAMille has greatly benefited from collaboration with data specialists at the Ratio DH centre (ULB).⁴⁵ **Sébastien de Valeriola** (Ratio DH, ULB) details the research method designed for the automatic reconstruction of newspaper newsrooms based on signature recognition in digitised newspaper corpora. A key challenge is the often poor quality of OCR in available data files, which results in errors in signature recognition and article segmentation. To address these issues flexibly, an

45. <https://quadihum.ulb.be/ratio-dh/>.

automatic classification model was designed to distinguish between lines on a page that are signatures and those that are not. This is a supervised task, where the model was trained on a set of manually labelled data extracted from specific years of a single newspaper. The results of this model have been promising, paving the way for its broader application in the future.

The digitisation of historical newspapers and periodicals has profoundly changed how these sources are accessed and studied. While this transition has brought immense benefits—enhancing accessibility, searchability, and preservation—it has also introduced new challenges. As digital access becomes the norm, heritage institutions must navigate the diverse expectations and needs of collection managers, researchers and the general public.

Balancing these demands with the practical constraints of digitisation, such as resource limitations and the complexity of historical materials, remains a significant challenge. While technological advances offer new possibilities for access and analysis, they also raise questions about long-term usability and the role of heritage institutions in shaping digital research practices. Ongoing innovation in digital preservation, artificial intelligence, and user engagement will further shape the field.

In Belgium, the field of newspaper digitisation has made significant progress. The examples of ongoing research show what tremendous scientific gains can be realised thanks to the availability of qualitative data and the development of new research methods. Important collections of historical newspapers and magazines have already been or will soon be digitised, while heritage institutions are increasingly treating their collections as data. However, these developments are progressing at different speeds and sometimes lack cohesion. Institutional fragmentation means that initiatives are sometimes spread across various institutions and institutional levels.

Collaboration between the different players is therefore crucial. By pooling expertise and resources, they can work toward more integrated platforms, standardised metadata, and sustainable long-term solutions. Universities and researchers also have a role to play, as technical solutions developed for research projects may also be a benefit for heritage institutions. This volume seeks to contribute to this ongoing dialogue by highlighting achievements to date, while also identifying remaining challenges and areas for further development.

PART 1. State of the art

1. Reflections on a Workshop: Research Challenges and Opportunities for Belgian Print Media in the Age of Digitisation

Kaspar Beelen

Abstract

This essay provides a personal recollection of the conference *Belgian Print Media in the Age of Digitisation*, highlighting various challenges and opportunities related to working with digitised sources at scale. Emphasising the fragmentary nature of the conservation and digitisation of the Belgian press, the essay sketches possible ways forward. It stresses the need for a comprehensive understanding of the newspaper landscape and, in general, advocates a *collections as data* approach to enhance accessibility and facilitate computational analysis of the historical press.

Author

Kaspar Beelen is a digital historian, who explores the impact of datafication and machine learning on humanities research. After obtaining his interdisciplinary PhD in history and linguistics (2014) at the University of Antwerp he worked as postdoctoral fellow at the University of Toronto, as researcher on the Digging into Linked Parliamentary Data (Dilipad) project. In 2016, Kaspar moved to the University of Amsterdam where he first worked as a postdoc and later became assistant professor in Digital Humanities (Media Studies). Between 2019 and 2023, he worked at the Turing Institute as research associate for the Living with Machines project, where his work focuses on understanding biases in large historical collections. Currently, he works as Technical Lead Digital Humanities at the School of Advanced Study, University of London, exploring the use multimodal and vision models for cultural heritage.

Keywords

Newspapers; digitisation; collections as data; metadata; data work

Introduction

In September 2023, I was invited to deliver the concluding remarks at a two-day workshop on *Belgian Print Media in the Age of Digitisation* (Université libre de Bruxelles, 4-5 September 2023). The workshop aimed to assess the current landscape of digitised print media in Belgium, focusing on the challenges and opportunities brought by massive digitisation, the scattered nature of archives, the evolving digital infrastructures, and assessing their impact on research methods and media history. This essay serves as a record of my ruminations and numerous reflections that followed the event.

This essay aims to highlight and synthesise the various debates surrounding the digitisation of the (Belgian) press. At the outset, I wish to emphasise that this essay offers a personal (and predominantly researcher and historian) perspective and might reflect more the author's recollections (and interests) than the presentations given at the workshop. Instead of a summary, I have attempted a re-interpretation that aims to: a) capture themes that resurfaced during various presentation and discussions, b) articulate a few ideas that might help inform the future digitisation and study of the Belgian press (I am carefully avoiding the presumptuous term 'solutions' here).

Metadata and the Newspaper Landscape

Throughout the workshop, participants repeatedly described the conservation and digitisation of that Belgian press as 'fragmentary' or 'scattered'. The problem is not confined to the lack of a common access point but, more fundamentally, the absence of a centralised record that could help us understand and depict the comprehensive press landscape—the totality of newspapers that have circulated. As digitisation efforts expand, it becomes increasingly important to understand the sample of digitised newspapers within the overarching informa-

tion landscape and come to grips with what digitised collections ‘represent’ as objects of historical inquiry.

These issues are even more important as digitisation enables historians to work on larger scales. Historians are accustomed to working with a small number of carefully selected sources. However, they are less comfortable when required to operate and think at larger scales, a situation particularly problematic when dealing with newspaper corpora which can be simultaneously immense, scattered and opaque. How can we reasonably interpret information extracted from digital newspaper data when we remain uncertain about what it is we are looking at, what has been digitised, and what is missing?

Confidence in research findings might be undermined if scholars fail to identify which voices or perspectives are missing (or amplified) because of the fragmented and partial digitisation of the press. The researcher must, therefore, remain mindful of absences when studying the digitised press on a large scale. Invisibility does not imply non-existence. What we observe when browsing and searching collections is, in many ways, just the tip of the iceberg (a metaphor repeated during the workshop). This not only relates to what information we encounter (i.e. the historical documents), but also how we access these documents (often through specific renderings and interfaces that hide lots of complex, preparatory work). The information we observe is the result of a long process involving selection, digitisation and multiple data transformations. While we might be aware of these limitations, understanding their implications is often challenging.

But what to do about this, and who can help out? One strategy to overcome this situation, is to put more value on academic outputs that are not journal articles. The challenges posed by fragmented data are compounded by a broader issue in academia which tends to undervalue ‘data work’—by which I mean the creation, enrichment but also critical interrogation and representation of data—as academic research. Often such outputs fall through the cracks of recognition and reward systems. Time spent on enhancing existing data or creating new datasets—through transcription, annotation, digitisation of new data or merging existing information sources—is still undervalued.

My main argument here is that digitising newspapers (and building interfaces on top of them) is an important but not sufficient condition to enable

the critical use of newspaper collections for research purposes. Researchers, libraries and other institutions would benefit from (further) promoting a *collections as data* approach: this would enhance accessibility at scale, facilitate computational analysis, and perhaps offer new research possibilities, it would also motivate scholars to think of collections as complex constructs that require their own type of source criticism.

A related suggestion for improvement concerns metadata, which offers, to me at least, an excellent starting point for the critical interrogation and construction of digital collections. For example, in the *Living with Machines* project—a large-scale investigation of mechanisation in the 19th century using immense historical collections, such as the British Newspaper Archive—we spent a lot of energy creating newspaper metadata from historical resources such as press directories, combining this information with records retrieved from the British Library catalogue. Doing this not only provided us with a rich and detailed picture of the newspaper landscape, for example by articulating the geographical and political diversity of the press, it also helped us to understand biases and priorities in digitisation and how these might affect historical research.⁴⁶

In a Belgian context, the Abraham catalogue⁴⁷ is an excellent example of a similar endeavour. This project recorded Belgian newspapers since 1800 based on descriptions and inventories held by libraries in Flanders and Brussels. The data enables us to analyse the press landscape. Also inspiring was the presentation by Katrien Weyns (KADOC) which questioned the extent to which automation can assist in the selection process for digitisation, and investigated how to make results more accessible by publishing bibliographic information on platforms such as Wikimedia (see the contribution by Roxanne Wyns and Katrien Weyns in this volume).⁴⁸ Such efforts to improve and standardise meta-

46. BEELEN, Kasper. Digitising Newspapers Press Directories to Understand the Landscape of Historical Newspapers. *Living with Machines* 2023 [Online]. Available at: <https://livingwithmachines.ac.uk/digitising-newspapers-press-directories-to-understand-the-landscape-of-historical-newspapers/>.

47. <http://www.krantencatalogus.be>.

48. WEYNS, Katrien. *Automation of journal selection for digitisation in a fragmented library landscape*. Conference lecture. 2023.

data of cultural heritage and the historical press more specifically, are an impressive step forward, potentially increasing confidence in the results gathered from the digitised press among historians and other scholars. These advancements in metadata management and standardisation provide a foundation for more nuanced and reliable analysis of digital newspaper archives.

Infrastructure: Transparency, Visibility and Adaptability

The digitised press provides incomplete, biased information but does so often in a very convenient, i.e. searchable, format. Reactions by historians range from ‘digiphobia’ (rejecting all digitised materials as inherently biased or incomplete) to ‘digitalaziness’ (simply accepting whatever the data presents). However, neither extreme is satisfactory. The *collections as data* approach mentioned earlier provides a viable and fruitful middle-way. We need critical and computational tools to make collections understandable and transparent, enabling us to describe vast amounts of newspaper information both qualitatively and quantitatively.

Transparency concerns not only the composition of the data—i.e. issues of bias and absence which we discussed in the previous section—but also its transformation and presentation, which brings us to the topic of interfaces. Transparency is crucial when designing interfaces for the analysis and exploration of digital heritage. The analysis of digitised newspapers—be it through code and application programming interfaces (APIs), or through graphic user interfaces (GUIs)—usually relies on data that emerged throughout a complex process of digitisation and transformation, a process which produced noisy and partial reproductions of the ‘original’ (i.e. the paper page).

These issues were highlighted in the keynote presentation on the *Impresso* app, an interface designed for exploring digitised historical newspapers.⁴⁹ The speakers Maud Ehrmann (École Polytechnique Fédérale de Lausanne) and Marten Düring (University of Luxembourg) discussed the challenges of designing research-oriented interfaces.⁵⁰ Controlling the flow of information from

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

50. DÜRING, Martin & EHRMANN, Maud. *Impresso - Media Monitoring of the Past*. Conference lecture. 2023.

backend to frontend requires a balancing of the need for transparency with usability requirements. Sharing every detail is impractical, as it would make navigating large newspaper collections unfeasible, especially for less digitally literate users. However, putting more effort in the co-development of tools, producing extensive tutorials, and of course writing extensive documentation, provides a major step forward to making computational research of newspapers more reliable, transparent and reproducible.

In emphasising a need for transparency, systems must strike a delicate balance between providing comprehensive information and maintaining an intuitive, user-friendly interface. This balance is crucial for ensuring that digital newspaper archives remain accessible and useful for researchers with varying levels of technical expertise.

Another topic frequently raised during the workshop concerned the use of computational tools for analysing the digitised press. Digital Humanities research is often tool-focused, yet the tools themselves are frequently transient. In other words, there exists no fixed set of models or tools that constitutes a core toolkit for DH researchers. This makes it challenging for libraries and other GLAM (Galleries, Libraries, Archives, and Museums) institutions to tailor their services to researchers. As Julie Birkholz (KBR and Ghent University) noticed in her presentation on digital workflows for newspaper and press collections, there exists ‘a tension between the services that holding institutions provide versus the users’/researchers’ needs’. This tension is partly the result of rapid technological changes as well as an incredible diversity in research requirements and needs which cannot simultaneously be accommodated by libraries and holding institutions. Possible ways forward she suggests are to encourage researchers to invest in term engagements with data and their collection managers, as well as collaboratively develop sustainable ways for ‘exchanging knowledge and getting feedback’.⁵¹

Relevant in this context is a recent study commissioned by the Dutch National Library regarding their intention to build a ‘text suite’ that bridges the gap between simple exploration and advanced textual analysis. Based on

51. BIRKHOLZ, Julie M. *Digital Workflows for Historical Newspapers & Periodicals: Why Both the Start and the End Matters*. Conference lecture. 2023.

discussions with researchers, the study's authors recommended against building such a research platform.⁵² The primary reason for this recommendation was that researchers' workflows and corpora are highly specific and tailored to individual research questions. Beyond the need to filter and select data, these workflows are too diverse to be effectively served by a single platform (which, moreover, would risk becoming outdated quickly as methods change dramatically over time). This finding underscores the complexity of developing tools for Digital Humanities research. While there is a clear need for computational tools designed to facilitate the analysis of the digitised press, creating a one-size-fits-all solution has proved challenging. Institutions must balance the desire to provide useful resources with the reality of the diverse and evolving research methodologies that are evolving in the field.

Newspaper Collections as Data

This connects to another critical observation: properly managed data tends to be more persistent than tools, which have been changing at an increasingly rapid pace in recent times. In this context, *collections as data*—a recurrent theme that resonated throughout the workshop and this essay—becomes a crucial concept for the future of digital newspapers in libraries.

The *collections as data* approach aims to develop cultural heritage resources in ways that enable computational research—a form of data stewardship still rarely undertaken by GLAM institutions.⁵³ This approach recognises that while tools may come and go, well-structured and well-documented data can remain valuable and usable over time.⁵⁴

For digital newspapers, adopting a *collections as data* approach might entail not just giving access to digitised images and OCR text, but also to structured

52. KEMMAN, MAX & CLAEYSSENS, STEVEN. User Demand for Supporting Advanced Analysis of Historical Text Collections. *DH Benelux 2022 - ReMIX: Creation and alteration in DH* (hybrid), Belval Campus, Esch-sur-Alzette, Luxembourg and online. Zenodo. DOI: <https://doi.org/10.5281/zenodo.6595769>.

53. *Always Already Computational: Collections as Data* [Online]. Available on <https://collectionsasdata.github.io/>.

54. ONB LABS [Online]. Available on: <https://labs.onb.ac.at/en/>.

metadata, named entity annotations, and other machine-readable formats. This would allow researchers to apply a wide range of computational tools and methods to the newspaper collections.

By focusing on creating robust, well-documented and accessible datasets, libraries and other GLAM institutions can provide a more flexible and enduring resource for digital humanities researchers. This approach aligns with the diverse and evolving needs of researchers, as discussed in the context of the KB Workbench and KBR Labs, while also ensuring the long-term value and usability of digitised newspaper collections.

The *collections as data* concept may also foster new relationships between libraries and researchers. Libraries frequently provide services that allow researchers to access and analyse newspaper data. However, the results of this research often do not flow back to enrich the library collections. We need to think more broadly about how research use of the digitised press can be leveraged to enhance and expand the original collections.

This perspective inverts the traditional relationship, suggesting a new motto: ‘Ask not what your library can do for you, but what you can do for your library’. By encouraging a more symbiotic, enduring and stable relationship, both libraries and researchers can benefit from shared knowledge and improved resources.

The presentation of Sébastien de Valeriola (Université libre de Bruxelles) focussed on the automatic recognition of signatures in Belgian newspapers in the context of the CAMille project (see his article in this volume). In CAMille’s attempt to reconstruct ‘the newsroom’ of historical French-speaking newspapers, the project developed computational methods to detect and extract the journalist’s signature often found at the bottom of an article. This type of research has excellent potential for contributing to and enriching library collections in meaningful ways.⁵⁵ Also interesting in this context was the presentation of Sébastien Hermans (KU Leuven and KBR), which showed how cataloguing of individual issues of digitised historical journals generated fine-grained and rich metadata which historians could explore (see the contribution by Sébastien Hermans and

55. DE VALERIOLA, Sébastien. *Automatic Recognition of Signatures in a Corpus of Belgian Newspapers*. Conference lecture. 2023.

Morgane Ott in this volume).⁵⁶

These contributions could significantly enhance the usability and searchability of digital newspaper collections. For instance, accurate genre classification could allow for more nuanced analysis of historical trends in journalism, while improved signature identification could aid in authorship studies.

By fostering such a reciprocal relationship, libraries can amplify their role as repositories of information, by encouraging collaboration and becoming dynamic platforms for research and knowledge creation. This approach aligns well with the concept of *collections as data*, as it encourages the continuous improvement and enrichment of digital resources through active use and contribution by the research community.

Access and Acquisition: The Things We Don't Talk About

Even for those researchers who are enthusiastic proponents of collection as data in theory, the reality of working with newspapers is often complex and riddled with hurdles. A major bottleneck for many research projects is data acquisition and access, especially when approaching newspaper collections as data. Various issues whether technical (such as handling and processing terabytes of data) or legal (including copyright restrictions) still complicate matters. Although many projects and researchers face similar challenges, solutions are often sought in isolation.

Solutions to these problems can be partly achieved by building trust between data users and providers, and partly by creating platforms for sharing strategies and solutions. One example is the Living with Machines project, which not only adapted a tiered data security approach but also reflected extensively and publicly on the challenges of data acquisition. This tiered approach allows for different levels of data access based on user needs and permissions, thus balancing security concerns with research requirements.⁵⁷

56. HERMANS, Sébastien. *Cataloguing Individual Issues of Digitised Historical Magazines*. Conference lecture. 2023.

57. ARENAS, Diego, ATKINS, John, AUSTIN, Claire and others. *Design Choices for Productive, Secure, Data-Intensive Research at Scale in the Cloud*. 2019 [Online]. Available on: <https://arxiv.org/abs/1908.08737>.

While we often write extensively about data, methods, and findings, the process that makes all of this possible often remains in the dark. Therefore, it is essential to share insights into the challenging and vital work of data acquisition. By openly discussing these processes we can collectively improve our approaches to accessing and utilising digital newspaper collections, ultimately enhancing the quality and scope of research in this field.

Errors are Bliss

While acquiring data is often hard, analysing them is, of course, also challenging. The most persistent issue is noise, primarily in the form of OCR errors and faulty page segmentation. These challenges won't disappear anytime soon. However, imperfect information does not mean useless information, even though humanities researchers tend to quickly discard data that doesn't meet their (sometimes unrealistic) expectations.

The solutions are partly attitudinal, requiring a nurturing of pragmatism, and partly methodological, as humanists need more training in dealing with uncertainty and error at scale. While the scale of newspaper data often mitigates the impact of noise on results, this relationship still warrants critical investigation. Errors are not randomly distributed; they appear at higher rates in certain types of newspapers, often those printed on lower-quality paper. This uneven distribution of errors could potentially skew research outcomes, particularly in studies focusing on specific newspaper types or time periods.

Therefore, errors are part of the digital workflow, and addressing this challenge presents an opportunity to showcase creativity and problem-solving abilities. Researchers might develop new methods to detect and correct errors, or design studies that are robust to the types of errors commonly found in digitised newspapers.

Digitisation amounts to more than collecting everything into one database. With access comes responsibility. As we continue to digitise and analyse historical newspapers, we must remain mindful of the limitations and challenges inherent in this process. By adopting a pragmatic approach and fostering a culture of transparency about data quality and bias, we can substantially improve the reliability and the research value of these historical resources.

The Digital is Still Material

The future of the digitised historical press lies in a more holistic approach that incorporates materiality into digital analysis. This evolution recognises the multimodality of digitised press, acknowledging that historical newspapers are not merely text but complex artifacts with various visual and physical elements. The materiality of original sources significantly influences the digitisation process, shaping how information is captured and presented in digital formats.

As we advance, there will be a greater emphasis on understanding historical reading practices and information organisation, allowing for more authentic digital representations. Importantly, the visual aspects of newspapers will be increasingly treated as crucial context for their textual content, rather than as separate or secondary elements. This integrated approach will provide researchers and readers with a richer, more nuanced understanding of historical press materials in their digital form.

Concluding Remark

Despite the importance of issues such as fragmentation, invisibility, and transparency in the realm of digitised press, the way forward remains unclear. A key observation that emerged from the discussions is that technical as well as institutional problems often necessitate social solutions. Workshops such as *Belgian Print Media in the Age of Digitisation* bring together diverse stakeholders and disciplines, and represent a crucial step in the right direction. The workshop emphasised the importance of increased coordination and collaboration between institutions, researchers, and international networks to more effectively utilise digitised historical newspapers and periodicals for academic and public purposes. A similar observation also applies to skills. Unlocking the (research) potential of the digitised press is, to a large extent conditional on training and collective upskilling and collaboration, and especially on improving critical and computational abilities to process and analyse collections at scale.

2. Digital Access to Historical Newspapers in Belgium and Abroad. From Endangered Heritage to a Gold Mine for Machine Learning Technology⁵⁸

Christoph De Spiegeleer

Abstract

This essay reflects upon some important trends in Belgium when it comes to digital infrastructure that offers access to digitised print media. In an ideal world, the digital accessibility of Belgian and, by extension, European newspaper heritage would adopt the form and approach of *Chronicling America*. This American initiative is considered a model example of an integrated umbrella platform that meets the wishes of different user groups, from those merely curious about the front-page news of a century ago to specialist digital humanities researchers. However, such a scenario has little chance of success in the near future, due to project-related financing models and the associated fragmentation of digital expertise-building at regional, national and international level in both the heritage sector and the field of digital humanities. Truly integrated (inter)national accessibility of digitised newspaper heritage for different user groups is still a thing of the future.

Author

Christoph De Spiegeleer is head of research at Liberas, a heritage centre for the history of the liberal movement and the freedom ideal in Belgium. He is also guest professor in History at the Vrije Universiteit Brussel (VUB) and Ghent University. He obtained his PhD in modern history at the VUB.

Keywords

Newspapers; digitisation; metadata; OCR; collections as data

58. A shortened, Dutch version of this essay has been published in *META. Tijdschrift voor bibliotheek en archief*, 2023, n° 8, p. 10-15.

The digitisation of physically fragile newspaper collections requires significant efforts and investment from a large number of heritage institutions. Researchers in the digital humanities are experimenting with artificial intelligence technologies to make the data of digitised newspaper collections more searchable across national borders. However, truly integrated (inter)national accessibility of digitised newspaper heritage for different user groups is still a thing of the future.

A Digital Eldorado?

Newspapers, i.e. regular newspaper publications that report (usually daily) on current affairs, are, as a historical source, a treasure trove for research in almost every social domain. However, newspapers were not made for eternity and the paper used for daily issues from the 1850s-1950s has now become very acidic. Archives and heritage libraries are working to give this vast physically endangered reservoir of historical knowledge a sustainable digital future.



Fig. 1. In January 1900, the Library of Congress in Washington opened a reading room where researchers could consult newspapers and magazines (source: Library of Congress Prints and Photographs Division).

Since 2021, Liberas, a cultural archive accredited by the Flemish government,⁵⁹ has been digitising its collection of newspapers and magazines to A3 format via a completely internal procedure. Internal staff are responsible for every step of the digitisation process. This ranges from scanning the pages and creating metadata for each page, to processing the digital images using OCR (optical character recognition) software and storing the full text and layout information of automatic text recognition in the ALTO XML standard. Finally, Liberas uploads these images and XML files to the collection management system. Between 2021 and 2023, Liberas was already able to digitise 120,000 pages from its collection of newspapers and magazines and make them searchable.



Fig. 2. Scanning pages as part of the internal digitisation procedure for newspapers and magazines up to A3 size in the collections of Liberas (source: Liberas).

However, the user should not forget that these digital collections only give a false impression of completeness and are, in reality, always heterogeneous, incomplete, and therefore not representative. Digitisation operations are often

59. <https://www.liberas.eu/en/>.

initiated at the request of specific ad hoc projects and because of the physical state of certain collections. This applies to both the local collections of municipal archives and the large collections of national libraries. Access to over 33 million digitised press pages via the online British Newspaper Archive, for example, makes it easy to forget that these millions of pages make up only 6% of the British Library's newspaper collection.

In order to compile a balanced digitised source corpus, users and researchers benefit from online databases of historical newspapers offering sufficient contextual information about the digitised titles within their holdings (circulation, ideological profile, publication period, etc.). *Chronicling America* offers such information via brief title essays.⁶⁰ *Chronicling America* is the Library of Congress' landing page for collections of digitised historical newspapers that have been and are supplied according to the same guidelines by institutions from all U.S. states. *Chronicling America* is also an exemplary model when it comes to visualisation of the collections. The platform regularly publishes updated and interactive visualisations (timelines, maps) that give users a picture of the geographical and chronological imbalances in the supply of digitised newspapers.

Digitisation according to professional standards is labour-intensive, and sustainable online access to the data of digitised newspapers is essential. Even the collections of smaller institutions demand the scanning of hundreds of thousands of pages. When Belgian and Dutch heritage institutions outsource the digitisation of original newspaper collections, their questions concerning compliance with the quality requirements of the *Metamorfoze Light* guideline⁶¹ – part of the Dutch programme for preserving the nation's paper heritage – are not always compatible with the options offered or the procedures undertaken by commercial digitisation companies. The internal quality control of externally

60. <https://chroniclingamerica.loc.gov/>.

61. VAN DORMOLEN, Hans. *Metamorfoze Preservation Imaging Guidelines*. The Hague: Koninklijke Bibliotheek, 2012. Available online: https://www.metamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf.

digitised newspaper pages is also very time-consuming for heritage institutions and requires the right tools. In addition, each institution depends on the capabilities and limitations of its collection management system, such as the possibility to upload the text-recognition output formats. OCR technology does not stand still, either, and new text recognition programmes can offer significant improvements in quality over the OCR of even ten years ago.

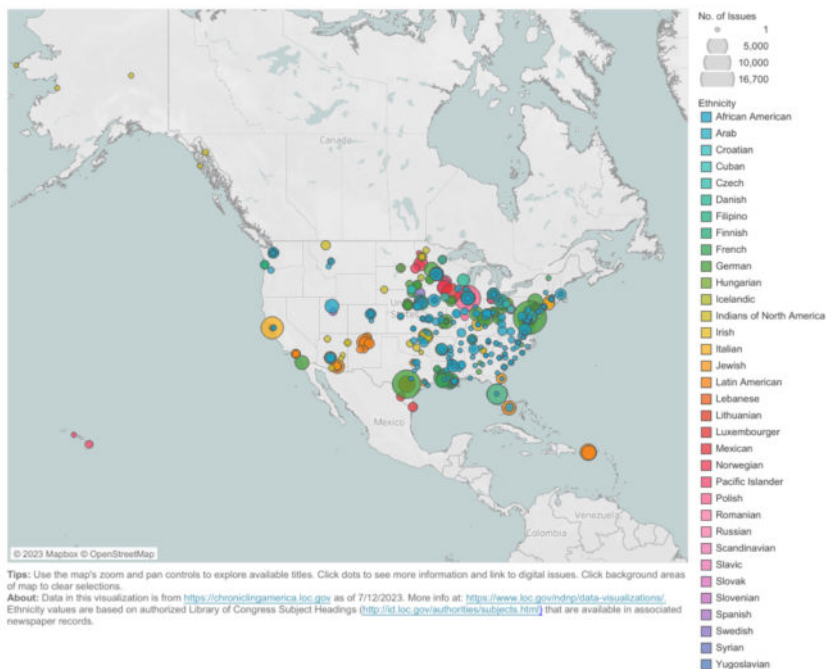


Fig. 3. Updated in July 2023, this interactive map shows the locations of historical U.S. newspapers targeting specific ethnic and immigrant communities and which are available digitally via Chronicling America (source: Library of Congress).

Confronting Fragmentation

In Europe, we see a fragmented landscape of digital infrastructure providing access to historical newspapers at regional, national, and international level.



Fig. 4. The journalist Piet Van Brabant looks through various newspapers in the editorial office of *Het Laatste Nieuws*, in 1962 (source: *Liberas*).

Belgium is also experiencing growth in access platforms with diverse functionalities at the level of individual organisations and regional partnerships. Heritage institutions operate primarily as independent entities at different policy levels with different funding authorities.

This variety of heritage institution platforms within a small country like Belgium, or even a region like Flanders, is valuable in itself. Mass digitisation programmes based on major investment programmes following the same standards and procedures – and the resulting umbrella platforms that provide access to digitised collections from dozens of custodians – require years of preparation. With their own platforms and digitisation programmes and based on their own collection profile, heritage institutions can respond more quickly and flexibly to opportunities and thus gradually expand or supplement their own digital offering through partnerships with other organisations. *Liberas*, a medium-sized Flemish heritage institution, currently offers digital access to 500,000 searchable pages of newspapers and magazines thanks to this approach. More than a quarter of these pages became accessible in June 2024 when *Liberas* completed up-

loading files containing 22,200 digitised editions of the local liberal newspaper *Journal de Gand* (1856-1923) and its predecessors. Another example is the wide range of local and regional Flemish newspaper databases that has been created through the many efforts of local heritage services (*erfgoedcellen*), city archives and heritage libraries.



Lees 380.000 pagina's nieuws uit WW1

80.374 nummers van Belgische kranten en tijdschriften die werden gepubliceerd tussen 1914 en 1918 zijn digitaliseerd en vrij online toegankelijk via drie open portalen:

Nieuws van de Grote Oorlog

904 titels
52.370 nummers
ca. 270.500 pagina's

Naar Nieuws van de Grote Oorlog ↗

The Belgian War Press

91 titels
22.115 nummers
ca. 84.300 pagina's

Naar The Belgian War Press ↗

BelgicaPress

19 titels
5.889 nummers
ca. 27.600 pagina's

Naar BelgicaPress ↗

Fig. 5. The large online collections of Belgian newspapers from the First World War are spread over three 'digital archives' (source: *Nieuws van de Grote Oorlog*).

Nevertheless, we should not be blind to the potentially adverse consequences of increasing fragmentation, especially when it comes to overarching access

platforms in the Belgian context. Over the last ten years, for example, major digitisation projects with changing content partners have made almost a thousand digitised Belgian newspaper titles that appeared during the First World War available via three newspaper platforms. For a complete overview, the user needs to check whether the issues of the Belgian war newspaper he or she is looking for are included in *Nieuws van de Grootte Oorlog*⁶² (a platform maintained by meemoo, the Flemish institute for archives),⁶³ The Belgian War Press⁶⁴ (a platform maintained by CegeSoma, the Study and Documentation Centre for War and Contemporary Society) or BelgicaPress (a platform maintained by KBR, the Royal Library of Belgium).⁶⁵ *Nieuws van de Grootte Oorlog* has recently ceased to exist as a separate platform. As of August 2025, the digitised war newspapers from Flemish collections are accessible through the renewed Flemish platform hetarchief.be (see below).⁶⁶

Users who are less experienced in searching different databases are best served by digitisation programmes and portal environments at national levels managed by national libraries. *Chronicling America*, *Trove*⁶⁷ and *Delper*⁶⁸ are good examples of this. Library of Congress-managed *Chronicling America* is funded through a central federal program that awards digitisation grants annually to university libraries, non-profit organisations and regional government agencies in order to add digitised newspaper collections from the various U.S. states to the central platform. *Delper*, managed by the National Library of the Netherlands (the KB), also functions as a gateway to the Dutch newspaper collections of numerous institutions at home and abroad and already integrates the search results of eight external regional newspaper databases on the platform. The *Trove* portal, operated by the National Library of Australia, provides access to digitised historical newspaper collections from Australia's main state libraries.

62. <https://nieuwsvandegrooteoorlog.hetarchief.be>.

63. <https://meemoo.be/>.

64. <https://warpress.cegesoma.be/>.

65. <https://www.belgicapress.be>.

66. <https://hetarchief.be/>.

67. <https://trove.nla.gov.au/>.

68. <https://www.delpher.nl/>.

Surpassing Borders and Project-Related Subsidies

The state structure and linguistic situation in Belgium make cultural heritage a language-based Community matter. This complicates the uniform and national management of the digitised newspaper heritage and makes a national and central platform that functions as a gateway, such as Chronicling America, Trove and Delpher, less likely to succeed in Belgium. The non-profit organisations *Vlaamse Erfgoedbibliotheken* (Flanders Heritage Libraries)⁶⁹ and meemoo are working with the *Nieuwe Tijdingen* project on the development of a digitisation programme for Flemish newspaper heritage (see Sophia Rochmes' contribution in this volume).⁷⁰ The aim of the preparatory project launched in 2021 is to have a programme ready to digitise historical newspapers in Flanders, to archive them in a sustainable way and to make them accessible in a centralised way via a renewed version of the platform hetarchief.be (deployed in August 2025). The focus for future digitisation is on historical newspapers published within the current borders of Flanders, or newspapers published outside of Flanders with Flemish people as their target audience.

Parallel to *Nieuwe Tijdingen*, meemoo also carried out the newspaper digitisation project *Primeur* (2021-2023) together with Flanders Heritage Libraries. Through this relaunch project, meemoo has already digitised 630,000 newspaper pages from the collections of eight heritage institutions. Under the *Nieuwe Tijdingen* programme, 65,000 pages of small and local newspaper collections have already been digitised as a test case. These overarching digitisation projects are separate from the further expansion of the existing BelgicaPress newspaper platform of KBR. KBR plans to double the number of pages available on BelgicaPress today by 2027. Ongoing consultation continues, of course, between the parallel digitisation programmes at the Flemish and federal level in relation to selection lists.

Although Flemish 'rescue programmes' designed to safeguard endangered

69. <https://bibliothecairerfgoed.be>.

70. *Nieuwe Tijdingen* [Online]. Available on: <https://meemoo.be/nl/projecten/nieuwe-tijdingen-redt-historische-kranten-van-het-verval>.

newspaper heritage, such as *Primeur* and *Nieuwe Tijdingen*, are valuable, they unfortunately suffer from dependence on project subsidies. The building of project-based expertise in the digitisation of newspapers in these rescue programmes is no guarantee that such expertise will be incorporated into the broad Flemish cultural heritage sector. In ideal circumstances, sustainable digital transformation takes place at the organisational level (“bottom-up”) through the regular financing of day-to-day operations. However, far from all institutions within the Flemish cultural heritage sector have the necessary resources and trained staff to follow the latest developments in the digitisation of newspapers, let alone to implement these developments in their own digital infrastructure.

In other words, not everything moves at the same pace in the Flemish cultural heritage sector. That is why meemoo’s service-oriented work in the area of newspaper heritage does not focus solely on the preparation and implementation of mass digitisation projects and the build-up of associated expertise. Meemoo also supports content partners of all sizes in setting up and carrying out their own digitisation procedures for their newspaper and magazine collections. This is done, for example, through sharing knowledge and expertise in relation to tendering and quality control. Meemoo has been organising the ‘Digitisation of Periodicals’ collegiate group together with Faro, the Flemish Institution for Cultural Heritage,⁷¹ and Flanders Heritage Libraries since 2020. Since June 2023, meemoo’s digital services have also been enshrined as a structural part of Flemish cultural policy by way of a separate decree.

Collections as Data

In addition to the problem of the differing speeds of digital transformation among heritage organisations within and across the language-based community boundaries in Belgium, Belgian heritage organisations are confronted with multiple expectations of user bases regarding the accessibility of newspapers through digital infrastructure. The size of the organisation does not materially change these user expectations that can be varied as the users themselves. Academic researchers sometimes have wishes and needs different to those of users with a search query (genealogists, teachers, students, politicians, journalists, curious people among the

71. <https://faro.be/>.

general public, etc.). Most users and many professional researchers and heritage experts usually search for and find information based on keywords through qualitative techniques of close reading (human analysis of singular texts as sources). Qualitative techniques of close reading techniques only require a JPG or PDF download of the pages or editions and a possibility to highlight and save search results. However, more and more academic researchers who together make up the research community and the field of ‘digital humanities’ also want to engage in distant reading (computer-controlled analysis of large machine-readable data volumes). Digital humanities researchers are a small but very active subgroup within the user groups of digitised newspapers collections.

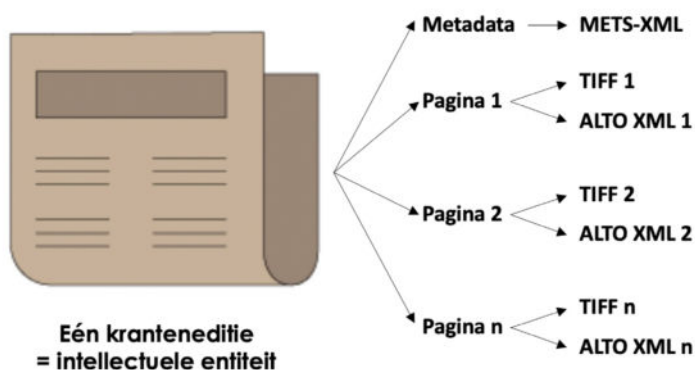


Fig. 6. The structural metadata relating to a digitised newspaper edition is best stored in the XML standard METS (source: meemoo, CC BY-SA 4.0).

As a result, there has been widespread friction in recent years between the sometimes high expectations of groups of academic researchers who want to apply various computational techniques to digitised newspaper collections and the (limited) possibilities of heritage institutions. When using online newspaper databases, researchers and data scientists active in the digital humanities want to be able to compile datasets themselves on the basis of search queries us-

ing APIs (application programming interfaces), regardless of the user interface. Furthermore, researchers in the digital humanities are calling for the possibility of bulk downloads of large, curated sets of data files. In the case of digitised newspapers, this comes down to the metadata, the images and the OCR results in machine-readable formats. This requires the heritage institutions to adopt an approach that focuses on the *collections as data* principle, i.e. access to digitised collections at the level of the underlying databases.

The digitised and searchable newspaper collections of European national libraries are, for the purposes of the digital humanities, still too often institutional ‘silos’. For example, BelgicaPress, which today already provides access to over 140 Belgian newspapers from the Dutch regime (1815-1830) up to the 1960s via OCR, does not yet grant access to the newspaper collections as datasets. KBR is currently working on the DATA-KBR-BE project (2020-2024), aimed at enabling access to KBR’s digitised newspaper collections at data level in the future.⁷² The KBR DATA Platform has yet to be publicly launched. Moreover, the new Flemish newspaper interface within hetarchieff.be does not provide an open data platform. In the neighbouring countries, the national libraries of Luxembourg and the Netherlands are the furthest along in granting access to APIs and releasing curated datasets of digitised newspapers. The *collections as data* movement comes from the United States. So it is not surprising that Chronicling America allows the collection of datasets in HTML, XML or JSON formats via the platform API based on search queries and offers the possibility to download bulk datasets – irrespective of the API.

Incoherent Datasets and OCR Noise

Even if newspaper databases release the underlying datasets at the request of researchers, the structural problem of the mutually differing quality of these databases still arises. When collecting the underlying databases of large, digitised newspaper collections from different countries, researchers often encounter varying and incomplete data (incoherent filling-in of fields and variation in terminology concerning the metadata, different formats, corrupted zip archives, etc.).

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

This explains the flawed capabilities of the newspaper interface on the Europeana platform. Europeana Newspapers⁷³ brings together the (duplicates of) digitised newspaper collections, which were produced from separate digitisation programmes of heritage libraries from ten European countries, in one virtual Europeana environment. In this environment, the focus is necessarily on ‘discovering’ and not on ‘researching’. As a portal environment, the Europeana newspaper interface does not offer advanced search functions, and filtering options are limited.

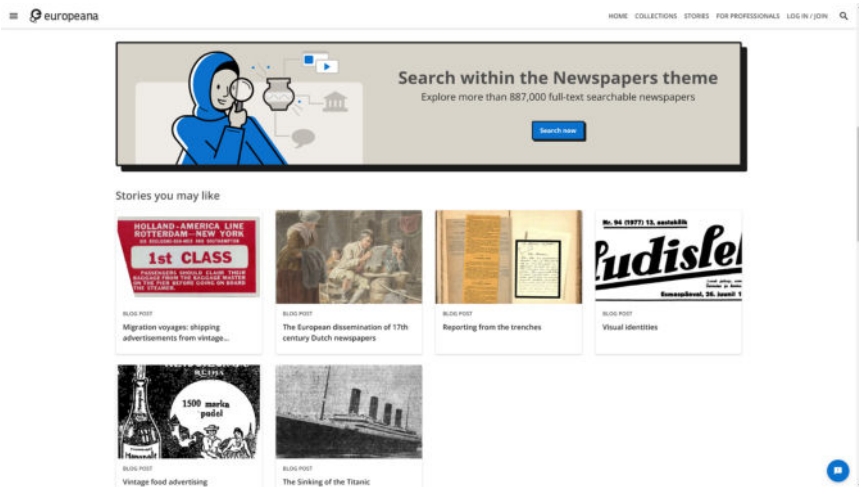


Fig. 7. In contrast to *Chronicling America*, *Europeana Newspapers* only addresses a wide audience of curious people and does not meet the needs of the professional researcher (source: Europeana).

Digital humanities researchers also warn of the impact of varying OCR quality (OCR noise) on the usability of certain digitised collections and certain periods in digitised collections for distant reading research. Tests based on collections from *Chronicling America* have made it clear that the precision of OCR can vary considerably within the collection of one digitised title, a fact of which users are often unaware. The physical state of a title’s source material can vary over time and the use of different OCR software in multi-year digitisation pro-

73. <http://www.europeana-newspapers.eu/>.

jects can also cause variations. In addition, we should not forget that, for budgetary and management reasons, *Chronicling America* is based on the digitisation of microfilms as standard, and not on the original newspapers. This exacerbates OCR noise. Also on the Australian portal Trove, most digitised newspapers have been/are scanned based on microfilm collections. Trove took the lead in 2009 by allowing users to correct OCR errors themselves. Crowd-sourced text correction, however, appears to improve a small percentage of the total collection at best, mainly with interventions based on genealogical questions.

Artificial Intelligence Through Machine Learning

From the perspective of the digital humanities, researchers are also requesting sophisticated user interfaces that offer the possibility of segmentation at the level of articles, automatically enrich the digitised newspaper texts with semantic annotations (topics, names of persons and events, genre, etc.) and provide advanced search options (content-related filters, suggestions of other keywords, etc.).

Within the digital humanities, there is active experimentation with machine learning techniques to allow researchers to set up resource corpora based on digital research methods. Data scientists develop algorithms to automatically recognise specific types of texts within defined newspaper collections (automatic text classification) and automatically segment information blocks (document layout analysis), such as articles, illustrations, titles. In machine learning technology, a computer system learns to find patterns in large quantities of text and data based on training sets.

For example, in the course of a collaborative project with KBR, the data scientists of IDLab at Ghent University developed algorithms to segment the front pages of the 1938 editions of the Belgian socialist newspaper *Le Peuple* at article level based on layout characteristics.⁷⁴ IDLab's data scientists also man-

74. ALI, Dilawar, MILLEVILLE, Kenzo, VERSTOCKT, Steven, VAN DE WEGHE, Nico, CHAMBERS, Sally & BIRKHOLZ, Julie M. Computer Vision and Machine Learning Approaches for Metadata Enrichment to Improve Searchability of Historical Newspaper Collections. *Journal of Documentation*. 2024, vol. 80, n° 5, p. 1031-1056. DOI: <https://doi.org/10.1108/JD-01-2022-0029>.

aged to extract a certain type of article, the feuilletons or literary supplements, from these front pages based on textual and visual similarities. At both the Vrije Universiteit Brussel (VUB) and the Université Libre de Bruxelles (ULB), research groups in digital humanities and mathematics are currently involved in similar projects. The VUB-project developed algorithms that distinguish texts from non-textual elements in digitised editions of post-war Flemish humanist magazines.⁷⁵ The ULB-project developed algorithms that detect signatures of articles in digitised editions of the French-language Catholic newspaper *Le Vingtième Siècle* (see Sébastien de Valeriola's contribution in this volume).



Fig. 8. Article segmentation of the front page of *Le Peuple* (source: IDLab, Ghent University).

International Research Infrastructure

In recent years, various international and interdisciplinary research projects have launched interfaces that use machine learning technology to provide digit-

75. <https://ai.vub.ac.be/dsph/>.

ised collections of newspapers with new layers of information. These platforms aim to take the study of demarcated, multilingual collections to a higher analytical level through advanced search functions and article segmentation.

In the Swiss Impreso/Media Monitoring of the Past project (2017-2021), computer linguists and historians worked together on an interface accessible to researchers that makes it possible to query 76 Swiss and Luxembourg newspapers in precise ways via semantic indexing. Based on the search term entered, the Impreso application automatically generates semantically and orthographically similar search terms to be included in the search results⁷⁶. Via topic modelling, the user can highlight or filter out specific themes in the text fragments found. The interface also recognises people, locations and organisations and automatically links them to the Wikidata pages of these entities via named entity recognition.

The NewsEye project (2018-2022) is a similar European digital humanities-project that was completed recently.⁷⁷ The team behind this project launched a test platform in 2020.⁷⁸ On this platform, users can test the article segmentation and semantic text enrichment tools (named entity recognition, topic modelling, keyword suggestions) developed by the NewsEye team on a collection of 6 French newspapers, as well as 4 Austrian, 7 Finnish and 3 Swedish newspapers.

Despite these promising developments, the problem of fragmentation also arises within the field of digital humanities at both national and international levels. This is reflected in parallel investments in the same type of advanced research infrastructure that, depending on the project, is more or less user-friendly and publicly accessible and which provides access to different newspaper collections. Because the project-related funding derives from different government authorities, these digital platforms have no assurance of a sustainable future once their initial research grants have expired. After all, sustainable online embedding requires constant maintenance and continuous investment in refining the functionalities and further expanding the still limited multilingual newspaper collections (in the case of Impreso and NewsEye). In response to the high

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

77. <https://www.newseye.eu/>.

78. https://platform2.newseye.eu/users/sign_in.

development costs, the National Library of France already makes subscribers pay for the specialised functionalities of its advanced newspaper interface, RetroNews.⁷⁹

A Glimpse into the Future

It is logical that heritage institutions throughout Europe are mainly focused on the quantitative and sustainable expansion of their range of digitised newspapers for the widest possible audience via their own existing access platforms. In an ideal world, the digital accessibility of Belgian and, by extension, European newspaper heritage would adopt the form and approach of *Chronicling America*. This American initiative is considered a model example of an integrated umbrella platform that meets the wishes of different user groups, from those merely curious about the front-page news of a century ago to specialist digital humanities researchers.

The instigation of an equivalent system - ‘*Chronicling Europe*’, in effect - would require fully centralised coordination of digitisation policy that would initially transcend the regional and linguistic boundaries within European states such as Belgium and later also those of the nation states through structural financing at national and/or European level. Such a scenario has little chance of success in the near future, due to project-related financing models and the associated fragmentation of digital expertise-building at regional, national and international level in both the heritage sector and the field of digital humanities. In addition, a platform such as *Chronicling America* is not a perfect solution, due to the inherent gaps in what is offered and the varying OCR quality. Manual research into paper newspapers (whether microfilmed or not) will continue to be important in the coming years if researchers are to fully exploit the potential of newspapers as historical sources.

79. <https://www.retronews.fr/>.

Part 2. Digitisation: Institutions, Collections, Workflows

3. Belgian Press Heritage Online. Digitisation of Printed Media Collections at KBR

Brecht Deseure, Frédéric Lemmers & Sally Chambers

Abstract

The Royal Library of Belgium (KBR) holds a unique position in Belgium's historical press landscape. It manages the world's largest collection of Belgian newspapers and periodicals, making it Belgium's reference collection. Moreover, KBR has been a pioneer in the digitisation and online accessibility of printed media through its online platforms BelgicaPress and BelgicaPeriodicals. This article outlines the origins and composition of the collections and presents KBR's policy for digitising historical print media and making them accessible to the public.

Authors

Brecht Deseure obtained his doctorate in history at the University of Antwerp in 2011. He publishes on various aspects of Belgian history in the revolutionary and contemporary eras, including the history of political and constitutional culture, iconography, and the press. He is the lead researcher of the research centre *CAMille, Centre d'archives sur les médias et l'information*, hosted by KBR and the Université libre de Bruxelles. At KBR he is responsible for the historical newspaper collections. His publications include *Onboudbaar verleden. Geschiedenis als politiek instrument tijdens de Franse periode in België* (2014), *Constitution belge. Drafts and Proposals for the Belgian Constitution of 1831* (2020) and *Revolutie in Antwerpen. De aquarellen van Pierre Goetsbloets, 1794-1797* (2021).

Frédéric Lemmers is a member of the Royal Library of Belgium (KBR) Management Board in charge of leading KBR's digitisation policy. He is also

a part-time lecturer at ULB's Information and Communication Science and Technology Department where he teaches cultural heritage management. With more than 10 years' experience in managing digitisation projects at national and international levels (e.g. Europeana Collections 1914-1918), Lemmers has participated in various networked research projects related to digitisation and the digital humanities. He is active at an international level, where he occupies several executive positions within professional and scientific organisations.

Sally Chambers joined KBR, Royal Library of Belgium in October 2020 to coordinate the DATA-KBR-BE project to facilitate data-level access to KBR's digitised and born-digital collections for digital humanities research. She combined this with her role as Digital Humanities Research Coordinator at the Ghent Centre for Digital Humanities. In March 2022, she was appointed as a member of the DARIAH-EU Board of Directors. In 2024, she joined The British Library as Head of Research Infrastructures Services.

Keywords

Digitisation; data; metadata; OCR; newspapers; periodicals; collections as data

The Royal Library of Belgium (KBR) holds a unique position in Belgium's historical press landscape. It manages the largest collection of Belgian newspapers and periodicals in existence, making it the reference collection in Belgium and beyond. Moreover, KBR has been a pioneer in the digitisation and online accessibility of Belgian printed media through its online platforms BelgicaPress and BelgicaPeriodicals.⁸⁰ After an introduction into the origins and composition of the collection, this article will explain KBR's digitisation and access policies and provide a glimpse into future projects.

Origins and Composition of the Collection

Belgium (and the historical territories which preceded it) played a signifi-

80. <https://www.belgicapress.be>; <https://www.belgicaperiodicals.be>.

cant role in the history of the press. The region witnessed the emergence of newspapers during the early modern period and, thanks to liberal press laws, experienced a flourishing newspaper press in the 19th and 20th centuries.⁸¹ This richness is reflected in KBR's collections. As the national library, its fundamental task is to compile the most complete collection possible of publications issued in Belgium and in the historical Belgian territory or related to Belgium.

Complete comprehensiveness is, of course, an illusion, especially when it comes to newspapers and periodicals. Thousands of newspapers and magazines have been published in Belgium over the last four centuries. The well-known titles of the national and regional media outlets, which were mainly aimed at informing and shaping public opinion, have often been preserved in complete collections in several institutions across the country. Alongside them, many local or subject-specific periodicals existed with varying lifespans, many of which are absent from KBR's collection. Often, only a single—sometimes incomplete—run has been preserved, or in some cases, no copies at all. As a result, we can only guess at the total scope of the historical press in Belgium. The Abraham newspaper catalogue of the Flanders Heritage Libraries lists more than 12,500 newspapers and related publications as having been published in Flanders and Brussels (see Sophia Rochmes' contribution in the volume).⁸²

Since its founding in 1837, KBR has held a collection of discontinued and ongoing newspapers and periodicals, although they were not actively collected at that time.⁸³ Initially, KBR mainly held publications to which the Ministry of the Interior, to which the library was subordinate, subscribed. The collection grew thanks to the introduction of copyright legislation (1817 and 1840) and the subsequent bilateral agreement for the international protection of copyright in

81. ARBLASTER, Paul. *From Ghent to Aix: How They Brought the News in the Habsburg Netherlands*. Leiden and Boston: Brill, 2014; VAN DEN DUNGEN, Pierre. *Milieux de presse et journalistes en Belgique (1828-1914)*. Brussels: Académie royale de Belgique, 2005; DER WEDUWEN, Arthur. *Dutch and Flemish Newspapers of the Seventeenth Century, 1618-1700*. Leiden: Brill, 2017.

82. www.krantencatalogus.be.

83. The following paragraphs about the history and composition of the collections are mainly based on: BRUCHER, Roger. *Les collections de périodiques et de presse de la bibliothèque royale Albert Ier. Historique, structure et gestion*. Brussels: Royal Library of Belgium, 1984; BRUCHER, Roger. *Les fonds de journaux de la Bibliothèque royale Albert Ier. Évolution, aménagement et gestion récente d'une structure bibliothéconomique spécifique*. Brussels: Royal Library of Belgium, 1990.

1852. However, it was mainly the establishment of the *Bureau de Librairie* within the Ministry of the Interior in 1854 that significantly increased the collections: all publications covered by the 1852 agreement were henceforth deposited in the library. Around the same period, the library also began collecting less prominent periodicals, including regional, local, and satirical journals. Under head curator Edouard Fétis (1887-1904), the collection of old political newspapers from the Chamber of Representatives was transferred to the library. From 1874, the Belgian Bibliography law ensured that publishers began depositing their publications, including a large number of ‘smaller’ newspapers and periodicals, with KBR. Finally, in 1966, legal deposit came into effect: since then, publishers have been obliged to provide KBR with a copy of every printed medium published in Belgium. A law concerning digital legal deposit, which will automatically include digital publications in the collection, is in preparation.



Fig. 1. Nieuwe Tijdinghen, 1 January 1622. Source: KBR.

In addition, KBR holds an important collection of newspapers and periodicals from the 17th and 18th centuries. These include, among others, the *Nieuwe Tijdinghen*, one of the oldest newspapers in the Low Countries, published in Antwerp by Abraham Verhoeven from 1620, and the *Courier véritable des Pays-Bas*, published in Brussels from 1649. KBR also received several important donations of samples, which are housed in separate collections. The collection of Philippe Vandermaelen (1795-1869) includes more than 600 samples of foreign newspapers. The collection of press historian and collector Gaston Mertens comprises tens of thousands of samples of Belgian newspapers and periodicals from the period 1820-1940, including many unique copies. Historical Belgian newspapers that have been deaccessioned by the British Library and other institutions were also added. KBR also holds a unique series of printed press releases from the news agency Belga (1954 – 1995), as well as from Associated Press (1961 – 1985).⁸⁴

Apart from the oldest and most valuable titles, which mostly belong to the Old and Rare Books department, the newspapers and periodicals belong to the department of Contemporary Prints. Bibliographically, most newspapers belong to the series JB (*journaux belges*) and JE (*journaux étrangers*). The distinction between newspapers and periodicals was based on format: the JB and JE series mainly contain publications in folio or plano-format, while the smaller formats are kept in the B and R series. As a result of this choice, the so-called *petite presse* has largely been classified as periodicals under the B-series. Around 1,200 titles from this series could also be considered newspapers, depending on the definition employed.⁸⁵

The JB collection includes ca. 2,500 Belgian newspaper titles (including regional editions of national newspapers), covering more than 5 linear kilometres. Although most of the newspaper titles in this collection are discontinued, it is still daily supplemented with the new editions of ongoing Belgian newspapers. As of 2024, KBR receives 64 different newspaper editions of 19 Belgian dailies.

84. This collection is the subject of a joint ULB-KBR digitisation and research project, funded by ULB, which has been launched in August 2025. See the contribution by Florence Le Cam, Brecht Deseure and Max De Wilde in this volume.

85. BRUCHER, *Les fonds de journaux*, p. 8.

The JE collection contains over 500 foreign newspaper titles from 60 countries. This collection is closed, except for *Le Monde* and *NRC Handelsblad*. KBR holds 63,411 Belgian periodicals, including 17,948 ongoing titles. Belgian periodicals mainly belong to the B series, and since the legal deposit came into effect, to the BD series (foreign periodicals are housed in the R series).

As to conservation, newspapers and periodicals are usually bound in volumes and kept under stable climatological conditions in the library's Book Tower. To make room in the storage building, newspapers that are digitally accessible are transferred to the library's external depot in Péronnes-lez-Binche. Since 2020, consultation of newspapers and periodicals (both on paper and on microfilm) takes place in the General Reading Room, as collection-specific reading rooms have been discontinued in the wake of the COVID-19 crisis. Since the launch of the digital platforms BelgicaPress⁸⁶ and BelgicaPeriodicals,⁸⁷ physical consultation of paper newspapers and periodicals has fallen in favour of digital consultation (185.164 consultations on BelgicaPress and 28.764 consultations on BelgicaPeriodicals in 2024), making them the most popular collections in terms of number of consultations.

Digitisation Initiatives

Newspapers and periodicals are notoriously fragile heritage. Among other things, intense consultation, overly tight bindings, and exposure to light pose an acute threat to many historical press collections. Ironically, modern newspapers are more vulnerable to decay than old ones. Wood pulp paper, which came into use around the mid-19th century for newspaper printing, is much less durable than the previously used rag paper (made from cotton). The progressive and irreversible acidification of the paper, caused by the oxidation of the lignin it contains, ensures that modern newspapers are inevitably subject to ongoing decay, even under ideal storage conditions. According to a report by the Royal Institute for Cultural Heritage, 95% of KBR's newspaper collection from the

86. <https://www.belgicapress.be>.

87. <https://www.belgicaperiodicals.be/>.

period 1841-1959 was considered weakened by 1992.⁸⁸

Awareness of this issue led to worldwide initiatives, beginning in the 1960s, to safeguard the press heritage.⁸⁹ From the middle of that decade, 35mm microfilm came into use for microfilming newspapers.⁹⁰ Microfilm offers a very cost-effective and space-efficient preservation method and can last up to 500 years under the right storage conditions. KBR began building a microfilm collection in 1959. Between 1964 and 1975, a commission for the microfilming of the Belgian press was active, consisting of a collaboration between KBR and Belgian universities. The effort then stalled but was resumed in 1989 after the approval of a large-scale preservation plan.⁹¹ This initiative led to the microfilming of many hundreds of the main newspaper titles. Moreover, collaborations were set up with other libraries to mutually complete collections through microfilm exchanges. Today, 605 Belgian newspapers from KBR's holdings are available on microfilm.⁹² Unlike many foreign libraries, KBR continues to microfilm its collection according to a priority list of rare or valuable newspapers. To help ensure their preservation, paper originals can no longer be consulted by readers when microfilms are available.

It goes without saying that the advent of digital technology brought about a significant shift. In 2005, the Federal Science Policy launched an initial campaign to digitise vulnerable federal collections.⁹³ Among the nine pilot projects was a project by KBR and CegeSoma (the Study and Documentation Centre for War and Contemporary Society)⁹⁴ to digitise 30 Belgian newspapers between 1831

88. D'HOORE, Marc. Un exemple de projet de numérisation de masse. Les journaux de la Bibliothèque royale de Belgique (KBR). *Archief- en bibliotheekwetenschappen in België*, 2011, vol. 83, n° 1-4, p. 125.

89. In the U.S., large-scale microfilming of newspapers began as early as the 1930s with the Foreign Newspaper Project at the Harvard University Library.

90. D'HOORE. Un exemple de projet, p. 124.

91. BRUCHER, *Les fonds de journaux*, p. 21.

92. The total microfilmed newspaper collection in Belgium was estimated at 15% in 2006. LUYTEN, Dirk & D'HOORE, Marc. Nationaal project voor de bewaring van Belgische kranten "Presse-Papier". *Bladen voor documentatie*. 2006, n° 2, p. 5.

93. METTENS, Philippe. Het Belgisch digitaliseringsplan voor de Federale Wetenschappelijke Instellingen en het Koninklijk Belgisch Filmarchief. *Archief- en bibliotheekwetenschappen in België*. 2011, vol. 83, n° 1-4, p. 15-30.

94. <https://www.cegesoma.be/>.

and 1950 (amounting to approximately 2.8 million pages), as well as 76 titles of censored press and 81 of the clandestine press from the periods of the two World Wars (Fig. 2).⁹⁵ Since then, the digital collection has systematically expanded.

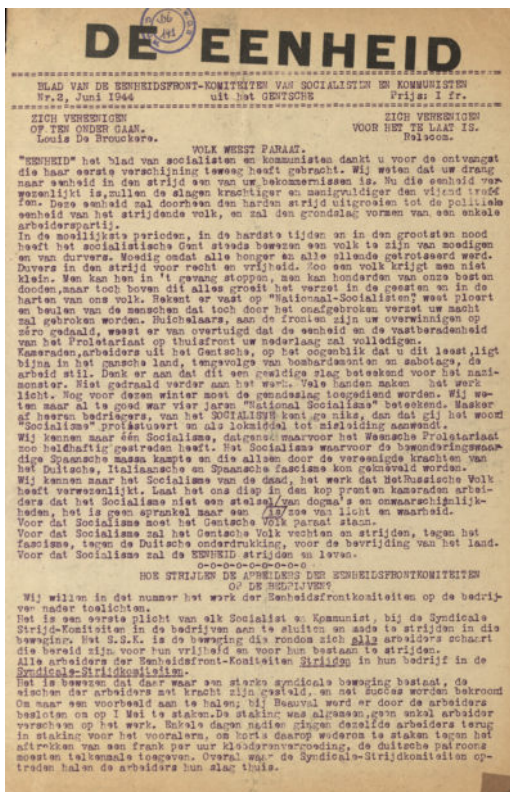


Fig. 2. De Eenheid. Blad van de Eenheidsfrontscomiteuten van Socialisten en Kommunisten, 2 June 1944. This journal from the clandestine press during WWII was scanned and made available via the project The Belgian War Press: <https://warpress.cegesoma.be/>. Collection CegeSoma – State Archives in Belgium (OD4).

95. GOOSSENS, Alain. Le projet fédéral belge de numérisation et coupures de presse. Une préoccupation quotidienne! *Archief- en bibliotheekwetenschappen in België*, 2007, vol. 78, n° 1-4, p. 185-196. The result can be consulted on <https://warpress.cegesoma.be>. See <https://belgianpressfromthegreatwar.be> for a complete overview of the digitised press from the First World War.

In 2011, KBR set up an internal general coordination for its digitisation policy and founded an in-house Digitisation Department (Digit) to implement it. This new dynamic, financed from the institution's permanent endowment, consolidated the project-funded initiatives of the Federal Science Policy Administration. On the one hand, it enabled KBR to continue digitising its newspaper collection based on microfilmed volumes and, on the other, to launch major mass digitisation projects by subcontracting. In 2024, KBR launched a parallel project to digitise more than 8,000 volumes of daily newspapers published between 1951 and 1989, creating a corpus of 4 million pages that will be made available online within the three following years.

Digital Platforms and Accessibility

To facilitate access to its digitised heritage, KBR has developed several digital platforms, each tailored to the specific nature of the documents they host. These platforms serve as gateways to Belgium's rich historical and cultural resources, making them accessible to a broad audience of researchers, educators, and the general public.

KBR started to make digitised documents available online in 2009 by launching a prototype of digital library, called Belgica.⁹⁶ At that time, the focus was on a few historical treasures including newspapers: *Vlaemsch België* and *Deutsche Brusseler Zeitung*. It was possible to consult the newspapers by scrolling through the pages in a viewer. In 2012, the beta version of BelgicaPress was launched to provide access to the results of the first major daily press digitisation campaign. This platform offered online access in calendar mode and full text. Consultation was limited to library reading rooms. At the same time, the censored press was put online in the portal The Belgian War Press, developed by CegeSoma.⁹⁷ An upgraded version of BelgicaPress was launched in 2016, offering access to the entire corpus of the rights-free press online. Access to the corpus after 1918 was restricted to computers in KBR's reading rooms and in the reading rooms of Belgian partner research institutions within the Belnet network. Since 2020, online access via BelgicaPress has been extended to the entire digitised press corpus.

96. <https://belgica.kbr.be/belgica/>.

97. <https://warpress.cegesoma.be/>.

At time of writing, the collection of digitised Belgian newspapers included over 150 titles published between 1814 and 1987, with a strong focus on the period 1830-1950, corresponding to a corpus of over 4,346,591 pages. KBR also launched the first version of a new platform dedicated to digitised periodical publications, called BelgicaPeriodicals, in 2020. This new platform was designed in the same way as BelgicaPress. It allows users to perform full-text searches in OCR or in calendar mode based on a specific date. This second platform offered access to 348 titles published between 1622 and 2015, amounting to 2,235,970 pages, at the time of writing.

BelgicaPress and BelgicaPeriodicals offer both calendar and full text access to the corpus. A user may select one or several titles from the menu, with the possibility to focus on a specific period. An ‘advanced search’ window allows queries based on several options (all of these words, one of these words, exact phrase, none of these words), through all, one or several titles selected from a drop-down list of titles, with the possibility of filtering by period (one year, one specific date, all years, a certain time period, etc.) and by language (French, Dutch, German).

WWW.KBR.BE

General catalogue Belgica BelgicaPress BelgicaPeriodicals EN

KBR BelgicaPeriodicals MY KBR

Back to simple search

Advanced search

Text

All of these words

One of these words

Exact phrase

None of these words

Date or period

Search on date Search on period

From

Language

French Dutch German

Fig. 4. Advanced search menu on the platform BelgicaPeriodicals.

When displaying search results, both platforms first show the number of pages corresponding to the search terms found in the corpus. The results may then be organised in terms of pages (most pages, least pages, most recent pages, least recent pages) and be filtered by year, edition, and type of access (totally free or restricted by copyright). The words corresponding to the research queries are highlighted in the text. Navigation options allow for the user to search an entire issue or a single page of the newspaper in question. Links are provided to the bibliographic catalogue description and to permalinks, which are useful for citations. There is no option to download material. This service will, however, be integrated in the near future within the context of the implementation of IIIF-services.

Naturally, access to the digitised press is restricted by copyright law. Belgian law does not provide a separate regime for the protection of collective works. For newspapers and magazines, this means that copyright remains protected until 70 years after the death of the last contributing author. This makes it difficult to determine precisely when a newspaper or magazine falls into the public domain. KBR opts to consider all press publications up to 1918 as public domain and therefore free of copyright. Many newspapers and magazines from 1918-1950 are possibly still in-copyright; for publications which appeared less than 70 years ago, this is the case. Digital access to publications before 1918 is therefore free and without conditions. The exception made in copyright law, allowing use for educational and scientific purposes, applies to publications post-1918. Access to these documents is only possible after creating a user profile (MyKBR), in which the user agrees to respect the terms of access.

Workflow and Technical Standards

KBR has been digitising its collections since the establishment of a Digit in 2011, in order to best preserve its collections while facilitating their online dissemination. Digitisation is the systematic reproduction in digital format of the documentary and artistic artefacts that are found within heritage institutions. The aim is to make it possible to view heritage items on a computer, and to do so in a way that is both understandable and scientifically reliable. This requires,

on the one hand, a systematic approach to capturing images and, on the other, the generalised creation of mutually interchangeable and structured metadata. Digitisation is a process which aims to preserve the original items by preventing them from being harmed by continued and repeated acts of consultation. This is so-called 'preventive preservation'. As the name suggests, KBR's Digital Data Strategy concerns itself primarily with the use of digital data.



Fig. 5. Le Catholique des Pays-Bas, 2 October 1830. Digitised from microfilm. Source: KBR.

KBR's newspaper digitisation is achieved primarily from its microfilms, for which Digit owns dedicated scanners. This choice is largely pragmatic: scanning microfilms is quick and cheap, allowing large volumes of newspapers to be brought online in a relatively short time. Microfilms, especially those of a more advanced age, can suffer from low visual quality, making reading, and thus engaging in OCR, difficult. They are also only available in monochrome (black and white). Since much of KBR's newspaper collection has not been microfilmed, there remains a significant number of items that still requires preservation. A solution will eventually need to be found for the part of the collection that only exists on paper. Additionally, Digit has state-of-the-art scanners that allow for in-house digitisation of the paper original. This is done both as part of the internal digitisation priorities and as part of the Digit on Demand service, where readers can request the digitisation of collection items for a fee. The acquisition of robot scanners has significantly increased the pace of the digitisation process, as they are fully automated and can scan up to 2000 full color pages per day. Due to their large format, most newspapers do not fit on the robot scanners. The latter are therefore mainly used for scanning books and magazines. Both microfilm and robot scanning are followed by a time-intensive post-production process which involves manual oversight.

Large-scale digitisation projects, often funded through additional budgets, are not carried out in-house due to the volume of material involved. In such cases, Digit coordinates the appointment of an external service provider and oversees the entire procedure, from issuing a public tender to checking the result. Of course, the same quality standards are imposed on digitisation firms as those customary within KBR. These standards are compliant with the *Matamorfoze Light* guidelines, whose objective it is to produce images in such a way that they can serve as a replacement for the original object.⁹⁸

98. VAN DORMOLEN, Hans. *Matamorfoze Preservation Imaging Guidelines*. The Hague: Koninklijke Bibliotheek, 2012. Available online: https://www.matamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf.

Selection of Titles

From the outset, the digitisation programme at KBR has aimed at providing a balanced and representative overview of the Belgian press, including national, regional and local titles, selected according to geography, period, language and ideology. The priority list of titles to be digitised is drawn up according to the same principle. It is important to note that, as digitisation of newspapers is largely carried out from existing microfilm, most newspapers that have not been microfilmed are, as yet, excluded from this list.

Initially, digitisation focused on the so-called *grande presse* between 1830 and 1950, e.g. the political opinion papers printed in large format. These are the main media outlets which have shaped political opinion in Belgium since the country's founding (*L'Indépendance belge*, *Le XXme Siècle*, *Het Laatste Nieuws*...). As more titles were added, this focus was expanded to include papers from the so-called *petite presse* (strongly opinionated, often radical journals, usually printed on smaller formats) and the popular press. Examples include satirical newspapers (*La Bombe*, *Le Charivari*, *Méphistophélès*) and sports dailies (*Sportwereld*, *Velosport*). The selection's chronological breadth has also been expanded, as newspapers from the old regime (*Nieuwe Tijdinghen*, *Relations véritables*, *Gazette des Pays Bas*) and from the Dutch period (*Le Courier des Pays-Bas*, *Le Belge*, *L'Emancipation*, *Le Politique*) have also been digitised. Extensive collections of specimens (i.e. individual issues) of newspapers and periodicals, which have often been been fragilised due to inadequate bindings, are currently also being scanned and brought online.

The broadening of the types of newspapers offered is partly due to project-based collaborations with external institutions. The digitisation of clandestine and censored war press has been mentioned before. Newspapers from the Dutch period became available online thanks to a collaboration with the University of Passau (2013-2017).⁹⁹ The *Gazette de Leyde* (1750-1789) was digitised thanks to the Université de Lyon 2 as part of a project on French-language political newspapers in the 18th century.¹⁰⁰

99. <https://www.reconfort.uni-passau.de/en/>.

100. <https://www.gazettes18e.fr/>

still running after 1950. This amounts to approximately 4 million pages and will almost double KBR's digital newspaper collection. The end date was chosen because newspaper publishers began switching to digital editions from ca. 1990, making it unnecessary to scan the paper versions. Funding for this first project has come from KBR's own budget.



Fig. 7. Courrier de l'Afrique, 31 December 1967-1 January 1968. Digitised as part of a campaign to bring the colonial press online. Source: KBR.

The second project is a collaboration between KBR and the AfricaMuseum to enable the digitisation of colonial newspapers and periodicals, funded by a subsidy from Belspo (the Belgian Federal Science Policy Office). This involves over 50 titles from the collections of KBR (especially the former collection of the Ministry of Foreign Affairs) and the AfricaMuseum, totaling approximately 500,000 pages. It includes titles like *La Croix de l'Afrique*, *L'Essor du Congo* and *L'Echo du Katanga*. Given the great contemporary interest in the colonial past, this project will make an important contribution to ongoing research.

Periodicals, unlike newspapers, were hardly ever microfilmed. Digitisation is therefore primarily done from the original media and in color. The selection of titles is often a function of ongoing research projects in which KBR is a partner. An important example is the Artpresse project, which led to the digitisation of ca. 140 titles of the artistic press from the Interwar period (see the contribution by Sébastien Hermans and Morgane Ott in this volume).¹⁰¹ Similarly, a collection of youth magazines will be digitised thanks to the FED-tWIN project Pop Heritage Lab.¹⁰² Collaborations have also been put in place with external partners to complete the periodicals collection. Thanks to a collaboration with the project CAMille KBR-ULB, and additional funding from ULB, the series of the professional journals *Le Journaliste/De Journalist* and the *Annuaire officiel de la presse belge* are being digitally compiled from the combined collections of KBR, ULB, Ghent University, KU Leuven, the *Association des journalistes professionnels* (AJP) and the *Vlaamse Vereniging van Journalisten* (VVJ)

Collaboration with Other Institutions

KBR has chosen to develop its digitisation policy in consultation with other Belgian heritage institutions. After all, KBR is by no means the only Belgian institution that digitises printed media. It is evident that institutions have an interest in avoiding multiple digitisations of the same publications and in aligning their digitisation policies. For example, KBR was one of the partners in the 2003 *Presse-Papier* project, the predecessor of today's Flemish newspaper catalogue

101. <https://www.kbr.be/en/projects/artpresse/>.

102. <https://www.kbr.be/en/projects/pop-heritage-lab/>.

Abraham.¹⁰³ Regular consultations also take place with other institutions that pursue an active digitisation policy, such as the Flanders Heritage Libraries (see the contribution by Sophia Rochmes in this volume). Priority lists for digitisation are updated accordingly.

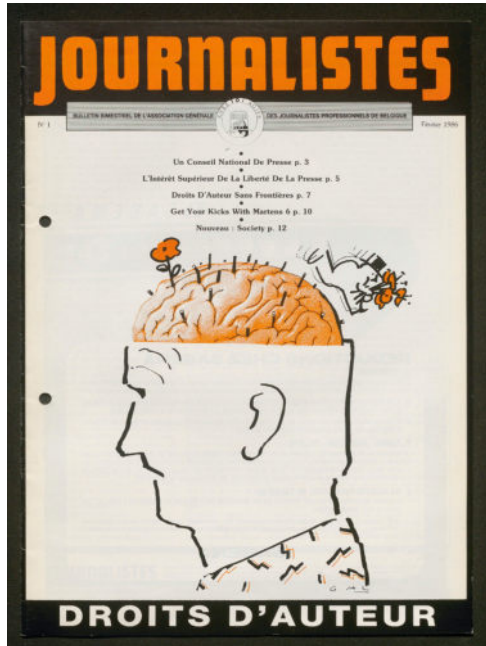


Fig. 8. Journalistes, February 1986. Scan after a paper original from the collection of the Association des journalistes belges (AJP). Digitisation funded by ULB.

Furthermore, active collaborations are entered into gaps in collections might be digitally filled. Digital files may be exchanged, collection items lent for digitisation, or KBR may digitise external collection items, depending on the circumstances of each case. Several newspapers series in BelgicaPress were

103. LUYTEN & D'HOORE, Nationaal project; COPPOOLSE, David. De krachten bundelen. De Vlaamse Erfgoedbibliotheek als expertisenetwerk digitalisering. *Archief- en bibliotheekwezen in België*. 2011, vol. 83, n° 1-4, p. 31-46.

completed in this way: *Le Soir* (Brussels City Archives),¹⁰⁴ *Le Drapeau rouge* (CArCoB - Centre des Archives du communisme en Belgique),¹⁰⁵ *Le Politique* (Bibliothèque Ulysse Capitaine),¹⁰⁶ *Vooruit* (AMSAB),¹⁰⁷ and *Vers l'Avenir* (Réseau namurois de lecture publique).¹⁰⁸ Similar collaborations have also been established with local archives and heritage institutions (including from Halle, Diksmuide, and Sint-Truiden). This process is also carried out for periodicals. Such collaborations require ad hoc agreements, which take into account the complementarity of the collections, the quality and compatibility of the files, and respect for copyright and property rights.



Fig. 9. St. Vith Volks-Zeitung, 29 September 1934. Scans produced by ZOG and hosted by KBR.

Thanks to these collaborations, KBR's digital offering now includes a large number of images whose paper originals are held by other institutions. This

104. <https://archives.brussels.be/>.

105. <https://carcob.eu/>.

106. <https://lesmuseesdeliege.be/les-fonds-patrimoniaux/>.

107. <https://www.amsab.be/>.

108. <https://www.namur.be/fr/loisirs/culture/lectures/bibliotheques/bibliotheques-du-reseau>.

evolution fits with KBR's mission to offer the most complete access possible to Belgian publications. With dozens of institutions today working on digitising and making their collections accessible, the digital offerings are highly fragmented. The advantages of scale that KBR offers because of its existing platforms and expertise can also be of interest to smaller institutions. For example, an important collection of German- and French-language newspapers from Eastern Belgium (including *Eupener Zeichnung*, *Neue Nachrichten* and *Grenz Echo*) is now available via BelgicaPress thanks to a collaboration with the *Zentrum für Ostbelgische Geschichte* (ZOG),¹⁰⁹ and a German language interface has been launched.

Publishing Newspapers as Data

The digitisation of cultural heritage collections, such as KBR's historical newspapers and periodicals collections, has led to the increased interest of these materials for digital humanities research. However, traditional ways of providing access to such collections, for example through digital library interfaces, are less than ideal for researchers who are looking to build datasets around specific research questions. Inspired by the *collections as data* movement,¹¹⁰ an approach for cultural heritage institutions to prepare their digital collections for analysis using digital methods, KBR embarked on a project in October 2020, which was funded by the Belgian Science Policy Office (Belspo) as part of the Belgian Research Action through Interdisciplinary Networks, BRAIN 2.0 programme.¹¹¹ The aim of the DATA-KBR-BE project is to facilitate data-level access to its digitised and born-digital collections for digital humanities research.¹¹²

Originating in the United States, the *collections as data* movement was established to encourage cultural heritage professionals to start thinking differently about how they provide access to their collections in order to facilitate analysis using digital tools and methods. In its first phase, *Always Already Computational*:

109. <https://www.geschichte.be/>.

110. <https://collectionsasdata.github.io/>.

111. <https://www.belspo.be/BRAIN-be/>.

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

Collections as Data (2016-2018)¹¹³ focused on exchanging experiences and sharing knowledge, while documenting this process.¹¹⁴ In its, second phase, *Collections as Data: Part to Whole* (2019-2021)¹¹⁵ supported the implementation and use of collections as data through a number of funded collaborative case studies that were jointly led by cultural heritage professionals and researchers.¹¹⁶ Until now, implementations of *collections as data* have been largely restricted to the United States,¹¹⁷ but they are gradually appearing in Europe as well.¹¹⁸

Inspired by this work, DATA-KBR-BE is an interdisciplinary research collaboration between cultural heritage experts, digital humanities researchers, and data scientists which aims to optimise KBR's existing ICT infrastructure in order to stimulate sustainable data-level access to KBR's digitised collections for digital humanities research. Data-level access means providing access to the underlying files of digitised cultural heritage collections to facilitate data analysis by means of tools and methods developed in the field of digital humanities.

For the project, research teams in Ghent, Antwerp and Brussels worked closely together with the digitisation, collections and ICT experts at KBR to co-design three interdisciplinary research scenarios focused on KBR's digitised historical newspaper collection: BelgicaPress. These research scenarios were conceived as initial case studies to demonstrate the scientific potential of providing data-level access to KBR's collections, as well as to understand how *collections as data* could be implemented at KBR.

113. <https://collectionsasdata.github.io/>.

114. PADILLA, Thomas, ALLEN, Laurie, FROST, Hannah, POTVIN, Sarah, RUSSEY ROKE, Elizabeth & VARNER, Stewart. *Final Report: Always Already Computational: Collections as Data*. 2019. DOI: <https://osf.io/mx6uk/wiki/home/>.

115. <https://collectionsasdata.github.io/part2whole/>.

116. <https://collectionsasdata.github.io/part2whole/cohortone/>.

117. WITTMANN, Rachel, NEATROUR, Anna, CUMMINGS, Rebekah & MYNTTI, Jeremy. From Digital Library to Open Datasets. *Information Technology and Libraries*. 2019, vol. 38, n° 4, p. 49-61. DOI: <https://doi.org/10.6017/ital.v38i4.11101>.

118. CANDELA, Gustavo, SÁEZ, María Dolores, ESCOBAR ESTEBAN, Mpilar & MARCO-SUCH, Manuel. Reusing Digital Collections from GLAM Institutions. *Journal of Information Science*. 2020, vol. 48, n° 2, p. 251-267. DOI: <https://doi.org/10.1177/0165551520950246>; AMES, Sarah & LEWIS, Stuart. Disrupting the Library: Digital Scholarship and Big Data at the National Library of Scotland. *Big Data & Society*. 2020, vol. 7, n° 2. DOI: <https://doi.org/10.1177/2053951720970576>.

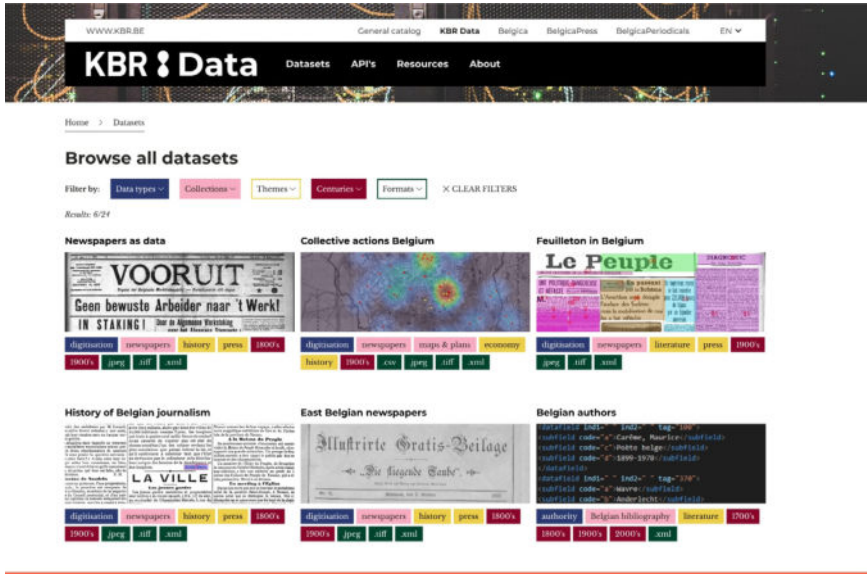


Fig. 10. KBR Data Platform with an initial selection of datasets.

The interdisciplinary research scenarios selected for the project were: Collective Action Belgium: led by GhentCDH,¹¹⁹ this focuses on social history in the Interbellum and World War Two periods and aims to trace the dynamics of contention, strikes, demonstrations and other forms of collective action in Belgium as reported in Belgian newspapers;

The Feuilleton in Belgium: led by ACDC (the Antwerp Centre for Digital Humanities and Literary Criticism),¹²⁰ this focuses on literary studies in the period 1830–1930 and aims to map the publication of literature in Belgian newspapers across the first century of the Belgian nation state;

The History of Belgian Journalism: led by CAMille, this focuses on media history from 1886 until now and aims to trace the history of Belgian journalism through the lens of critical discourses about journalism in Belgian newspapers.¹²¹

119. <https://www.ghentcdh.ugent.be/>.

120. <https://www.uantwerpen.be/en/research-groups/digitalhumanities/>.

121. <https://www.kbr.be/en/projects/camille/>.

The digital humanities research undertaken in this project collaborates closely with KBR's Digital Research Lab,¹²² which facilitates text and data mining research using KBR's digitised and born-digital collections and the KBR Data Science Lab,¹²³ which focuses on the application of Artificial Intelligence for the analysis of digital cultural heritage. Furthermore, the project harnesses the expertise of data scientists for the semi-automatic extraction and classification of articles from historical newspapers.¹²⁴

A key focus of the DATA-KBR-BE project was to facilitate data-level access to KBR's digitised and born-digital collections. In Summer 2022, the International Galleries, Libraries, Archives and Museums (GLAM) Labs Community¹²⁵ explored the challenges that cultural heritage institutions are experiencing when attempting to publish their collections as data. This activity led to the emergence of the *Checklist to Publish Collections as Data in GLAM Institutions*,¹²⁶ which is intended as an easy to apply method to encourage small and medium-sized organisations to publish their digital holdings as collections as data. Later, this checklist was transformed into a *Workflow for the Social Sciences and Humanities Open Marketplace*, in the context of the common European Data Space for Cultural Heritage.¹²⁷ The DATA-KBR-BE team used both the checklist and the workflow for publishing KBR's collections as data.

While the DATA-KBR-BE project focused on providing curated datasets of digitised historical newspapers or newspapers as data, such collections as data

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

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

124. ALI, Dilawar, MILLEVILLE, Kenzo, VERSTOCKT, Steven, VAN DE WEGHE, Nico, CHAMBERS, Sally & BIRKHOLZ, Julie M. Computer Vision and Machine Learning Approaches for Metadata Enrichment to Improve Searchability of Historical Newspaper Collections. *Journal of Documentation*. 2024, vol. 80, n° 5, p. 1031-1056. DOI: <https://doi.org/10.1108/JD-01-2022-0029>.

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

126. CANDELA, Gustavo, GABRIÉLS, Nele, CHAMBERS, Sally, PHAM, Thuy-An, AMES, Sarah, FITZGERALD, Neil, HOFMANN, Katrine, HARBO, Victor, POTTER, Abigail, FERRITER, Meghan, MANCHESTER, Eileen, IROLLO, Alba, VAN KEER, Ellen, MAHEY, Mahendra, HOLOWNIA, Olga & DOBREVA, Milena. *A Checklist to Publish Collections as Data in GLAM Institutions*. 2023. Available at: <https://arxiv.org/abs/2304.02603v1>.

127. <https://pro.europeana.eu/page/common-european-data-space-for-cultural-heritage>; CANDELA, Gustavo, CHAMBERS, Sally & IROLLO, Alba. *A Workflow to Publish Collections as Data: the Case of Cultural Heritage Data Spaces*. 2023. Available at: <https://marketplace.sshopencloud.eu/workflow/I3JvP6>.

datasets can include any data that is related to KBR's collections. For example, as part of KBR's emerging Digital Data Strategy, nine 'data types' related to KBR's collections have been identified: Bibliographic data; Authority data; Heritage data; Digital data; Collections as data; Born-digital data; Electronic resources; Research data, and Lab data (code). The possibility of publishing one dataset for each data type is being explored.

A key outcome of the DATA-KBR-BE project is to develop a sustainable platform providing access to KBR's collections as data, which is expected to be launched shortly. The KBR Data platform will be fully integrated into the KBR website, alongside KBR's core digital services such as Belgica, BelgicaPress, BelgicaPeriodicals and the General Catalogue.¹²⁸ As the DATA-KBR-BE Platform is intended to be a long-term initiative extending beyond the end of the DATA-KBR-BE project itself, this first phase focuses on the creation of a simple, user-friendly website, where *curated datasets* based on KBR's digitised and born-digital collections will be published. This phase is inspired by existing data platforms such as the National Library of Luxembourg's Open Data Platform¹²⁹ and the National Library of Scotland's Data Foundry.¹³⁰ In a later phase, a *searchable repository* of curated datasets could be anticipated, inspired by the British Library's Research Repository,¹³¹ where the British Library's Collection Datasets¹³² have been published or the Digital Library of the Caribbean (DLoc)'s DLoc as a Data repository.¹³³ A final phase could focus on providing computational access to KBR's digitised and born-digital collections via Application Programming Interfaces (APIs) or data on demand services. As part of the Belspo, ESFRI-FED Programme,¹³⁴ a proposal to develop

128. <https://opac.kbr.be/>.

129. <https://data.bnl.lu/data/>.

130. <https://data.nls.uk/>.

131. <https://bl.iro.bl.uk/>.

132. <https://bl.iro.bl.uk/collections/64e3804a-788a-4c4b-962c-ae180d955455?locale=en>.

133. <https://dataverse.fiu.edu/dataverse/dloc-as-data>.

134. https://www.belspo.be/belspo/organisation/call/ESFRI-FED_2021_en.stm.

the *KBR Virtual Lab: e-infrastructure for facilitating access and research of KBR's collections as data*¹³⁵ as a Belgian contribution to DARIAH,¹³⁶ the Digital Research Infrastructure for the Arts and Humanities has been funded. The inspirations for this phase include, for example, the National Library of the Netherlands' Data Services.¹³⁷

The availability of collections as data in general, or newspapers as data in particular, paves the way for the development of interdisciplinary scholarly workflows to analyse such cultural heritage datasets using digital methods¹³⁸ or using such datasets for the training of computational models in the areas of Machine Learning, Computer Vision or Large Language Models (LLMs).¹³⁹ Such efforts contribute to the recognition of cultural heritage data as humanities research data¹⁴⁰ within the broader context of initiatives such as the European Open Science Cloud (EOSC).¹⁴¹ In the same spirit, KBR participates in the Swiss-Luxemburgish research project *Impresso - Media Monitoring of the Past*,¹⁴² a cutting-edge laboratory for the semantic enhancement and exploration of an international and transmedial corpus of text sources. By sharing an important part of its digitised newspapers and periodicals, KBR takes a first step towards implementing its emerging newspapers as data policy.

135. https://www.belspo.be/belspo/ESFRI-FED/project_en.stm#KBRVirtualLab.

136. <https://www.dariah.eu/>.

137. <https://www.kb.nl/en/research-find/for-researchers/data-services-apis-and-downloads>.

138. OBERBICHLER, Sarah, BORÓŠ, Emanuela, DOUCET, Antoine, MARJANEN, Jani, PFANZELTER, Eva, RAUTIAINEN, Juha, TOIVONEN, Hannu & TOLONEN, Mikko. Integrated Interdisciplinary Workflows for Research on Historical Newspapers: Perspectives from Humanities Scholars, Computer Scientists, and Librarians. *Journal of the Association for Information Science and Technology*. 2022, vol. 73, n° 2, p. 225-239. DOI: <https://doi.org/10.1002/asi.24565>.

139. ALKEMADE, Henk, CLAEYSSENS, Steven, COLAVIZZA, Giovanni, FREIRE, Nuno, LEHMANN, Jörg, NEUDECKER, Clemens, OSTI, Giulia & VAN STRIEN, Daniel. Datasheets for Digital Cultural Heritage Datasets. *Journal of Open Humanities Data*. 2023, vol. 9, n° 1, p. 17. DOI: <https://doi.org/10.5334/johd.124>.

140. TASOVAC, Toma, CHAMBERS, Sally & TÓTH-CZIFRA, Erzsébet. 2020. *Cultural Heritage Data from a Humanities Research Perspective: a DARIAH Position Paper*. <https://hal.science/hal-02961317/>.

141. <https://open-science-cloud.ec.europa.eu/>.

142. <https://www.impresso-project.ch/>.

Conclusion

393 years separate the oldest print media issue present in KBR's digital platforms (Abraham Verhoeven's newspaper *Nieuwe Tijdinghen* of 1 January 1622, relating the safe arrival of the Spanish silver fleet at Seville) from the youngest one (the 2015 issue of the of *Revue belge d'archéologie et histoire de l'art*) in late 2025. Between them lies an immensely fascinating and multi-faceted history, of which the daily and periodical press is an eminent witness. Since 2011, ca. 150 newspapers and 348 periodicals, totaling over 6,5 million pages, have been made freely accessible on BelgicaPress and BelgicaPeriodicals, with new titles regularly added. While numbers do not tell everything, these numbers do show KBR's commitment to digitally preserve and disseminate Belgian print media heritage at a massive scale.

What the numbers fail to show is the enormous investment of time and energy that drives them. Providing digital access to historical print media is an expensive and time-consuming undertaking, and one for which no standard procedure exists. Financial means are limited, and choices need to be made. Moreover, current results may be quickly surpassed or outdated by technological evolutions. OCR performance of text scanned from the oldest generation of microfilms for example scores poorly compared to OCR undertaken on recent scans done from paper. Providing access to digitised print media also comes with its challenges: copyright law imposes limits on what can be shared with the public, while data level access for researchers requires the development of an internal policy as well as a dedicated technological infrastructure. Reconciling these varying and often complicated demands is by no means an easy task.

The future of historical print media collections is however, digital. The need to create digital copies of the fragile paper originals is imperative from the point of view of conservation. The gains in terms of access are, moreover, immense: never before have KBR's collections of newspapers and periodicals been so intensely exploited. Data level access will provide the next step in that evolution, and KBR prioritises collaboration with other institutions in this endeavour. Thanks to the scale of its collections and to its existing digital technical infrastructure, KBR can play a central position in the landscape of Belgian digital

print media heritage, fostering exchanges and partnerships with other institutions. Moreover, collaborations with research projects on an international level allow it to monitor the technical evolutions in the field. In the end, KBR's digitisation policy aims not only to preserve Belgian print media heritage but also to foster new forms of scholarly engagement and exploration.

4. *Nieuwe Tijdingen*: Towards a Newspaper Preservation Programme for Flanders

Sophia Rochmes

Abstract

More than 100 organisations across Flanders hold collections of historical newspapers. These range from large organisations such as university libraries and national cultural archives to smaller organisations such as local libraries, city and municipal archives and local historical societies. More often than not, the newspapers in these collections are incomplete, with series fragmented across multiple organisations and often only few specimens available. To preserve our newspaper heritage through digitisation and at the same time increase findability and usability, it is essential to coordinate efforts across institutions.

Flanders Heritage Libraries and meemoo, the Flemish institute for archives, have therefore been working together with cultural heritage organisations and experts on the building blocks of a newspaper programme for Flanders. Building upon experience from previous initiatives, the goal is to move away from stand-alone projects and instead develop a centrally coordinated programme for the digitisation of Flemish newspaper collections that will be complementary to initiatives at the national level by KBR. The proposed programme centres around three essential and related pillars: digitisation, digital archiving and online access.

We estimate that around 17 million newspaper pages in total in Flemish collections await digitisation. Of these, collection managers have already indicated that 3.34 million pages from 549 titles are in urgent need of digitisation. In this article, we give an overview of past initiatives by Flanders Heritage Libraries and meemoo, provide a sketch of the current landscape of newspaper collections in Flanders and outline our approach and the objectives of the proposed newspaper programme.

Author

Sophia Rochmes is an expert on periodicals and digitisation at Flanders Heritage Libraries (since February 2026 known as *BibliothecairErfgoed.be*), a network organisation for library heritage collections in Flanders. She manages the Belgian newspaper catalogue Abraham and coordinated the projects *Nieuwe Tjdingen* and *Primeur* (GIVE) in partnership with meemoo, the Flemish institute for archives. The article in this volume was developed together with colleagues at meemoo.

Keywords

Newspapers, periodicals; digitisation; archiving; OCR

More than 100 organisations across Flanders hold collections of historical newspapers. These range from large organisations such as university libraries and national cultural archives to smaller organisations such as local libraries, city and municipal archives and local historical societies. More often than not, the newspapers in these collections are incomplete, with series fragmented across multiple organisations and often only few specimens available. To preserve our newspaper heritage through digitisation and at the same time increase findability and usability, it is essential to coordinate efforts across institutions.

Flanders Heritage Libraries and meemoo, the Flemish institute for archives, have therefore been working together with cultural heritage organisations and experts on the building blocks of a newspaper programme for Flanders. Building upon experience from previous initiatives, the goal is to move away from stand-alone projects and instead develop a centrally coordinated programme for the digitisation of Flemish newspaper collections that will be complementary to initiatives at the national level by KBR. The proposed programme centres around three essential and related pillars: digitisation, digital archiving and online access.

We estimate that around 17 million newspaper pages in total in Flemish collections await digitisation. Of these, collection managers have already indicated that 3.34 million pages from 549 titles are in urgent need of digitisation. In this article, we give an overview of past initiatives by Flanders Heritage Libraries and

meemoo, provide a sketch of the current landscape of newspaper collections in Flanders and outline our approach and the objectives of the proposed newspaper programme.



Σαπιπ

Fig. 1. Newspapers from the Limburgensia Collection of Bibliotheek Hasselt Limburg.

Past and Ongoing Initiatives

We are working towards a cooperative newspaper programme for Flanders by building on our collective expertise, accrued through past and ongoing initiatives.

Both Flanders Heritage Libraries and meemoo are non-profit organisations supported by the Flemish government.

Flanders Heritage Libraries connects and supports organisations that manage library heritage collections in Flanders and Brussels. We aim to strengthen these organisations by building expertise networks, increasing visibility of and knowledge about the collections, and working together on shared solutions.

Meemoo is committed to supporting the digital archive operations of cultural, media and government organisations. Together with our partners, we bring the past back to life and prepare it for the future. We safeguard their digitally archived content, and make it accessible and usable.

Abraham. Belgian Newspaper Catalogue

2007 - present | Flanders Heritage Libraries

Abraham is an online catalogue of Belgian newspapers kept in Flemish libraries and other collecting organisations.¹⁴³ The catalogue was first established in a project that ran from 2007 to 2009 and is named after the Antwerp publisher Abraham Verhoeven (1575-1652), who published one of the world's first newspapers in 1605. At present, the catalogue brings together more than 12,500 titles spread over more than 25,000 paper, microfilm and digital holdings. With more than 100 participating organisations, the catalogue's coverage of extant Flemish newspaper heritage is high.¹⁴⁴

With the help of participating organisations, Flanders Heritage Libraries has worked further to expand the catalogue since its launch. This has included both

143. <https://krantencatalogus.be/>.

144. See *Abraham — Participating institutions (Abraham — Deelnemende instellingen)*. Antwerp: Flanders Heritage Libraries. Available on: <https://manual.krantencatalogus.be/p/BNC:Lists/Collections> (in Dutch).

expanding the scope of the catalogue as well as providing links to digitised newspapers, making the Abraham catalogue the central access point for digital newspaper collections in Flanders. In 2019-2020, the catalogue interface was significantly improved and the contents enriched with additional data thanks to a project grant.

News of the Great War

2013 - 2015 | meemoo and Flanders Heritage Libraries

News of the Great War (*Nieuws van de Grootte Oorlog*)¹⁴⁵ was a two-year project resulting in the digitisation and online presentation of 270,000 newspaper pages (52,370 editions from 904 titles) from World War I. The project brought together newspapers from 13 heritage organisations in Flanders. The project was implemented in three phases: preparation (inventory and selection, with the Abraham catalogue as starting point), digitisation (registration, packing, transport and digitisation) and presentation (platform design and development, valorisation). News of the Great War proved the usefulness of a collaborative approach, and the processes developed in this project have since been updated, fine-tuned and re-implemented in later projects.

Het Archief

2019 - present | meemoo

Meemoo launched hetarchief.be in late 2019 as a public metadata catalogue to showcase its partners' materials stored in the meemoo digital archives. This catalogue covered audio-visual content. With the metadata — descriptions, titles, keywords and technical data — the user could search through a wealth of information. In 2023, meemoo developed and launched a revised version of the platform, which brings together several functionalities, such as access requests and on-site appointment bookings. Since August 2025, newspaper collections (including the results of the projects News of the Great War, *Primeur* and *Nieuwe Tijdingen*) are offered as well via hetarchief.be. In 2024, in preparation for this

145. <https://meemoo.be/nl/projecten/digitalisering-eerste-papieren-drager-frontpers-uit-wo-i>.

addition, adjustments were made to the catalogue to ensure that all users are able to successfully search through and consult this newly added material

*GIVE-Primeur*¹⁴⁶

2021 - 2023 | meemoo and Flanders Heritage Libraries

The GIVE project was part of the digital transformation focus of the Flemish Government's Resilience Recovery Plan. The project also received support from the European Regional Development Fund (ERDF) under the REACT EU initiative, a recovery programme that was part of the European Union's response to the COVID-19 pandemic. This influx of support allowed for several important digitisation initiatives to be realised, including *Primeur*, a newspaper initiative under the GIVE project.

630,000 newspaper pages (91,490 editions from 110 titles) were digitised as part of *Primeur*. The newspapers themselves came from 8 different organisations: ADVN Archives and Research Centre,¹⁴⁷ AMSAB - Institute of Social History,¹⁴⁸ Bruges Public Library,¹⁴⁹ Bibliotheek Hasselt Limburg,¹⁵⁰ Hendrik Conscience Heritage Library,¹⁵¹ KADOC Documentation and Research Centre,¹⁵² KU Leuven Libraries¹⁵³ and University Library Ghent.¹⁵⁴ The focus was on newspapers with an urgent need for digitisation, with physical condition being the most important of the criteria considered. The most fragile newspapers (those which have suffered, for example, from acidification and/or high frequency of consultation) were considered to be in most urgent need of digitisation.

146. <https://meemoo.be/nl/projecten/give-krantenproject-primeur>.

147. <https://advn.be/>.

148. <https://www.amsab.be/>.

149. <https://brugge.bibliotheek.be/erfgoed>.

150. <https://hasselt.bibliotheek.be/>.

151. <https://consciencebibliotheek.be/>.

152. <https://kadoc.kuleuven.be/>.

153. <https://bib.kuleuven.be/>.

154. <https://lib.ugent.be/>.



Fig. 2. Condition assessment and preparation of newspapers for digitisation during the Primeur project, Limburgensia Collection of Bibliotheek Hasselt Limburg.

The central goal of the *Nieuwe Tijdingen* project was to lay the groundwork for a cooperative newspaper digitisation and preservation programme in Flanders. This involved developing a proposal for the programme, creating basic infrastructure and implementing processes on a small scale through pilot projects. *Nieuwe Tijdingen* was a multifaceted project with activities clustered under four rubrics.

1. Digitisation: defining criteria for digitisation; compiling a list of newspapers to be digitised; developing a proposal for a long-term structural programme as well as a five-year plan; implementing a pilot digitisation with small newspaper collections.

2. Existing content: evaluating current OCR quality of digitised newspapers and the potential for improvement; implementing a pilot OCR improvement project; identifying digitised newspaper collections and creating standards for integrating them into a central digital archive.

3. Interface: defining user needs for a newspaper interface and creating a ranked list of desired features and functions; researching copyright issues and proposing a re-use policy; researching and proposing communication and valorisation best practices.

4. Expertise: working with an advisory committee throughout the project; coordinating a professional network group for sharing expertise; communicating project findings and recommendations.

The Landscape of Newspaper Collections in Flanders

The catalogue Abraham provides an overview of newspaper collections in Flanders (see Figure 3).¹⁵⁶ Data from the catalogue shows that very large collec-

155. <https://www.bibliothecairerfgoed.be/projecten/nieuwe-tijdingen>.

156. A select number of collections in Brussels, including that of the KBR, are included in Abraham. For the purpose of sketching the newspaper landscape in Flanders, we focus here only on the collec-

tions are few and far between, while small collections are numerous.¹⁵⁷ Of the more than 100 collections in Flanders, over half hold fewer than twenty newspaper titles (see Figure 4).

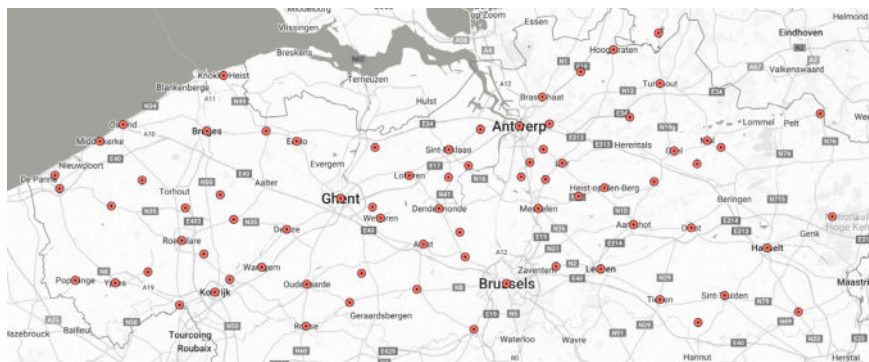


Fig. 3. Map of newspaper collections in Flanders. Source: *krantencatalogus.be*, February 2024.

SCOPE OF NEWSPAPER COLLECTIONS IN FLANDERS

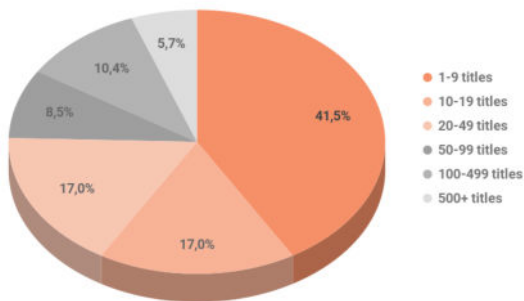


Fig. 4. Overview of the number of titles per newspaper collection in Flanders. Almost 60% of the collections in *Abraham* preserve fewer than 20 titles each; 41.5% have fewer than 10 titles. Source: *krantencatalogus.be*, February 2024.

tions in Flanders, plus the Archive and Museum for the Flemish living in Brussels (AMVB).

157. Data in the catalog is available as exports, updated weekly. See *Abraham — Data. Antwerp: Flanders Heritage Libraries*. Available on: data.krantencatalogus.be.

Collections

The smaller collections are often located in municipal or city archives, at local historical societies or in museums (see Figure 5). In other words, these are diverse organisations that may not consider newspaper preservation as central to their operations. Yet their collections contain many interesting — and often unique — titles.

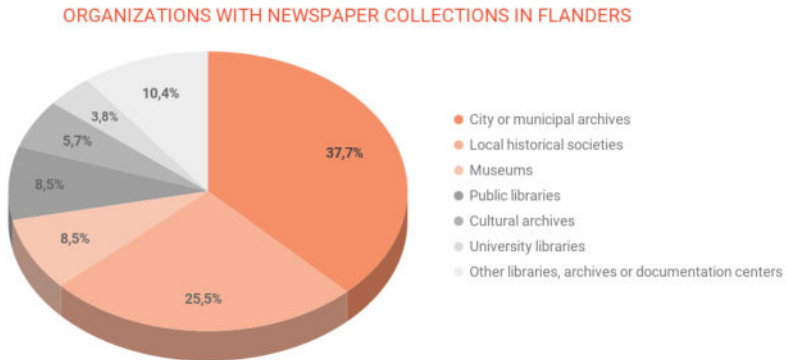


Fig. 5. Overview of collections in Abraham by type of organisation. More than 60% are local archives or historical societies. Source: krantencatalogus.be, February 2024.

In fact, approximately half of the newspaper titles in Flemish collections are only preserved by one organisation (see Figure 6). These unique titles are not found in Belgium's national library, KBR. Although KBR has the largest and most important newspaper collection in Belgium, one should not underestimate the number of titles missing from its stacks. After all, Belgium has only had a legal deposit requirement since 1966. Further, even historical newspapers with a high circulation cannot be guaranteed to survive in large numbers or in their entirety. Whether preserved in only one collection or across several collections, those newspapers that do survive are often incomplete, and in many cases only scattered specimens remain.

PRESERVATION OF NEWSPAPER TITLES IN FLANDERS

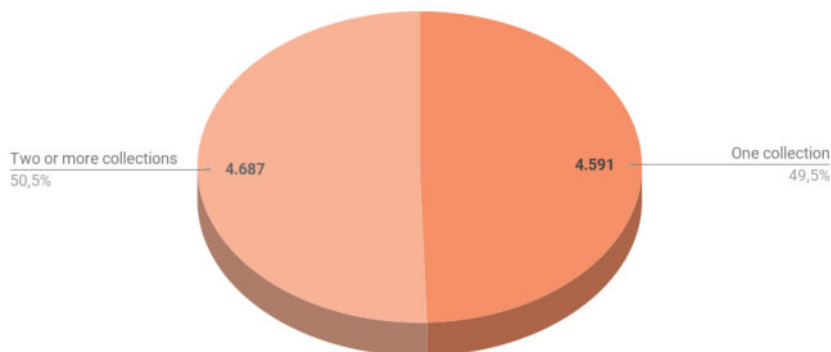


Fig. 6. Frequency of preservation in Flemish collections per newspaper title. Approximately half of all titles are preserved in only one collection. Source: *krantencatalogus.be*, February 2024.¹⁵⁸

The Challenge

Newspapers are not created to be everlasting. Acidification of the paper makes them very unstable and frequent consultation causes additional damage. More and more newspapers can no longer be consulted because they have become too fragile. With good conservation practices, collection managers can try to slow their decay, but acidification is unfortunately an inevitable and irreversible process. Digitisation is the only way to preserve the contents of this valuable heritage.

The dispersion of newspapers over a large number of collections, many of which are small and found in organisations without a specific focus on library heritage, creates logistical challenges to digitisation. Digitisation requires central coordination and cooperation across these many collections.

Another major challenge to the digitisation of newspaper collections in Flanders is their sheer volume. Although the exact quantity of historical news-

158. The list of newspaper titles in Flemish collections was cross-checked against titles in the following collections in Brussels: KBR, CegeSoma, Dacob, City Archive, State Archive, and the Royal Military Museum. Titles counted as present in only one collection are not present in any of these Brussels collections.

papers yet to be digitised in Flanders is difficult to determine, the data contained in the Abraham newspaper catalogue suggests that approximately 17 to 18 million pages would be a reliable estimate.



Fig. 7. Examples of fragile newspapers in Flemish collections. Photos: Yvo Wennekes; Ghent University Library; ADVN; Heemkundige Kring Meerbouts Patrimonium via Kempens Erfgoed.

Not all of these newspapers can be digitised at the same time. We therefore work with gradations of urgency. In 2021, Flanders Heritage Libraries developed *Courant*, an online selection tool which allows both urgency and quantity to be recorded per title.¹⁵⁹ In 2022, collection managers were asked to submit their digitisation priorities using this tool. They were asked to take specific criteria into account: selection criteria define which titles we consider eligible for digitisation; and prioritisation criteria define which titles can be marked for urgent digitisation. Both the physical condition of the newspaper — especially the risk of information loss — and its demonstrated use or potential for use are considerations for urgent digitisation. As of early 2024, 29 collection managers had earmarked 549 titles in *Courant* for urgent digitisation, with an estimated total of 3.34 million pages (see Figure 8).

These estimates take into account other completed and ongoing digitisation initiatives, including those of KBR and of Flemish organisations. BelgicaPress, KBR's online newspaper platform, now allows users to consult approximately 4 million newspaper pages, the tip of the iceberg of our national library's

¹⁵⁹. <https://courant.krantencatalogus.be/>.

enormous newspaper collection.¹⁶⁰ Local initiatives (such as *Erfgoed Brugge*,¹⁶¹ *Kempense Kranten*,¹⁶² *Historische Kranten Leper/Poperinge*)¹⁶³ and thematic projects (such as News of the Great War) have also already digitised millions of newspaper pages. A future digitisation programme for Flemish newspapers should be complementary to other initiatives and avoid unnecessary duplicate digitisation.

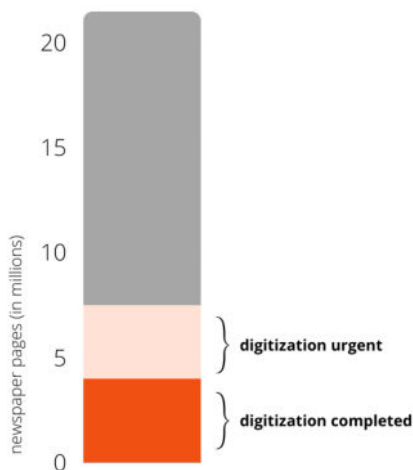


Fig. 8. Scope of digitisation in Flanders. We estimate approximately 17-18 million pages await digitisation, of which at least 3 million pages are in urgent need of digitisation. Sources: *krantencatalogus.be* and *courant.krantencatalogus.be*, February 2024.

A Strategy for the Future of Newspaper Digitisation in Flanders

The need for newspaper digitisation in Flanders is great: millions of pages await digitisation and dozens of collecting organisations are in search of a solution. Such a large-scale challenge demands a structural solution. We propose a

160. <https://www.belgicapress.be>.

161. <https://erfgoedbrugge.be/>.

162. <https://www.kempenserfgoed.be/>.

163. <https://historischekranten.be/>.

Flanders-wide newspaper programme as the answer.

When researching the possibilities and feasibility of such a programme, we looked into established programmes in other countries for inspiration. While newspaper digitisation programmes such as those in the Netherlands, Switzerland, Austria, Denmark and Germany each have their own approach and characteristics, they share two primary features. First, they digitise newspapers on a large scale: close to and sometimes exceeding one million pages per year. Second, each country has a central platform for online presentation of the newspapers. These programmes provide not only examples of how a newspaper programme can be structured, but also reference points for what it can accomplish.

These examples suggest that a total volume of approximately 20 million pages is a realistic estimate. They also suggest that a digitisation rate of about 1 million pages per year is feasible. However, we should note that foreign programmes are usually embedded in the workings of national libraries and start from their own collections, whereas in Flanders we have to consider the distribution of the newspapers over dozens of organisations, including small organisations, organisations for whom the newspaper collections are not central to their operations, and those who do not have experience with newspaper digitisation or similar digitisation projects.

As part of the *Nieuwe Tijdingen* project, Flanders Heritage Libraries and meemoo developed a proposal for a newspaper programme specific to the circumstances and needs in Flanders.

Three Pillars

While the digitisation of fragile and endangered newspapers forms the core of the proposed programme, this is just the first step. Both newly and previously digitised newspaper collections need to be preserved in sustainable digital archives. Digitised newspapers should also be made accessible and usable for a larger audience.

The proposed programme would thus centre around three essential and complementary pillars: digitisation, digital archiving and online access.

Digitisation. The programme would digitise newspapers held in collections in Flanders and in Dutch-language organisations in Brussels. The titles would have to meet the selection criteria: they must be newspapers, published in Flanders, and not yet digitised. Detailed selection criteria outline these rules as well as the exceptions. Eligible newspapers would be prioritised based on urgency. Several criteria would be used to determine relative urgency, with the physical condition of the material and need for conservation being the most important. Newspapers selected for urgent digitisation would then be divided into batches. The batches are meant to go through the digitisation workflow step-by-step, from preparing for digitisation by recording metadata and securely packing the papers for transport, all the way through quality control of the resulting digital images, OCR and metadata files.

We estimate that with sufficient funding, we could expect to digitise 3 million newspaper pages within an initial five-year period: approximately 700,000 pages per year with the exception of the first year.

Archiving. Newspapers digitised within the programme would be permanently digitally archived by meemoo. Newspaper collections that have already been digitised could also be included in the meemoo digital archive. An intake procedure for such collections has been developed by meemoo, including a Submission Information Package (SIP) specification.

We identified eleven organisations in Flanders that are interested in centrally archiving their digital newspaper collections. In 2024, meemoo started to survey its partners about their existing digital newspaper collections in order to create an inventory of these collections, and get a more accurate picture of their extent and the interest in central archiving.

Access. Once digitised, newspapers are presented online via the central platform hetarchieff.be. Meemoo has developed a newspaper interface and integrated it in this platform. This interface takes into account the specific functionalities and features that are needed when presenting digital newspapers. During the *Nieuwe Tijdingen* project we identified over 200 of these functionalities and ranked them based on importance ('must have', 'should have', 'nice to have'). Although it is unrealistic for any one platform to fulfil every last one of these requirements, and certainly not all of them can be realised in a first iteration of

a new interface, the aim is to meet as many needs as possible and to continually improve the platform in the coming years.

Hetarchief.be is accessible to the public. We wish to make the material on the platform as widely accessible as possible according to open data principles, with respect for copyright, potential contractual agreements and protection of personal data. Further, the programme must actively encourage the widespread (re)use of the digitised papers by a broad audience.

Benefits to a Centralised Approach

There are already several initiatives to digitise newspapers in Flemish collections either ongoing or completed. By proceeding in such a fragmented and often project-based manner, however, we miss the opportunity to increase efficiency, reduce costs, standardise output and pool expertise. Organisations that digitise newspapers independently may encounter many challenges, including insufficient infrastructure, lack of practical experience and technical expertise, and inadequate funding. Not every organisation with a newspaper collection has experience drafting a tender for digitisation, for example, or has an appropriate online platform for presentation or a suitable digital archive within which they might preserve the newspapers permanently.

A Flanders-wide programme with central coordination would solve many of the current problems and provide an overarching, systematic and sustainable approach to the three pillars of digitisation, archiving and access. Moving away from separate projects and toward a centralised approach allows us to make improvements for both the collecting organisations and the end users.

These ten core objectives serve as the foundation for the proposed programme:

1. **Collaboration.** The programme will support as many Flemish and Dutch-speaking Brussels collecting organisations as possible in the preservation of their Flemish historical newspapers through digitisation and digital archiving, and making these digital collections publicly accessible and usable.
2. **Preservation.** The programme will digitise historical newspapers in order to secure the content for the future in the face of the physical deterioration

of the collections.

3. High quality. The programme will ensure a uniformly high digitisation quality, in accordance with common international norms and standards.

4. Usability. The digitised newspapers will be made findable, searchable and thus usable through the application of techniques such as Optical Character Recognition, Named Entity Recognition and Named Entity Linking.

5. Cost efficiency. Economies of scale will ensure that newspaper heritage is digitised, digitally preserved and made accessible in a cost-effective manner.

6. Sustainable preservation. The digital reproductions of newspapers, their metadata and OCR texts will be preserved in a sustainable manner, in accordance with common international norms and standards for digital archiving.

7. Central online access. The programme will provide a newspaper interface that builds on a central online platform through which digitised historical newspapers will be searchable.

8. Accessibility. The digital reproductions of newspapers, their metadata and OCR texts will be made available as freely as possible via the central online access platform according to open data principles and with respect for copyright, contractual agreements and the protection of personal data.

9. Valorisation. The programme will actively encourage the widespread (re)use of the digitised papers by a broad audience.

10. Expertise. Expertise will be centralised and we will continually build upon accumulated knowledge, including technical expertise and practical experience. The programme will strive to share this knowledge and expertise with a variety of stakeholders, especially the collecting organisations involved.

Conclusion

How can such an ambitious and widespread programme be realised? Many of the building blocks are already in place.

Through the previous projects *News of the Great War* and *Primeur*, processes for large-scale and high-quality digitisation have been developed, tested and improved. Alongside these projects, a standard protocol for the intake of

digital newspapers has also been developed and tested. The central access platform hetarchief.be has been redesigned and a new interface has been developed which brings digitised newspaper collections together. Finally, a large and active network of collecting organisations who want to participate in a centralised newspaper programme has been established. Working together, and building on many years of experience, the existing infrastructure and processes are ready to be scaled up to create a structural newspaper programme suitable for the whole of Flanders.

As established by similar initiatives abroad, the success of such a large-scale programme requires structural financial resources. We are committed to keeping the urgent need for the centralised digitisation, archiving and presentation of this important historical heritage on the agenda, and we look forward to future endorsement of the proposed programme by the Flemish government and sufficient funding for its realisation.

Σαράη

100



5. Transforming ULB Libraries Digitisation Service: Going Beyond Scanning Our Own Collections to Also Become a Data Provider for Our Researchers

Anthony Leroy & Ingrid Mayeur

Abstract

The ULB Libraries' Digitisation Service has evolved from a basic scanning unit into a comprehensive mass-digitisation infrastructure. Initially focused on preserving and providing access to its own collections, the service expanded to include large-scale projects, such as digitising Belgian press and literary journals. Lessons from outsourcing challenges led to the development of an in-house workflow, increasing efficiency and quality control. Today, the service aligns with the collections as data paradigm, offering machine-readable digital corpora for researchers. Collaboration with digital humanities initiatives ensures data accessibility and interoperability. Preservation efforts are supported by the SAFE distributed network, ensuring long-term digital sustainability. Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) further enhance accessibility. The service now integrates lean management principles to optimize processes and meet researchers' evolving needs. A Service Level Agreement (SLA) ensures transparency and efficiency. This transformation positions ULB Libraries as a key research data provider, fostering innovation and academic collaboration.

Authors

Ingrid Mayeur is a lecturer in information and communication didactics and digital media education at the Université de Liège. Her research interests include computerised knowledge mediation and open science discourses. In addition to several articles published in these fields, she recently took part in editing the

volume *Communication scientifique et science ouverte. Opportunités, tensions et paradoxes* (De Boeck, 2023) and is a member of the *Dé_montages* collective (ULiège), which aims to rethink approaches to media education.

Anthony Leroy has worked as a software engineer at the Libraries of the Université libre de Bruxelles (Belgium) since 2011. He is in charge of the digitisation infrastructure and the distributed digital preservation program. He coordinates the SAFE distributed preservation network, an international LOCKSS network operated by seven partner universities. He is also actively involved in various research data management activities at ULB. Anthony is an engineer in electronics and telecommunications with a PhD in microelectronics (ULB) and has been a researcher for almost ten years in collaboration with several industrial partners.

Keywords

Digitisation; data; digital humanities; collections as data; preservation; OCR

The ULB Libraries' Digitisation Service has undergone a remarkable transformation over the past two decades; what began with the acquisition of a single book scanner eventually led to the design of a complete in-house professional mass-digitisation facility that has already made possible the digitisation of tens of thousands of items, totalling millions of pages, largely accessible through Open Access.

This chapter traces the history of our digitisation service, highlighting its key achievements. It places a particular emphasis on our Belgian press and literature digitisation project, which has proven to be not only a great learning experience but also a turning point in the history of our service.

Looking ahead, our ambition is to also provide digitisation services to our researcher community, embracing the principles of collections as data while adopting best practices in research data management to ensure our production is FAIR-compliant and machine-actionable.¹⁶⁴ In concrete terms, meeting this

164. PADILLA, Thomas, ALLEN, Laurie, FROST, Hannah, POTVIN, Sarah, RUSSEY ROKE, Elisabeth & VARNER, Stewart. *Santa Barbara Statement on Collections as Data. Always Already Computational: Collections*

new challenge means providing digitised corpora as structured data in formats adapted to quantitative analysis methods derived from the Digital Humanities, thus enabling these collections to be apprehended and explored from a new and broader angle.

First steps in Digitisation (2005-2008)

ULB Libraries started developing their digitisation service in 2005 with the acquisition of their first A0 book scanner.

During the first stage of their deployment, digitisation projects mainly involved copyright-free, degraded or no longer published books to provide or restore access to those important collections.

Binarised (black and white) PDF files were made available online via the *Digithek*, a website providing direct access to the files and contextualising information about the collections.¹⁶⁵ At first, PDFs were not searchable as OCR was not introduced until 2011.

First Mass-Production Project: Digitising the Belgian French-Language Literary Journals (2008-2013)

Our first major digitisation project, aimed at digitising Belgian French-language literary journals spanning the period from 1920 to 1960, was launched in 2008. This ambitious 5-year project was scientifically directed by Prof. Paul Aron and logistically coordinated by Ingrid Mayeur.¹⁶⁶ It was funded by various financial instruments: a Concerted Research Action (ARC) and the *Plan de préservation et exploitation des patrimoines* from the Wallonia-Brussels Federation (Pep's). It has also given rise to fruitful partnerships with other heritage conservation establishments (KBR, other university libraries, Mundaneum, Archives de la Ville de Bruxelles, etc.).

as Data. Zenodo (online), 20 May 2019, <https://doi.org/10.5281/zenodo.3066209>.

165. <https://digitheque.ulb.ac.be/>.

166. MAYEUR, Ingrid. *ARC 'Presse et littérature en Belgique francophone' (1920-1960) : Note générale au sujet des numérisations*, 3 October 2013 [internal report].

This programme has its origins in the *Presse et Littérature en Belgique francophone* (1920-1960) project (directed by Prof. Paul Aron), which was designed to study the key role of periodicals in Belgian literary life, and the connections between the literary and journalistic spheres, as evidenced by the joint activities of authors both in the press and in literary/artistic journals.¹⁶⁷ In other words, the aim was to examine the social uses of literature as seen through the relationship between the press and literature in French-speaking Belgium. Digitising a corpus composed of a selected number of journals was intended to facilitate the study of how literary texts were disseminated through periodicals. Its exploitation was effected on one hand by integrating it into an existing database resulting from a collaboration with ULiège focusing on the time frame 1920-1960 (the CIEL database)¹⁶⁸ and, on the other hand, by creating a section dedicated to Belgian literary journals on the ULB Libraries' Digithèque website, taking into account a wider temporal spectrum.

Much effort was put into identifying materials before they were digitised, identifying and checking for gaps in the collections and evaluating the conditions of the original materials. The first step was to draw up an inventory of the Belgian periodical collections held by the ULB from which an outline plan for digitisation could be developed. Priority was given to those ULB collections in good condition which were either complete or easily supplemented by borrowing missing volumes from other institutions. This led to a collaboration with the Royal Library of Belgium (KBR) for a first phase of digitisation in 2009. This institution was itself also involved in a large-scale project to digitise the Belgian daily press: BelgicaPress.¹⁶⁹

Later, the digitisation of periodicals published before 1914 was conducted as part of the Pep's plan of Wallonia-Brussels Federation, with KBR and AML agreeing to lend missing volumes to ULB.

Over the course of successive digitisation campaigns, several issues concerning the legal, technical and methodological aspects of the process became apparent.

167. ARON, Paul & SOUCY, Pierre-Yves. *Les Revues littéraires belges de langue française de 1830 à nos jours*. Brussels: Labor, 1998.

168. *Collectif interuniversitaire d'étude du littéraire*, available at <http://ciel.philo.uliege.be/cielems/>.

169. <https://www.belgicapress.be>.

Lessons Learned

Compliance with copyright regulations was particularly challenging, as a large part of the collections was still under copyright. The latitudes presented by legislation specific to these copyrights and the exceptions provided around exploitation for research and teaching purposes had to be considered. As a result, it was agreed to provide only restricted access to collections published after 1914. Simultaneously, however, it was decided that the policy of other conservation institutions regarding periodicals published before this date was to be followed. It was deemed reasonable to provide free access to heritage works that are sufficiently old not to prejudice the rights holders, as far as we know, and leaving them the possibility of coming forward if necessary.

In other respects, the volume of the material to be digitised and our limited internal capabilities at the time meant that most of the collections thus processed were digitised by external service providers. The varied and uneven quality of their services resulted in a perceptible heterogeneity in the quality of digitised materials with some deliverables containing recurring errors: missing pages, black and white covers when full colour was expected, etc.

This first experience with mass digitisation taught us that delegating the digitisation process to third parties has severe drawbacks and securing the services of qualified people who are familiar with the issues involved in digitisation for heritage conservation purposes is vital.

The cost of outsourcing digitisation may seem very competitive at first glance, but such contracts also require extensive and time-consuming quality control measures to be undertaken after digitisation — these can only be carried out in-house and their cost and inconvenience is often underestimated.

In 2023, at the request of KBR, ULB transferred all digitised content corresponding to these collections to KBR (see dissemination strategy *infra*). The pooling of these important collections at the national level appeared to be particularly relevant, as it greatly facilitates discovery, accessibility and usability, paving the way for new scientific projects.

Bringing Mass-Digitisation Projects In-House (2013-2016)

In 2013, we launched our first in-house mass digitisation project.

The aim of the project was to digitise our complete collection of PhD theses awarded since the creation of the University in 1834, a total of over 8,000 volumes and 1.8 million pages.



Fig. 1. One of the three high-volume scanner stations operated by a student worker.

After requesting quotes from commercial digitisation service providers and taking into account our previous experience, we decided not to outsource this work but to carry it out internally with the help of a team of student workers coordinated and supervised by Anne Leroy from our digitisation staff.

Despite our very limited experience in the field at the time, this ambitious undertaking captured the bold vision and innovative spirit of our then Head of Library, Prof. Jean-Pierre Devroey, who strongly supported the initiative. Using overhead book scanners to digitise such a large collection would have been prohibitive in terms of both time and money — we chose instead to invest in high-volume production scanners with automatic document feeders. These

machines are designed to digitise up to 100,000 pages per day (duplex scanning at over 100 pages per minute, up to 500 pages at a time). They are typically used by commercial service providers for high-throughput mail scanning as a service to large companies.

The expected increase in speed for the digitisation step was by a factor of at least ten, as only the most prestigious theses (less than 5% of the collection) were to be digitised on a traditional book scanner, while the rest of it would go through the production scanners.

Of course, production scanners require the binding of each thesis volume to be trimmed to fit into the automatic feeder. After digitisation, the trimmed volumes could not be reintegrated in the library but they are kept in our silos. This way, if a problem is detected in the digitised version of a volume made available online, it is always possible to retrieve the original paper volume and rescan it.



Fig. 2. Overhead A2 book scanner station for books that cannot be trimmed.

Development of a Complete Digitisation Workflow

While it is obvious that having efficient scanners is essential, the importance of scanning software is often underestimated. It is just as important to have powerful, customisable software to ensure optimum quality and maximise productivity throughout the entire workflow.

Our whole digitisation process has been automated to maximise productivity with the support of the Libraries IT department headed by Benoit Pauwels.

Barcode identification is at the heart of the digitisation process. A barcode uniquely identifies every object to be digitised at the required granularity (typically, a volume).

All generated and derivative files are then automatically assigned a name based on this identifier. After digitisation, the boxing of the physical volumes is also supported with barcodes to facilitate retrieval.

Prior to the digitisation process, descriptive metadata of the object was collected in our institutional repository either manually or after automatic migration from an existing catalogue. The barcode identifier for each object to be digitised was included in the institutional repository bibliographic record, enabling automatic file integration after digitisation.

In-house software was initially developed to support quality control.¹⁷⁰ This web application enabled quality reporting with the original in hand for comparison as well as text extraction after OCR. However, it quickly became challenging to maintain and improve this software, with the limited development resources available to the department.

Two years ago, therefore, this software was replaced by a state-of-the-art commercial software which provided advanced image quality enhancement features, exhaustivity checks, and better reporting features. Unfortunately, no comparable open source alternative could be found, although some open source software have improved a lot in the last years and could be a good match in the future.

170. LEROY, Anthony & PAUWELS, Benoit. Ph.D. Theses Mass Digitization at ULB, unpublished poster session presented at *ETD 2016 Data and Dissertations 19th International Symposium on Electronic Theses and Dissertations* (11-13 July 2016): Lille (France), <http://hdl.handle.net/2013/ULB-DIPOT:oi:dipot.ulb.ac.be:2013/231956>.



Fig. 3. The image enhancement processes improve both OCR performance and the compression ratio of dissemination formats.

Working with Students

A total of 32 students took part in our digitisation projects, working on our production scanners under the coordination of a staff member.

Coordinating students for such projects can be challenging and requires a great deal of flexibility, as they have limited availability and complicated schedules, especially during the exam periods. Furthermore, they are usually available for two or three years only – turnover is relatively high, and the training of new students to use the machines and software is a regular job. Fortunately, cooperation between students facilitates know-how transfer but regular supervision and reminders of instructions are indispensable.

Scaling Up

Our first internal mass-digitisation project, encompassing more than 8,000 volumes and 1.8 million pages, took less than 3 years to complete with only one staff FTE dedicated to the project and over 5,000 hours of student work. Of course, the main advantage of carrying out digitisation ourselves was that we could reuse our workflow for many other projects. Following the successful completion of the PhD collection, we digitised a large collection of offprints (25,000 volumes), a collection of lecture notes that had accumulated since 1834

(more than 12,000 volumes) and we are about to start the digitisation of the master's theses. In the meantime, we continued to invest in the digitisation machinery, acquiring two more production scanners, two new book scanners and a plan scanner.

ULB Libraries Digitisation Scope

ULB Libraries strategy is to focus on the digitisation and the dissemination of our own collections, making it available through open access as much as is possible.

Several criteria are taken into account to evaluate if a collection needs to be digitised and what level of priority it might reach:

Can the digitised collection be made available in open access or is it under embargo?

Are the items part of our own collections?

Are the physical objects in poor condition?

Is the volume of objects to be digitised small or can the items be trimmed to fit in production scanners?

Is the expected frequency of consultation significant?

Already convinced by the principles underpinning the recent Vancouver statement on collections as data, we decided not to outsource our material to commercial suppliers but to keep full control over our digitisation, dissemination and preservation workflows.¹⁷¹

ULB Libraries Dissemination Strategy: Breaking Silos and Fostering Collaboration

The vast majority of the collections we digitise consists of our own scientific collections. Therefore, the digitised content in the form of text-under-image PDF objects are submitted to our institutional repository, and for the most part in open access. The content is made discoverable publicly as metadata are regularly harvested by OpenAIRE, BASE (Bielefeld Academic Search Engine), Worldcat and Google Scholar.

171. PADILLA, Thomas, SKATES KETTLER, Hannah, VARNER, Stewart, & SHORISH, Yasmeeen. *Vancouver Statement on Collections as Data*. Zenodo (online), 13 September 2023, <https://doi.org/10.5281/zenodo.8342171>.

Whenever possible, digitised collections that are not directly in our custody are made centrally available to facilitate access, thus sharing digitisation efforts between public institutions.

In line with this approach, we transferred our complete digitised press collections to KBR in 2023.

Sharing and mutualising non-owned digitised collections at the national level is important both for the advancement of science and to ensure that the material will be made easily available to the general public, thus avoiding unnecessary data silos.

ULB Libraries Preservation Strategy: Community-based Digital Preservation Network

The primary goal of digitisation is to provide easy access to surrogates of analogue objects, capturing their key features. However, it is important to remember that digital replicas cannot replace physical objects. By essence, the digitisation process inevitably results in a loss of information, due to limited spatial resolution, imperfect colour reproduction and other digitising artefacts (noise, skew, moiré...). Moreover, the preservation of digitised material is generally much more expensive and resource-intensive than that of born-digital objects; raw images in particular require considerable storage capacity and energy consumption over the long term.

As a consequence, digitising objects to ensure their preservation is generally not the best strategy. Of course, there are a few exceptions, such as unique works in a very advanced state of deterioration.

That is why we decided to preserve only the dissemination derivatives (PDFs) in our digital preservation solution, with digitised objects being considered only as last-resort facsimiles of the original.

As already stated, we have a strong commitment to our missions of providing access to and preserving the collections of which we are the custodians.

Our digitised collections are required to be preserved in a digital service that we fully control. We did not want to outsource the digital preservation mission to commercial actors and remain an actor of our digital preservation.¹⁷²

172. SKINNER, Katherine & SCHULTZ, Matt. (eds.), *A Guide to Distributed Digital Preservation*. Atlanta,

A university library acting on its own is unlikely to have the required human and technological resources to build an efficient distributed preservation solution allowing it to maintain several archive copies of digital assets, spreading them in multiple locations that are considered safe from natural hazards.

But if a group of institutions with similar needs come together, they can create a network filled with copies of all of their archives, with each network node maintaining copies for the others. This collaborative approach inspired ULB and six other university libraries to found the SAFE PLN network to preserve their digitised and born-digital collections.¹⁷³

SAFE PLN is a digital preservation network founded by seven universities with the aim of archiving and preserving their scientific output over the very long term (several decades). Over the past ten years, the SAFE PLN has proven its viability and its capability to safeguard the valuable digital resources of its members: Lunds Universitet; Universität Bielefeld; Université catholique de Louvain; Université de Genève, Université de Montréal, Universiteit Gent, and Université libre de Bruxelles which acts as network coordinator.



Fig. 4: Current members of the SAFE PLN distributed digital preservation network.

Each member maintains a preservation node that collects the digital objects from all members' institutional repositories. Objects are thus replicated in seven distributed copies across two continents and five countries.

GA (USA): Educopia Institute, 2010. Available on: https://metaarchive.org/wp-content/uploads/2017/03/A_Guide_to_Distributed_Digital_Preservation_0.pdf.

173. LEROY, A. LOCKSS Distributed Digital Preservation Networks. *Proceedings of the 11th Conference on Grey Literature and Repositories* (online). Prague: National Library of Technology, 2018, ISSN 2336-5021. Available on: <https://nusl.techlib.cz/en/conference/conference-proceedings>.

The assets are then automatically verified and repaired by the state-of-the-art LOCKSS open-source software, a peer-to-peer bit-level preservation solution developed at Stanford University.¹⁷⁴

Optimising Access and Preservation to Ensure Sustainable and Responsible Digitisation

Digital sobriety is also an essential factor, one taken into consideration in both our dissemination and our preservation strategies. Our digital footprint is reduced by using the best compression algorithms (e.g. visually lossless JPEG2000 compression instead of TIFF LZW, providing a reduction of 50% size for the same perceived image quality), and by finding a suitable compression ratio resulting from a trade-off between file size and image quality for dissemination formats that will be transmitted on the web as well as improving the overall compression efficiency by smoothing page backgrounds.

Scanner manufacturing is also a major contributor to the global environmental footprint of digitisation. Therefore, we maximise the use of our existing scanners before considering the acquisition of new ones. For example, our 20-year-old A0 book scanner is still fully operational today and regularly used.

Lastly, as mentioned earlier, our policy prioritises the scanning of collections whose digitisation brings demonstrable added value and is expected to be put to frequent use for research or education purposes, not just to fill a quota.

Embracing the Collection as Data Paradigm

Nowadays, most humanities research projects need to access digital material, not only to facilitate access and storage but, more importantly, to allow them to use machines to support their research workflows, and to improve and guide their analysis.

Unfortunately, humanities researchers are often overlooked in this digital

174. ROSENTHAL, David S.H., ROBERTSON, Thomas, LIPKIS, Tom, REICH, Vicky & MORABITO, Seth (2005). Requirements for Digital Preservation Systems: A Bottom-Up Approach. *D-Lib Magazine*, n° 11. Available on: <https://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.

era. They usually do not have the digital infrastructure dedicated to their specific needs at hand and sometimes they also lack the advanced skills necessary for the efficient handling of digital data.

As a result, researchers generally have to call on a variety of external commercial companies to carry out their projects. This can entail significant additional costs, rarely meets their needs exactly, and leads to great discrepancies in the digitisation practices and data formats.

Libraries, on the other hand, have a crucial role to play in supporting humanities researchers via their mission of supporting research data management. Consequently, it makes sense for libraries' digitisation services to collect and deliver research data to researchers, and to provide them with access to storage and digital preservation solutions.

Our role also extends to helping researchers unfamiliar with this type of material by making it more accessible, understandable and reusable by others by following the FAIR principles.

Conversely, data analysis requires very specific skills that we do not have at library level, so close collaboration with digital humanities experts is essential.

This is why our digitisation service has joined forces with Praesto,¹⁷⁵ an interfaculty digital humanities support platform headed by Prof. Sébastien de Valeriola, who holds the Chair in Digital Humanities at ULB.

By combining the capabilities of ULB Libraries' digitisation infrastructure with support in using state-of-the-art digital humanities analytical methods, this platform will act as a one-stop shop for researchers to help them adopt a macroscopic quantitative approach to their corpora, opening the way to new research questions.

PhD Theses as Data for the Project (*Dé*)chiffrer la recherche à l'ULB

This project, carried out by postdoc researcher Matthieu Pichon and under the supervision of Prof. Kenneth Bertrams and Prof. Sébastien de Valeriola, is the first digital humanities project in which we are collaborating as a provider of collections as data.

The project concerns our entire collection of digitised and born-digital

175. <https://quadihum.ulb.be/praesto-dh/>.

doctoral theses, no longer considering them as individual items but as a whole corpus, a single dataset to be studied using a wide range of quantitative methods derived from digital humanities (natural language processing, network analysis, etc.).

In particular, topic modelling will help automatically determine the research themes covered in PhD theses, in order to provide a global overview of the history of ideas at ULB.

The research project will also contribute to the analysis of the evolution of the multidisciplinary nature of research, the evolution of the place of women in STEM, the reconstitution of schools of thought, and the way in which the work of other researchers is cited.

It represents an innovative and promising approach to our collections, opening the door to exciting discoveries in both the history of science and the history of our University.

ROPB as Data for the Ordonum Project

Another digital humanities project in which we are collaborating as a data provider concerns the *Recueil des ordonnances des Pays-Bas - série 1506-1700 (ROPB)*.¹⁷⁶

Designed primarily for teaching purposes, and led by Prof. Marie Kervyn and Prof. Sébastien de Valeriola, this project is intended to showcase the many ways in which such a corpus of historical documents can be explored using various quantitative analysis methods. By using the very concrete example of ROPB with which they are familiar, history students can get a deeper understanding of what these cutting-edge quantitative methods can achieve.

At a later stage, the exploitation of this corpus will itself be the subject of research projects.

176. <https://quadihum.ulb.be/ordonum/>.

Converting Digitised Items into Machine-Actionable Data with OCR and HTR

Once digitised, Optical Character Recognition (OCR) can be performed on the images to extract the text from typewritten and printed documents.

Despite considerable progress in recent years, OCR results are still far from perfect. In particular, performing accurate layout and segmentation analysis (recognising the lines and text blocks in a page prior to character detection) can be a challenge, in particular for multi-column press articles. Some uncommon fonts are also particularly problematic for OCR.¹⁷⁷

While a Character Error Rate (CER) of 2% is usually considered as acceptable accuracy, it is also important to take into account the Word Error Rate (WER).¹⁷⁸ The OCR could indeed accurately recognise all the characters in a text leading to a low CER, but nevertheless produce many word segmentation errors making the text output difficult to use without additional processing.

While even imperfect OCR undeniably facilitates document searches, special care must be taken for other use cases, for example in the field of digital humanities.¹⁷⁹

Fortunately, various post-OCR processing pipelines approaches exist to enhance the raw data obtained from OCR.¹⁸⁰

Moreover, to keep pace with the rapid evolution of OCR technology over the years, we had planned from the start that we needed to be able to regenerate all derived objects when required. For example, while PhD theses had initially been OCRed using ABBYY Finereader 9, we have later fully reprocessed all theses using ABBYY Finereader 11. If better results can be achieved in the future with new software, such processing could be carried out once more.

Currently, we use two OCR softwares: Google Tesseract and ABBYY Finereader 12. The choice depends on the type of material, the language and the purpose.

177. HOLLEY, Rose. How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitization Programs, *D-Lib Magazine*. 2009, vol. 15, n° 3-4.

178. Ibidem.

179. See Sébastien de Valeriola's contribution in this volume.

180. NGUYEN, Thi Tuyet Hai, JATOWT, Adam, COUSTATY, Mickael. & DOUCET, Antoine. Survey of Post-OCR Processing Approaches, *ACM Comput. Surv.* 2022, vol. 54, n° 6, article 124, <https://doi.org/10.1145/3453476>.

Post OCR data is provided to the researchers in several forms: plain text, ALTO-XML (a structured document format integrating the original text layout), the images after enhancement processing, the raw images and a text-under-image PDF.

Many other projects are in the pipeline, including Handwritten Text Recognition (HTR) - which would help researchers save a lot of time in transcribing manuscripts to allow for greater focus on corpus analysis, using quantitative methods derived from digital humanities.

For handwritten material, we have started to deploy an internal solution based on eScriptorium and Kraken called TAMI (*Transcription Automatique de Manuscrits et Imprimés*).¹⁸¹

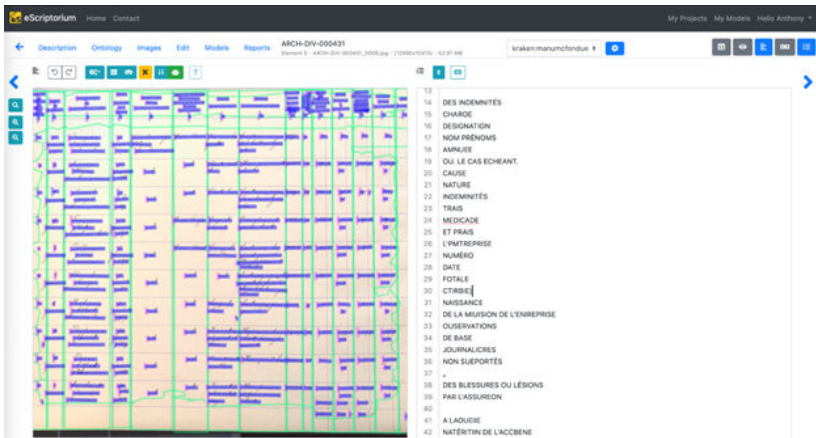


Fig. 5. TAMI web interface (based on eScriptorium) allows end users to easily collaborate by providing a reference segmentation and transcription that can later be used as ground truth to retrain (fine-tune) HTR models.

181. KIESSLING, Benjamin. Kraken—A Universal Text Recognizer for the Humanities, *DataverseNL* (online). (2019) V2, <https://doi.org/10.34894/Z9G2EX>; KIESSLING, Benjamin, TISSOT, Robin, STOKES, Peter & STÖKL BEN EZRA, Daniel (2019), eScriptorium: An Open Source Platform for Historical Document Analysis, International Conference on Document Analysis and Recognition ICDARW Workshops. (online). Sydney, NSW, Australia, 2019, <https://doi.org/10.1109/ICDARW.2019.10032>.

eScriptorium can be seen as an open-source alternative to Transkribus running on our own servers. We use eScriptorium as a user-friendly interface for non-tech-savvy users to provide a manually corrected transcription that we will later use as reference data to train the models (ground truth) for a specific handwriting.

Kraken, the engine under eScriptorium, can also be used as a command-line interface tool to train models to exert better control on the training parameters.

Future developments also integrate the support of the IIF dissemination scheme in our repositories. This will enable on-the-fly creation of image derivatives tailored to researchers' needs. Thanks to the standardised image access API provided by IIF, researchers will be able to assemble machine-accessible collections by seamlessly combining images from our repositories and other IIF data sources which they will be able to process and analyse in a single unified workflow.

Transforming from an Internal Digitisation Service to an On-Demand Service Tailored to Our Researchers' Needs

Transforming an internal library service into an outward-looking service for our whole research community is a challenging task.

Fortunately, we can count on the invaluable experience of our colleague Alexandra Jaumouille to work on improving the workflow and processes in line with the Lean principles.

One of the most recent developments aimed at guaranteeing service quality is the drafting of a Service Level Agreement (SLA) template to ensure clear service expectations, clarify responsibilities and guarantee smooth communication with researchers. To ensure that we will consistently meet our SLA commitment, we have formalised and streamlined our digitisation procedures. We are also equipping our entire workflow with status monitoring software to enable us to track and manage several large-scale concurring digitisation projects simultaneously.

Finally, we are precisely evaluating and optimising our digitisation costs to provide competitive prices to researchers.

Conclusion

Our digitisation service has evolved considerably over the last two decades.

As well as continuing the digitisation of our own collections to facilitate access to our most valuable items, our ambition is now to also become a data provider for our researchers. This new role poses considerable challenges, but will enable us to offer important services to our researchers, freeing them as much as possible from tedious and time-consuming tasks so that they can focus on their core research questions. This exciting transformation is still in its early stages and will probably require many years to complete but it holds great promise.

Acknowledgments

We extend our gratitude to everyone who supported and contributed to the projects that have been highlighted in this article.

In particular, our deepest appreciation goes to Prof. Paul Aron, Muriel Baguet, Renaud Bardez, Véronique Delannay, Prof. Jean-Pierre Devroye, Bruno Leclercq, Benoit Pauwels, who have been essential supporters of all these projects.

The work of Anne Leroy, who coordinated and supervised the team of student workers, was absolutely crucial to the success of all these projects.

Xavier De Keijser's expertise in book scanners, especially in complex cases, was also essential to many projects.

A heartfelt thank you to Sarah Bourazma, Youssef Boutahar, Alexandra Jaumouille, and Nicolas Hachez who were involved in our most recent projects.

We would also like to express our thanks to Christophe Algoet, Cécile Gass, Wendy Hendrickx and Françoise Vandooren for their involvement in the valorisation of the digitised objects.

We are immensely grateful to Prof. Sébastien de Valeriola, without whom the ongoing transformation of our service to better meet the needs of researchers would not be possible.

Last but not least, our thanks go to all the student workers without whom these projects would not have been possible: Antoine Vandermosten, Artuur

Grauwels, Camille Xiang, Caroline Campion, Catherine Thomas, Chloé De Wolf, Eliot Harney, Emilie Antoniol, Estelle Didi, Grimbert Pauwels, Julie Toussaint, Julie Tran, Joséphine Antoniol, Laurent Vanalbada, Léon Rubbens, Lola Grignard, Lucile Didi, Manon Grauwels, Marin Capelle, Nathan Vaiser, Olivier Hamende, Pauline Lagace, Quentin Tran, Reinout Pauwels, Rémi Priem, Romain Van der Vennet, Ségolène André, Sophia Azzagnuni, Soufiane Rahmoun, Tibeert Pauwels, Thomas Griffet and William Druet.

6. From Paper to ‘Paper’. Facilitating Structural Digitisation and Reuse of Print Media Preserved in a Heritage Institution

Roxanne Wyns & Katrien Weyns

Abstract

The growing demand for digital data is prompting heritage institutions and research infrastructures to explore new approaches to the large-scale digitisation of analogue media. To facilitate both this research and data preservation, large series of periodicals are systematically digitised by KADOC-KU Leuven. In the *Journals in the Picture* project, this archival institution investigated the automation of the selection process within that digitisation workflow. By enriching bibliographic descriptions, a more efficient process was obtained. Experiences during the project prompted reflection on the use of linked (open) data in the heritage sector: do we need more linked data to support digitisation work processes and to facilitate access? Conclusions can also be drawn from the experiences of the iCANDID project, a research infrastructure managed by Libis for collection, normalisation and access to big data. This FAIR data infrastructure facilitates easy access for researchers to big datasets of modern (born-digital) and historical (digitised) media. iCANDID is also influenced by the latest technological advances, and it is therefore used to enrich data by means of artificial intelligence.

Authors

Roxanne Wyns is innovation manager at LIBIS, a digital service provider that is part of KU Leuven Libraries. As innovation manager she specialises in FAIR data infrastructures and works on several domain-specific research infrastructure projects related to FAIR management of data collections. Roxanne is co-promotor and technical lead of the FWO medium scale infrastructure project iCANDID 3.0 SSH FAIR Data Hub.

Katrien Weyns was head of the heritage library of KADOC, the Interfaculty Documentation and Research Centre on Religion, Culture and Society at KU Leuven until July 2025. After a master's in history at KU Leuven she deepened her knowledge of heritage management during a masters in Archival science. At KADOC she focused on strategies to collect, preserve and provide access to born-digital and digitised collections. Now she is head of Collection at the Flemish Architecture Institute (VAi).

Keywords

Linked open data; LOD; FAIR data; periodicals; digitisation

A project initiated by researchers at KU Leuven aimed to safeguard archives and publications of Catholic political, cultural, socio-economic and religious organisations led to the founding of a Catholic Documentation and Research Centre (KADOC) at the University of Leuven in 1976. With financial support from the Flemish government, the University of Leuven, and Catholic organisations, KADOC evolved into an accredited research and heritage institution and interfaculty centre.¹⁸² Over time, the scope of the collection broadened from Catholic heritage to documenting the relationship between religion, culture, and society in Belgian, European, and global contexts since 1750.

As of 2023, the collection encompasses more than 30 running kilometres of paper archives, 214,000 physical publications, 14,000 periodicals, and nearly 4 million digital objects. The digital collection is expanding rapidly, featuring born-digital heritage such as web archives, e-books, and digital records, as well as digital copies of physical heritage objects. KADOC manages these extensive collections and makes them available through multiple specialised catalogues and databases, applying international standards for management, access, and use. To develop and maintain this infrastructure, KADOC collaborates with LIBIS, a digital service provider of KU Leuven. LIBIS is responsible for the deployment and management of systems and processes that support a FAIR (Findable, Accessible, Interoperable, and Reusable) approach to various collec-

182. KADOC-KU Leuven [online]. Leuven (Belgium): KADOC-KU Leuven. [consulted on 28/03/2024]. Available on: <https://kadoc.kuleuven.be>.

tions and research data.¹⁸³ LIBIS assists its partners in the digital transformation of their collection management and has been closely working with KADOC in streamlining their collection and digitisation workflows, in response to the significant increase in the volume of (born-)digital materials it is handling.

The landscape of information flows for heritage institutions has undergone significant changes. Content circulates more rapidly than ever and we are inundated with ever-increasing volumes across many formats and platforms, including print media, digital media, and social media. In this era of rapid technological advancements and their concomitant research and valorisation possibilities, a *digital-first* approach is becoming a mantra. As researchers can distinguish themselves by accessing and analysing comprehensive and accurate datasets, the pressure for heritage organisations to provide data in an interoperable and reusable manner has grown.

These changes necessitate a shift from manual appraisal and selection to automated selection, via automated data collection and enrichment of data for efficient analysis and selection. Furthermore, heritage organisations and digital service providers must reconsider how to offer access to these vast collections in an efficient and practical manner, enabling researchers, heritage professionals, or anyone with an interest and legitimate right to access the data to handle it effectively. This transition presents challenges at the organisational, social, and infrastructure levels. This case study provides an overview of the efforts of KADOC and LIBIS in facilitating FAIR access to data collections from print, digital, and social media sources. It also identifies areas that require further discussion and action within the heritage community to achieve sustainable and FAIR access to collections.

The Data Collection Process

Gathering data for research can be a time-consuming endeavour. As a researcher, you must identify where to locate the data, how to access it, the standard (or proprietary) format in which it is stored, and how to use it in your

183. Libis [online]. Leuven (Belgium): Libis. [consulted on 28/03/2024]. Available on: <https://libis.be/en/home>.

research. To address these challenges in data usage and sharing, the FAIR principles — making data Findable, Accessible, Interoperable, and Reusable — have been introduced. These principles are also influencing heritage organisations by promoting a shift from object-centric findability and access to more interoperable and reusable data collections. To promote this transition, the Flemish government is encouraging GLAM (Galleries, Libraries, Archives, Museums) institutions to enhance the availability of curated data through the dissemination of Linked Open Data (LOD).

Conceived by Tim Berners-Lee, the LOD concept focuses on interconnecting data to create a network of datasets that enhances their discoverability, accessibility, and reusability, both directly and indirectly, for both machines and humans.¹⁸⁴ While data can be linked without being ‘open’, Berners-Lee argues that publishing data under an open licence is crucial for facilitating reuse. In 2010, he introduced a 5-star ranking method to promote the adoption of LOD. One star is awarded for making data freely available online irrespective of format, two stars for publishing data in a machine-readable format and three stars are granted for making data available in a machine-readable and non-proprietary format. Four stars are achieved by using open W3C standards to publish the data, enabling seamless referencing of datasets. Compliance with the 5-star model involves linking one’s own data with that of others and publishing one’s data using W3C standards.

The requirements for each star or step may be fulfilled using various methods, and there are no specific guidelines regarding the inclusion or exclusion of specific content within this semantic web framework. In a heritage and research context, for example, an image of a historic newspaper page can be linked to the corresponding collection in the hosting library, an authority record detailing information about the publisher, the machine-readable text layer stored in a repository, or even to the original drawing of the depicted comic strip. In the heritage context, the benefits extend beyond making material available for reuse by visitors in the reading room. Adopting more FAIR and linked data practices will also streamline internal work processes, such as digitisation. Digitisation plays an

184. BERNERS-LEE, Tim. *Linked Data*. [online] 27 July 2006 [consulted on 28/03/2024]. Available on: <https://www.w3.org/DesignIssues/LinkedData>.

important role in improving the findability and reusability of paper collections. As previously mentioned, the demand for digitised sources for research and valorisation purposes is continually increasing. The popularity of techniques for processing machine-readable content is a contributing factor to this growth.

Optical Character Recognition (OCR) facilitates easier analysis of sources, enabling researchers to search for topics and names, even if the recognition process itself is not flawless. Additionally, Natural Language Processing techniques offer new insights based on big data. Furthermore, digitisation broadens access to new groups of researchers worldwide. This enables them to access relevant sources without the need for travel and facilitates the virtual reconstruction of collections dispersed across disparate institutions. Digitation also serves as an instrument in the preservation strategy of GLAM institutions by enabling the creation of digital replicas, particularly for physically delicate or cumbersome objects such as large maps or newspapers. When digitising in these cases, the aim is not only two-dimensional reproduction of a legible text, but also proper colour and depth that mimic the tactility of the piece digitally as much as possible. Additionally, artificial intelligence technologies such as object recognition or named entity recognition provide a promising method for generating additional metadata to enhance access to heritage held in libraries and archives. However, the essential prerequisite for this is the conversion of physical artefacts into digital formats. That is why KADOC invests in systematic digitisation of periodicals.

Digitisation priorities at KADOC are determined by various criteria, including the availability of digital copies elsewhere. If a copy is already accessible in another library, KADOC will not undertake its digitisation. The heritage library of KADOC houses around 14,000 different historic periodicals, many of which are unique specimens. However, some are also held in other libraries as extensive series or as single issues. Each of these libraries has different digitisation priorities and policies, and an issue or series that is not digitised today may be in the future. Consequently, the process of checking whether a title or issue is already available as digital copy elsewhere needs to be conducted repeatedly, consuming a significant amount of time.

During the cultural heritage project *Journals in the Picture. Digital disclosure of*

periodicals via Wikimedia platforms and preparation for future digitisation — funded by the Flemish Government and EFRO between 2021 and 2023 — KADOC explored how to automate this verification process.¹⁸⁵ A workflow was developed with the aim of enriching the catalogue by adding references to digital copies housed elsewhere. This enrichment initiative also assists researchers in locating relevant digital copies that are not held in KADOC. Throughout the project, it became evident that there is a significant lack of accessible and reusable bibliographic data in Flanders. Initially, the scarcity of open access bibliographic datasets posed a challenge to data collection efforts.

The periodical collection at KADOC primarily comprises issues printed in Belgium during the twentieth century. Consequently, large international initiatives such as Google Books or OpenLibrary offered only limited relevant information. Similarly, aggregation platforms like Unicat (Union Catalogue of Belgian Libraries) and Antilope (Belgian union catalogue for periodicals), which contain collections that are similar to the KADOC periodical collection, could not immediately provide pertinent information about available digital copies. The Abraham dataset, a union catalogue of Belgian newspapers, published all the necessary information in open access but focused solely on newspapers.¹⁸⁶ This represented only a small subset of the 14,000 titles in the collection that required enrichment with references to digital copies housed elsewhere.

As a result, KADOC had to approach individual libraries across Belgium. While some Flemish libraries had published their entire datasets in open access, they were not as reusable as anticipated. The downloadable file format, specifically a MARC sequential file, was difficult to work with, and the desired information concerning digital copies was often missing from the set. Although all libraries would eventually provide an export of their catalogue upon request, this required additional steps and manual effort, while the objective of KADOC

185. *Journals in the picture. Digital disclosure of periodicals via Wikimedia platforms and preparation for future digitisation* [online]. Leuven (Belgium): KADOC-KU Leuven. [consulted on 28/03/2024]. Available on: https://kadoc.kuleuven.be/8_projecten/2022/tijdschriften_picture.

186. *Abraham. Catalogus van Belgische Kranten* [online]. Antwerp (Belgium): Vlaamse Erfgoedbibliotheken. [consulted on 28/03/2024]. Available on: <https://krantencatalogus.be>.

was to automate as much of the process as possible. For our specific needs, it would have been beneficial if the bibliographic data were published in open access in a usable format compliant with international standards and ontologies like MARC21, BIBFRAME, or FRBRoo, or if they were accessible via an API. These recommendations were disseminated, and all institutions involved in the project will incorporate them into their roadmap for future infrastructure development.

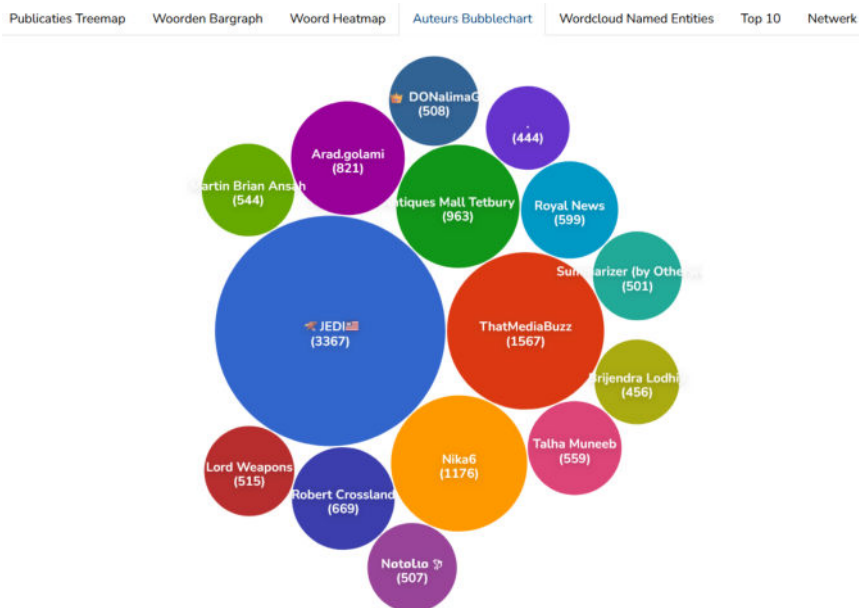


Fig. 1. Spot the Bots. iCANDID Visualisation of main authors represented in Twitter dataset #AbolishTheMonarchy’.

Achieving FAIR access to extensive data about its collection is an ongoing effort at KADOC. To that end, the centre is pursuing affiliation with iCANDID.¹⁸⁷ The iCANDID infrastructure was developed to streamline the big data collection process, and provide researchers with a single access point to aggregate

187. iCandid [Online]. Available on: <https://icandid.t.libis.be/>.

gated collections. This Research Infrastructure (RI), funded by FWO, started in 2018 with a focus on printed, digital, and social media, with the initial large datasets of Flemish press data and Twitter social media data. The current project, iCANDID 3.0 – SSH FAIR Data Hub, extends the RI to encompass GLAM collections. As such the KADOC collections, beginning with the digitised parish journals and poster collections, are being integrated into the iCANDID data hub. While iCANDID offers features such as searching and filtering collections and downloading results, it is not a digital repository for heritage. Digitised issues are ingested and preserved in Teneo, the digital repository of KADOC-LIBIS. They can be viewed as collections or individual items in the catalogue, or picked up for download using iCANDID.

Additionally, ongoing efforts include the continued collection of social media data. Challenges in this area arise from the access restrictions imposed by big tech companies. While the Twitter API had previously been integrated, new management and the subsequent change of the service as X resulted in the termination of academic API access. In response to the European Digital Service Act (DSA), X has recently introduced a new application form for researchers wishing to access the service. However, this form has significant shortcomings which are already recognised by the European Commission.¹⁸⁸ Other social media providers, such as TikTok and Meta (still in Beta phase), are taking steps to comply with the DSA.¹⁸⁹ The TikTok research API application process is transparent and has proved successful.¹⁹⁰ iCANDID has integrated this and is able to collect TikTok data for the researchers granted access.

188. *Twitter moderation research consortium application* [online]. (US): X. 2024. [consulted on 25/03/2026]. Available on: https://blog.x.com/en_us/topics/company/2022/twitter-moderation-research-consortium-open-researchers; *Commission opens formal proceedings against X under the Digital Services Act* [online]. Brussels (Belgium): European Commission. 18 December 2023. [consulted on 28/03/2024]. Available on: https://ec.europa.eu/commission/presscorner/detail/en/IP_23_6709.

189. CLEGG, Nick. *New Tools to Support Independent Research* [online]. (US): Facebook. 21 November 2023. [consulted on 28/03/2024]. Available on: <https://about.fb.com/news/2023/11/new-tools-to-support-independent-research/>.

190. *Research API* [online]. TikTok. 2024. [consulted on 25/03/2026]. Available on: <https://developers.tiktok.com/doc/research-api-get-started>.

The Data Normalisation and Enrichment Process

As a Data Hub infrastructure, iCANDID normalises all aggregated data using a Schema.org-based data model, which has proven sufficiently rich to encompass various data types from media, social media, GLAM, and other online sources requested by various research groups. If necessary, the application profile can be expanded to include other ontologies. The platform enables researchers to search for and access data from various sources through a single query and to export the data in a standardised format of their choice, facilitating easy analysis with domain-specific tools and resulting in substantial time savings. The integration of KADOC's data into iCANDID makes it easier for researchers to access and work with bulk data from the archive, as well as to combine it with various other sources available via iCANDID.

iCANDID also incorporates artificial intelligence (AI) and machine learning (ML) processing, such as machine translation, OCR, handwritten text recognition (HTR), named entity recognition (NER), Speech-to-Text, and image recognition. Initial tests using Google AI and ML services on KADOC's periodicals, poster collections, and photo album collections yielded mixed results. For the periodicals, the outcomes were consistent with previous tests conducted as part of the *Nieuwe Tijdingen* project,¹⁹¹ with illogical segmentation of the periodicals resulting in poor results.¹⁹² Improving the segmentation would require the development of additional training material, but the collection's diversity is such that it is uncertain whether such development would lead to significantly better results. The OCR and HTR processes performed on the poster and photo album collections rendered usable results. However, the image recognition was too generic to be of practical use, identifying animals but not the specific category, for example. This issue could easily be mitigated by creating a training set that maps to a predefined classification. Since this improvement requires the involvement of domain experts to define the classification and a workforce to create ade-

191. <https://www.bibliothecaierfergoed.be/projecten/nieuwe-tijdingen>.

192. ROCHMES, Sophia, DENYS, Montaine & COPPOOLSE, David. Nieuwe OCR voor bestaande digitale krantencollecties: de moeite waard? *META Tijdschrift voor bibliotheek & archief*. 2023, n° 3, p. 10-16.

quate training material representative of the dataset, its implementation depends largely on KADOC's future strategy regarding the use of machine learning to enrich the collections (a strategy yet to be developed).

Part of this strategy involves determining how to incorporate and present the enriched data to users, particularly as it is important for data integrity that a distinction be made between original and curated data and the results of automatic enrichments. This distinction is important because the enrichments depend on specific algorithms that may produce biased results and are likely to improve over time. In iCANDID, a clear distinction between the original data (from the provider) and the enriched (derived from ML processing) is maintained by implementing JSON nested objects. These nested objects contain the enriched data and indicate the provenance (e.g. Google AI Cloud Translation). The enrichments are both discoverable and available in the JSON-LD export formats. Due to the complexity of these records, it is not possible to provide the same information in the CSV export format. This also means that users who wish to work with datasets including the enrichments must be capable of understanding and processing JSON-LD. New features are being planned to enhance accessibility, such as making it easier to include or exclude enrichments from the search results.

The enrichments of the library catalogue, in the aforementioned Journals in the Picture project, were not integrated in iCANDID but were incorporated directly within the catalogue itself to ensure visibility and accessibility for users. Given that iCANDID was designed to facilitate data normalisation, this normalisation process presented significant challenges, particularly in reconciling bibliographic data sourced from various catalogues in multiple formats. Interlinking periodicals was even more challenging due to their inherent complexity. Firstly, periodicals can span over a century and titles may evolve over time. This means that there are datasets with early titles that need to be matched to datasets containing subsequent titles. Some libraries register these as part of the same record or series, while others create multiple records for these subsequent titles. Secondly, the structure of periodicals presents additional challenges because a title consists of multiple issues. Information about digitisation can be included at the individual issue level or can be found at the overarching title level.

Furthermore, few paper series are complete within a single institution, and few libraries have digitised complete series of any periodical to date. This results in a multitude of datasets that show disparities in the granularity of information and the comprehensiveness of series coverage. In order to keep the project feasible, KADOC decided to match titles and include links to digitised issues at the bibliographic entry level, focusing on the title as a whole rather than on individual issues.

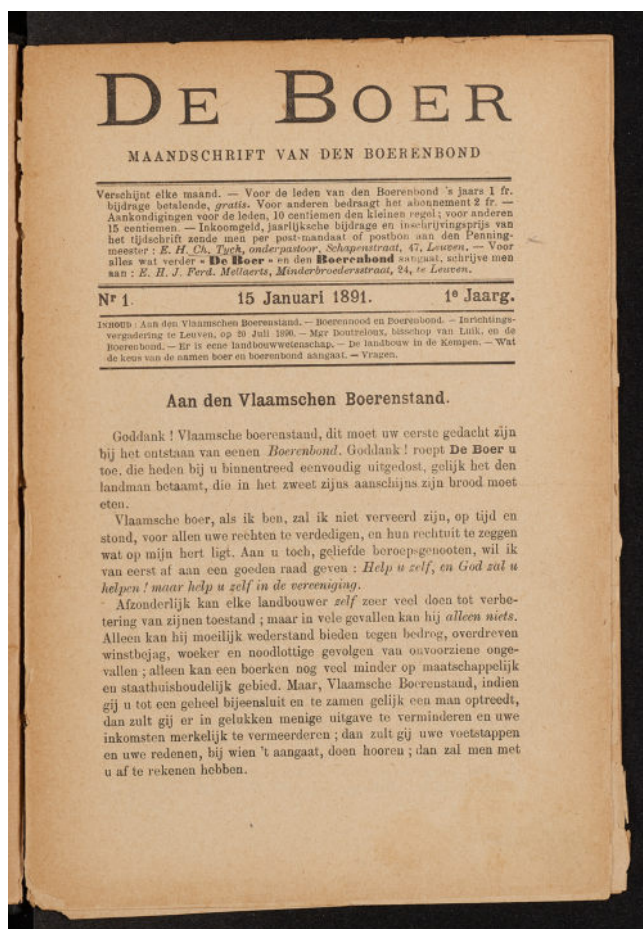


Fig. 2. First page of the first edition of *De Boer* (1891), Catholic periodical for farmers.

Ideally, all received exports should include a shared unique identifier to unambiguously identify a title and link it to an item in the KADOC collection. To address this need, the ISSN (International Standard Serial Number) was introduced and implemented in 1973 as an ISO standard for the unique identification of (modern) serials.¹⁹³ A new ISSN number is generated with each title change. For example, *De Boer: orgaan van de Belgische Boerenbond*, published under this title from 1966 to 1975, has the ISSN 0006-5595. After the title changed to *De Boer en de tuinder: weekblad van de Belgische Boerenbond* in 1976, it was assigned the ISSN 0772-7054. Periodicals published after 1973 are assigned this number upon deposit at the Royal Library of Belgium (KBR). Historic serials printed in the 19th or early 20th century can obtain this number through the local ISSN agency.¹⁹⁴ In Belgium, KBR manages this process for free, but it is not tailored for bulk registration. Libraries can alternatively use the ISSN Portal to access and import ISSN numbers, but this service is not free and is therefore not adopted by many GLAM institutions. As a result, a significant number of historic serials lack an assigned ISSN number.

Fortunately, another regional identifier, one created to aggregate periodicals held in KBR, university, research and heritage libraries in Belgium into a union catalogue called Antilope, was included in all datasets. This Antilope ID is assigned and managed by Anet, the library service of University of Antwerp.¹⁹⁵ New record titles imported into this catalogue that lack an Antilope ID are assigned one based on Antilope's guidelines. If a record already exists, a duplicate record is generated. Anet periodically cleans up these duplicates. To meet the project objectives, automating manual checks using the Antilope ID proved to be the most efficient method for matching the majority of the titles. A more labour-intensive approach to matching registered titles and subtitles was explored, but it yielded less comprehensive and lower-quality results. Around 40% of the

193. *The International Centre for the registration of serial publications – CIEPS* [online]. Paris (France): ISSN. 2024. [consulted on 28/03/2024]. Available on: <https://www.issn.org/the-centre-and-the-network/our-mission/the-international-centre-for-the-registration-of-serial-publications-cieps/>.

194. *ISSN* [online]. Brussels (Belgium): KBR. 2024. [consulted on 28/03/2024]. Available on: <https://www.kbr.be/en/issn/>.

195. *Anet projecten* [online]. Antwerp (Belgium): UAntwerpen. Since 2026: Wander. [consulted on 25/03/2026]. Available on: <https://www.wander.be>.

set was correctly matched, while the remaining 60% required manual verification with this last approach.

```

<?xml:namespace prefix="x" nsURI="http://www.w3.org/1999/xhtml" />
<string MC="0.20333333313465118" CONTENT="nu" HEIGHT="52" WIDTH="135" VP05="1832" HP05="1271"/>
<SP MIDIII="289" VP05="832" IP05="40"/>
<string MC="0.82499998897987164" CONTENT="15" HEIGHT="38" WIDTH="49" VP05="1841" HP05="1888"/>
<SP MIDIII="25" VP05="840" IP05="718"/>
<string MC="0.4699999988879871" CONTENT="1januari" HEIGHT="43" WIDTH="182" VP05="1837" HP05="764"/>
<SP MIDIII="27" VP05="837" IP05="947"/>
<string MC="0.8238888147819516" CONTENT="1891." HEIGHT="46" WIDTH="113" VP05="1835" HP05="975"/>
<SP MIDIII="118" VP05="812" IP05="1208"/>
<string MC="9." CONTENT="1" HEIGHT="38" WIDTH="35" VP05="1812" HP05="1229"/>
<SP MIDIII="42" VP05="822" IP05="1265"/>
<string MC="0.55888881102093865" CONTENT="11" HEIGHT="37" WIDTH="17" VP05="1834" HP05="1368"/>
<string SYLL="superscript" ML="0.2599999986925694" CONLRI="0" ILLIGI="26" MIDIII="16" VP05="817" IP05="1325"/>
<SD WIDTH="17" VP05="831" HP05="1346"/>
<string MC="0.2566666895444121" CONLRI="Jaong" ILLIGI="47" MIDIII="150" VP05="831" IP05="1364"/>
</TextLine>
<?xml:namespace prefix="x" nsURI="http://www.w3.org/1999/xhtml" />
<string MC="0.53166669889162475" CONTENT="TINHOLO" HEIGHT="49" WIDTH="160" VP05="931" HP05="1861"/>
<SP MIDIII="9" VP05="832" IP05="378"/>
<string MC="9." CONTENT="1" HEIGHT="38" WIDTH="34" VP05="936" HP05="1880"/>
<SP MIDIII="8" VP05="828" IP05="385"/>
<string MC="0.469999998642372131" CONTENT="Aan" HEIGHT="49" WIDTH="55" VP05="929" HP05="1894"/>
<SP MIDIII="8" VP05="827" IP05="438"/>
<string MC="0.50866668583883789" CONTENT="aan" HEIGHT="48" WIDTH="58" VP05="927" HP05="1880"/>
<SP MIDIII="18" VP05="827" IP05="389"/>
<string MC="9." CONTENT="1" HEIGHT="42" WIDTH="158" VP05="924" HP05="1828"/>
<SP MIDIII="11" VP05="824" IP05="582"/>
<string MC="0.53758889384185791" CONTENT="Roemenland." HEIGHT="42" WIDTH="178" VP05="922" HP05="1891"/>
<SP MIDIII="8" VP05="831" IP05="1878"/>
<string MC="9." CONTENT="1" HEIGHT="38" WIDTH="34" VP05="931" HP05="1888"/>
<SP MIDIII="14" VP05="822" IP05="913"/>
<string MC="9." CONTENT="Roemenland" HEIGHT="38" WIDTH="153" VP05="918" HP05="928"/>
<SP MIDIII="12" VP05="818" IP05="1892"/>
<string MC="0.71499997377339583" CONTENT="we" HEIGHT="46" WIDTH="31" VP05="923" HP05="1805"/>
<SP MIDIII="8" VP05="818" IP05="1127"/>
<string MC="0.82454545492894267" CONTENT="Roemenland." HEIGHT="42" WIDTH="169" VP05="916" HP05="1158"/>
<SP MIDIII="12" VP05="826" IP05="1396"/>
<string MC="0.55888881102093865" CONTENT="1" HEIGHT="42" WIDTH="119" VP05="938" HP05="1819"/>
<SP MIDIII="16" VP05="816" IP05="1318"/>
<string MC="9." CONTENT="1" HEIGHT="38" WIDTH="153" VP05="914" HP05="1856"/>
</TextLine>
<?xml:namespace prefix="x" nsURI="http://www.w3.org/1999/xhtml" />
<string MC="0.50636364221572979" CONLRI="vorgeleding" ILLIGI="26" MIDIII="158" VP05="911" IP05="1295"/>

```

Fig. 3. First page of the first edition of De Boer (1891), *Catholic periodical for farmers*. Extract of the OCR text in Alto XML format.

This raises questions about how researchers and institutions can address interoperability challenges. Both Antelope and ISSN are not fully accessible for reuse; either a fee is required or an export needs to be requested to use the available data in bulk. In recent decades, new platforms have emerged, each using their own identifiers. This has resulted in many systems developing sets of unique identifiers to reference the same item. To advance linked open data and promote FAIR access to GLAM data, it would be beneficial to leverage free and open hubs like Wikidata to centralise all these identifiers.

Open Data and FAIR Access

After consulting with the Wiki community, KADOC took an initial step by adding all bibliographic data related to its periodical collection in Wikidata and ensuring the data was as LOUD (Linked Open and Usable Data) as possible.¹⁹⁶ However, it is important to distinguish between Open Data and FAIR as two distinct concepts. Making data Findable, Accessible, Interoperable, and Reusable does not necessarily mean the data can be used without conditions, such as GDPR and third-party rights.¹⁹⁷

For bibliographic data, there are no restrictions. However, as iCANDID also aggregates data from various commercial platforms for research purposes (including press, social media, and online databases), it must adhere to the terms of service of these platforms when extracting data via API.¹⁹⁸ If scraping is used, text and datamining regulations apply.¹⁹⁹ In both scenarios, the data must be protected and disclosed according to the specific conditions set for research or heritage institutions.

GLAM collections often contain content protected by copyright, other third-party rights, and GDPR. Only documentary sources older than 140 years can be safely published as open data. iCANDID has therefore implemented Identity and Access Management (IAM) to ensure controlled access to the environment and the datasets it contains. The collection registry informs researchers about the data collections available in iCANDID and what they already have access to, and enables them to apply for access to restricted collections. Organisations like KADOC can thus determine who gains access to sensitive

196. WEYNS, Katrien. Kom dat zien! Tijdschriften nog even in the picture. *KADOC-e-Nieuwsbrief* [online]. 2024, n° 1. [consulted on 28/03/2024]. Available on: https://kadoc.kuleuven.be/6_nieuwsbrieven/enbr/enbr_2024_01.

197. *FAIR principles* [online]. Hamburg (Germany): GO FAIR. 2024. [consulted on 28/03/2024]. Available on: <https://www.go-fair.org/fair-principles/>.

198. *TikTok Research API Terms of Service* [online]. TikTok. 2024. [consulted on 28/03/2024]. Available on: <https://www.tiktok.com/legal/page/global/terms-of-service-research-api/en>.

199. *Europese richtlijn over auteursrechten en naburige rechten in de digitale eengemaakte markt – omzetting in Belgisch recht* [online]. Brussels (Belgium): FOD Economie. 27 September 2024. [consulted on 28/03/2024]. Available on: <https://economie.fgov.be/nl/themas/intellectuele-eigendom/intellectuele-eigendomsrechten/auteursrecht-en-naburige/auteursrechten/europese-richtlijn-over>.

data or copyrighted materials and draft the necessary agreements.

The IAM of iCANDID enables researchers to log in using their university credentials. For those whose organisation is not a part of the EduGAIN identity federation, support for other identity providers (IdP) such as ORCID or Google login will soon be available.²⁰⁰ An application form is provided for users to describe their research case and their need for access to a specific dataset. The collection holder can then evaluate and grant access accordingly. While this level of protection is necessary to ensure legal compliance, it does result in additional costs at both the organisational and infrastructural levels. Research infrastructure initiatives such as iCANDID provide scalable solutions for a broader network of partners. Implementing such technology individually would pose challenges due to the technical complexities of setup and continuous maintenance. The evolving practices of IAM and IdP, which increasingly incorporate security measures such as two-factor authentication to prevent unauthorised access, underscore the advantages of a centralised approach.

The FAIR data principles, which originally stemmed from a research context, are now increasingly recognised as best practices in data management and publication within the GLAM sector.²⁰¹ While it can be argued that the sector has long prioritised making their collections findable and accessible, the aspects of interoperability and reusability, especially in a machine-actionable manner, have been lagging behind. FAIR introduces a renewed emphasis on disclosing collections as data, shifting from object-centric visualisation and single-record downloads to providing access to large datasets through bulk downloads and APIs. Platforms such as iCANDID enable users to download information in their preferred format for easy processing and analysis in research-specific environments.

There is a misconception that researchers always seek the most semantically rich formats, such as RDF (RDF/XML, N-Triples, etc.) or JSON-LD. In the case of iCANDID, the most commonly used export formats are CSV and text.

200. EduGAIN Entities [online]. *GEANT*. 2024. [consulted on 28/03/2024]. Available on: <https://technical.edugain.org/entities>.

201. *Publishing GLAM data as FAIR data* [online]. Den Haag (Netherlands): Europeana Foundation. 2024. [consulted on 28/03/2024]. Available on: <https://pro.europeana.eu/event/publishing-glam-data-as-fair-data>; <https://arxiv.org/abs/2304.02603>.

Of course, the choice of format depends on the specific use case. CSV has its limitations, such as the inability to structure complex (hierarchical), interlinked (linked data), or enriched (as nested objects to differ from the original) information. JSON-LD and API are also preferred due to their relative simplicity compared to RDF and SPARQL, which require a higher level of technical expertise.²⁰²

All formats and endpoints have their advantages, but from an infrastructural perspective, triple stores and SPARQL endpoints demand more technical resources (skills, hardware, software) for setup and maintenance. When advocating for specific technologies and formats, funders should consider these factors. Ultimately, having access to a well-documented and standardised data export in a format that can be easily used by various audiences (heritage professionals, researchers, and civilians) and is compatible with popular tools should be prioritised.

The experimental setup of Linked Open Data infrastructure, including triple stores and SPARQL endpoints, has been a focal point in numerous cultural heritage projects. However, years after these experiments, only a few productive environments have been sustained. It is challenging to maintain these environments in a productive manner, especially when funding for such projects has ended. The broad applicability of JSON and API has made them more popular alternatives. Efforts to interlink data by gathering identifiers and enriching data on data hubs like Wikidata can facilitate interoperability.

iCANDID operates with specific use cases, focusing on providing access to various data collections, including GLAM, for research purposes. The infrastructure is built following the ‘API first’²⁰³ and ‘FAIR by design’ principles.²⁰⁴ One significant advantage is the system’s capability to convert data into different export formats. This flexibility allows KADOC data can be transformed to oth-

202. RDF AND JSON-LD UseCases [online]. *W3C*. 15 September 2024. [consulted on 28/03/2024]. Available on: https://www.w3.org/2013/dwbp/wiki/RDF_AND_JSON-LD_UseCases.

203. TRIELOFF, Lars. Three Principles of API First Design. *Adobe Tech Blog* [online]. (US): Adobe. 2 June 2017. [consulted on 28/03/2024]. Available on: <https://blog.developer.adobe.com/three-principles-of-api-first-design-fa6666d9f694>.

204. *Fair by Design*. Blog [online]. (Netherlands): UNLOCK. 2024. [consulted on 28/03/2024]. Available on: <https://fairbydesign.nl/>.

er standards when necessary and made available via API.

However, challenges persist, particularly concerning the diverse types of supported-use cases and the resulting complexity of data handling. Another issue is the volume of data. While it is relatively straightforward to provide researchers with extensive exports of textual data, handling images and videos places a greater strain on the system. Strategies need to be developed to manage such requests effectively, while limited resources necessitate practical decisions, leading to the possibility that large sets of image data, for example, would be processed in a different manner. Collaboration and continued investment are essential to ensure the long-term functionality of advanced research infrastructures such as iCANDID. Business plans are necessary to achieve sustainability once project funding has ended. Altogether, there are numerous factors to consider when providing access to open and FAIR collections as data.

Conclusion

The growing demand for digital data is prompting heritage institutions and research infrastructures to explore new approaches to the large-scale digitisation of analogue media. Among other things, this paper highlighted how linked open data could aid in supporting these digitisation processes. Data normalisation is a labour-intensive task. In a semantic web where shared identifiers can serve as nodes, this workload could be significantly reduced.

However, it is important to recognise that linked open data is not a panacea for providing data to researchers or heritage professionals. Many researchers lack the necessary skills to work with semantic data and often prefer simple text files for their research needs. Similarly, heritage professionals often lack expertise and may find that a combination of a text file and API services meets their requirements. Moreover, setting up and maintaining a linked open data infrastructure presents its own set of challenges. The future of the semantic web for heritage and research data, as envisioned by pioneers like Tim Berners-Lee and funding entities such as the Flemish Government, remains a matter of speculation.

7. News from Around the Church. The Participatory Digitisation and Contextual Disclosure of Local Parish Journals

Joris Colla

Abstract

The parish journal *Kerk en Leven* and its predecessors were/are among the most widely read periodicals in Flanders. They are valuable historical sources for a variety of research questions. KADOC, the Documentation and Research Centre on Religion, Culture, and Society of KU Leuven, aims to ensure their proper preservation and accessibility. Since 2014, the centre is carrying out the project *News from Around the Church*, which focuses on the digitisation and disclosure of collections of local parish journals held throughout Flanders and Brussels. Its goal is to create a chronologically and geographically representative digital collection which may serve as a source for innovative historical research. A five-step workflow is applied. It involves (1) identifying interesting collections, (2) concluding temporary deposit agreements, (3) scanning collections, (4) ingesting the digital copies in the digital repository Teneo, and (5) disclosing the digital copies. The digital collection in Teneo is linked to information clusters on the historical development of the parish landscape in ODIS, an online database on the history of civil society. As a result, users consulting the local parish journals can easily access historical contextual information on parishes and on other topics covered in the parish journals, such as the parish clergy, church buildings, processions, etc. The current *ODIS renewal project 2022-2027* (FWO) will offer new possibilities to explore these information clusters and strengthen the contextual disclosure of KADOC's digital collection of parish journals.

Author

Joris Colla (Hasselt, 1983) studied History and International Relations at KU Leuven. Since 2009, he works at KADOC-KU Leuven. Since 2025, he is head of Collection Management of ODIS and heritage expert specialised in Catholic heritage, educational heritage, and intangible heritage. He also coordinates the *ODIS renewal project 2022-2027*. His research focuses on the religious and cultural history of Belgium in the long nineteenth century.

<https://www.kuleuven.be/wieiswie/en/person/00064754>

<https://orcid.org/0000-0003-3826-9593>

joris.colla@kuleuven.be

Keywords

Catholic Church; contextual disclosure; digitisation; local history; parish journals

Parish Journals in Flanders: An Introduction²⁰⁵

The history of parish journals in Flanders begins around 1900. At that time, the Catholic Church viewed modern mass media as an appropriate channel for the dissemination and deepening of faith. A poster of the *Sint-Pauluswerk/ Œuvre de Saint-Paul* ('Saint Paul's Work'), an association for the promotion of the Catholic press in Belgium founded in 1884, shows Désiré-Joseph Cardinal Mercier (1851-1926), primate of Belgium and archbishop of Mechelen from 1906 till 1926, stating that he wished 'to see a press committee formed in every parish'. Indeed, some parishes started their own journals. Additionally, several central initiatives involved the publication of more general journals that also

205. I would like to thank Guido De Bergh, Roeland Hermans, Peter Heyrman, Winand Van Meerbeek, Hanne Vermeiren, Katrien Weyns, and the organisers of the conference *Belgian print media in the age of digitisation. Current challenges of research into the history of media and journalism in Belgium*.

offered space for local parish news. The Norbertine fathers of the Abbey of Averbode played an important role in this respect. From the 1940s on, the myriad of publications offering parish news that had come into existence was gradually rationalised, leading to the appearance of the parish journal *Kerkelijke Leven* ('Ecclesiastical Life'). In 1967, it was rebaptised *Kerk en Leven* ('Church and Life'). The Dominican fathers had a pivotal role in this process of rationalisation. Later, their responsibilities regarding the publication of the journal were taken over by the dioceses.²⁰⁶

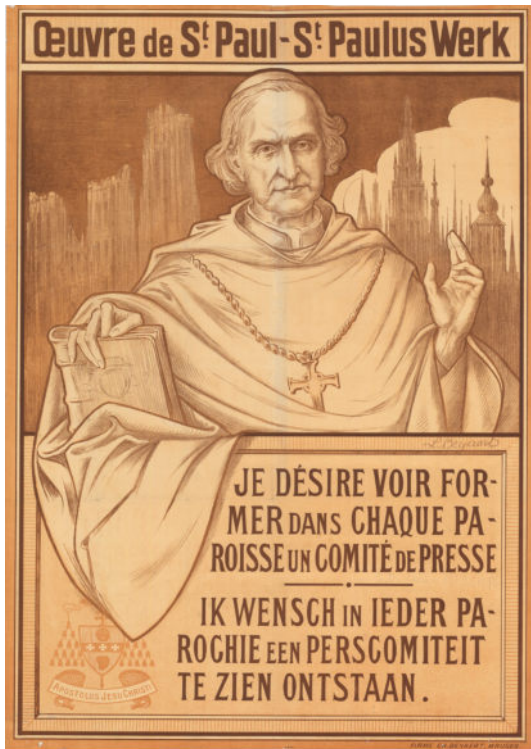


Fig. 1. Poster of the Saint Paul's Work depicting Cardinal Mercier calling for a press committee to be formed in every parish (KADOC-KU Leuven, poster collection, KCB336).

206. On the history of *Kerk en Leven* and its predecessors, see *De Kerk in het midden. Een halve eeuw parochieblad*. Antwerp (Belgium): Uitgeverij Halewijn, 1993; WEYNS, Katrien. Nieuws vanonder de kerktoeren. *META. Tijdschrift voor bibliotheek & archief*. 2021, n° 9, pp. 10-11.

Kerkelijke Leven/Kerk en Leven was and is one of the most widely read periodicals in Flanders. Its peak was around 1980, with 772,754 subscribers. In the decades that followed, that figure dropped, but this weekly parish journal, offering a combination of national, diocesan, and local news, remains an important medium. In 2022, *Kerk en Leven* had nearly 132,000 subscribers, while the number of readers was estimated at 367,000.²⁰⁷

The wide distribution and influence of parish journals in the Catholic community in Flanders makes them valuable historical sources. The content of *Kerk en Leven* and its predecessors reflects the development of the Catholic Church in the Dutch-speaking part of Belgium and demonstrates its vision on a wide variety of religious, social, and cultural topics. The local pages provide a chronicle of parish life and illustrate how these local faith communities operated and evolved. Furthermore, parish journals can be significant sources in other areas of research, providing much material relevant to the history of advertising or, to take a more left-field example, of comic strips. Indeed, the role of these journals in the development of the so-called ‘ninth art’ in Flanders/Belgium should not be underestimated. Comics by several well-known artists, such as Hergé (Georges Remi, 1907-1983), Willy Vandersteen (1913-1990), and Jef Nys (1927-2009), appeared in the parish journal. One of the most popular comic series in Flanders, *Jommeke* by Jef Nys, began as a gag strip in *Kerkelijke Leven*. It was there that the eponymous comic hero with the famous haircut made his first appearance in 1955.²⁰⁸

Given their importance as historical sources, it is crucial that parish journals are preserved appropriately and made accessible for researchers. KADOC, the Documentation and Research Centre on Religion, Culture, and Society of KU Leuven, plays an important role in this respect.

207. *De Katholieke Kerk in België 2023* [online]. S.l.: 2023, p. 88 [consulted on 26/03/2024]. Available on: <https://www.kerknet.be/sites/default/files/Jaarrapport%20Katholieke%20Kerk%202023.pdf>; *De Kerk in het midden*. p. 183; NEIRYNCK, Pieterjan. *Kerk & Leven-uitgever wil 21 mensen ontslaan*. *De Tijd*. 2024, vol. 54, n° 176 (8 September), p. 17; WEYNS. *Nieuws vanonder de kerktoren*. p. 11.

208. HERMANS, Roeland & VINTS, Luc (eds.). *Hemel en aarde. Religie in Vlaanderen sinds 1750*. Leuven (Belgium): Leuven University Press, 2017. pp. 108-109; WEYNS. *Nieuws vanonder de kerktoren*. p. 12. For the first appearance of the comic hero *Jommeke* in *Kerkelijke Leven*, see https://repository.teneo.libis.be/delivery/DeliveryManagerServlet?dps_pid=IE6079877 [consulted on 01/04/2024].

KADOC-KU Leuven and Parish Journals

KADOC was founded in 1976 and is an interfaculty centre of the Human Sciences Group of the University of Leuven. It is recognised as a cultural archive and heritage library by the Flemish authorities and is funded under the *Cultureelerfgoeddecreet* ('Cultural Heritage Decree'). Both in its collections, its fundamental and applied research projects, its public outreach (such as publications and exhibitions), and its public services (reading room, consultancy), KADOC aims to highlight the relationship between religion, culture, and society in Flanders within its Belgian and international context, from 1750 onwards.²⁰⁹

KADOC keeps a more or less complete collection of the general and diocesan pages of *Kerk en Leven* and its predecessors in its heritage library. Since each journal issue contains one or several pages with local news, a fragmented set of local pages is also available. These pages are not, however, collected systematically, not least as lack of space precludes their physical preservation by KADOC. In 2023, for instance, *Kerk en Leven* had approximately 340 local editions (with one edition typically covering several parishes). In the past – when most parishes had their own edition – that number was even much higher (e.g. 1,200 in 1993). Fortunately, digitisation offers a viable alternative.²¹⁰

In 2013, KADOC instigated a digitisation project for parish journals in collaboration with *Kerk en Leven's* publisher, Uitgeverij Halewijn. The collection of parish journals held in the KADOC heritage library – both the complete collection of national and diocesan pages and the fragmented set of local pages – was scanned. This digitisation process was primarily carried out by an external digitisation company.

A second project, also in cooperation with Uitgeverij Halewijn, began in 2014 and is ongoing. News from Around the Church (*Nieuws vanonder de kerktoeren* in Dutch) is focused on the digitisation of collections of local parish journals. These collections are held throughout Flanders and Brussels by various entities,

209. More information on KADOC can be found on the centre's website: <https://kadoc.kuleuven.be/english/index> [consulted on 26 March 2024].

210. *De Kerk in het midden*, p. 183. When referring to 'local parish journals', we primarily mean the local pages of issues that also include national and diocesan news. However, there are also some journals that are entirely locally produced.

including parishes and *kerkfabrieken* (“church fabrics”),²¹¹ local historical associations, and private persons, as well as by other professional heritage organisations such as the Belgian State Archives. Consequently, the project relies on a large network of partners. Its goal is to create a chronologically and geographically representative collection of local parish journals from Flanders and Brussels, which may serve as a source for historical research. Another important objective is the preservation of vulnerable documentary heritage, given that the current realignment of parish structures – which includes the merger of parishes and the closure of churches and presbyteries – can pose serious risks to collections preserved *in situ*.



Fig. 2. First issue of the *Parochieblad van Onze Lieve Vrouw - Halle* (1908), one of the oldest local parish journals in KADOC's digital collection (KADOC-KU Leuven, heritage library, KYE001531).

211. In Belgium, a *kerkfabriek* or *fabrique d'église* is a public entity associated to a parish. It manages the material aspects of worship.

Digitising and Disclosing Local Parish Journals: A Workflow in Five Steps

To implement the News from Around the Church project, a five-step workflow is applied. The first step involves *identifying interesting collections* of local parish journals up to 2012.²¹² This is done in several ways. By making the project widely known, KADOC wants to give as many organisations and individuals as possible the opportunity to participate and submit their collections for digitisation. The project is regularly highlighted by announcements in *Kerk en Leven* and through KADOC's communication channels, as well as those of other heritage organisations. The regional heritage cells (*erfgoedcellen*) in Flanders for instance play an important role in tracing collections from underrepresented regions, which helps to ensure that the digital collection being built is geographically representative. As for locating collections, there is also a close synergy with the KADOC consultancy service for the archives of parishes and *kerkfabrieken*.²¹³ While the centre does not house these archives in its collections, it supports parishes and *kerkfabrieken* in Flanders and Brussels in the management of their documentary heritage, working in close collaboration with the dioceses and the Belgian State Archives. This service has helped to uncover many interesting collections of local parish journals. Furthermore, databases, catalogues, and inventories of archives are systematically screened to track collections. Lastly, the role played by word of mouth ought not to be underestimated.

When the owner of a collection is ready to have it digitised, the next step involves the *conclusion of a temporary deposit agreement*, which defines the modalities of the digitisation. KADOC takes custody of the parish journals for a certain period, typically around three months, and agrees to digitise them at no cost to the owner. Afterwards, the paper originals will return to their owner. While the project does not intend to replace the analogue collections, some owners do not wish to retain their ownership post-digitisation, for example due to lack of space. In these cases, KADOC will work to find a suitable reallocation.

The third step of the workflow involves the actual *scanning of the collection*.

212. The volumes since 2012 are provided digitally by Uitgeverij Halewijn.

213. On this service, see COLLA, Joris. Voor de eeuwigheid. Zorg dragen voor parochiearchieven. *META. Tijdschrift voor bibliotheek & archief*. 2018, n° 2, pp. 26-30.

This takes place in the KADOC Digital Lab in Leuven, which is equipped with a book scanner of professional quality. Since there is no dedicated funding for the News from Around the Church project, the digitisation is carried out by a team of skilled volunteers.

In the fourth step, the *digital copies are ingested* into Teneo, the digital repository with preservation facility used by KADOC. Teneo is part of the Leuven's Integral Archiving System (Lias), developed by LIBIS-KU Leuven. For each scan, both a JPEG and a PDF file are ingested, with an OCR layer added. The digital copies are bundled by journal issue, but the national, diocesan, and local pages (though combined in a single paper journal issue) are presented as separate series. Concurrently with the ingestion process, bibliographical metadata records (including geographical keywords) are created in the library catalogue Alma (Ex Libris).

The fifth and final step is the *disclosure of the digital copies*, which is done via the collection platform Limo. End users can access the full collection of parish journals without prior registration. There are two ways to access the collection. The first is by conducting a search query using Limo's general search function.²¹⁴ The system offers full-text searches, thus facilitating big data research. From the list of search results – which includes both records of entire journals and records of specific journal issues – users can click through to the digital repository. The second way to access the parish journals is through a specific 'curated collection presentation' in Limo, enabling users to browse through the entire collection.²¹⁵ For the local pages, users can call up a list of available titles for each municipality. By selecting a particular title, they can navigate to Teneo, where they first see an overview of available volumes (years of publication) for that title. Clicking on a volume reveals the issues for which a digital copy is available. The digital copies can be opened immediately and downloaded as PDF files.

Fig 3. An issue of the journal of the parish of Our Lady Mediatrix in Turnhout in the digital re-

214. See <https://kadoc.limo.libis.be> [consulted on 29/03/2024]. A search guide (in Dutch) can be found on https://kadoc.kuleuven.be/8_projecten/2021/2021_06_parchiebladen, section 'Hoe op zoek gaan in de digitale collectie parochiebladen?' [consulted on 29/03/2024].

215. https://kadoc.limo.libis.be/discovery/collectionDiscovery?vid=32KUL_KADOC:KADOC&collectionId=815834000001480 [consulted on 29/03/2024].



pository Teneo. In the menu top right links to contextual information in the online database ODIS are included (KADOC-KU Leuven, heritage library, KYE002337).

In local parish journals, one can often find personal data revealing religious beliefs, such as lists of baptisms and marriages. Those data can be seen as a 'special category of personal data', as defined by Article 9 of the European General Data Protection Regulation (GDPR). Therefore, before the digital copies were disclosed, a data protection impact assessment was conducted in collaboration with the Privacy Team of KU Leuven. Since parish journals are publications intended for public dissemination, there appeared to be no objections to their online public disclosure. As far as copyright is concerned, the bonds of trust between KADOC, Uitgeverij Halewijn, and the parish communities responsible for drafting the local parish journals are essential. Indeed, because thousands of parish employees and volunteers (not always unambiguously identifiable) wrote articles over the years, it is impossible to proactively clear all rights.

In early 2024, the digital collection of local parish journals contained approx. 1.1 million scans/pages. For 407 parishes, a digital set comprising at least twenty volumes/years was available. In the coming years, the collection will continue to expand. It is difficult to predict the pace of such expansion, however, as it depends on how many new collections suitable for digitisation can be traced. In any case, special efforts will be made to enhance the geographical representativeness of the digital collection. Concurrently, KADOC aims to promote its use as a source for innovative research. This is done through the channels and in collaboration with the partners mentioned under step 1.²¹⁶

216. An example of the collaboration with regional heritage cells is the article COLLA, Joris. Het Sint-Mauritiusklokje online. *Erfgoedkrant Haspengouw*. 2023, vol. 10, n° 1, p. 7, published in the magazine

Parishes for which a digital collection of local parish journals spanning twenty years or more is available, by actual number and percentage (January 2024)		
Diocese	Number of parishes with a collection of twenty+ years	Percentage of parishes with a collection of twenty+ years ²¹⁷
Antwerp	59	19.7%
Bruges	75	20.7%
Ghent	48	11.2%
Hasselt	85	27.2%
Mechelen-Brussels (Archdiocese) – Vicariate of Flemish Brabant and Mechelen	130	33.4%
Mechelen-Brussels (Archdiocese) – Vicariate of Brussels	10	9.3%
Total for Flanders and Brussels	407	21.5%

The Contextual Disclosure of the Digital Collection of Local Parish Journals: ODIS

KADOC's digital collection of local parish journals is systematically linked to the contextual information clusters on the historical development of the parish landscape in ODIS, a bilingual (Dutch-English) contextual online database on the history of civil society.²¹⁸ ODIS is used by a growing number of heritage and research organisations in Flanders and Brussels. It aims to stimulate cross-fertilisation between the custodians of cultural heritage collections and researchers. The database was set up in 2000-2003 thanks to a grant from the Research Foundation Flanders (FWO). In 2013, a new version was launched. ODIS is administered by the non-profit association ODIS vzw, which includes representatives of the universities of Antwerp (UAntwerpen), Brussels (VUB), Ghent (UGent), and Leuven (KU Leuven). KADOC is responsible for its day-to-day management.²¹⁹

of *Erfgoed Haspengouw*, the heritage cell in the south of the province of Limburg.

217. The percentages are based on the number of parishes extant in 2010, before the major restructurings of the parish landscape began.

218. <https://www.odis.be> [consulted on 29/03/2024].

219. For more information on ODIS, see https://www.odis.be/hercules/_en_overODIS.php [consulted on 27/03/2024]; COLLA, Joris & HEYRMAN, Peter. Het middenveld en zijn erfgoed in context: het veelzijdig gebruik van de onlinedatabank ODIS (2000-2017). In: LEYDER, Dirk and ISTASSE,

The members of the ODIS partnership use the system for the storage, validation, online publication, and analysis of comprehensive data sets on organisations, persons, families, buildings, and events, as well as basic information on documentary heritage (archives, periodical publications), with (deep) links to primary catalogues and digital repositories. End users can approach ODIS as an encyclopedia, a heuristic instrument, an authority database, or an analytical tool. The latter relates to the database's advanced search functions, which are based on the use of controlled vocabularies and thesauri for data input.

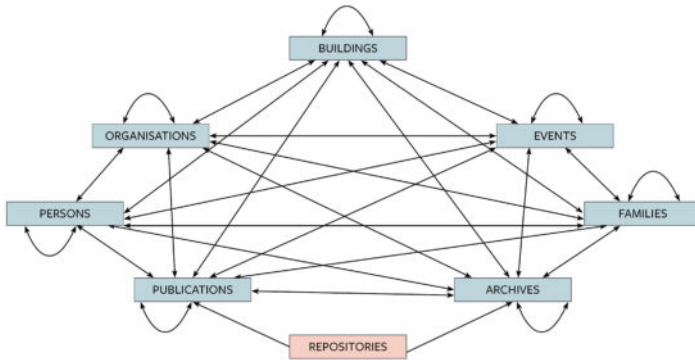


Fig. 4. Entity-relationship diagram (ERD) of ODIS. The database consists of seven interrelated main modules, used to describe organisations, persons, families, buildings, events, archives, and publications (© ODIS νζπν).

ODIS actually contains over 308,000 records, of which nearly 48% is published in the public catalogue (OPAC) and available under a Creative Commons license (CC BY-NC-SA 4.0). The diversity of the partnership results in a data-

Nathael (eds.). *Ten dienste van de gebruiker. Over de zoekinstrumenten ontwikkeld door archivariissen en bibliothecarissen. Akten van het colloquium van 8 december 2017*. Brussels (Belgium): Archief-en Bibliotheekwezen in België, 2019. pp. 99-111; MARYL, Maciej & GRABORSKA, Klaudia (eds.). *How to Facilitate the Cooperation between Humanities Researchers and Cultural Heritage Institutions. Guidelines* [online]. Warsaw (Poland): Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences, 2019. Available on: <https://zenodo.org/records/2587481>. pp. 24-25 [consulted on 28/03/2024].

base content with a broad thematic scope: politics, social organisations, art and culture, national movements, local civil society, education and care, church and religion, etc.

Issues of local parish journals in the digital repository Teneo are systematically linked to the ODIS organisation records of the parishes involved. These records – available for all Catholic parishes in Belgium – are the cornerstone of the contextual data sets on the parish landscape in ODIS. They are linked to other records within ODIS:

— Other organisational records, e.g. of the diocese to which the parish belongs, the deanery, etc.

— Biographical records of the parish clergy, belonging to the *Biographical repertory of the clergy and the members of religious institutes in Belgium, 1750-present*, which is being developed within ODIS. At the moment (2024), the database offers a complete repertory of (deceased) diocesan priests of the archdiocese of Mechelen-Brussels and the dioceses of Antwerp, Bruges, and Ghent. For the dioceses of Hasselt, Liège, and Tournai, data input is ongoing.

— Building records on parish churches and chapels. ODIS contains records on all parish church buildings in the Flemish region and over 14,000 records on chapels in Flanders and Brussels.

— Event records on processions. An inventory of circa 1,000 (disappeared and existing) processions in Flanders and Brussels is available.

— Archival records offering basic information on parish archives, guiding end users to these source materials.

In reverse, deep links to the bibliographical metadata records on local parish journals in Limo are included in the ODIS organisation records of parishes.

By offering links to ODIS records on parishes, users consulting the digital

collection of parish journals in Teneo can easily access historical contextual information on the parish and click through to ODIS records on other organisations, persons, events, etc. covered in the local parish journals.

In 2022, ODIS embarked on a transformative journey. Thanks to a grant for medium-scale research infrastructure from FWO and with LIBIS-KU Leuven as technical partner, an in-depth modernisation of the database is being carried out. A main goal of the *ODIS renewal project 2022-2027* envisions the technical development of a new triple store Linked Data authority database (Virtuoso) and of a new OPAC.²²⁰

The new database will offer its users innovative ways to query, analyse, and visualise the ODIS data sets, including tools for network and geographical visualisation. The development of an interoperability and discovery layer is the cornerstone of a durable open access policy, based on the FAIR principles. It will facilitate more and better connections with other catalogues, research instruments, platforms, and Linked Open Data resources. ODIS will also be equipped with a statistical data warehouse. These developments will offer new possibilities to explore the information clusters on the parish landscape and link them to other relevant data sets, strengthening the contextual disclosure of KADOC's digital collection of parish journals.

220. The supervisors of the renewal project are Prof. Dr. Kim Christiaens (KADOC), Dr. Peter Heyrman (KADOC), and Jo Rademakers (LIBIS). The project is organised into five key areas, focusing on (1) the mission and goals of ODIS, (2) data input and quality care, (3) participation and volunteers, (4) communication, and (5) technical developments. More information on the project can be found on <https://www.odis.be/vernieuwing/index.html> [consulted on 01/04/2024] and in HEYRMAN, Peter. Online tools and instruments supporting the study of transnational elites: the functionalities of the contextual database ODIS. In: CIAMPANI, Andrea and KROLL, Thomas (eds.). *Transnational Encounters? European Elites, International Associations and National States* (1882-1914). Berlin (Germany)/Boston (United States of America): De Gruyter/Oldenbourg, 2025, pp. 117-128.

8. Newspapers in the Mundaneum's Collection

Jacques Gillen

Abstract

This article focuses on the extensive collections of newspapers and periodicals housed at the Mundaneum, a private archival centre in Mons, Belgium. These collections reflect the founders Paul Otlet and Henri La Fontaine's goal of creating a comprehensive repository of global knowledge. The largest collection is that of the International Press Museum, established in 1907. It comprises 1,000 running metres (of which 250 are from the Belgian press), and includes over 100,000 titles from around the world, dating from the 18th century to the mid-20th century. This collection represents a significant portion of the Mundaneum's archives and provides valuable insights into the evolution of the press. Newspapers and periodicals are also present in other of the Mundaneum's collections, particularly those relating to the themes in which the Mundaneum specialises: pacifism, feminism and anarchism. The article touches on the progress of the inventory and digitisation of these collections and how they are being made accessible to the public.

Author

Jacques Gillen is a historian. He is archivist and deputy director of the Mundaneum, a private archive centre in Mons. He is responsible in particular for the collections relating to anarchism and pacifism, as well as the press collections. A scientific collaborator of the *Centre d'histoire et de sociologie des gauches* at the Université libre de Bruxelles, his research focuses on the history of anarchist and pacifist movements in Belgium.

Keywords

Documentation; digitisation; newspapers, periodicals; bibliography

Based in Mons since 1993, the Mundaneum is a private archival centre of the Wallonia-Brussels Federation. It holds no fewer than 6 running kilometres of documents in various formats and types. Because of the encyclopaedic desire of its founders, Paul Otlet (1868-1944) and Henri La Fontaine (1854-1943), to bring together all the knowledge in the world, researchers might expect to find information on virtually any subject and in any medium (archives, publications, documentary files, iconographic documents). As well as the archives of the founders and the collections they assembled, along with the additions made to them over time, the Mundaneum holds a variety of archives, most of which focus on three main themes: pacifism, feminism and anarchism.



Fig. 1. The Universal Bibliographic Repertory, around 1900 (Coll. Mundaneum).

The Mundaneum finds its origins in the International Institute of Bibliography (IIB) founded in Brussels in 1895 by Paul Otlet, considered to

be the father of documentation and one of the precursors of the information sciences and the Internet, and Henri La Fontaine, winner of the Nobel Peace Prize in 1913.

The main objective of the IIB was to set up the Universal Bibliographic Repertory (UBR), designed to build an inventory of all the world's publications on bibliographic cards, regardless of their topic, their subject, date or place of publication, or place of preservation. This repertory, which was included in UNESCO's *Memory of the World* register in 2013, has been lining the walls of the Mundaneum's exhibition space since 1998. The UBR is made up of two main sections, one being onomastic (answering the question 'What did this author write?'), the other being thematic (classified by subject).

From the outset, the IIB attached particular importance to periodicals, considering that they should be treated in a similar way to books. In addition to the onomastic and thematic sections, the RBU therefore included a Repertory of periodical titles. The IIB also endeavours to encourage libraries to develop collections of newspapers and periodicals.

The bibliographical cards in the RBU include basic information such as the author's name, title, place and date of publication, but also, where this information is available, the place where the publication can be found. In a way, the universal character of the UBR, sometimes referred as the 'paper internet', is that of an immense bibliographic database. This is why some consider it to be one of the first search engines in history. Between 1895 and the mid-1930s, thanks to the contributions of a network of bibliographers and librarians from every corner of the globe, Otlet and La Fontaine collected some 18 million records.

Otlet and La Fontaine also developed the Universal Decimal Classification (UDC) for the thematic classification of bibliographic cards. Based on the decimal classification system devised in the 1870s by the American librarian Melvil Dewey (1851-1931), the UDC works by dividing knowledge into 10 classes numbered from 0 to 9 (for example, all works dealing with history will be assigned to class 9). Each class is in turn sub-divided into 10 groups, each group into 10 divisions and each division into 10 sub-divisions, so that the subject of a book can be defined with great accuracy. The advantage of this method lies in the fact that it replaced key words with classification numbers. Doing so did

away with interpretation problems of the key words chosen and any obstacles presented by the language being used, as numbers by definition are universal.

In 1895, with the help of scholars and academics, Otlet and La Fontaine embarked on the task of working up the subject divisions in order to make Dewey's classification - which they felt was insufficient to reflect the world's cultural diversity - universal. To their minds, the CDU was to be a comprehensive summary of knowledge. Its function was to provide an exhaustive definition of the subject of any given publication or document. Thanks to a system of symbols and punctuation marks which was developed to associate numerical elements, the CDU makes it possible to define not only the main subject of a publication or document, but also associated subjects, dates, places, links, information on the type of document, etc.

Otlet and La Fontaine were driven by universalist and pacifist ideals. The idea that knowledge should be accessible to all runs like a thread through all the projects they ran within the IIB or in its wake. They were convinced that knowledge is a major tool (alongside international law) for establishing lasting peace.

With bibliography as their starting point, they gradually extended their scope of activities with the aim of promoting access to knowledge and its dissemination. At the outset of the 20th century, Otlet and La Fontaine were convinced that books were no longer the sole carriers of knowledge, and they set up several documentary units and specialised organisations designed to act as extensions of the IIB, giving rise to the concept of *documentation*. Their ambition was to provide access to all knowledge, regardless of the format in which it had been produced: "[...] the document in all its forms (books, journals, newspapers, photographs, archival documents, scientific and administrative reports, minutes of meetings, industrial publications, etc.) [...] becomes, for science, a source of information and a transmitter of its thoughts: it is its indispensable tool".²²¹

221. « [...] le document sous toutes ses formes (livres, revues, journaux, photographies, pièces d'archives, rapports scientifiques et rapports administratifs, procès-verbaux d'assemblées, publications industrielles, etc.) [...] devient, pour la science, source d'information et transmetteur de sa pensée : il s'affirme son indispensable outil ». OTLET, Paul. *Conférence internationale de bibliographie et de documentation. Mouvement sociologique international*. 1908, vol. 9, n° 4, p. 5.



Fig. 2. Board of the *Encyclopedia Universalis Mundaneum*, 1937 (Coll. Mundaneum).

By way of example, here are some of the extensions created between 1900 and 1910, all of which apply the methods established by the IIB:

— the International Institute of Photography, set up in 1905 with Ernest de Potter (the editor of the *Revue belge de photographie*), with the aim of creating a vast encyclopaedia of images. The IIP's collections of photographs, postcards, posters, glass plates and magic lantern slides form the core of the iconographic collection housed at the Mundaneum. De Potter also set up a *Répertoire iconographique universel*, consisting of folders containing photographs and illustrations on cards or standard-sized sheets;

— the Universal Repertory of Documentation, started in 1907, consisting of thematic (classified in accordance to the CDU), biographical and geographical files containing brochures, press cuttings and extracts from publications;

— the International Press Museum, set up in 1907 to collect all Belgian newspapers and periodicals and at least one specimen of all periodicals published around the world;

— the Central Office for documentation on Women, set up in 1910 by Léonie La Fontaine, Henri La Fontaine's sister, to collect documentation on women and their demands;

— the *Bibliothèque collective des sociétés savantes* ('Collective Library of Learned Societies'), later known as the *Bibliothèque Internationale* ('International Library'): set up in 1906 to bring together the libraries of the scientific institutions based in Brussels.

In 1920, the IIB and all the institutions created in its wake, previously housed in various locations in the centre of Brussels, were brought together in the *Palais du Cinquantenaire* to form the Palais Mondial, later known as the Mundaneum. In the spirit of Otlet and La Fontaine, this complex dedicated to knowledge was to become a global intellectual centre and the nucleus of a vast utopian project: a world city conceived as a universal centre of knowledge devoted to understanding between peoples.

The newspapers and periodicals currently held by the Mundaneum are largely drawn from the collections built up between the creation of the IIB in 1895 and the deaths of La Fontaine and Otlet in 1943 and 1944 respectively. Newspapers and periodicals are present not just in these collections (which were constantly enriched by Otlet and La Fontaine's successors) but also in the various archives acquired by the Mundaneum since it was set up as a private archival centre in Mons.

The largest collection is that of the International Press Museum. This alone represents one-sixth of the Mundaneum's collections, or 1,000 running metres, of which 25% is Belgian and 75% foreign press. It contains more than 100,000 titles of newspapers and periodicals from all over the world, published from the 18th century to the present day, though most of them cover the period from the end of the 19th century to the Second World War.

These collections, whose richness makes it possible to follow the evolution of the press from its beginnings, continue to arouse great interest.

The idea of creating a museum dedicated to the press was born at the International Exhibition of Ancient and Modern Press organised in Brussels in 1893 by the *Union de la presse périodique belge* ('Union of the Belgian Periodical Press' or UPP, founded in 1890) and the *Cercle des collectionneurs de journaux* ('Newspaper Collectors' Circle): "The usefulness of such a museum is indisputable. Who, in the exercise of their profession, has not had to consult Belgian or foreign newspapers? Journalists, writers, politicians and industrialists have all drawn useful information from the 'life of the newspaper', because a newspaper is the mirror of an era: everything is reflected in it and nothing is lost in it. Where can you find such information? The newspaper alone is the true document, because it is written under the impression of the moment".²²²

A committee was set up to create this museum, "which would collect one issue of each newspaper and be universal, with a section devoted to autographs, portraits of famous publicists, press books, etc".²²³

However, it was not until 1907 that the International Press Museum (IPM) came into being. The decision was taken on the initiative of the journalist and bibliophile Gaston Mertens at the Third Congress of the Periodical Press, held in Spa from 31 August to 3 September.

From the outset, the museum was envisaged as being international in scope: "We believe that the Museum to be created should be universal so that it is possible to describe the importance of the press in different countries as the first

222. « L'utilité d'un pareil musée est incontestable. Qui, dans l'exercice de sa profession, n'a eu à consulter des journaux belges ou étrangers ? Journalistes, écrivains, hommes politiques, industriels, tous ont, dans la « vie du journal » puisé d'utiles renseignements, car le journal est le miroir d'une époque : tout s'y reflète et rien de s'y perd. Où trouvera-t-on de pareils renseignements ? Le journal seul est le vrai document, car il est écrit sous l'impression du moment ».

223. « qui collectionnerait un numéro de chaque journal et qui serait universel et qui comporterait une section de destinée aux autographes, aux portraits de publicistes célèbres, aux ouvrages de presse etc ». MAUS, Octave & MERTENS, Gaston. *Création d'un musée de la presse à Bruxelles. Rapport présenté à l'Administration communale de Bruxelles, par le Comité de l'Union de la presse périodique belge.* (Extrait du n°2 de *La Presse universelle*), Brecht, 1896.

demonstration of the moral and material development of these countries”.²²⁴ This position was closely linked to the personality of Paul Otlet, then vice-president of the UPP (three months later, he became its president, succeeding Octave Maus). The project was fully in line with his willingness to organise knowledge and documentation on a global scale.

The Museum was quickly set up: meetings were held at the *Cercle des collectionneurs de journaux* to discuss the project, and in May 1908, Otlet presented the draft programme and statutes of the Museum,²²⁵ which began its activities in earnest in that same year.

The IPM was administered by a committee made up of members of its constituent associations: the International Institute of Bibliography (to which it was attached and whose premises it occupied), the *Cercle des collectionneurs de journaux* and the *Union de la presse périodique belge*. The chairman was Jean Vanden Broeck, president of the *Cercle des collectionneurs de journaux*, the curator Albert de Fonvent, a member of the same Cercle.

The IPM was intent on collecting all the newspapers and periodicals in Belgium and at least one specimen of all the periodicals published in the world, “preferably the first and last issues as though collecting the initial profession of faith and the last cry of those who disappear carried off by the torment of ideas”.²²⁶ Its wider goals were to draw up a bibliography of the press and of works and periodicals devoted to the press; to compile documentary and iconographic files on questions relating to the press in all its aspects; to establish rules for building up collections of periodicals; to encourage the creation of local collections in collaboration with public administrations; to establish

224. « Nous pensons que le Musée à créer doit être universel pour qu'il soit possible de décrire l'importance de la presse dans les différents pays comme première démonstration du développement moral et matériel de ces pays ». MERTENS, Gaston. *Projet de création à Bruxelles d'un Musée international de la presse. Rapport présenté par M. Gaston Mertens, au nom du bureau de l'Union de la presse périodique belge*. Union de la presse périodique belge. Bulletin officiel mensuel. 1907, vol. 16, n° 10-12, p. 259.

225. Personal papers of Paul Otlet, correspondence relating to the International Press Museum (box 1038).

226. « de préférence le premier et le dernier numéro, de manière à recueillir la profession de foi initiale et le dernier cri de ceux qui disparaissent emportés par la tourmente des idées ». *Le Musée international de la presse. Section de l'Institut international de bibliographie. Notice-catalogue*. Brussels: Institut International de Bibliographie, 1908, n° 108, p. 6.

relations between collectors in all countries, in particular to facilitate exchanges and standardise documentary methods.²²⁷ The IPM would thus be a link in the chain of universal documentation as conceived by Otlet and La Fontaine.



Fig. 3. Photograph of the International Press Museum, 1900-1910 (Coll. Mundaneum).

The Museum was complemented by a library devoted to the press and, from 1911, by a museum displaying specimens from all over the world. It was also involved in compiling a general repertory of the Belgian press, which formed part of the Universal Bibliographic Repertory.

The museum's collection initially consisted of the collection of periodicals held by the International Institute of Bibliography and those in André Warzée's collection. A lifelong collector of newspapers, André Warzée (1816-1898) was a civil servant at the Ministry of Public Works, who devoted himself to the study

227. *Le Musée international de la presse. Recueil publié par l'Association auxiliaire du Musée international de la presse.* 1912, vol. 1, n° 1-2.

of journalism and wrote numerous notes on the state of the press in different parts of the world. He published several works and catalogues, including an *Essai historique et critique sur les journaux belges* in 1845. His press collections and manuscripts were bequeathed to the *Union de la presse périodique* on his death in 1898, and duly transferred to the International Press Museum. They contributed to the creation of the *Répertoire bibliographique général de la presse belge* ('General bibliographical repertory of the Belgian press'), which forms part of the Universal Bibliographic Repertory).

In 1910, the IPM collections included 6,500 items,²²⁸ and were further enriched following the death of Jean Vanden Broeck,²²⁹ who had bequeathed his collection to the Museum. It includes 35,000 specimens of newspapers from all countries and all periods, complete collections of ancient periodicals, more than 400 works relating to the press, notes and manuscripts on the bibliography of the Belgian press and 5,000 duplicate specimens that can be used for exchanges with other institutions.²³⁰

Subsequently, the collections continued to grow, thanks in particular to collaboration with Belgian and foreign libraries, printers, journalists' associations and collectors. By 1912, they contained some 60,000 items,²³¹ which were classified by place of publication and entered into a catalogue comprising four series: by inventory number, by place, by subject (according to the universal decimal classification system) and by alphabetical order.

In 1910, the IPM consolidated its status by taking part in the documentation and press section of the exhibition organised, as part of the Brussels World Fair, by the Union of International Associations (UIA):²³² this section explained

228. 'Musée de la presse', handwritten note, [1910] (Personnal papers of Paul Otlet, box 1038).

229. Jean Vanden Broeck (1841-1910) was a newspaper collector. In 1890, he contributed to the foundation of the Cercle des collectionneurs de journaux.

230. *Le Soir*, 20/12/1910, n° 353.

231. *Le Musée international de la presse. Recueil publié par l'Association auxiliaire du Musée international de la presse*. 1912, vol. 1, n° 6-8.

232. The UIA was founded in 1907 by Paul Otlet and Henri La Fontaine with the aim of facilitating cooperation between international organisations, starting with the field of documentation. The UIA subsequently became Otlet and La Fontaine's 'political' lever when they tried to turn the institutions they had created into an international centre dedicated to science and intellectual cooperation.

how the press works and provided examples of newspapers. At the same time, the second International Congress of the Periodical Press was held (the first having been held in 1894) with the aim of developing relations between periodicals in all countries, and of studying comparatively and from an international point of view all questions connected with the writing, printing, administration, distribution and use of periodicals, while seeking ways and means of achieving international co-operation in this field.²³³

Other exhibitions were held throughout the life of the IPM, including an exhibition on the Hainaut press organised in 1911 at the occasion of the Charleroi International Exhibition, and the first exhibition of periodicals held at the Palais Mondial in 1923 to coincide with the 4th Brussels International Trade Fair, which showcased some 1,000 active periodical publications.



Fig. 4. Advertising poster for the newspaper *La Dernière Heure*, Brussels, s.d. (Coll. Mundaneum).

233. *Deuxième Congrès international de la presse périodique – Bruxelles (Belgique) 24-25-26 juillet 1910. Documents préliminaires* (Personal papers of Paul Otlet, box 1038).

While the Belgian Government had supported the Palais Mondial, where all the institutions and collections founded by Otlet and La Fontaine had been assembled since 1920, it decided, in 1934, to close it down. Although the work continued (notably at Otlet's home), its collections were no longer accessible. When the Palais du Cinquantenaire was requisitioned by the German occupiers in 1941, the collections were moved to the Parc Léopold. From the end of the 1930s, the IPM no longer existed as an institution. Its collections continued to be maintained thanks to the work of Georges Lorphèvre²³⁴ and André Colet,²³⁵ two of Otlet's close collaborators, who ensured the survival of the Mundaneum after Otlet's death. From the early 1970s onwards, the collections were moved several more times to different locations in Brussels, before finally arriving in Mons in 1993.²³⁶

Newspapers and periodicals are also present in other Mundaneum collections, particularly those relating to the three themes on which the Mundaneum specialises: pacifism, feminism and anarchism. The most important of these, after the International Press Museum, is the *Bibliothèque collective des sociétés savants* ('Collective Library of Learned Societies'). This library, created in 1906 and officially inaugurated in 1908, was intended to bring together the libraries of the other scientific institutions based in Brussels around the library of the International Institute of Bibliography. Once again, this project reflected a willingness to organise knowledge and documentation: "Seen from the point of view of scientific societies, the Collective Library is an attempt at concerted work that cannot fail to bear fruit. It responds, in fact, to the tendencies of modern science to organise work more and more and to unify methods in order

234. Georges Lorphèvre (1912-1997) was trained as a librarian. He met Paul Otlet in 1927 at a conference at the Palais mondial and offered to help him on a voluntary basis. He later became his secretary. As Otlet's testamentary executor, he was a key figure of the Mundaneum and succeeded Otlet as president, secretary or representative of Belgium to the International Federation of Documentation, the International Federation of Librarians and Archivists and the International Standardisation Association. He also founded the Belgian Bibliography Commission in 1950.

235. André Colet (1896-1978) was a teacher and a librarian. He taught in several schools in Brussels. In the 1920s, he contributed to the work of the Mundaneum. After the death of Paul Otlet, he became General Secretary of the Palais Mondial-Mundaneum.

236. To find out more about the Mundaneum and its history, see: GILLEN, Jacques, MANFROID, Stéphanie, MONTIGNIE, Aurélie & STRAGIER, Justine. *Le Mundaneum, machine à penser le monde*, Mons: Mundaneum, 2022.

to advance more rapidly along the path of discovery and progress”.²³⁷

Later known as the *Bibliothèque internationale* (‘International Library’), the *Bibliothèque collective des sociétés savantes* brought together the libraries of 62 societies. It included a large number of technical and scientific publications and periodicals from the end of the 19th century and the first half of the 20th century. By 1909 it had already collected 2,472 periodicals.²³⁸

Inventories of the press held at the Mundaneum are still in progress. As far as the International Press Museum is concerned, several inventories have been finalised (namely those of Belgium, France, Italy, Africa, Canada, Central America, and the anarchist, feminist and pacifist presses), while others are in progress at the time of writing (those of Great Britain, United States, Germany, etc.). All available inventories can be consulted in the online catalogue.²³⁹

The Mundaneum’s digitisation policy focuses primarily on pure archive collections (in particular the personal papers of Paul Otlet and Henri La Fontaine). These digitisations are carried out on a one-off basis (such as in response to requests from researchers preparing a publication or an exhibition) or on a more systematic basis (for example, as part of a research project or a digitisation campaign). These digitisations are most often carried out in-house, but, such as in the case of large-scale digitisation campaigns, sometimes carried out by external service providers (these larger projects are financed by the Pep’s plan set up by the Wallonia-Brussels Federation to support heritage digitisation projects in the French-speaking Community of Belgium).

Some 20,000 pages of the press collections are available in digital format (TIFF or PDF/A), some of which can be consulted in the online catalogue using a full-text search facility. Newspapers and periodicals are either digitised in response to specific, one-off requests, or on a larger scale: examples of the latter being the local press of Molenbeek, Dixmude, and Andesse, which were

237. « Envisagée au point de vue des sociétés scientifiques, la Bibliothèque collective est un essai de travail concerté qui ne peut manquer de porter ses fruits. Il répond, en effet, aux tendances de la science moderne d’organiser de plus en plus le travail et d’unifier les méthodes en vue de marcher plus rapidement dans la voie des découvertes et du progrès ». *Bibliothèque Collective des Sociétés Savantes*. Brussels: Institut International de Bibliographie, 1908, n° 89, p. 3.

238. *Union de la presse périodique belge. Bulletin officiel mensuel*. 1909, vol. 13, n° 7-8.

239. <https://catalogue.mundaneum.org/>.

digitised at the request of the municipalities concerned, as well as the Belgian anarchist press and the pre-1830 'Belgian' press held at the Mundaneum, which were the subject of in-house digitisation campaigns. Other titles have been digitised as part of a scientific research project: for example, the Mundaneum took part in the *Press and literature in French-speaking Belgium* project to preserve most of Belgium's literary heritage, which was run by the Université libre de Bruxelles from 2008 to 2013. Over the next few years, the Mundaneum plans to digitise the Belgian feminist press that it preserves.

9. *Historische Kranten*: 20 Years of Digitising and Unlocking Newspapers in the Westhoek Region

Nick Coorevits

Abstract

Twenty years ago, the city of Ypres started an ambitious project called *Gebeugen van Ieper* in which a heritage bank of historical newspaper material was developed. Since then, the heritage bank has expanded and today it provides access to newspaper material from the entire Westhoek region. Today, the heritage database is at a crossroad: the growth of the project and additional challenges are pushing it to its limits. Meanwhile, preparations are being made to prepare the platform for the future.

Author

Nick Coorevits (°1996) holds a Master's degree in Art Sciences (University of Ghent) and a Master's degree in Heritage Studies (University of Antwerp). Since 2021, he has been working as Cultural Heritage Consultant at CO₇.

Keywords

Newspapers; heritage; digitisation; database; OCR

Twenty years ago, the city of Ypres started an ambitious and innovative project to develop a heritage database that made historical newspapers from the Ypres district accessible. The website has grown over the years into a heritage database where more than 70 different newspaper titles from the Westhoek region can be consulted digitally. The growth of the project is now reaching its

limits. To meet today's challenges, the database is being prepared for the future under the coordination of CO₇, the partnership for culture and heritage between the seven municipalities of Heuvelland, Ypres, Langemark-Poelkapelle, Messines, Poperinge, Vleteren and Zonnebeke.²⁴⁰

From *Geheugen van Ieper* to *Historische Kranten van de Westhoek*

The story of *Historische Kranten* ('Historical Newspapers') started on 1 March 2004 when the city of Ypres concluded a heritage covenant with the Flemish Community. The conclusion of the covenant made it possible to set up an *erfgoedcel* or local heritage service that could perform a service role in support of cultural heritage, its managers and the city's cultural heritage communities.²⁴¹ The application included a dozen projects proposed for implementation within the covenant. One of the proposals was the development of this digital heritage database with historical newspapers from the Ypres district, inspired by *Geheugen van Nederland* ('Memory of The Netherlands').²⁴² Funds from the cultural heritage covenant enabled Ypres to realise this ambitious plan.²⁴³

The plan was certainly both ambitious and innovative, even if *Erfgoedcel Ieper* was not the first local heritage service in Flanders to prioritise making heritage available digitally through a database. At the start of *Erfgoedcel Ieper*, other *erfgoedcellen* in Bruges, Ghent and Mechelen were already actively developing a single database which served to make historical images accessible to the public through a dedicated website. But unlike these other image banks, *Geheugen van Ieper* focused on the digital opening-up of the city of Ypres' collection of historical newspapers. The decision of the project partners of *Geheugen van Ieper* to focus on historical newspapers was not accidental. Three arguments contributed

240. <https://www.co7.be/>.

241. An *erfgoedcel* is a local heritage service working for a regio or larger city.

242. <https://geheugenvannederland.nl/>.

243. WUYTS, Eva. *Geheugen van Ieper: de digitale ontsluiting van lokale krantenarchieven*. *Bibliotbeek & Archiefgids*. 2007, vol. 83. n° 3, pp. 23-28.

to the decision to make these collections accessible digitally.²⁴⁴



Fig. 1. Ruins of the belfry and cloth hall of Ypres during World War I. The belfry tower housed the historical archives of the city of Ypres (Source: Westboek verbeeldt, private collection).

First, the city's newspaper collection is and remains a valuable source for historical research on Ypres to this day. During the destruction of World War I, many of the historical resources of Ypres were lost. For instance, the city's archives, which at that time were kept in the belfry tower of the Cloth Hall, went up in flames on 22 November 1914 following German artillery bombardments. The city museum also lost a large part of its collection. Only the city library was able to remove its collection to safety in time. The book and newspaper collection found temporary shelter in France, only eventually returning several years later. Since then, the librarians have kept this collection as treasures. The newspapers are a unique resource providing insights into the social life of Ypres, especially in the 19th century. They highlight many aspects of contemporary daily life contained in town council reports, market prices, sports results, obituaries and other reports.

244. *Ibidem.*

A second reason for digitising historical newspapers was that these sources are highly endangered. The increasing acidification of the material and frequent consultation of the physical newspapers are both pernicious influences that hinder their preservation. In the absence of suitable precautions being taken, this degradation process will eventually lead to the destruction of this heritage.

Finally, *Erfgoedcel Ieper* hoped to be experimental and innovative with the newspaper digitisation project. The search for possible solutions to the project's specific challenges (format and storage capacity, the different levels of disclosure and so on) was in line with this ambition.²⁴⁵

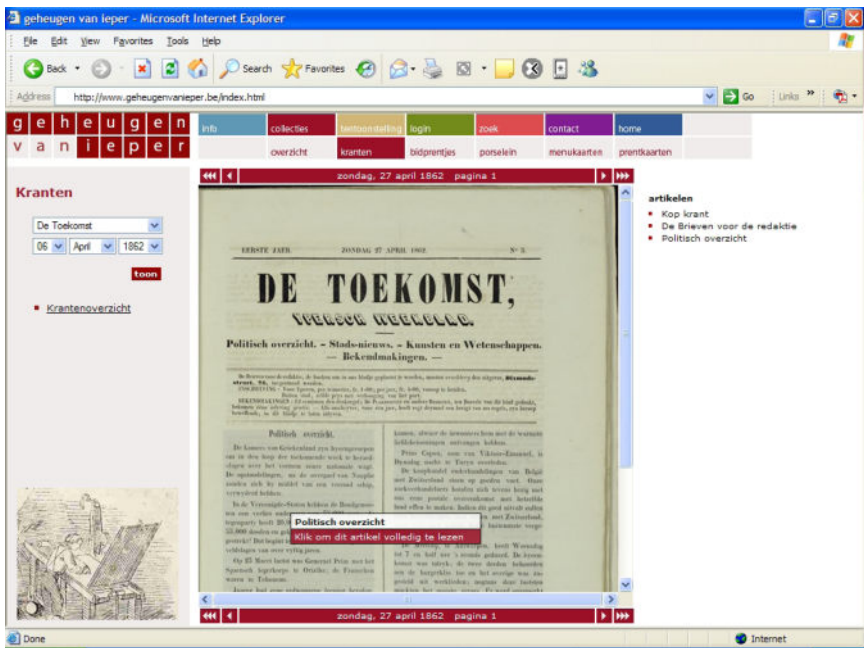


Fig. 2. View of the front page of the newspaper *De Toekomst* of 27 April 1862 on *Geheugen van Ieper*, predecessor of *Historische Kranten* (Source: CO₂).

As *Geheugen van Ieper* began, the municipality of Heuvelland almost simul-

245. Ibidem.

taneously entered into its own initiative called *Heuveland verbeeldt* ('Heuveland depicts') to register visual material in a database. The aim was to digitise historical visual material of the municipality and make it available to a wide audience through an accessible website. The aim was to preserve the daily life of the past for the future. The set-up of the project was simple but very effective: volunteers were to track down potentially valuable visual material from their fellow villagers, digitise and add explanatory material to it. In 2005, *Heuveland verbeeldt* ended up at the newly established intermunicipal partnership *CultuurOverleg Zeven*, or CO₇ for short.²⁴⁶ Heuveland's model was adopted and applied in the other municipalities of the Southern Westhoek. In the period 2005-2007, an image bank operation was launched in the other municipalities of the Southern Westhoek, namely Ypres, Langemark-Poelkapelle, Messines, Poperinge, Vleteren and Zonnebeke. *Heuveland verbeeldt* was renamed *Westhoek verbeeldt* ('Westhoek depicts').²⁴⁷

From the beginning of *Westhoek verbeeldt*, *Erfgoedcel Ieper* supported CO₇ in developing the regional image bank. What was written down in the text from newspapers could now also be visualised with images from the regional image bank.²⁴⁸

Geheugen van Ieper had ambitions to become a database where all kinds of heritage (mainly two-dimensional heritage) could be interconnected.²⁴⁹ This ambition could only be realised from 2009 onwards when a new heritage covenant was concluded with the Flemish Community. *Erfgoedcel Ieper* joined the project association CO₇ and extended its operations to the other municipalities of the Southern Westhoek. Thus, the projects of *Erfgoedcel Ieper* also joined CO₇ and were continued regionally alongside *Westhoek verbeeldt*.

Geheugen van Ieper entered a new phase. In the first year of *Erfgoedcel CO₇*, the heritage bank was renamed *Historische Kranten* ('Historical Newspapers'). Those newspapers already available (then over 44,000 pages) were supplemented with

246. <https://www.co7.be/>.

247. <https://westhoekverbeeldt.be/>.

248. SIMOENS, Tine & THIERS, Liesbeth. 'Westhoek verbeeldt': een buitenbeentje onder de beeldbanken. *FARO: Tijdschrift over cultureel erfgoed*. 2009, vol. 2, n° 1, pp. 28-33.

249. WUYTS, Geheugen van Ieper, p. 24.

new digitised newspapers from the collections of the Ypres City Archives and the Poperinge City Archives. The new website went online on 18 November 2009.

True integration of the two regional databases had to wait until 2014. After the expansion in 2009, problems surfaced with the then provider X-Cago over the hosting of the digital newspapers. The ongoing problems eventually led to the termination of cooperation this company's contract in 2013. To find a new provider, a tender process was launched. The process of putting the job out to tender finally allowed for true integration and a new contract was issued accordingly. Potential suppliers were asked to house the historical newspapers in *Westboek verbeeldt*, and to also make them searchable through the same search function. A collaboration was eventually launched with the Dutch company Picturae, where the image database had previously been housed.

In the meantime, the operation of *Westboek verbeeldt* had also experienced geographical growth. After 2010, collaborations were established in the Northern Westhoek with the intermunicipal collaborations *Hout & Blooteland* and *5-art*. In this way, the operation of *Westboek verbeeldt* expanded to the municipalities of Alveringem, De Panne, Diksmuide, Houthulst, Koekelare, Koksijde, Kortemark, Lo-Reninge, Nieuwpoort and Veurne. In each municipality, a core of volunteers was set up to track down valuable heritage locally.

These new collaborations opened the doors for archival institutions in the North of the Westhoek to join the platform *Historische Kranten*. In 2018, talks were initiated with the city archives of Diksmuide and Veurne, which for the first time expressed interest in making newspapers from their collections accessible through the website. The motives for joining the project were the same as in the early days of *Geheugen van Ieper*. The historical archive of Diksmuide, for example, was destroyed, lost or dispersed because of the First World War. The platform *Historische Kranten* provided an opportunity for the institution to once more house its archive under the same roof, albeit a digital one.²⁵⁰

Due to some difficulties and issues in delivering the digital files, the addition of new newspaper material took rather longer than expected. Eventually, the addition of these newspapers was finalised in 2021. Another major expansion

250. Testimony by the late Chris Vandewalle, former archivist of the Diksmuide city archives.

followed in 2023 with the incorporation of newspaper material from Diksmuide and Poperinge. Even today, the platform is actively adding to its holdings. New additions of newspaper material are being prepared, mainly in the North of the Westhoek.

Respect for Newspaper as Witness to Its Time

At the beginning of the digitisation project, there was consensus among the project partners that digitisation in itself should never be the main goal. For the heritage institutions involved, the optimal preservation of the newspaper material was important, while the local heritage service in turn saw it as its mission to strengthen the understanding of cultural heritage in society by making the newspapers easily accessible. The development of a digital portal met the wishes of all partners. The website removed the need for the original pieces to be consulted physically in the archives and also made it possible for interested parties to consult the newspapers from home.

When digitising newspapers, the vision was to consider newspapers as heritage in their own right and not merely as carriers of text. Approaching the look and feel of the original material was paramount. Newspapers are seen as witnesses of their time and digitisation had to be done with respect for this heritage. The colour and texture of the paper was allowed to be clearly seen, as was any discolouration of the paper, folds, tears and, in rare cases, bullet holes. The original archival experience of consulting these newspapers was to be approximated in the digital editions wherever possible. The project partners strongly believed in the disarming charm of these old newspapers.²⁵¹

This intention had implications on the methodology used for digitisation: 19th-century bindings, for example, were not detached, as they are an integral part of the collection. This was at the expense of the text of the newspapers. Indeed, due to tight binding, certain pieces of text could not be digitised. The newspapers were also digitised in their current state, which meant that gaps were not filled, and folds were respected even if they caused a particular line of text to remain unreadable.

251. WUYTS, *Geheugen van Ieper*, p. 24.

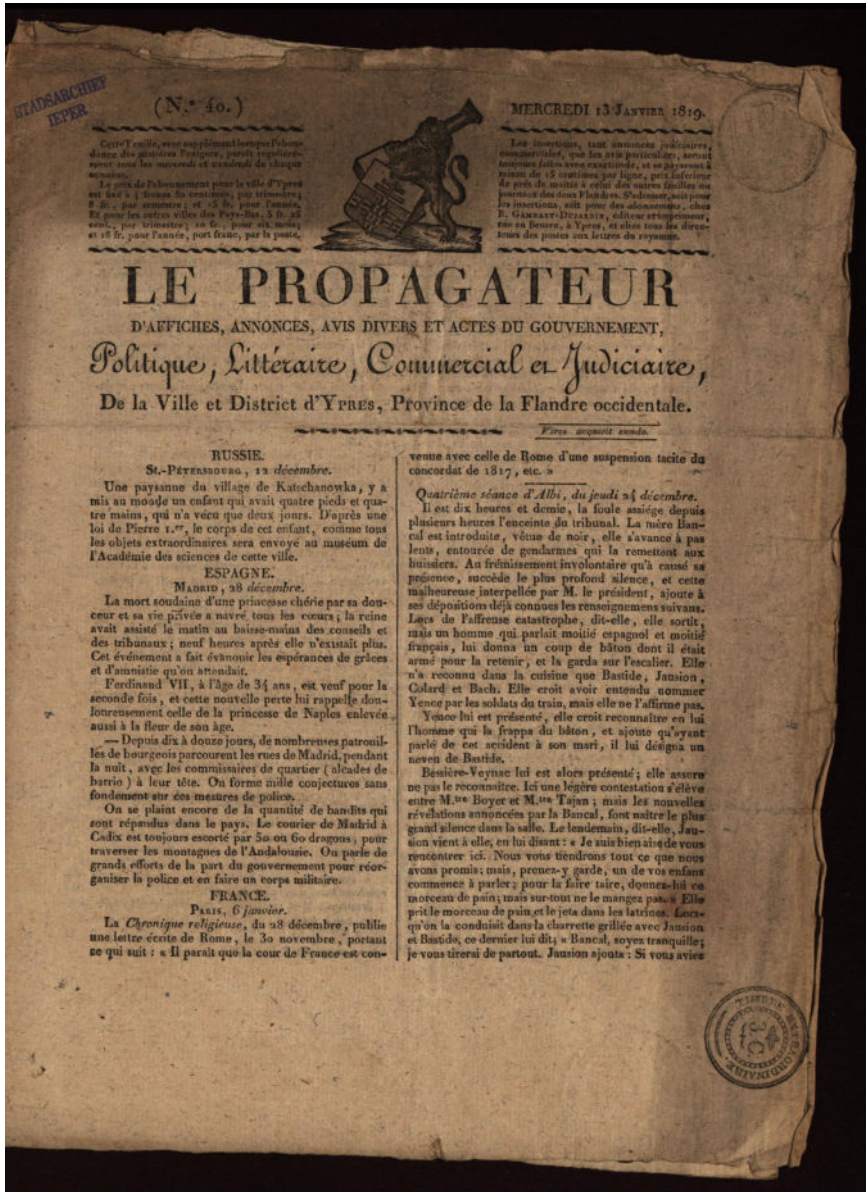


Fig. 3. The newspaper *Le Propagateur* of 13 January 1819 is the oldest newspaper available for consultation on the platform *Historical Newspapers* (Source: *historischekranten.be*).

The digital copies were manipulated as little as possible to preserve the original colour and texture of the newspapers. After all, the digital reproduction had to approach the original material faithfully. Frayed edges were not cut off, but text that was crooked or convex due to the bindings of the newspapers was fixed software-wise.



Fig. 4. Scanner used to digitise newspapers among others at the Ypres City Archives (Source: Ypres City Archives).

Unlike other newspaper digitisation projects at the time, the scanning of newspapers was done from the original print in colour. A conscious decision

was made not to use microfilm digitisation, although this option was cheaper.²⁵²

In the early years, the digitisation of newspaper material was handled differently than it is today. Initially, scanning was done manually at article level with a video scanner. As the methodology had its limits, it was gradually changed to one in which entire newspaper pages were scanned using specialised scanning equipment.

The digital preservation of the files has remained the same over the years. Scans made of newspapers are saved in TIFF format. These master files are kept by the archives themselves. For online display, these files were copied into JPG format. Although initially it was not a goal to make the newspapers searchable, it was eventually decided to use the available technology to facilitate targeted consultations and searches. To make the newspaper pages searchable online Optical Character Recognition (OCR) was applied to the digital files. This application makes the newspapers searchable at word level.

To this day, archives digitise newspaper material themselves as much as possible with available scanning equipment. The city archive in Diksmuide, for example, invested in new equipment in recent years to carry out digitisation in-house.²⁵³ This avoids high costs of outsourcing digitisation to commercial players. However, digitising large newspaper formats is not for every archival institution. For this reason, the Poperinge city archive transferred to the Ypres city archive to digitise larger formats there.²⁵⁴

Today digitisation is done indoors by the archives themselves. However, such work was not always carried out in-house. The first digitisation of the Ypres newspapers was outsourced to Dutch software company X-Cago in 2004. The company was then awarded all four sub-tasks of the project's tender: digitising the newspaper collection, building the database, building the public website and hosting the digital files. To make all of this possible, it was necessary to have a large budget. To stay within budget limits, it was necessary to work only with a limited selection of newspapers. The final selection was based on several criteria: age, language, state of preservation, public interest, diversity of content,

252. Ibidem.

253. Testimony by the late Chris Vandewalle, former archivist of the Diksmuide city archives.

254. Testimony by Tijs Goethals, archivist of the Poperinge city archives.

size and completeness of the series.²⁵⁵

When Poperinge joined *Historische Kranten* in 2009, the digitisation of newspapers was again outsourced to X-Cago, which was then also commissioned to extend the *Gebeugen van Ieper* platform to *Historische Kranten*. This outsourcing once more cost a large sum of money and was only possible by a project subsidy obtained by the City of Poperinge for a project on 100 years of the newspaper *Het Wekelijks Nieuws* ("The Weekly News"). The archive had previously received the entire collection as a donation from the former printing works company Sansen, which was located in Poperinge. Here, as well, choices eventually had to be made to stay within the project's budget, but thanks to the grant, the Poperinge newspapers could be digitised in large quantities and added to the updated portal.²⁵⁶

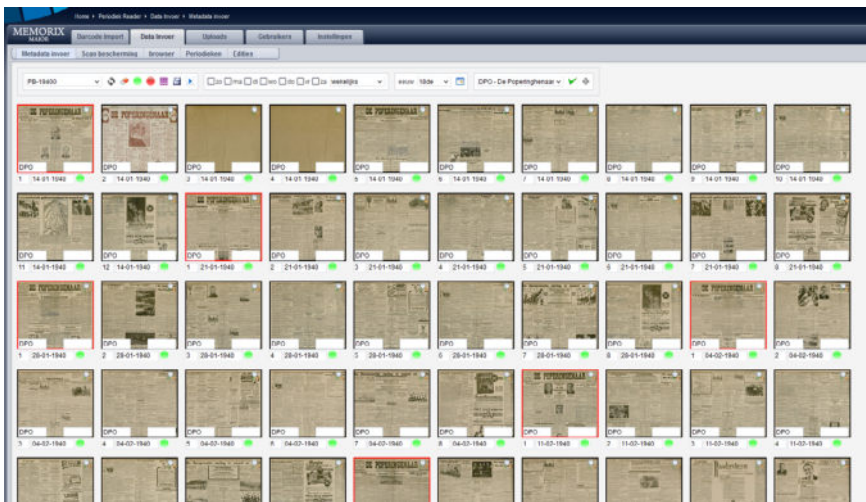


Fig. 5. Back office of the Memorix Maior system, provided by Picturae. The system allows metadata to be adjusted easily and conveniently with each uploaded digital file (Source: CO₂).

255. WUYTS, *Gebeugen van Ieper*, p. 25.

256. Testimony by Tijs Goethals, archivist of the Poperinge city archives.

A Westhoek-Wide Collaboration

Over the years, cooperation within the platform *Historische Kranten* had also begun to expand. Whereas initially only city services of Ypres were involved in the project, the digital platform also started to expand geographically, from *Erfgoedcel Ieper* to *Erfgoedcel CO₇*. New collaborations with institutions in the rest of the Westhoek region followed.

Today CO₇ works together with the intermunicipal collaborations *IJzervallei* and *Viertoren* and the city archives of Diksmuide, Ypres, Poperinge-Vleteren and Veurne for *Historische Kranten*. Within this grouping CO₇ assumes a coordinating and supporting role. Together with *Westhoek verbeeldt*, the platform is coordinated by CO₇, which is the first contact for the provider Picturae.

Unlike the archives, the intermunicipal partnership CO₇ does not manage its own heritage collection. The material made available on *Historische Kranten* remains the property of the archives. However, CO₇ does contribute financially to the storage and hosting of the digital files and is responsible for maintaining the platform. As general coordinator of the platform, the local heritage service maintains the network of project partners and convenes them on a regular basis to discuss the future of the platform and structure future additions. Furthermore, like the other project associations involved (*Viertoren* and *IJzervallei*), CO₇ supports the archives in getting the digitised newspapers online: checking the metadata of the files, check if the file names are written in the right way, etc. CO₇ acts as an intermediary between the various partners and the software supplier Picturae, which uploads the supplied files into the system itself after performing OCR on the files.

However, evaluation and adjustment remain necessary. For instance, the current way of working is being reviewed with the project partners. Since only small quantities of newspapers are being added to the platform nowadays, we are investigating whether archives might upload the files themselves in the *Memorix Maior* database. This way, archives will not have to wait long before their digital files come online and intervention by CO₇ or Picturae will not be necessary. In addition, it would mean that adding the newspapers to the database would no longer have to be outsourced to Picturae, which in turn could save costs.

Limits to the Growth

Historische Kranten today forms part of the heritage bank *Westhoek verbeeldt*. The ambitions of the early years of the project have since been realised: the heritage bank *Westhoek verbeeldt* tells and visualises an integral story of the daily life of the Westhoek region at that time. By entering a search term, three databases can be searched simultaneously for results: old image material, historical newspapers and heritage collections unlocked through the heritage database *Erfgoedinzicht*.²⁵⁷

Today, 217,353 pages of digitised newspaper pages from 70 titles originating from the entire Westhoek can be consulted on *Historische Kranten*. With the latest additions of newspapers from Diksmuide and Veurne, the platform offers a unique insight into daily life in the Westhoek from the mid-19th century onwards. The oldest newspapers on the platform date from around 1850, while the youngest newspapers date from 1990.



Fig. 6. Display of the front page of the newspaper *Gazette van Dixmude* of 5 February 1871 on the platform *Historische Kranten* today (Source: CO₂).

257. <https://erfgoedinzicht.be/>.

The platform still has a lot of growth potential. In principle, municipalities and other heritage institutions could have their newspaper collections added to the platform if there is interest and a willingness to contribute financially to its maintenance. The addition of newspapers from the Nieuwpoort city archive is currently being discussed. As the municipality itself does not have the necessary equipment to carry out the digitisation, it is collaborating with the Diksmuide city archive. The newspapers from the collection were digitised in Diksmuide and the digital files are being prepared for addition to the database. Arrangements are being made with the municipality regarding the storage and hosting costs of the Nieuwpoort newspapers. The municipality of De Panne has also expressed interest in making its newspapers available through the platform.

Meanwhile, archives continue to seek opportunities to add newspapers to their digital newspaper archives. The city archive of Diksmuide has a limited newspaper archive due to destruction/damage sustained during the First World War. Since 2009, the archive has been actively acquiring newspapers from the district of Diksmuide and its wider surroundings with the intention of digitising them and adding them to its own digital newspaper archive as well as to *Historische Kranten*.

In turn, the Poperinge city archives digitised the periodical *De Hopboer* in 2021-2022 and had it added to the Historical Newspapers platform. The archive received the periodical on loan from *De Boerenbond* from Leuven to digitise the newspapers in-house and in the process also received permission to publish the newspaper online.

Although the platform has potential to keep growing, it is also gradually reaching its limits. Administrative, financial and logistical challenges are currently preventing the project from growing sustainably. The size of the project is also gradually exceeding CO₇'s financial capacity and scope. Costs are rising, not only because of the increased volumes, but also because of costly services that are of course also indexed. The high dependence of the provider and the chosen software makes this problem all the more pressing. The shelf life of software systems is limited, and each new upgrade entails higher costs. Like many locally and regionally based heritage databases, CO₇ is watching developments within other heritage databases in Flanders with interest.

Looking Ahead: Towards a Flemish Cooperation?

Historische Kranten, together with *Westhoek verbeeldt*, is still a much-visited website today. It is consulted for a variety of different reasons: for nostalgia and private use, but equally for scientific research at universities and even by artists looking to working creatively with the material. Applications to consult newspapers younger than 70 years old are no longer an exception.²⁵⁸

After almost 20 years of newspaper digitisation in the Westhoek, almost all the historical newspaper collections in the Westhoek archives have been digitised and can now be accessed through *Historische Kranten*. Occasionally, new editions or annuals of periodicals still turn up through accidental finds or donations. Younger newspapers in particular are currently waiting to be digitised. On the platform, the youngest newspapers are editions of *Het Wekelijks Nieuws* from 1990. Newspapers more recent than this have not yet been digitised due to constraints of resources.

An additional concern with more recent newspapers is, of course, the applicable copyrights. Although there is great demand for digital access to more recent newspapers, it cannot yet be met. Since clearing the copyrights of historical newspapers is not easy, it was decided to blur newspapers younger than 70 years when consulting the website publicly. They will, however, be stored in the database if they have already been digitised. Interested parties can consult these newspapers in the archives and libraries of the Westhoek. In rare cases, online access can be granted to researchers who do not live in the Westhoek or who might experience difficulties in travelling to the region. For this, however, a request must be submitted to CO₇ with an explanation of the research for which such a consultation is necessary.

To continue meeting the current challenges and to give the platform *Historische Kranten* a sustainable future, CO₇ is looking with great interest towards the developments around Flemish heritage databases. Meemoo, the

258. As copyrights for newspapers younger than 70 years on the platform have not been cleared yet, it was chosen to blur these newspapers on the website until they reach the age of 70. The choice was made to comply with Belgian copyright legislation.

Flemish institute for archives,²⁵⁹ has become an essential partner in the digital story not only for CO₇, but for all cultural heritage organisations. Back in 2017, CO₇ entered into a cooperation agreement with the Flemish Institute for Digital Archiving and Access to Audiovisual Heritage (VIAA), a predecessor of meemoo. Through this agreement, CO₇ became a content partner of the VIAA. In 2022, this cooperation agreement was renewed following the merger of VIAA, PACKED and Lukas into meemoo.

The agreement makes it possible for CO₇ to call on the services and expertise of meemoo regarding digitisation of cultural heritage since then, and for it to permanently store its digital collections at meemoo. Storage at meemoo and access via *Het Archief*²⁶⁰ makes it possible to reduce the high costs of storage and hosting of digital material. Furthermore, it enables making audiovisual material available on *Westhoek verbeeldt* in the future. In early 2024, a trajectory of digital inflow at meemoo was started. The focus is now on the visual material of *Westhoek verbeeldt*. At present, it is not yet possible to have newspaper material transferred to meemoo through the digital inflow.

To join other Flemish cultural heritage organisations in giving newspaper databases a sustainable future, CO₇ joined the three-year project *Nieuwe Tijden: Landelijk Programma voor het behoud van het Vlaamse Krantenerfgoed*, a project of the Flanders Heritage Libraries in collaboration with meemoo (see Sophia Rochmes' chapter in this volume).²⁶¹ The objective of the project was to build a large-scale and overarching programme at the Flemish level for digitising, unlocking and archiving endangered newspaper collections. Within this project, CO₇ shared the knowledge and experiences it has gained in the field of newspaper digitisation.

One part of the project was the development of a newspaper interface within *Het Archief* of meemoo that allows newspapers to be consulted digitally. The newspaper interface is coming in the next few years. CO₇ will realise the inflow of newspaper material at meemoo as soon as is possible.

Furthermore, CO₇ continues to closely monitor developments around dig-

259. <https://meemoo.be/>.

260. <https://hetarchie.be/>.

261. <https://meemoo.be/nl/projecten/nieuwe-tijden-redt-historische-kranten-van-het-verval>.

ital platforms and applications in the cultural heritage sector. There is active participation in the collegial group with the Flemish and Brussels local heritage services. Within this peer group, shared problems are discussed and addressed, and opportunities are identified. The meeting allows for a network to develop around this theme, allowing experiences to be shared and learning from the experiences of other local heritage services regarding digital heritage databases.

CO₇ sees following the policy in Flanders and the developments around digitisation of heritage in order to be able to translate this to the partners in the region as its task, one which also includes many non-professional heritage organisations such as local heritage societies and churchwardens. These organisations are also increasingly looking to digitise their own archives and heritage collections. CO₇ is often called upon to guide them in this process.

Conclusion

Historische Kranten was an ambitious and innovative project when it started out 20 years ago. Like many other innovative projects, it originated locally through cooperation between ambitious partners who found in each other a shared strong commitment and support for the elaboration of the project. Thanks to motivated people at *Erfgoedcel Ieper*, CO₇ and nowadays also within a network in the Westhoek, the platform has been able to grow into a database containing hundreds of thousands of pages of digitised newspapers. The project lived up to its ambitions and still receives a lot of interest, both from the inhabitants of the Westhoek and from professional researchers.

At its 20th anniversary, *Historische Kranten* finds itself at a crossroad. Making the project sustainable is no longer possible without a solid Flemish framework, both financially and in terms of expertise. The Flemish Government's decision to transfer the Flemish heritage databases to meemoo and the clear role recently given to meemoo to further support the digital transformation of the cultural heritage sector offer perspectives for the future of this innovative project from the Westhoek region.

Σαράη

182



10. The Digitisation of Historical Newspapers in Bruges Public Library: Experiences and Perspectives

Alexander Soetaert, Hilde Van Parys

Abstract

Since 2014 the Bruges Public Library (*Openbare Bibliotheek Brugge*) has digitised around 700,000 pages from local newspapers published between 1792 and the early 2000s. Since 2016 the digitised newspapers have been accessible on the *Erfgoed Brugge* platform. The article deals with the various steps that were taken to digitise the newspapers and to make them accessible online. Next to some technical aspects, also the use of the digitised newspapers and some challenges concerning the storage of the digital assets and text recognition are discussed.

Authors

Alexander Soetaert is Digital development and innovation expert at the Bruges Public Library.

Hilde Van Parys is a historian and works at the Bruges Public Library. She is responsible for the *Brugensia* and the historical newspaper collection and coordinated the newspaper digitisation project.

Keywords

Newspapers; heritage; digitisation; OCR; libraries

The Press in Bruges

The fact that the press is a crucial source for historiography needs no introduction. During the 1960s, the Bruges historian Romain Van Eenoo was one of the first to consider newspapers and magazines as worthy source material. It was also at that time that the first endeavours were made to map the Belgian press landscape. The Interuniversity Centre of Contemporary History (ICCH) published a series of press directories dealing with several cities and regions in Belgium.²⁶² Van Eenoo drafted a directory of the press in Bruges between 1792-1914 collecting data on around 150 Bruges newspapers, which were found in several Flemish libraries.²⁶³

A significant amount of the newspapers included and discussed in the directory are part of the heritage collection of the *Openbare Bibliotheek Brugge* (Bruges Public Library).²⁶⁴ The oldest newspapers date back to the 17th century, however 1792 is the starting point of a continuous newspaper production in Bruges which culminates in the second half of the 19th century. At least twenty new papers were established during that time span.

The collection is preserved in the best possible conditions. The storage space is climate controlled, and the newspapers are stored horizontally in bespoke acid-free boxes. The collection has gained great renown since the 1960s — partially thanks to Van Eenoo's research — and was consequently consulted quite often in the library. This regular and not always careful handling of the newspapers has inflicted a significant amount of damage. During the 1990s the library opted to transfer a great amount of the collection onto microfilm and to no longer make the physical newspapers available to the public.

Digitisation

In 2014 a project of digitisation of the Bruges historic newspapers was launched to improve the accessibility of the collection and to make it available

262. <http://commissionroyalehistoire.be/en/reference-works/publications-of-the-icch/>.

263. VAN EENOO, Romain. *De pers te Brugge 1792-1914: bouwstoffen*. Leuven: Nauwelaerts, 1961.

264. <https://brugge.bibliotheek.be/erfgoed>.

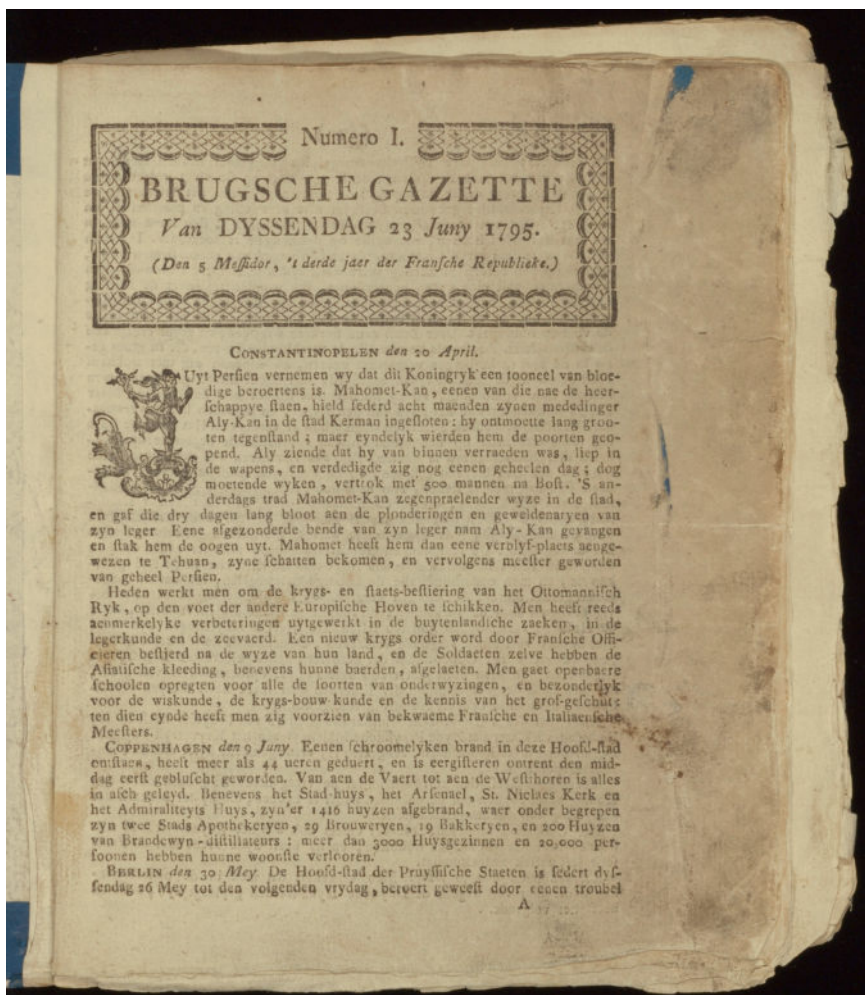


Fig. 1. Brugsche Gazette, 23 June 1795.

to a wider audience. *Het Geheugen van Ieper* ("The Memory of Ypres"), a project from 2007 by the Heritage department of Ypres, was seen as exemplary (see Nick Coorevits' contribution in this volume).²⁶⁵

265. Historische kranten [online]. Poperinge: CO, [consulted on 21/03/2024]. Available on: <https://historischekranten.be/>.

Every kind of consultation of the newspapers causes damage. The very act of digitising a newspaper can cause damage — neither drawing up an inventory or the scanning process itself are neutral activities. While it is possible — and economically efficient — to digitise existing microfiches, this option was not chosen. This was because it would have resulted in inferior text recognition while the black and white images would have lacked the look and feel of heritage materials.



Fig. 2. Gazette van Brugge, 1 January 1893. Jubilee issue following 100 years of the Gazette van Brugge.

The nature of the binding process implies that some of the text margins disappear into the spine. Breaking up these bindings, however, would have caused irreparable damage to the newspapers and it was considered best to keep the newspapers in their current state. To spare the spines as much as possible during digitisation and to keep the pages at an equal height, a scanner with a book cradle was necessary. The allotted budget did not allow for extensive restoration of the collection, which meant that the newspapers were digitised in the conditions they were found in at the project's beginning.

The budget did not only fail to allow for restoration, but it also meant that, initially at least, only a selection of those titles eligible for digitisation could be processed. It was decided that, in terms of temporal, political and religious relevance, as broad a selection as possible was to be made. Digitising the *Gazette van Brugge* was an obvious choice. The newspaper was established in 1792 – under the title *Vaderlands Nieuws-Blad* – and was published until 1919. As with most newspapers, the *Gazette van Brugge* remained largely non-partisan until the mid-19th century, when it took a more church-minded stance. The first phase of digitisation, which processed some 250,000 pages, was completed with the addition of *La Patrie*, a French-language liberal-catholic newspaper (1848-1940), and the liberal *Journal de Bruges* (1837-1953).

Creating the inventory of the newspapers was a time-consuming aspect of the project, but it allowed for the library to estimate how many pages needed to be digitised, a crucial feature of the public tender. It also allowed the library to seek missing issues. It was understood that the manhandling of the newspapers inherent in the digitisation process would inevitably cause some damage, which led to the physical condition of the collection being registered. By noting tears, stains, folds and other types of damage, the project sought to avoid accusations of rough treatment of the collection after the fact, as well as drawing attention to future restoration needs. The public tender was published in August 2014. The library made use of the expertise of PACKED (2011-2019, now meemoo) and *Vlaamse Erfgoedbibliotheken* (since February 2026 *BibliothecairErfgoed.be*)²⁶⁶ to draw up the specifications. Three digitising firms contributed over the course of the project (2014-2022). In the meantime, a partner was found to make the

266. <https://vlaamse-erfgoedbibliotheken.be/>.

digitised newspapers accessible online. On 11 October 2016 www.historischekrantenbrugge.be went live: 250,000 pages of 9 Bruges newspapers were now digitally available to everyone. The newspaper *Het Laatste Nieuws* dedicated an entire page to the newspaper database on Wednesday 27 November 2016.²⁶⁷ At the latest count, 43 Bruges newspapers have been made available by the project and 9 Bruges periodicals were added recently.

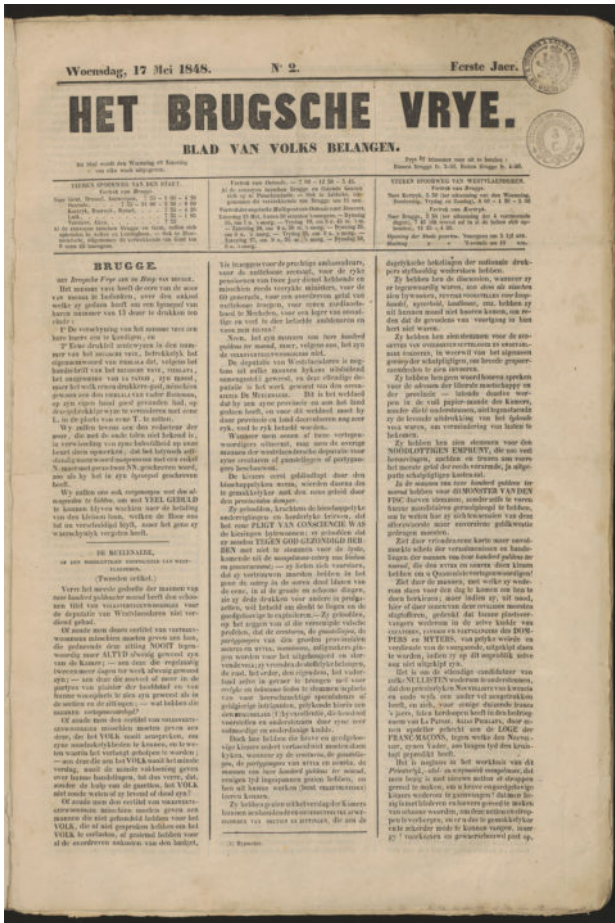


Fig. 3. Het Brugsche Vrye, 17 May 1848.

267. HUYSENTRUYT, Bart. Zo maken ze geen gazetten meer. *Het Laatste Nieuws*, 27/11/2016, p. 20.

Determining copyright status is a very difficult task for newspapers. For newspapers published more than 150 years ago, it is assumed that no copyrights apply anymore. For newspapers from the late 19th century and first half of the 20th century, determining the exact copyright status is next to impossible. Copyrights may apply to the publisher of the newspaper, the authors of individual texts or photographers. Meanwhile, most publishers/printers have not been active for several decades. Moreover, in many cases it is not clear who wrote a text and thus who owns the copyright: the author of the text or the publisher of the newspaper. For the pages for which it is virtually certain that copyrights still apply (published less than 70 years ago), the choice was made to offer them anyway for scientific and educational purposes. For the *Brugsch Handelsblad*, which is still in circulation, this was done in consultation with Roularta, its publisher.

Online Access through Heritage Bruges (*Erfgoed Brugge*)

How do you make more than 700,000 newspaper pages searchable online? This question arose after the library decided to digitise the first part of the newspaper collection. The usual answer is to develop a distinct platform, such as national newspaper databases, e.g. Deutsches Zeitungsportal²⁶⁸ or the Austrian ANNO.²⁶⁹ Newspapers are sometimes enriched with digitised books as was the case with Delpher in the Netherlands.²⁷⁰

In Bruges we opted for a different route by joining the *Erfgoed Brugge* ('Heritage Bruges') portal.²⁷¹ This portal is designed to bring together all digitised Bruges heritage collections in one searchable platform. In *Erfgoed Brugge*, you will not only find newspapers, but also photographs, pictures and museum collections. In a first phase the Bruges Public Library joined forces with *Beeldbank Brugge*, a project dating back to 2001 coordinated by the Bruges City Archives.²⁷² *Beeldbank Brugge* showcased tens of thousands of digitised pho-

268. <https://www.deutsche-digitale-bibliothek.de/newspaper>.

269. <https://anno.onb.ac.at/>.

270. <https://www.delpher.nl/>.

271. <https://erfgoedbrugge.be/>.

272. <https://www.brugge.be/archief>.

tos, picture postcards, pictures, drawings and posters, mostly from the nineteenth and twentieth century.²⁷³ When *Erfgoed Brugge* was launched the separate *Beeldbank Brugge* website was discontinued.

The new portal's purpose consisted of breaking down the barriers between city departments in a digital environment. A researcher looking for a Bruges street name, for example, will find historical photographs and newspaper pages pertaining to that particular street. Someone looking for paintings by Hans Memling should obviously find those in their search results, but will also encounter photos of the visit by the Danish queen Margarethe (who viewed the paintings at the *Sint Janshospitaal Museum* in 1976) or a newspaper article about Memling's paintings.

However, this multi-genre richness was not always self-evident due to the technical architecture of *Erfgoed Brugge*. The digitised newspaper pages and the corresponding OCR-files were directly uploaded into the system on which the new public environment was built. The newspaper collection was given a hierarchic layout with three levels: newspaper collections (by title), issues (by edition/date) and pages. The newspaper's title, the issue date and the page number were derived from the file names and added as metadata. Other metadata (e.g. issue place, publisher, ideological background of the newspaper, etc) were not added.

Another collection management system was used to store and catalogue the images of *Beeldbank Brugge*. No hierarchical or chronological layout is used, instead the emphasis is on a substantive and thematic description of the images. For each image you will find a title, date, mention of pictured places and themes and, whenever possible, the name of the photographer or publisher. A longer text description of what and who can be seen on an image is mostly available (usually 50 to 300 characters, but more in some cases). The metadata is delivered through OAI-PMH²⁷⁴ to the system behind *Erfgoed Brugge*, which is also where

273. SOETAERT, Alexander & VIAENE, Dieter. Beeldbank Brugge: twenty years of making historic images accessible online [online]. Peccioli (Italy): *Photoconsortium*. 25/03/2021 [consulted on 21/03/2024]. Available on: <https://www.photoconsortium.net/beeldbank-brugge/>.

274. *Open Archives Initiative Protocol for Metadata Harvesting* [online]. S.I. s.d. [consulted on 21/03/2024]. Available on: <https://www.openarchives.org/pmh/>; *Metadatauitwisseling met OAI-PMH. Protocol en implementatie* [online]. Hilversum: Beeld & Geluid. s.d. [consulted on 21/03/2024]. Available on: https://publications.beeldengeluid.nl/pub/451/BenG_Metadatauitwisseling_met_OAI-PMH.pdf.

the digitised newspaper pages and the OCR files can be found. Because of the different database structure and metadata model, the images from *Beeldbank Brugge* ended up in a separate module. Later on separate modules were added for e.g. museum collections, archaeological collections and genealogical data. This did create separate data and image repositories.

Consequently, the joint searching of the digitised newspapers and the *Beeldbank* collection was not as simple as hoped. The different metadata fields for instance resulted in different search filters for the newspapers than the ones for the photos. When a search entry generates results in both the newspaper collection and the *Beeldbank* collection, a user will not have an easy experience navigating due to the sometimes substantial number of results.

The large amount of OCR data which is part of the newspaper collection made the relevance sorting of the search results a challenging task. The OCR data generated results for almost every possible Dutch search entry — the relevance of those results, however, may not always be obvious to the user. A search entry found in the metadata descriptions of *Beeldbank Brugge* usually had a more significant relevance for the user. These metadata were made by contributors and volunteers of *Beeldbank Brugge*: the terms used are the results of deliberate human choices.

In order to avoid search results from the *Beeldbank* being drowned under a flood of less relevant OCR hits, they were always prioritised. Results from the newspaper collection were shown further down the list of search results. A search for the street name ‘Peperstraat’, for example, yielded 7260 results, of which 5420 will originate from the newspaper pages, 1307 from genealogical data and 380 from *Beeldbank Brugge*. You would only have found the first result from the newspaper collection on position 1858, which is around page 150 of the search results.²⁷⁵ The results of the other collections were ranked higher in the results list. It was possible to limit your search to the newspaper collections of *Beeldbank Brugge*, however that did not quite coincide with the original intention to make all collections jointly searchable.

275. Usually, 12 results are shown on a page.



Fig. 4. De Halletoren, 30 August 1874.

The development of a new version of *Erfgoed Brugge*, launched in November 2024, allowed for some adjustments. Inspired by the European Data Model, all collections now use the same set of metadata fields. In the case of *Erfgoed Brugge*, however, we limited the number of fields. This limitation of searchable fields was designed to enhance the friendliness of the portal to those who lack specialist knowledge. If someone is looking for very specific data, such as dimensions

of objects, they will be able to find those on the external platforms of the institutions managing the collections, such as Musea Brugge or the archaeological department *Raakvlak*.²⁷⁶

The information architecture of the portal was redesigned as well. The starting point is no longer a particular storage facility or a (past) project, but the type of document or object. The intention is to be able to make searching for a newspaper, photograph or painting much easier, regardless of which institution the object is stored or through which system its data are delivered. We also hope to make progress regarding relevance sorting, even though this remains a technical challenge.

The Use of the Digitised Newspapers

Since 2020 we have recorded the user statistics of *Erfgoed Brugge* as diligently as possible. They tell us that Bruges newspapers are searched intensively. Views of newspaper pages more than doubled between 2020 and 2022, from 185,000 to 384,000.²⁷⁷ This increase can be at least partially attributed to the additional uploading of some 300,000 pages to the portal in 2022. Newspapers are by far the most intensively viewed collection on *Erfgoed Brugge*, representing 70% of the total number of views of detail pages made on the portal in 2022.²⁷⁸

In 2020, a user survey sought to analyse the ways in which users performed searches through the portal. While 60% of searches involved newspapers, fully 26% of all searches were limited to the newspaper collection, with many searching for a specific date. 80% of the newspaper-limited searches using the advanced search form were carried out with a specific date (dd/mm/yyyy). In other words, users are very clearly looking for information which was published in a newspaper on a specific day or in a specific time frame.

Our conversations with users tell us that the newspaper collection is often

276. For example: *Collectie Musea Brugge* [online]. Brugge: Musea Brugge. S.d. [consulted on 21/03/2024]. Available on: <https://www.museabrugge.be/collecties/ontdek>; *Collectie Raakvlak* [online]. Brugge: Raakvlak. S.d. [consulted on 21/03/2024]. Available on: <https://collectie.raakvlak.be/>.

277. We do not have exact numbers for 2023 because of the transition phase of Erfgoed Brugge.

278. By a detail page we imply a page where you can find the images and metadata of a document or object (e.g. newspaper page, photograph, painting...).



Fig. 5. Burgerwelzijn, 28 April 1900.

searched systematically. A classic example is the search for all mentions of a certain person, family name or street name. A group of volunteers from Oostkamp, a municipality south of Bruges, systematically searches the newspapers for information concerning (disappeared) businesses in their village. They were able to find a lot of ads which they share with their fellow villagers on the Facebook

page of the local history society (*Heemkring Oostkamp*). References to the digitised newspapers also appear very regularly in the footnotes of (local) history magazines,²⁷⁹ books on the 19th and 20th century history of Bruges²⁸⁰ and also feature on quite a number of Wikipedia pages on subjects linked to Bruges.

Challenges

In addition to making (and keeping) hundreds of thousands of digitised newspaper pages available online, archiving the digital assets in a sustainable way is also a significant challenge. Because the physical state of the newspapers themselves continues to deteriorate, the digital archive is essential for preserving this vast source of information about Bruges' past for future generations. The entire digitised newspaper collection contains about one and a half million files (high resolution TIFF files and OCR files in XML format) that amount to more than 50 terabytes in storage space.

The files were delivered on several external hard drives by the digitisation firms. But nearly ten years on, a more sustainable and safer solution is urgently required. Network drives or cloud storage does not offer enough storage space, demands too much of the organisation or will prove a financial burden in the long run. In 2026, we will launch a pilot project with meemoo (the Flemish institute for archives)²⁸¹ for those very reasons. This project will design a process whereby the newspaper pages, OCR data and metadata can be moved into the meemoo system. Meemoo is an organisation supported by the Flemish government and its objective is the sustainable archiving of digital assets.²⁸² It offers the structural framework necessary to guarantee a long-lasting conservation of the newspaper files. Beyond the simple quantity of files that must be moved, the project is concerned with the different folder structures on the external hard

279. E.g. *Biekerf. West-Vlaams Archief voor Geschiedenis, Archeologie, Taal en Volkskunde; Handelingen van het Genootschap voor Geschiedenis te Brugge; Brugs Ommeland...*

280. E.g. PEVERELLI, Tjmen. *De stad als vaderland. Brugge, Leeuwarden en Maastricht in de eeuw van de natiestaat, 1815–1914*. Nijmegen: Vantilt, 2019.

281. <https://meemoo.be/>.

282. *Het meemoo-archiefsysteem* [online]. Gent: meemoo. S.d. [consulted on 21/03/2024]. Available on: <https://support.meemoo.be/hc/nl/categories/6039894771857-Het-meemoo-archiefsysteem>.

drives and the different standards of the OCR data in XML format. Both the folder structure and the OCR standards employed differ depending on the firm that digitised the newspapers.²⁸³

Almost ten years after the first pages were digitised, the quality of the OCR data is also being called into question. The text recognition process happened right after digitisation and was executed by the different digitisation firms. Recent research by the Flanders Heritage Libraries showed that the quality of older OCR data can be significantly improved thanks to new evolutions in text recognition software.²⁸⁴ An even better OCR quality would evidently mean added value to users of the digitised collections. But again, this will require effort and investments. In the current situation the new OCR data would have to be re-uploaded to the systems at the back end of *Erfgoed Brugge* and there connected to the right images. These efforts will slow down the digitisation of other collections. Furthermore, the text recognition algorithms will keep on developing, which does not make it easy to decide on the opportune time to update the OCR data. A software system that can both store the images and periodically execute the OCR process, without the addition of a new uploading process, would solve a part of this conundrum in the future.

283. ABBYY for the first cycle of digitised newspapers, afterwards ALTO was used, see: *ALTO. Technical Metadata for Layout and Text Objects* [online]. Washington DC: The Library of Congress. 08/06/2022 [consulted on 21/03/2024]. Available on: <https://www.loc.gov/standards/alto/>.

284. *Nieuwe tijdingen: landelijk programma voor het behoud van het Vlaamse krantenerfgoed* [online]. Antwerp: bibliothecairerfgoed.be [consulted on 25/03/2026]. Available on: <https://www.bibliothecairerfgoed.be/projecten/nieuwe-tijdingen>.

Part 3. Research Projects

11. Constructing a Digital Corpus of Interwar Belgian Magazines to Investigate the Representations of Fine Arts. An Artpresse Case Study

Morgane Ott & Sébastien Hermans

Abstract

This study examines the Artpresse digitisation and research project, focused on interwar illustrated periodicals within the KBR collections (2020–2024). The article first addresses various challenges associated with the creation and expansion of an extensive digitised corpus of mass media magazines. These efforts establish a foundation for a comprehensive analysis of Belgian magazine culture, with particular emphasis on the development of rotogravure printing techniques and the interconnected networks of printers and publishers. The discussion then transitions to an intermedial analysis, utilising an internally generated dataset, conceptualised as a content catalogue, to investigate the representation of Belgian art and artists in selected family magazines from the late 1920s to the early 1930s. By tracing these interconnected threads, this contribution aims to shed light on Belgian art-historical narratives and networks of publishers and journalists as seen through the lens of middlebrow magazines.

Authors

Morgane Ott holds a master's degree in Art History from ULiège and is pursuing a doctoral thesis under the supervision of Julie Bawin, focusing on the reception and representation of art in Belgian mass-market magazines during the interwar period, as part of the Brain-be 2.0 Artpresse project (KBR-KU Leuven-ULiège).

Sébastien Hermans coordinates the digitisation of magazines and periodicals collections for the Artpresse project in KBR, having previously fulfilled the same function with serialised photonovels in postwar Belgian magazines for the Photo-Lit project. He holds a master's degree in History and Cultural Studies from KU Leuven.

Keywords

Mass digitisation, middlebrow magazines, Belgian periodical press, art representation, interbellum

Artpresse is an interdisciplinary intermedial exploration of Belgian art in the Interbellum, examined through the lens of mass-market magazines—commercial, non-specialised periodicals that have largely been overlooked in periodical and cultural studies.²⁸⁵ Through the digitisation of an extensive corpus of periodical print heritage in the Royal Library of Belgium (KBR), the project delves into Belgian interwar magazine culture, particularly at the turn of the 1930s, in order to investigate from an art historical perspective the broad spectrum of representations and discourses related to the domain of the fine arts in interwar Belgium. Artpresse focuses on a segment of mainstream print culture which can be referred to as ‘middlebrow’ family magazines, i.e. mass-produced illustrated weekly publications aimed at men, women and children, covering art news alongside miscellaneous, visually engaging content. After discussing the creation and expansion of the digitised corpus on various levels, this contribution will try to give a brief overview of Belgian interwar magazine culture in the light of the emerging rotogravure printing technique and networks of printers and publishers. Subsequently we will explore how fine arts were featured in a selected number of family magazines during the early 1930s, drawing on collected data in the form of an internal illustrated and indexed catalogue to examine the portrayal of Belgian artists (whether alive or not) in relation to Return to Order discourses.²⁸⁶

285. ARTPRESSE is a BELSPO-funded Brain-be 2.0 interdisciplinary research project kicked off in 2020 and coordinated by KBR in partnership with KU Leuven (Faculty of Arts, Literary Theory and Cultural Studies) and ULiège (Faculty of Philosophy and Letters, Departments of Contemporary Art History and Modern Languages): <<https://www.kbr.be/en/projects/artpresse>>.

286. By ‘Return to Order discourses’, we refer to certain tastes and narratives in art history that emphasise works from the near or distant past, national or regional cultural values, and an iconography rooted in figurative arts, as outlined in the following key reference works: LAUDE, Jean. *Retour et/ou rappel à l'ordre*. In: *Le Retour à l'ordre dans les arts plastiques et l'architecture, 1919-1925*. Saint-Étienne: CIEREC, Université de Saint-Étienne, 1975, p. 7-45; LAUDE, Jean. *La crise de l'humanisme et la fin des utopies (Sur quelques problèmes de la peinture et de la pensée européennes 1929-1939)*. In: *L'Art face à la crise. L'art en Occident, 1929-1939*. Saint-Étienne: CIEREC, 1980, p. 295-391;

Drawing solely from KBR's patrimonial print media collections, the in-house digitisation campaign of the Artpresse project has to date produced a digital corpus of nearly 25,000 issues (counting over 750,000 pages) from about 130 illustrated periodicals published in Belgium during the first half of the 20th century (ca. 1910-1950). However, this campaign did not unfold like most interinstitutional or outsourced digitisation projects conducted in the field of print media, which are traditionally defined by detailed selections and precise figures of editions and pages preceding the predetermined output.²⁸⁷ Artpresse developed a research-driven, small-scaled mass digitisation workflow in close interaction with the intermedial study of the fine arts through the lens of family magazines. Where possible, the digital collections have been published in BelgicaPeriodicals without delay.²⁸⁸ In the challenging balance between digitisation efforts and research timelines, we managed to establish a dynamic and open approach towards the continuous process of selection and prioritisation within the considerable number of available materials. This process was further empowered by technical innovations much welcomed in KBR's digitisation department over the course of the project. The construction of this digital corpus has ultimately been characterised by a tension between the objective to digitise with the available means as much relevant content as possible (covering interwar Belgian print magazine culture) on the one hand, and the progressive lines of inquiry into the reception and representation of fine arts in 'middlebrow', mass-market periodicals on the other.

DEVILLEZ, Virginie. *Le retour à l'ordre. Art et politique en Belgique 1918-1945*. Brussels: Labor, 2002; DEVILLEZ, Virginie. The Avant-Garde on the Reworking of Tradition. In: AUBERT, Nathalie, FRAITURE, Pierre-Philippe & MCGUINNESS, Patrick (eds.). *From Art Nouveau to Surrealism, Belgian Modernity in the Making*. London: Legenda, 2007, p. 142-151. A comprehensive history of Belgian art that departs from the avant-garde, modernist perspective needs to incorporate the largely overlooked middlebrow magazine culture. Artpresse thus seeks to further deepen the contextualisation of the ideological, 'reactionary' shift that occurred in interwar Belgium, drawing on family magazines as valuable historical and visual sources.

287. Such as the digitisation of Belgian newspapers in KBR holdings, described by D'HOORE, Marc. Un exemple de projet de numérisation patrimoniale de masse: les journaux de la Bibliothèque royale de Belgique (KBR). *Archives et Bibliothèques de Belgique. La numérisation du patrimoine écrit: du virtuel à la réalité*. 2011, vol. 82, n° 3, p. 117-134.

288. At the time of writing, the BelgicaPeriodicals platform (KBR) hosts over 1.7 million pages in more than 300 periodical titles [Consulted on 01/10/2024]. Available on <<https://www.belgicaperiodicals.be>>.

Scouting the Interwar Periodical Press in KBR

The initial extent of the to-be-digitised research corpus followed projections for a small set of Belgian interwar mass media magazines of which representative coverage (ideally their entire interwar production) had been identified and inspected in KBR collections. The main challenge that the project faced in its initial phase was the organisation of a digitisation workflow that allowed for scanning, metadating, archiving and disclosing different types of magazines per issue in a smooth and documented manner. In order to overcome these problems, we based our workflow on those that already existed for the processing of digitised newspapers in KBR, which we adapted to take account of the media-specific characteristics of non-newspaper periodicals. Furthermore, the digitisation process demanded a permanent interaction between Artpresse and the KBR public catalogue. The digitisation of any complete series often shines a light not only on the precise state of the collections from which it comes, but also on their relationship with other collections, both internal and external. For example, the first title entered into the digitisation workflow for the Artpresse project was *Pourquoi Pas?* (1910-1988), a well-known Brussels based conservative weekly which is amongst the most consulted of the interwar periodicals held in KBR. During the process it became apparent that the bound KBR collections showed considerable traces of wear and use. In parallel a digitised version based on unbound copies in a reasonably good condition became available on the platforms of the library of the University of Liège. The Artpresse digitisation was then reduced to focus only on the few gaps in the Liège collection and negotiations concerning a possible file exchange to produce a complete digital collection of this important title in BelgicaPeriodicals were entered into.²⁸⁹ This highlights how the combination of a large amount of material potentially in need of processing and the limited technical and human resources available demand that a rather economical approach be taken to the digitisation process. A similar economy

289. *Pourquoi Pas?* on the ULiège platform Donum (downloadable PDFs): <<https://donum.uliege.be/expo/pourquoiipas/>> versus the current state of the digitised KBR collection in BelgicaPeriodicals : <<https://www.belgicaperiodicals.be/link/opac/15257142>>.

of resources was behind the adoption of the tabloid size of approx. 42 x 30 cm as the maximal dimensions for bound documents to be processed in the project's first phase: such were the dimensions of the several i2S Copybook manual scanners available to the project. The emergence, however, of material such as the 1928 and 1929 series of *Le Soir Illustré* as one of the primary sources for the fine arts and art discourse research to which Artpresse is dedicated made the digitisation of some larger formats — via the much slower Metis flatbed scanner — unavoidable.

In late 2021, the project's digitisation workflow was further refined and combined with semi-automatic scanning and postproduction introduced in the KBR Digitisation department, the result of which was the output capacity exceeding the project's initial digitisation goals, despite size limitations remaining the same as in the manual scanning process.²⁹⁰ Consequently, we found ourselves in the opportunity of extending the research corpus by continuously exploring and querying KBR collections on interwar periodicals and vice versa. The result of this process of mutual interrogation was that we were able to both frame the digitisation process and develop our research questions more firmly within a broader publishing and editorial context that marked magazine production in interwar Belgium. In practice, this meant the incorporation within the Artpresse corpus of a number of popular interwar magazines released by some of the leading Belgian mass media magazine publishers such as *Patria* (Antwerp), *Pim* (Liège) or *Jean Meuwissen* (Brussels). The assortment of digitised interwar Belgian periodicals currently available has evolved through multiple stages of collection assessment, and reinterpretation of selection criteria. The available resources can be structured into four concentric, methodologically hierarchical levels for analytical purposes and research focus (Fig. 1):

290. For a full discussion of the impact of semi-automatic scanning on the Artpresse digitisation of interwar Belgian magazines, see: LEMMERS, Frédéric, OTT, Morgane & HERMANS, Sébastien. Printed Mass Media and Automatic Digitisation: the Case of Belgian Illustrated Magazines from the Interbellum. In *Monte Artium* [online]. 2023, 15. Available on: <<http://journals.openedition.org/ima/427>>.



Fig. 1. Schematic representation of the Artpresse corpus.

The enhanced digitisation infrastructure in KBR also allowed for taking into account the output of some major Belgian magazines beyond the strict chronological boundaries of the interwar period but up to ca. 1950, revealing (dis)continuity in 20th century Belgian magazine culture and paving the way for the digitisation of the massive postwar magazine production held in KBR collections. However, this multifaceted and gradual corpus expansion required continued bibliographic research since a clear panoramic view on Belgian interwar magazine culture is lacking so far. The few occasions in which the generally scarce, selective and fragmented scholarly attention paid to 'popular' print media has crystallised into a national-level survey or synthesis are typically prompted by anniversaries or special commemorations.²⁹¹

Consulting contemporaneous bibliographical resources, secondary liter-

291. A fine example is the exhibition catalogue by STRUYE, Johan et al. (eds.). *150 ans à la une: un siècle et demi d'information illustrée en Belgique. Exposition organisée par le Crédit Communal de Belgique du 19-9 au 2-11 1980*. Brussels: Crédit Communal de Belgique, 1980. Historiographical attention for magazines seems to surface on the occasion of press exhibitions and anniversaries, as is also the case for two other valuable works, although on magazine culture in The Netherlands: OVINK, Gerrit Willem. *Het aanzien van een eeuw: de periode 1856-1956 weerspiegeld in 950 illustraties uit de voornaamste Nederlandse familiebladen*. Haarlem: De Spaarnestad, 1959; VAN DELFT, Marieke et al. (eds.). *Magazine! 150 jaar Nederlandse publiektijdschriften*. The Hague: Koninklijke Bibliotheek, 2006.

ature, and digitised newspaper collections in BelgicaPress²⁹² both has been and remains essential for understanding the intertwined publishing histories and phenomenologies of the two ostensibly distinct categories of periodical print media: the newspaper and the magazine. The Artpresse project prioritises interwar illustrated weeklies with a middle-brow profile, commercial products aimed at broad and mixed audiences (such as families), within the proper ideological pillar (e.g. catholic families) or beyond ('all Belgian families'). The relationships and interdependence with newspapers played a significant role in the interwar evolution of Belgian print media and thus for retrieval of interwar magazines for digitisation. By the First World War, the daily press in Belgium had reached saturation point, with only a few new titles appearing during the interwar period while, when compared to the late 19th century, there was an overall decline in their number. It was during this same period that the 'non-newspaper' periodical press, alongside new media such as film and radio, came to full development in that same period.

At the second exhibition of the Belgian periodical press in Brussels in 1923 (the first had been held in 1893), circa 1200 active 'press organs' were presented in six categories: the daily press from the capital and the provinces; the local press in French and in Dutch; the illustrated press; the financial press; the war-time press, and the specialised press (Fig. 2). The visionary bibliographer Paul Otlet (1868-1944), co-founder of the Mundaneum where the exhibition took place (see Jacques Gillen's contribution to this volume), remarked that 'all these newspapers, bulletins, magazines, and illustrated magazines [...] ceaselessly produce the ideas that will guide the masses, and one cannot think without dread of the great black holes that their disappearance would leave [...]. The periodical press completes the mission assigned to the book and the newspaper. Thanks to it, all areas of science, social life, and practical activities are constantly kept up to date with new ideas and facts. This press is developing in quantity, specialisation, number of pages, print run, illustration, and more frequent periodicity, and one could hardly exaggerate its usefulness as an agent of general progress'.²⁹³ This

292. <https://www.belgicapress.be>.

293. OTLET, Paul. *La presse périodique. Annuaire de la presse périodique belge 1923-1924*. Brussels: Maison du livre, 1924, p. 2. Our translation.



Fig. 2. Photographs from the Second National Exposition of the Periodical in the grand auditorium of the World Palace (Mundaneum) in Brussels in 1923. From top left to bottom right: two exhibition overviews; the financial newspapers; the French-language weekly press; the commercial and industrial periodicals; the fashion periodicals; the wartime front journals (1914-1918), and the sports publications. Source: *Annuaire de la Presse Périodique Belge 1923-1924*. Brussels: *Maison du Livre*, 1924. Online: <https://uurl.kbr.be/2189357>.

approach to periodicals underlines the complementarity of newspapers and magazines as well as how fluid the categorical demarcation between them can be in terms of content and periodicity. These assumptions are also implicit in the 1928 directory of Belgian periodicals by Bacha and Dupierreux, who attest the ‘discovery’ of 2954 running titles of Belgian ‘periodicals, newspapers and official documents’, 2074 of which published in French and 803 in ‘Flemish’ (i.e. Dutch), 13 in ‘Walloon’ (a dialect of French) and 64 in other languages.²⁹⁴ Gaston Mertens, in his 1930 report as president of the Union of the Belgian Periodical Press, counts 3360 periodicals having originated in Belgium since November 1918, ‘of which barely a hundred continue to appear at this moment’.²⁹⁵ A decade later, the association of Belgian and foreign journalists put forward the round number of 2500 periodicals appearing in Belgium,²⁹⁶ confirming that the Belgian periodical press during the 1920s and 1930s was marked by a strong and continuous dynamism, diversification and a multiplication of titles, many of which were short-lived.

The proliferation of serial publications during the interwar years led the Royal Library of Belgium to reorganise its periodicals collections, notably the ‘R’, ‘II’, ‘III’, and ‘VI’ series which had been established since the late 19th century. This process resulted in the creation of the ‘B’ series for Belgian periodicals and the ‘J.B.’ series for Belgian newspapers in the early 1930s. According to KBR, a total of some 8,000 periodical titles were available for consultation by 1939.²⁹⁷ Most of the Belgian illustrated magazines we drew from these collections to create the digital Artpresse corpus are hosted in the ‘B’ series, although many crucial publications related to the core selection have been drawn from other series. These include some trailblazing magazines as *Voir & Lire* (direct

294. BACHA, Eugène & DUPIERREUX, Richard. *Périodiques belges. Répertoire par titres et par sujets*. Brussels: Albert Dewit, 1928. Supplément, 1929.

295. MERTENS, Gaston (ed.). *Union de la presse périodique belge. Bulletin officiel*. Brussels: Union de la presse périodique belge, 1930, vol. 40, n° 3, p. 62-63.

296. DE KEYSER, Léon. *La presse périodique. Annuaire de la presse périodique belge 1939-1940*. Brussels: Association des journalistes périodiques belges et étrangers, 1940, p. 5.

297. VAN HOVE, Julien. De algemene verzamelingen van gedrukte werken. In: *Koninklijke Bibliotheek. Liber Memorialis 1559-1969*. Brussels: Koninklijke Bibliotheek van België, 1969, p. 138-156. BRUCHER, Roger. *Les collections de périodiques et de presse de la Bibliothèque royale Albert Ier, historique, structure et gestion*. Brussels: Bibliothèque royale Albert Ier, 1984, p. 12-17.

precursor of *Le Soir illustré*),²⁹⁸ *Antwerpen 1930* (a daily magazine in rotogravure published by De Vlijt during the world exhibition which paved the way for the weekly *Zondagsvriend* before the end of that year),²⁹⁹ *Echos* (a short-lived French-language family magazine published by Averbode in parallel with the Flemish *Ons Land* in 1932),³⁰⁰ or *Le Peuple illustré* (a short-lived illustrated weekly launched by socialist newspaper *Le Peuple* right after the war in 1919).³⁰¹ We also scoured the KBR collections to include in the digital corpus the range of mass media magazines produced in both languages by publisher Patria in Antwerp since its founding in 1928 (Fig. 3).³⁰² Fitting technical preconditions (such as its size) as well as research interests also led us to digitise the first instalments of Patria's 'neutral' newspaper *De Dag*. This newspaper is one of those few Belgian titles originating in the interbellum years, for which the publisher relied heavily on the formats, editors and graphic style of its tabloid sized magazines, notably corpus C titles *Tout* and *Wereldrevue*, which were respectively discontinued and sold after the successful launch of *De Dag* in November 1934.³⁰³

298. *Voir & Lire* (1926-1928), VI 7.660 B. Online: <<https://www.belgicaperiodicals.be/link/opac/13661335>>.

299. *Antwerpen 1930* (1930), III 99.896 C. Online: <<https://www.belgicaperiodicals.be/link/opac/12270105>>.

300. *Echos* (1932-1933), III 99.924 C. Online: <<https://www.belgicaperiodicals.be/link/opac/11745129>>.

301. *Le Peuple illustré* (1919-1920), III 91. 305 C. Online: <https://www.belgicaperiodicals.be/link/opac/11667253>>.

302. Patria magazines digitised from the KBR collections include, in order of appearance from left to right on the letterhead (Fig. 2): *Wereldrevue* (1931-1936), <https://www.belgicaperiodicals.be/link/opac/11624228>; *Patria magazine* (1933-1940), <<https://www.belgicaperiodicals.be/link/opac/12171977>>; *Piccolo* (1928-1944), <<https://www.belgicaperiodicals.be/link/opac/12133410>>; *De Vrouw* (1930-1944), <<https://www.belgicaperiodicals.be/link/opac/11605132>>; *De Film* (1932-1940), <<https://www.belgicaperiodicals.be/link/opac/11823539>>; *Le Film* (1934 -1940), <<https://www.belgicaperiodicals.be/link/opac/12304421>>; *Femme* (1931-1934), <<https://www.belgicaperiodicals.be/link/opac/11828468>>; *Mon Copain* (1930-1940), <<https://www.belgicaperiodicals.be/link/opac/13484546>>; *Radiopost* (1931-1936), <<https://www.belgicaperiodicals.be/link/opac/11714380>>; *Tout* (1932-1934), <<https://www.belgicaperiodicals.be/link/opac/11703899>>. *Sport* (in Dutch and French), *Kindervriend* and *Optimiste* have unfortunately not been tracked down in KBR collections.

303. *De Dag* (1934-1944), J.B. 1065. Online: <<https://www.belgicapress.be/link/opac/17131445>>.



Fig. 3. Letterhead (ca. 1935) of the Patria publishing and printing company displaying the distinctive covers of its Flemish magazine titles on the left and the French-language titles on the right. Source: Letterenhuis, Antwerp, K1A18D5-D6 (photograph by the author). Patria letterbeads kept frontlining *Wereldrevue* and *Tout* long after they had ceased publication.

The Patria back catalogue reflects the emerging dynamism of commercial magazine production in the 1930s, in this case leading to the publishing of a newspaper (*De Dag*). This dynamism resembles the trailblazing function performed by the Flemish catholic weekly *Ons Volk Ontwaakt* since its inception in 1911 to the founding of newspaper *De Standaard* in 1914.³⁰⁴ The digitised Patria magazines also demonstrate how expanding the Artpresse research corpus for purposes of completeness or context has prompted a reevaluation of the core selection criteria leading to a more focused selection of materials suitable for the art discourse analysis. Although offering miscellaneous content, including, albeit to a limited degree, fine arts related topics, magazines targeting explicitly a female middle-class readership such as Patria's *Femme* and *De Vrouw* have been confined to corpus level B (Fig. 1), a decision which was consistently applied for magazines by other publishers. Here we note the well-known *Het Rijk der Vrouw*³⁰⁵ published by Julius Hoste and its French-language counterpart

304. *Ons Volk Ontwaakt*, shortened to *Ons Volk* in 1932, lived on after the Second World War until its disappearance in 1985. Online: <<https://www.belgicaperiodicals.be/link/opac/15431975>>.

305. *Het Rijk der Vrouw* (1931-1990). Online: <<https://www.belgicaperiodicals.be/link/opac/16029287>>.

Femmes d'Anjourd'hui,³⁰⁶ created by the Dutch fashion magazine entrepreneur Jan Meuwissen (1883-1954). Meuwissen played a vital role in Belgian interwar magazine culture as founding publisher or originator of numerous family magazines such as the catholic inspired *Ons Land in Woord en Beeld* and its French-language counterpart *Notre Pays* in the 1920s, or the socialist inspired magazines *ABC* in Dutch and *A-Z* followed by *Bonjour!* in French in the early 1930s.³⁰⁷

Interwar Magazines and the Rise of Rotogravure Printing in Belgium

The layered approach in the corpus construction and the needs of digitisation and research development have revealed an unintended selection bias favouring richly illustrated general interest magazines printed in rotogravure over those using letterpress. This preference has emerged gradually rather than being predetermined, being largely driven by the presence of fine arts subjects in the former and their general absence in the latter. Refraining from rigid typological classification of magazines — apart from physical constraints such as height limitations for mass digitisation of full runs as mentioned above — it is the obvious combination of rotogravure printing quality with heterogeneous visual content including fine arts that explains the eventual adoption or omission of magazines in level C of the Artpresse corpus.³⁰⁸ Furthermore we could not find any mass produced rotogravure magazines from the interwar period in the KBR collections exceeding a height of 42 cm (the limitation for mass digitisation),

306. *Femmes d'Anjourd'hui* (1933-). Online: <<https://www.belgicaperiodicals.be/link/opac/15410455>>.

307. *Ons Land in Woord en Beeld* (1919-1932). Online: <<https://www.belgicaperiodicals.be/link/opac/15431902>>; *Notre Pays* (1919-1923). Online: <<https://www.belgicaperiodicals.be/link/opac/21014828>>; *ABC* (1932-1973). Online: <<https://www.belgicaperiodicals.be/link/opac/11703822>>; *A-Z* (1932-1937). Online: <<https://www.belgicaperiodicals.be/link/opac/18488044>>. *Bonjour!* (1937-1940). Online for the period 1939-1940: <<https://www.belgicaperiodicals.be/link/opac/12564577>>. For the involvement of Jean Meuwissen in these titles, see document 184 in his prosecution file (Military Court, June 6th, 1947) held in the Belgian State Archives.

308. Illustrated weeklies which are at disposal in KBR collections but have not (yet) been digitised, partially because of varying dimensions exceeding mass digitisation capacity, include the (cheap) letterpress magazines *Le National illustré*, *De Zweep*, *Het Huisgezin* en *De Nieuwe illustratie*, as well as a number of general interest letterpress weeklies by publishers that also produced rotogravure magazines (such as *La Semaine d'Averbode/De Week van Averbode*).



Fig. 4. Relative dimensions of the interwar rotogravure magazines published by Rossel (Brussels). Source: KBR.

except for the two first series of *Le Soir illustré* (1928-1929), which act as an exception to the rule as this weekly had, by December 1929, been dressed to a standard tabloid format to remain unaltered until long after the Second World War (Fig. 4). The example of *Le Soir illustré* shows that a large format rotogravure magazine, even one with relatively high print runs was simply not economically viable in as small a market as Belgium.

In fact, *Le Soir illustré* appears as the largest in size of the 17 Belgian magazines withheld from corpus level C (1929-1936), all of which were entirely or partially shaped by rotogravure printing, except for *Ons Volk*, one of the eldest Flemish magazines (°1911) that eventually shifted to rotogravure print as late as 1938 (Fig. 5).

Without elaborating here on the technical aspects in detail, it is essential to point out that rotogravure printing, which became the dominant process for the production of illustrated magazines in the interwar years, is basically a photo-mechanical reproduction practice which applies the principles of the photogravure process developed in the late 19th century on an industrial scale by using engraved copper cylinders mounted on a rotation press machine, hence the term ‘rotogravure’ (‘roto-photogravure’ would be a better term).³⁰⁹ Rotogravure com-

309. For a deeper technical and historical understanding of halftone printing and the photogravure/rotogravure process, we rely on and recommend the contemporaneous work by the Belgian print



Fig. 5. Relative dimensions from big (top left) to small (bottom right) of the 17 Belgian family magazine titles grouped in corpus level 'C', represented by single pages described in the *Artpresse catalogue* (level 'D'). The smallest (*Soirées*) is approximately half the size of the biggest (*Le Soir illustré*). *Ons Volk*, in letterpress but on luxury paper, is the second smallest (a newsprint quality edition existed also, which would have come out less distinctive in this comparison).

binés two distinct but highly compatible innovations relevant to the study of fine art discourses and visual remediation in mass print. Firstly, the photogravure process produces halftone reproductions of far superior quality compared to any relief (letterpress) printing method, making it especially suitable for including large photographic reproductions of high-tone images, such as paintings. Secondly, the rotogravure process allows for an almost limitless graphic design liberty as page layouts are composed on transparent substrates with separate photo positives on celluloid film and typographic positives on cellophane paper, including manual writing, all placed on top of a light table, allowing for cutting and pasting images, the placing of text and graphic elements in tilted positions, superposition (such as text over background silhouettes), and ultimately complex montages. All of these elements challenge, and often disrupt, the traditional dominance of text columns in traditional print, potentially elevating page design to an artistic endeavour in its own right.³¹⁰ Often recognisable as a result

expert SEYL, Antoine. *La rotogravure. Héliogravure en creux rotative*. Brussels: Imprimerie scientifique et littéraire, 1928, as well as the more recent works by STULIK, Dusan & KAPLAN, Art. *Halftone*. Los Angeles: J. Paul Getty Trust, 2013, and Idem. *Photogravure*. Los Angeles: J. Paul Getty Trust, 2013.

310. FRIZOT, Michel. Photo/graphismes de magazines: les possibles de la rotogravure, 1926-1935. In: *Photo/graphisme. Actes du colloque « Photo / Graphisme » tenu au Jeu de Paume le 20 octobre 2007*. Paris:

of these design features or the subtle tonality in the images, rotogravure printing in magazines is in most cases easily detected by the colour of the printing inks, which in contrast to relief printing need to be very fluid and lean in order to fill the innumerable microscopic recessed cells that are etched in varying depths in the copper cylinder (or plate) while the excess ink has to be easily scraped off the cylinder before printing the paper. As rotogravure results in a ‘velvety’ sensation of the printed page, the use of black ink produces a very dull or ‘dead’ impression. It is for this reason that printing inks in early rotogravure are almost always coloured (ranging from light sepias, oranges and pinks to all shades of blue, green, brown and purple). Yet, as is the case for the physical dimensions of the magazines, the printing process defining their overall visual and graphic identity is not always determinable when solely relying on digital facsimiles in 300 dpi. This is particularly the case when different printing processes are combined within the same magazine (rotogravure, letterpress, litho/offset), such as in special editions, or when rotogravure is applied without process-specific design features while printed in darker inks, or when the halftone relief cliché of a large picture is made with a high definition screen and printed on luxury paper (as can occasionally be seen in *Ons Volk*).

Rotogravure printing in mass media began with its first successful application in the 1910 Easter special of the *Freiburger Zeitung* and the first complete pictorial newspaper section in rotogravure in the 1912 Christmas edition of *The New York Times*, the cover of which holds a full-page reproduction of ‘the first Correggio painting shown in America’.³¹¹ Until further evidence emerges, we assume the photogravure process was first introduced in Belgian illustrated magazines with the special Christmas 1913 joint edition of *Le Patriote illustré*, *Le Globe illustré* and *L’Illustration européenne*, printed in Brussels by Victor Ernult-Doncq, all three of which contain 4 bound-in supplement pages printed in *béliogravure* (as the French call it), with reproductions of master paintings curated in Belgian

Editions de Jeu de Paume, 2008, p. 5-12; ZERVIGÓN, Andrés Mario. Rotogravure and the Modern Aesthetic of News Reporting. In: HILL, Jason & SCHWARTZ, Vanessa (eds). *Getting the Picture. The Visual Culture of the News*. London: Routledge, 2015, p. 197-205.

311. <<https://graphicarts.princeton.edu/2017/12/25/first-american-rotogravure-section-december-1912/>>.

museums.³¹² The first weeklies in Belgium printed entirely in rotogravure, however, appear to have emerged to report on the outbreak and developments of the war since August 1914.³¹³ Once under German censorship, which began in 1915, these magazines were printed by Constant Van Cortenberg in Brussels, who later would print the Rossel magazines *Voir & Lire* (°1926) and *Le Soir illustré* (°1928).³¹⁴ As argued above, the First World War was disruptive for Belgian periodicals as many titles, and most of the rotogravure weeklies which appeared under censorship, did not survive. Few rotogravure weeklies seem to have originated or persisted in 1920s Belgium, although *Ons Land in Woord en Beeld* (°1919), published by Jean-Baptiste Felix³¹⁵ remained in publication, unlike its French-language counterpart *Notre Pays*.³¹⁶ In France, rotogravure magazines apparently disappeared for several years after the First World War until the founding of the well-studied, innovative and highly influential weekly

312. Until now only the edition of *Le Globe illustré* 28:51 (21-12-1913) has been digitised: <<https://uurl.kbr.be/2153742>>. This extraordinary edition, also featuring different high-quality color prints, is announced as a novelty in the previous issue 28:50 (14-12-1913), p. 9: <<https://uurl.kbr.be/2153740/p9>>.

313. *1914 illustré* (1914-1918). Online: <<https://www.belgicaperiodicals.be/link/opac/13162817>>.

314. STRUYE, 1980, p. 71. *L'Événement illustré* (1915-1920). Online: <<https://www.belgicaperiodicals.be/link/opac/17463758>> and its short-lived Flemish counterpart *De Week* (1915). Online: <<https://www.belgicaperiodicals.be/link/opac/11498048>>. Another example is the short-lived *Le Temps présent* (1914). Online: <<https://www.belgicaperiodicals.be/link/opac/11487470>>.

315. *Ons Land* was published by Jean Felix and later on by N.V. Ons Land and Nationale Uitgeverij and was eventually printed by Charles Bulens in Brussels. The history of the catholic publisher-printer Felix, the precise evolution of the publication of *Ons Land*, situated until 1932 at the address 9-11 Rue Sterckx in Brussels, as well as the precise nature of the relationships which Felix held with notably Jan Meuwissen and the *Société anonyme de rotogravure* (S.A.R.) remains highly unclear. For an introduction on this subject, see: LIESEN, Bruno. Sur les traces des imprimeurs bruxellois dans l'entre-deux-guerres: l'imprimerie J. Felix et fils. *Histoire et civilisation du livre*. 2018, vol. 14, p. 81-93. In the very early 1920s Felix also printed the first two series of the youth weekly *La Jeunesse* in rotogravure (1920 to July 1922), online: <<https://www.belgicaperiodicals.be/link/opac/17488280>>, as well as the Flemish catholic weekly *De Vlaamse Werkman* (1921-1922). Online: <<https://www.belgicaperiodicals.be/link/opac/15242899>>. In this period Felix also published two sports magazines in rotogravure, *l'Illustration Sportive* in French from 1921 to 1923 (KBR III 99.961 C) and *Sport Illustratie* in Dutch from 1921 to 1924 (III 99.961 C), both not yet digitised from the seemingly complete collections held in KBR.

316. *Notre Pays: revue panoramique belge* disappeared in 1924 when Felix started the rotogravure monthly *L'Illustration congolaise* (1924-1940). Online: <<https://www.belgicaperiodicals.be/link/opac/17364680>>.

VU by Lucien Vogel in 1928.³¹⁷ In the Netherlands, conversely, the rotogravure technology found an early pioneer in Levie Levisson (1878-1948) who established the Nederlandsche Rotogravure Maatschappij (N.R.M.) in Leiden in 1912, printing *Panorama* for publisher Sijthoff since 1913, followed by catholic magazine publisher De Spaarnestad, who acquired a rotogravure press machine in 1915 to print *Katholieke illustratie*.³¹⁸

Arranged in order of appearance (Fig. 6), the set of 17 titles in the Artpresse corpus level C highlights the interwar rise of illustrated magazines printed entirely or partially in rotogravure from the end of the 1920s. When Rossel launched *Le Soir illustré* in 1928, the Dutch publisher De Spaarnestad entered the Flemish magazine market with the 24 page rotogravure magazine *De Stad Antwerpen* marketed as a 'regional' edition of *De Stad Amsterdam*.³¹⁹ That same year also marks the establishment of the Anderlecht-based printing company Société Anonyme de Rotogravure d'Art (S.A.R.), which would print many youth magazines of the Catholic Action in the following decade.³²⁰ In 1930, the Catholic newspa-

317. FRIZOT, Photo/graphismes, p. 8.

318. OVINK, *Het aanzien*, p. 19. VAN TRIEST, Jaap. Vormgeving. In: VAN DELFT, Marieke et al. (eds.). *Magazine! 150 jaar Nederlandse publiekstijdschriften*. The Hague: Koninklijke Bibliotheek, 2006, p. 50-51. *Katholieke illustratie* 50:1 (1915), p. 2 provides a photo and report on the acquisition by De Spaarnestad of the rotogravure installation, commenting that the colour of the ink to be used is not yet determined. Online: <https://resolver.kb.nl/resolve?urn=MMKDC09:017318002:00002>. Although The Netherlands were spared from the war, Ovink and others have considered the period 1914-1934 as an artistic and technical low point for Dutch illustrated magazines.

319. *De Stad Antwerpen* introduced a free accident insurance for subscribers, a commercial feature to attract readers that was swiftly taken over by most Flemish and even some French-language Belgian family magazines in the early 1930s. The first series of *De Stad Antwerpen*, from March 1928 to March 1929, is unfortunately absent from KBR's collections.

320. The Catholic Action, as a broad international lay movement within the Catholic Church inspired by the papal encyclical *Rerum Novarum* (1891), was determined to fight secularisation and win back the masses after World War I. Therefore, a whole range of organisations was created since the 1920s and multiple mass media rotogravure magazines aimed at youth and young labourers saw the light of day. Among the interwar Catholic Action youth magazines printed in rotogravure by S.A.R. (directed by Oscar Boute) and digitised in the Artpresse corpus level B, we mention *JOC* (1930-1940). Online: <https://www.belgicaperiodicals.be/link/opac/12241947>>; *KAJ* (1933-1940). Online: <<https://www.belgicaperiodicals.be/link/opac/11860624>>; *En Route* (1932-1935). Online: <<https://www.belgicaperiodicals.be/link/opac/11710617>>; *Mon Avenir* (1932-1933). Online: <https://www.belgicaperiodicals.be/link/opac/11754066>>; *Op Weg* (1932-1933). Online: <<https://www.belgicaperiodicals.be/link/opac/13041970>>; *Onze Toekomst* (1931-1933). Online: <<https://www.belgicaperiodicals.be/link/opac/11630301>>; *De Klaroen* (1933-1940). Online: <https://www.belgicaperiodicals.be/link/opac/12834345>>.

per publisher De Vlijt (*Gazet van Antwerpen*) installed a rotogravure press for its *Antwerpen 1930* exposition newspaper, which led to the start of *Zondagsvriend* in that December.³²¹ A few weeks later De Spaarnestad launched a 32 page Flemish edition of its rotogravure flagship magazine *Katholieke illustratie* at the price of 2 Belgian francs, quickly aligning its ‘neutral’ *De Stad Antwerpen* to a similar format and thus setting rather high standards for the competition. In February 1931, the photo reportage weekly *Wereldrevue* was launched as the Antwerp-based publisher Patria acquired proper rotogravure printing facilities, and the end of that year witnessed the arrival of two French-language magazines which while in smaller dimensions, held 32 or more rotogravure pages augmented with letterpress radio supplements. First came *Soirées*, published by Éditions Rex and the Catholic Action, followed within less than three months by the neutral *Hebdo* published by Pim (*La Meuse*) in Liège which was also exported to France.³²² Early in 1932 the press affiliated to the Belgian Workers Party started the 32-page rotogravure magazines *ABC* in Dutch and *A-Z* in French, both of which relied on the commercially driven initiative of Dutch magazine entrepreneur Jan Meuwissen who had them printed throughout the 1930s by the N.R.M. in Leiden.³²³ At the same time Patria in Antwerp launched *Tout* as a French-language counterpart of *Wereldrevue*, both of which were printed on alternating rotogravure and letterpress pages. In the meantime, the monastic Averbode publishers also made a significant investment in rotogravure printing machinery,³²⁴ which then formed the main motivation behind their acquisition of *Ons Land*. Averbode seamlessly continued the publication of *Ons Land* in June 1932 on 32 instead of 16 roto-

321. The 50th issue of *Antwerpen 1930* (17 June 1930), dedicated to the pavilion of *Gazet van Antwerpen* on the world exposition site, features an interesting ‘making of’ photo report on the different stages in the rotogravure process: <<https://uurl.kbr.be/1943359>>.

322. Pim also launched the small sized and cheap 16 page rotogravure weekly *Pim illustré* in 1934 (Online: <<https://www.belgicaperiodicals.be/link/opac/12130836>>), which has not been withdrawn in the corpus level C as Belgian fine arts are not a subject of interest. It was soon continued as *Paris Monde* (1934-1936), clearly illustrating the international ambitions of its publisher. Online: <<https://www.belgicaperiodicals.be/link/opac/22104794>>.

323. Also a Luxembourgish edition *A-Z Luxemburger illustrierte Wochenschrift* was released as from December 1933. Online: <https://luxemburgensia.bnl.lu/info1_2/azillust.html>.

324. GHESQUIÈRE, Rita & QUAGHEBEUR, Patricia. *Averbode, een uitgever apart 1877-2002*. Leuven: Universitaire pers, 2002, p. 49.

gravure pages, followed a few weeks later by the launch of a French-speaking counterpart *Echos* which proved not to be viable and disappeared the next year, after which Averbode started printing *Soirées*. Before the end of 1932, Pim added *Vie Heureuse*, a title conceived in the vein of *Hebdo* but offering more artistic and literary content, to its publishing catalogue. The emergence of rotogravure weeklies from catholic, neutral, independent, and socialist signatures did not appear to prompt the Belgian Liberal Party and its press organs to respond competitively until the launch of *Libra illustré* in 1933. Similarly, liberal newspaper publisher Julius Hoste (*Het Laatste Nieuws*; *Het Rijk der Vrouw*), faced stiff competition from Patria, particularly when the latter launched the neutral newspaper *De Dag* in November 1934. Despite this, Hoste did not introduce a rotogravure weekly alongside his ‘cheap’ letterpress weekly *De Zweep* (°1869) until the release of *De Illustratie* in April 1936. However, like *Libra illustré*, it was short-lived and had disappeared by 1937.

Magazine title	first (>1928)	last (<1937)	Publisher	Printer	Printing process(es)	Nn	Np	# items
<i>Le Soir illustré</i>	05/01/1929	26/12/1936	Rossei	Van Cortenberg	Rotogravure	415	11212	317
<i>Ons Land (I)</i>	05/01/1929	28/05/1932	Nationale uitgeverij	Charles Bulens	Rotogravure	180	2916	114
<i>La Patriote illustré</i>	06/01/1929	27/12/1936	La Libre Belgique	Ernult-Donoq	Rotogravure/Relief	416	13326	88
<i>Ons Volk</i>	06/01/1929	27/12/1936	De Standaard	De Standaard	Relief	417	15858	105
<i>De Stad</i>	22/03/1929	25/12/1936	De Spaarnestad	De Spaarnestad	Rotogravure	406	11704	110
<i>Zondagsvriend</i>	14/12/1930	25/12/1936	De Vrijt	De Vrijt	Rotogravure	314	7668	65
<i>Katholieke illustratie</i>	07/01/1931	21/03/1935	De Spaarnestad	De Spaarnestad	Rotogravure	220	7032	69
<i>Wereldrevue</i>	05/02/1931	26/12/1936	Patria / Averbode	Patria / Averbode	Rotogravure/Relief	307	9544	64
<i>Soirées</i>	10/10/1931	06/05/1936	Rax	S.A.R / Averbode	Rotogravure/Relief	234	8040	57
<i>Hebdo</i>	11/12/1931	25/12/1936	Pim	Pim	Rotogravure/Relief/Offset	264	23062	36
<i>ABC</i>	13/01/1932	27/12/1936	Het Licht/Ontwikkeling	N.R.M.	Rotogravure	257	8348	90
<i>Tout</i>	15/01/1932	09/12/1934	Patria	Patria	Rotogravure/Relief	152	4804	65
<i>A-Z</i>	27/03/1932	27/12/1936	Meuwissen	N.R.M.	Rotogravure	250	8832	139
<i>Ons land (II)</i>	04/06/1932	26/12/1936	Averbode/Altora	Averbode/Altora	Rotogravure	229	7366	90
<i>Echos</i>	31/07/1932	23/07/1933	Averbode	Averbode	Rotogravure/Relief	52	1092	8
<i>Vie heureuse</i>	14/12/1932	25/12/1934	Pim	Pim	Rotogravure/Relief/Offset	106	6118	55
<i>Libra illustré</i>	15/04/1933	14/06/1935	Libra illustré	?	Rotogravure	110	2788	50
<i>De illustratie</i>	01/04/1936	30/12/1936	Hoste	?	Rotogravure	40	942	11

Fig. 6. List in chronological order of appearance (1929-1936) of the 17 magazines withheld in Artpresse corpus level C (1929-1936): publishers; editors; print technology used; number of issues and pages. The column in blue shows how many items on fine art related subjects were drawn from these titles for the level D.

While the digitised Artpresse corpus reflects the shift to rotogravure in Belgian magazine culture during the interwar years, the digitisation from the bound physical

collections at KBR requires careful interpretation and ideally additional metadata on material characteristics. The current efforts can only pull back part of the curtain to the medium-specific technological, editorial and commercial processes underlying the present corpus of digitised magazine runs. Coherent, reliable information within the resources themselves as well as contemporaneous documentation and archives on these factors are rare as hen's teeth, when combined with the subsequent lack of scholarly attention paid to commercial family magazines of the pivotal interwar period, leaves many aspects of this voluminous print heritage hidden from view. Although progress has been made in identifying creative and editorial actors—such as authors, editors, publishers, and artists—to establish intermedial relationships for analysing and comparing fine art discourses, further efforts are essential if we are to comprehensively map the socio-material networks and commercial and creative interactions at play between these similar publications. This also concerns basic questions like the actual circulation of the highlighted family magazines in the 1930s. Print run figures, as found in diverse sources, indicate strong fluctuations over the decade. They are of varying levels of reliability, however, but support the general estimation of approximately a dozen magazine titles in Belgium attaining average print runs of ca. 50,000 copies or more in 1934.³²⁵ This level may have served as a threshold at which substantial investments in rotogravure presses and cylinders became economically profitable. These 'big' magazines would include *ABC* and *A-Z* (ca. 70,000),³²⁶ *De Stad* (claiming a 'guaranteed' run of 61.500 copies on its covers), *Ons Land*,³²⁷ *Zondagsvriend* (attaining 100,000 copies in 1934),³²⁸ *Le Soir illustré*³²⁹ and probably also *Hebdo* (which claimed a 'normal' cir-

325. LINEPHTY, Maurice Guillaume. A propos de la presse périodique belge. In: *Le musée du livre*, 1934, 29:7, p. 50. Online: <<https://uurl.kbr.be/2186283>>.

326. *ABC* would maintain this average until 1936: *ABC*, 1936, 5:1, 3. Online: <<https://uurl.kbr.be/1900775/p3>>. According to the documents in the Meuwissen files (Military Court, June 6th 1947), *A-Z* reached an average of 75,000 copies in 1933, which declined in the late 1930s, when the name changed to *Bonjour!*, to a mere 20,000 and less.

327. GHESQUIÈRE, *Averbode*, p. 505.

328. GHIJS, *100 jaar Gazet van Antwerpen*. Antwerp: De Vlijt, 1991, p. 57.

329. Such could be assumed by the fact that newspaper *Le Soir* attained print runs over 300,000, the highest in Belgium according to LUYKX, Theo. *Evolutie van de communicatiewetenschap*. Brussels: Elsevier, 1978, p. 509 and DE BENS, Els & RAEYMAECKERS, Karin. *De pers in België*. Tiel: Lannoo, 2010, p. 41.

culatation of 45,000 in 1932)³³⁰ and *Soirées*.³³¹ This list can be completed with *Femmes d'Aujourd'hui* and *Het Rijk der Vrouw*.³³² Some of the many rotogravure periodicals spawned by the Catholic Action may have well attained print runs of over 70,000 copies.³³³ As for the other titles in the Artpresse corpus level D, we either did not find any information or have only circumstantial evidence, such as short-lived existence (*Echos*; *De Illustratie*; *Libra illustré*), comparative publicity prices³³⁴ or estimations based on affiliated newspaper print runs, such as for *Ons Volk*, which may not have attained 50,000 copies, the print run of *De Standaard* in 1935, until its switch to rotogravure in 1938. For the partially rotogravure magazines *Tout* and *Wereldrevue* we did not find any concrete numbers, but their publisher Patria claimed a total of 1.5 million copies for its entire production in 1931 (magazines and other publications), of which presumably the boulevard letterpress weeklies *Mon Copain* and *Piccolo* reached over 100,000 copies already in the early 1930s.³³⁵

Fine Arts in Family Magazines

Content identified as related to the fine arts most often appeared as captioned pictures in news sections — to mark the unveiling of a statue, for instance — and as dedicated articles about an artist (usually accompanied by a photo portrait and reproductions of his works). Added to these were a variety of ‘diverse articles’ on group exhibitions, public collections, museums, monuments... which showcased the many explicit or implied interpretations of ‘fine arts’ in the interwar magazine culture. We aim to assess their significance as well as to determine related events, artists and iconography, through a critical selection of such content within the 17 magazine titles listed in Fig. 6. The content descriptions now form corpus level D (Fig. 1) and

330. *Hebdo*, numéro spécial Noël, 1932. Online: <<https://uurl.kbr.be/1915307/p11>>.

331. DENIS, Jean. *Bilan. Soirées 109* (1933), p. 3. Online: <<https://uurl.kbr.be/1927412/p3>> claims an exceptional print run of 85,000 for the special issue on the Rex movement (n° 99).

332. The Meuwissen file (Military Court, June 6th 1947) document 184 mentions 92.000 copies for *Femmes d'Aujourd'hui* and 45,000 for *Het Rijk der Vrouw* in 1936.

333. For a contemporaneous overview from the mid-1930s, see: *De persmacht der Katholieke Arbeidersjeugd*. KAJ, 1934, 2:1, p. 8-9. Online: <<https://uurl.kbr.be/1970041/p8>>.

334. *Annuaire publicitaire et commercial de Belgique*. Liège: Pim, 1932; 1933.

335. GEUKENS, Eric. *Het Dagblad «De Dag» (1934-1944)*. Rijksuniversiteit Gent, 1975, p. 26-27.

function as a limited database which we call the Artpresse Catalog. In a further stage we intend to publish this catalogue in the form of a multi-page PDF and make the dataset available on SODHA.³³⁶ While not included in the initial objectives of the project, this catalogue emerged as a response to evolving research needs driven by the substantial volume of material, aiming to facilitate more efficient navigation of art historical data. Creating this document has required continuous adjustments and careful consideration, as each publication offered a unique perspective on current (art) events and, more broadly, on modern life in the interwar years, and involved a close-reading methodology based on page-by-page reviews and text-based queries in BelgicaPeriodicals. The descriptions are linked to the digital portal and provided with an index of authors and artists mentioned as subjects which allows for easy organising and filtering of the data. Making use of Wikidata and analogue research instruments, the artists index plays a crucial role in correcting and standardising names, given frequent spelling errors, language variations between French and Dutch, and potential misattributions. In parallel, all relevant digitised magazine pages (jpg) were downloaded from KBR's storage and integrated in the Artpresse catalog as visual support of the 1,531 entries, grouping 2,173 magazine pages.³³⁷ Additionally, several pages are devoted to each magazine, addressing both editorial and material aspects while briefly discussing the selected content in a didactic effort to improve access to these magazines, given the lack of reference works on Belgian periodical print heritage during the interwar years.

The study commences in 1929, a pivotal year in historiography, where the stock market crash serves as a crystallising event for the dismantling of modernist networks during the early 1930s. Although a “tendency to withdraw into oneself, towards nationalism” was already perceptible during the 1920s as suggested by Devillez, this financial collapse precipitated the liquidation of avant-garde galleries and collections (mostly in Brussels), fostering a climate of mistrust and derision toward living art that was amplified by the contemporary (francophone) press.³³⁸ As

336. SODHA is the Belgian federal data archive for social sciences and the digital humanities: <<https://www.sodha.be/>>.

337. This represents a mere 1.5% of the total number of pages in corpus level C (ca. 150,000).

338. DEVILLEZ, Virginie. Les peintres belges dans la tourmente. Du krach économique à la seconde guerre mondiale. *CHTP-BEG*. 1997, vol. 2, p. 36-38. Online: <https://www.journalbelgianhistory.be/fr/system/files/article_pdf/chtpt2_002_Devillez.pdf>; NAHON, Valérie. Profil d'une critique

the controversies took primarily place in newspapers and specialised magazines (*Le Rouge et le Noir*, *Les Beaux-Arts*, *L'Art belge*), references can be found throughout illustrated periodicals, primarily opinion magazines or those offering humorous and satirical content (*Pourquoi pas?*, *Moustique*, *Koekoek*). Conversely, family magazines depicted a vibrant cultural scene driven by both public and private initiatives (retrospectives, group shows, world fairs, publications), often promoting 'eclecticism' and 'fine craftsmanship' ('*beau métier*' in French) in a visually engaging and non-polemical tone (albeit more polemical accounts occasionally surface). Alongside efforts to revive the economy, the celebration of the centenary of Belgium's independence served to consolidate nationalist, regionalist, and traditionalist discourses within (print) media. The interbellum period provides an opportunity to reexamine the so-called 'crisis years' and manifestations related to the Return to Order in mass-market print culture, considering this pivotal moment in the history of Belgian art in relation to the surge in family magazines occurring in the late 1920s/early 1930s. A recent hypothesis based on re-examination of KBR's periodical collections also suggests that advancements in rotogravure for large-scale illustrated print production prompted Belgian printers and publishers to expand their art coverage by leveraging the enhanced ability to reproduce detailed photographs and illustrations. As for the ending point, 1936 marked a period of political and visual transition with the re-launch of *Life* which, though originally founded in 1883, established itself that year as a seminal archetype for the modern newsmagazine. This shift also mirrored the shockwaves caused by the electoral breakthrough of the far-right Rexist movement in Belgium, alongside rising international tensions in Europe.

Initially centred on Belgian living, modern artists and their reception in popular magazines, the scope of the research has been expanded to more accurately capture mainstream discourses related to the fine arts. It now includes, for instance, 'old [Flemish] masters', like Rogier van der Weyden (1399-1464), Quinten Metsys (1466-1530), Pieter Bruegel (1526-1569), Peter Paul Rubens (1577-1640), in relation to current events (such as exhibitions, restorations, new attributions,

moderne. Charles Bernard et la défense de l'art vivant dans l'entre-deux-guerres. *Textyles*. 2010, Vol. 39, p. 63-84. Online: <<https://journals.openedition.org/textyles/101>>. DEVILLEZ, Virginie. The Avant-Garde on the Reworking of Tradition. In: AUBERT, Nathalie, FRAITURE, Pierre-Philippe & MCGUINNESS, Patrick (eds.). *From Art Nouveau to Surrealism, Belgian Modernity in the Making*. London: Legenda, 2007, p. 146.

or thefts), and ‘recently deceased’ artists, like Jakob Smits (1856-1928), Auguste Oleffe (1867-1931), Louis Thevenet (1874-1930), whose names and legacies were quickly commemorated by retrospective exhibitions. The treatment of world expositions resulted in numerous reproductions of artworks from public and private collections, along with aerial and monumental views of pavilions and installations, as part of narratives designed to spark the curiosity and emotions of readers as potential visitors.³³⁹ The palaces and sections dedicated to ‘ancient’, historic arts, such as *Oude Vlaamsche Kunst* in Antwerp in 1930 and *Five Centuries of Art* in Brussels in 1935, curated by Paul Lambotte (1862-1939), were part of a tradition of exhibiting and celebrating late medieval and baroque arts in the former Spanish Netherlands, one of the first major events being *The Flemish Primitives* in Bruges in 1902, which attracted 35,000 visitors.³⁴⁰ This interest was revived in subsequent exhibitions, such as *Arts of the 17th Century* at the Brussels World’s Fair in 1910, and the retrospective of Belgian art in London in 1927-1928, whose importance has been highlighted in *Le Soir illustré*.³⁴¹ Brussels 1910 was renowned for its extensive print communication campaign,³⁴² an approach that was reiterated on an industrial scale during Brussels 1935, likely with the help of publishers-printers (Marcel van der Donck, Pim, C. Van Cortenbergh, Weissenbruch) and coordinated by René Lyr, pseudonym of René Vanderhaeghe (1887-1957), and the Press and Information Service for publishing the official bulletins and guides to the World Fair.³⁴³ René Lyr was a central figure who

339. In this regard, examples can be found notably in publications related to the fairs (*Antwerpen 1930* was printed in the pavilion of the publisher Gazet Van Antwerpen; *1935 Bulletin...*), as well as in *De Stad*, *Le Patriote illustré*, *Le Soir illustré*, and others.

340. PAUMEN, Vanessa. *The Flemish primitives: What’s in a name?* Online: <<https://vlaamseprimitieven.vlaamsekunstcollectie.be/en/research/webpublications/the-flemish-primitives-whats-in-a-name/>>.

341. LAMBOTTE, Paul. L’exposition d’art italien à Londres. *Le Soir illustré* 3:101 (25-01-1930), p. 6-7. Online: <<https://uurl.kbr.be/1952352/p6>> (LSI-0040); LIEBRECHT, Henri. Cinq siècles d’art. Maîtres de jadis. *Le Soir illustré* 8:392 (24-08-1935), p. 16-18. Online: <<https://uurl.kbr.be/1954193/p16>> (LSI-0286).

342. JOURDAIN, Virginie. Promouvoir l’Exposition: le règne de l’image et de la presse. JAUMAIN, Serge & BALCERS, Wanda (eds). *Bruxelles 1910: de l’exposition universelle à l’université*. Brussels: Dexia-Racine, 2010, p. 94-98.

343. *Livre d’or de l’Exposition universelle et internationale de Bruxelles 1935*. Brussels: Comité exécutif de l’Exposition, 1935, p. 127-129.

has sparked numerous Artpresse investigations, having contributed to various publications and mass media during the interbellum, whether printed or not.

The ‘Old Masters’, who enjoyed institutional and public acclaim, were prominently featured in dedicated Belgian sections, based on their connection to the host city: Quentin Metsys, Hans Memling, Antoon Van Dyck, Pieter Bruegel, Peter Paul Rubens in Antwerp 1930; Rogier Van der Weyden, Bernard Van Orley, Pieter Bruegel, David Teniers, Hyppolite Boulenger, Constantin Meunier in Brussels 1935. They have been invoked in discourses reflecting various intentions, ranging from the celebration of the ‘Flemish people through their arts’,³⁴⁴ to the reaffirmation of the existence of a Brabant school of landscape painting and a rich Brussels tradition of decorative arts (particularly in tapestry).³⁴⁵ Antwerp 1930 had ‘ancient art’ as one of its central themes; the event was titled *International Colonial, Maritime, and Flemish Art Exhibition*. As for Brussels 1935, the press presented the *Five Centuries of Art* exhibition as one of the highlights of the World Fair.³⁴⁶ Bruegel has been particularly celebrated, notably through a monument project in his honour, which was solemnly inaugurated at Heysel in August 1935, featuring inscriptions in both French (‘À Bruegel’) and Dutch (‘Aan Brueghel’), intended for the Place de la Chapelle, but which ultimately never came to realisation.³⁴⁷ The confusion between the works of Bruegel the Elder and the copies by his son, today at-

344. ANON. De tentoonstelling van de oude Vlaamsche kunst. Het levensgevoel bij de Vlaamsche meesters. *De Stad* 3:23 (21-08-1930), p. 529-533; 540-541. Online: <[https://uurl.kbr.be/1921563](https://uurl.kbr.be/1921563>)> (DST-0014).

345. FLOUQUET, P.L. Cinq siècles d’art bruxellois à l’exposition de 1935. *1935 Bulletin officiel de l’exposition universelle* 1:4 (15-11-1933), p. 139-142. Online: <[https://uurl.kbr.be/2125406/p21](https://uurl.kbr.be/2125406/p21>)>; L.R. L’art ancien. *1935 Bulletin officiel de l’exposition universelle* 3:30 (01-06-1935), p. 519-521. Online: <[https://uurl.kbr.be/2125458/p9](https://uurl.kbr.be/2125458/p9>)>.

346. DEVILLEZ, Virginie. *Le retour à l’ordre* (2002), p. 70-71.

347. ANON. Actualités A-Z. (À l’exposition de Bruxelles, un monument vient d’être élevé à la gloire de Pieter Breughel). *A-Z* 4:22 (18-08-1935), p. 14 (AZ-0116). Online: <[https://uurl.kbr.be/1901667/p14](https://uurl.kbr.be/1901667/p14>)>; ANON., L’inauguration du monument Brueghel l’Ancien à l’exposition. *Le Patriote illustré* 51:33 (18-08-1935), p. 1029 (LPI-0066). Online: <[https://uurl.kbr.be/1953826/p5](https://uurl.kbr.be/1953826/p5>)>; ANON., Onthulling van een Brueghel-Gedenkteeken op de tentoonstelling te Brussel. *1935 Bulletin officiel de l’exposition universelle* 3:41 (17-08-1935), p. 1042. Online: <[https://uurl.kbr.be/2125480/p36](https://uurl.kbr.be/2125480/p36>)>; <<https://www.belgiansculptures.be/product/a-wansart-pieter-brueghel>>.

tributed as Pieter Brueghel the Younger³⁴⁸ (not to be mistaken for Jan Brueghel the Younger (1568-1625)) was common, likely fuelled by the widespread circulation of reproductions ‘after Brueghel’ in the interwar years. Without clear distinction between the two, enthusiasm for Bruegel during the period seems to have emerged in 1924 with the celebration of his supposed 400th anniversary,³⁴⁹ and was subsequently renewed in Christmas special issues of Flemish Catholic illustrated magazines up to the 1940s. He has been instrumentalised by authors and art historians within the context of the Flemish cultural movement, as well as being seen as a major influence for Belgian artists, and a key argument in the defence of living art.³⁵⁰

Moreover, Artpresse provided an opportunity to take a fresh, more neutral look at the artistic production of the 1930s, which was dominated by late forms of realism and impressionism, more ‘representative’ paintings sometimes described in scholarship as ‘mediocre’ and ‘uninspired’.³⁵¹ Major interwar art historical movements in Belgium (constructivism, surrealism) found little echo in family magazines. Several reasons can be cited for this lack of

348. CURRIE, Christina, ALLART, Dominique. *The Brueg(h)el Phenomenon*. Brussels: Royal Institute for cultural Heritage (KIK-IRPA), 2012, 28. Online: <<https://www.kikirpa.be/fr/publications/the-brueghel-phenomenon>>.

349. BIERMÉ, Maria. Le quatrième centenaire de Pierre Bruegel. *Le Patriote illustré* 40:21 (25-05-1924), p. 328-329. Online: <<https://uurl.kbr.be/1938356/p8>>; ANON. Bij het werk van Pieter Breugel. *Ons Volk* 10:37 (14-09-1924), p. 466-469. Online: <<https://uurl.kbr.be/1936064>>.

350. Artpresse’s extensive corpus is filled with references to Bruegel at all levels. Such a broad representation is difficult to trace using traditional, analogue methods. Therefore, gaps in the catalogue are highly probable, particularly due to the spelling variations and early investigations focused on living artists and ‘modern’ works. It highlights the ‘vitality’, the relevance of his art, deeply embedded in the interbellum with aesthetic, symbolic, and even political dimensions. In this regard, some notable examples are TIMMERMANS, Felix. Vlaamsch volksleven gezien door Peter Brueghel. *Ons Land*, p. 1358-1359. Online: <<https://uurl.kbr.be/1937085/p16>>; MULS, Jozef, De betlehemse kindermoord. *Zondagsvriend* 14:52 (25-12-1946), p. 27. Online: <<https://uurl.kbr.be/2136714/p27>>; ANON. Kerstmis bij Breughel. *Ons Volk* XXI:51 (22-12-1935), p. 810-813 (ONV-0103). Online: <<https://uurl.kbr.be/1948132/p4>>; SAROLEA, R. [Leo Arras], Vlaamsche koppen. Evarist De Buck. *De Stad* 2:36 (22-11-1929), p. 779 (DST-0005). Online: <<https://uurl.kbr.be/1921485/p15>>; Z.L. [Lode Zielens], De Vlaamsche schilder Albert Saverijs. *De Stad* 4:33 (30-10-1931), p. 795 (DST-0034). Online: <<https://uurl.kbr.be/1921688/p19>>; ANON., Ca représente quoi? *Le Centaure* 1:4 (01-01-1927), p. 75-77. Online: <<https://uurl.kbr.be/2202289/p7>>.

351. HOOZEE, Robert. *Vlaams expressionisme in Europese context 1900-1930*. Ghent: Snoeck-Ducaju, 1990, p. 15; LAUWAERT, Helke. Néo-réalisme. In: HOOZEE, Robert (ed.). *L’art moderne en Belgique 1900-1945*. Brussels: Mercatorfonds, 1992, p. 266.

representation: the collapse of the modernist art market, which reinforced the perception of an overrated, outdated art, causing a withdrawal from artistic life for some; or a deliberate intention by avant-garde artists to distance themselves from official networks and mass distribution channels? As Devillez has stated, these movements were ‘seen as reflections of a predominantly theoretical and internationalist movement in opposition to Flemish Expressionism, which is rooted in regional art logically supported by the administration of Fine Arts’.³⁵² Family magazines attest to this ‘artistic and commercial’ triumph and its many supporters in the Flemish movement, both socialist and catholic, which is not exclusively limited to the ‘artists of Laethem’.³⁵³ An observed tendency wherein certain artists were associated with specific Belgian regions and cities can be seen as a regionalist approach to fine arts, thus familiarising the public with local artists. Examples include Smits the painter of the Campine; Paulus, the painter of the *Pays Noir* (‘Black Country’ referring to a coal mines region in Hainaut); Barthélemy, the painter of the Ardennes; Lier and its three leading artists Isidore Opsomer, Felix Timmermans, and Louis Van Boeckel.

While public statues and monuments were commonly featured on covers and in news sections, some sculptors, like Karel Aubroeck (1894-1986) and Jozef Cantré (1890-1957), found favour within Flemish illustrated weeklies in relation to the theme of the Yser pilgrimage and the centenary of the composer Peter Benoit. As a result, approximately 1,200 artists were indexed, including 852 ‘Belgian’ ones with James Ensor (1860-1949), Isidore Opsomer (1878-1967), and Pierre de Soete (1886-1948) as the top 3 most recurring names in the selected content of the Artpresse catalog. An examination of their portrayal by family magazines shows different aspects of the Return to Order as it manifested in visual arts and criticism in Belgium during the interbellum: James Ensor, the painter of light and of the North Sea, attached to Ostend (see below); Isidore Opsomer, the portraitist of Belgian sovereigns and political figures, director of the Antwerp Institute of Fine Arts and painter of Lier; Pierre De Soete, sculptor of, for instance, the monuments to Marshal Foch in Spa or to the victims of

352. DEVILLEZ, Virginie. *Le retour à l'ordre* (2002), p. 25-26.

353. Ibidem.

the Dinant massacre³⁵⁴ in addition to effigies of celebrities.

Prominent critics and authors such as Paul Fierens (1895-1957), Roger Avermaete (1893-1988), René Lyr and Lode Zielens (1901-1944) gave Ensor notable attention in the 1930s as one of the leading artists of the so-called “Belgian school” in dedicated illustrated articles³⁵⁵ and linked to specific events, for instance: the unveiling of his bust in 1930 in Ostend, and the ‘Parisian consecration’³⁵⁶ with his 1932 exhibition at the Musée du Jeu de Paume, and the subsequent award of the French *Légion d’Honneur* in 1933. These articles have in common their emphasis on Ensor’s painted work, lauding him as the ‘country’s foremost colorist’ (L. Z.), the ‘painter of shells, fish, dunes, beaches, houses, the fiery yet pensive, taciturn yet tense life of the people of the coast’ (René Lyr), a ‘*franc-tireur* [who] started with a mood painting, a subdued palette, and a focus on direct observation, leading to a very clear, imaginative painting free from any strong attachment to reality’ (Roger Avermaete), his family house-shop (echoed in a rich and similar imagery),³⁵⁷ and his late recognition; elements of a myth still upheld today. His numerous speeches in which he glorified himself while denouncing critics and artists,³⁵⁸ and his activity as a music composer also greatly contributed to his success in the interbellum.

354. HOTON, Edmond. Furore Teutonico Diruta. *Libra illustré* 1:23 (30-09-1933), p. 4-5. Online: <<https://uurl.kbr.be/1902425/p4>> (LBI-0007); ANON. Le monument national aux victimes civiles de la guerre. *Le Soir illustré* 9:445 (29-08-1936), p. 2. Online: <<https://uurl.kbr.be/1955283/p2>>; DE SOETE, Pierre. *Les incidents de Louvain. Furore Teutonico... Ses origines et ses polémiques. Le droit artistique et spirituel*. 1914-1929. Brussels: Imprimerie scientifique et littéraire, 1929.

355. Z.L., Vlaamsche koppen. De prins der Vlaamsche schilders James Ensor. *De Stad* 3:35 (14-11-1930), p. 831 (DST-0019). Online: <<https://uurl.kbr.be/1921587/p15>>; EECKELS, Guido. Vlaanderen leeft. James Ensor. *Oms Land* 15:14 (02-07-1932), p. 300-301 (ONL-0117). Online: <<https://uurl.kbr.be/1942462/p28>>; FIERENS, Paul. L’exposition James Ensor à Paris. *Hebdo* 1:31 (08-07-1932), p. 17 (HBD-0007). Online: <<https://uurl.kbr.be/1915261/p17>>; René LYR [Vanderhaeghe, René]. La noble aventure de James Ensor. *A-Z* 2:25 (10-09-1933), p. 3; 6-7 (AZ-0046). Online: <<https://uurl.kbr.be/1901465/p3>>; AVERMAETE, Roger. Beaux-Arts. James Ensor à Paris. *Tout* 1:26 (08-07-1932), p. 27 (TOU-0017). Online: <<https://uurl.kbr.be/1903230/p27>>.

356. FIERENS, Paul. L’exposition James Ensor à Paris. *Hebdo* 1:31 (08-07-1932), p. 17 (HBD-0077). Online: <<https://uurl.kbr.be/1915261/p17>>.

357. ANON. James Ensor. *ABC* 3:44 (25-11-1934), p. 6-7 (ABC-0071). Online: <<https://uurl.kbr.be/1900653/p6>>; ANON., Bij James Ensor. *Wereldrevue* 3:29 (17-08-1933), p. 18 (WLR-0036). Online: <<https://uurl.kbr.be/1903987/p18>>.

358. FLORIZOONE, Patrick & COPPEJANS, Willem. James Ensor tussen avant-garde en traditie. *James Ensor een online Museum*. Online: <<https://jamesensor.vlaamsekunstcollectie.be/nl/bronnen/webpublicaties/james-ensor-tussen-avant-garde-en-traditie/>>.



Fig. 7. Ensor's representation in a selection of items from corpus level D. A photo portrait of the artist published on the cover of *Ons Land* on 09-02-1929 when René Lyr was *bestuurder* ('director') of the publication, together with selected news of the unveiling of the bust. ONL-0003: [Anon.], *James Ensor*. *Ons Land* 11:45 (09-02-1929), p. 705; 715. Online: <<https://uurl.kbr.be/1941640>> – ONL-0053: [Anon.], *Onthulling van het borstbeeld James Ensor te Oostende*. *Ons Land* 13:4 (26-04-1930), p. 59. Online: <<https://uurl.kbr.be/1942145/p11>> – ONV-0018: *Lefèvre, G., James Ensor*. *Ons Volk* XVI:17 (27-04-1930), p. 256-257. Online: <<https://uurl.kbr.be/1947183/p12>> – LPI-0005: [Anon.], *Inauguration du buste d'Ensor à Ostende*. *Le Patriote illustré* 46:16 (20-04-1930), p. 488. Online: <<https://uurl.kbr.be/1951446/p8>>.

Their personal correspondence suggests that Ensor and then director of *Ons Land*, René Lyr, shared a close friendship,³⁵⁹ and we believe we have found within these archives the original photograph used for the cover of *Ons Land*, annotated: 'J'offre à René Lyr mon meilleur portrait. James Ensor 1925'. The same medallion-portrait appeared a year later in *Ons Volk* on the unveiling of Ensor's bust, which suggests that this image was likely circulated by the artist

359. The painter-engraver also drew several portraits of the writer-journalist, one of which having been digitised and published online by KBR: <<https://opac.kbr.be/Library/doc/SYRACUSE/15801762>>.

himself or press or photo agencies, though these networks are difficult to trace. René Lyr appeared alongside Marcel Nyns (1887-1972), then Secretary General in the Belgian Ministry of Public Education, with whom he also corresponded notably regarding the acquisition of artworks for the Belgian state, in a photograph published in the news section of *Ons Land* in April 1930 at the unveiling event of the bust in the garden of the Kursaal in Ostend.³⁶⁰

The decisions shaping the treatment of fine arts, including which artists and works to feature, related to current cultural events and their overall visual presentation, were likely taken by magazine directors and editors-in-chief. Information on these Belgian writer-journalists, their motivations, personal tastes, and relationships with artists can be found, scattered throughout various Belgian archives.³⁶¹ Despite the considerable proportion of anonymous texts in the Artpresse catalogue (ca. 60%), a total of 241 authors could be identified in corpus level D. These are notably: Henri Liebrecht (1884-1955) as host of the fine arts section of *Le Soir illustré*; Charles Conrardy (1893-1957) in *Vie Heureuse*; Roger Avermaete in *Tout*, and Lode Zielens in *De Stad* (using his own name, initials or pseudonyms).³⁶² Other identified contributors include Guido Eeckels (1912-1986) in *Echos*, *Ons Land*, and *Soirées*; Antoine De Decker (?) in *Ons Land* and *Soirées*; Joris Verdin (?) in *De Stad*; Lucie Van or De Maldere (?) in *Libra* and *Le Soir illustré*, and Jef Crick (1890-1965) in *De Stad* and *Ons Volk*. Beyond shedding light on the circulation of certain magazines, archival research has helped us to clarify the identities and roles of specific authors. For instance, consulting the archives of Roger Avermaete and Pierre Chatelain-Tailhade (1904-1977), then director of *Tout*, allowed us to understand the reasons behind the discontinuation of the fine arts section in favour of

360. This marble bust was created by Edmond de Valeriola (1877-1956) and was later replaced by a bronze copy and moved to the Leopold Park following the demolition of the Kursaal. ANON. Onthulling van het borstbeeld James Ensor te Oostende. *Ons Land* 13:4 (26-04-1930), p. 59 (ONL-0053). Online: <<https://uurl.kbr.be/1942145/p11>>; CASIER, Micheline. Monument au peintre James Ensor (Oostende). *Be Monumen. Répertoire participatif des bronzes et fontes belges*, March 2024. Online: <<https://be-monumen.be/patrimoine-belge/monument-au-peintre-james-ensor-oostende-ostende/>>.

361. Namely the Archives et Musées de la Littérature (AML), Brussels; Archives Générales du Royaume (AGR) - Dépôt Cuvelier, Brussels; Archives de l'Art contemporain de Belgique (AACB), Brussels; Amsab-Instituut voor Sociale Geschiedenis, Ghent; Letterenhuis, Antwerp.

362. Carolus, Brabo, Scaldis according to Letterenhuis. 'Lode Zielens'. Online: <<https://collectie.letterenhuis.be/doc/au::21790>>.

‘picturesque’ content.³⁶³

There are still many unknowns regarding Belgian journalists, as can be deduced from the unknown birth and death dates of the above-mentioned authors, and most of the graphic designers and layout artists as well. We acknowledge the inevitably biased nature of the *Artpresse* catalogue, starting with the inductive method guided by art history assumptions and derived questions.³⁶⁴ The selective approach tailored to each magazine has demonstrated that art-specific content and what would today be referred to as ‘cultural’ cannot be easily classified or extracted from its original context. The close-reading methodology relies on full-text searches within *BelgicaPeriodicals*, a process inherently constrained by the technical limitations of OCR accuracy. To mitigate these constraints, an illustrated and indexed catalogue provides a vital visual framework that, despite potential omissions, supports *Artpresse* research by facilitating more efficient navigation through the digitised corpus. The Belgian art history conveyed through family magazines is, as the development of this catalogue has revealed, predominantly commemorative. Public events and their associated iconography played a pivotal role in these media, which were intrinsically designed for visual consumption. By focusing on narratives of the ‘great artists of the past’, these publications aligned with the aesthetic and political agendas of the Belgian Administration of Fine Arts, set within the context of a young, centenarian state.

Conclusion

To establish a representative corpus of digitised Belgian illustrated weeklies capable of addressing the research questions posed by the *Artpresse* project, it was first necessary to engage deeply with the Belgian magazine culture of much of the

363. Letterenhuis, Archief van Roger Avermaete, A 493/B2.

364. One of these methodological biases arises from the predominant focus on fine arts (painting and sculpture) at the expense of applied and graphic arts, a recurring issue in art historical studies. Additionally, questions emerge regarding certain Belgian artists for whom we conducted in-depth investigations while others were overlooked - choices also influenced by a particular conception of art history. This is particularly evident in the case of René Magritte, whose relative absence of relevant results in full-text searches raises concerns about the visual underrepresentation of avant-garde works and movements in mass-market periodicals (cfr. *supra*). On the other hand, these same biases and choices allow us to reflect on how art historical narratives are constructed, with progressive attention to representations and networks linked to, yet distinct from, the fine arts domain.

first half of the twentieth century. The digitisation initiative had to adopt a broad foundation, beginning with an inquiry into the general development of middlebrow magazines within the Belgian print market during the interwar period, as well as undertake a detailed examination of KBR's periodicals collections.

Supported by the introduction of semi-automated digitisation technologies at KBR, this approach enabled us to discover key collections, identify major actors and make a substantial part of French and Dutch language interwar magazines accessible via BelgicaPeriodicals. By focusing on family weekly magazines from the period 1929–1936, we were able to construct a coherent digital corpus from which a clearly defined subset of content was filtered using conventional research methodologies. The resulting Artpresse catalogue thus attempts to describe, contextualise and visualise a very fragmented, rather marginal and thematically well-defined multimodal content within an extensive digital archive in a critical, comprehensively selective but by no means exhaustive way.

As this output remains isolated and closed, it poses new challenges on the integration and progressive accumulation of enriched metadata in the existing data infrastructure. As a manually engendered dataset, the Artpresse catalogue is positioned between the digital archives on the one hand and the analytical promises of *collections as data* on the other. By indexing artists and authors in varying types of articles and news items, it also negotiates rudimentary connections between different categories of linked entities on the one hand and the current possibilities of linked data on the other. In creating this catalogue, we have developed meaningful visual, intermedial and materiality-based ways of viewing family magazines as historical sources even though navigating the resources in BelgicaPeriodicals remains firmly text-based and thus technically and methodologically restricted from these perspectives. The lens of middlebrow mass-market magazines and the generated art historical data have paved the way for new readings centered on news and visual culture, circulation of art reproductions and media representations of Belgian artists, also calling for further case studies and broader applications in Digital Humanities.

12. Studying Comics in the Belgian Press through Digitised Corpora: Challenges and Opportunities

Erwin Dejasse

Abstract

Institutions responsible for the preservation of printed cultural heritage have developed significant digitisation programmes that have profoundly renewed research on comics. As a result, huge corpora that were previously largely unknown have been brought to light. These initiatives have not only uncovered substantial new material, but are also transforming research practices through the introduction of methods drawn from the digital humanities. To highlight these ongoing changes and identify key challenges, this article reports on experiments conducted in this field at the Royal Library of Belgium (KBR).

Author

Erwin Dejasse is doctor in Art History, researcher, teacher and exhibition curator. He oversees BeComicsLab, a FED-tWIN research project on the study and enhancement of Belgian comics heritage, jointly supported by the Université libre de Bruxelles (ULB) and the Royal Library of Belgium (KBR). Moreover, he is lecturer at the Université de Liège (ULiège) and Professor at Saint-Luc School of Arts in Brussels. In 2008, he has co-founded ACME, an interdisciplinary research group on comics. His research focuses on comics, Art Brut and punk visual culture. He has written the books *Morris, Franquin, Peyo et le dessin animé* (Morris, Franquin, Peyo and animated cartoon, with Philippe Capart in 2005), *Art Brut et bande dessinée* (Art Brut and Comics, 2023) et *La Musique silencieuse de Muñoz et Sampayo* (The Silent Music of José Muñoz and Carlos Sampayo, 2024) for which he was awarded the SoBD/9e Art prize in 2025. E-mail: erwin.dejasse@kbr.be.

Keywords

Comics; digitisation; press; heritage; digital humanities

Introduction

“While extensive corpora are being made available to researchers in the digital humanities, the modalities of understanding comics are undergoing transformation”,³⁶⁵ states French researcher Sylvain Lesage.³⁶⁶ Massive digitisation campaigns initiated by public or private actors are bringing back to light the key role of the press in the history of comics. Recent books such as *La bande dessinée en France à la Belle Époque: 1880-1914* by Thierry Groensteen, an anthology of French comics published in the press, or *Dans l’ombre du professeur Nimbus* by Antoine Sausverd, an historical study about a famous but poorly studied comics character, have both benefitted greatly from the huge set of magazines now available on the Gallica portal at the Bibliothèque nationale de France (BnF).³⁶⁷ This trend is likely to develop further as the number of documents available online increases. However, these public digital libraries were preceded by private initiatives, some of which were barely legal. Websites on which scanned pages can be viewed and often downloaded are widely used by master students and researchers. These websites make it easy to locate specific publications but also facilitate research on particular titles due to their online availability. In this article I will highlight the importance of the periodical press as a source for the study of the history of comics and discuss the challenges and opportunities of using

365. « Tandis que de vastes corpus sont mis à disposition des chercheuses et chercheurs en humanités numériques, ce sont les modalités de connaissance de la bande dessinée qui sont en passe de se modifier ».

366. LESAGE, Sylvain. Le canon et l’écran. Le patrimoine de la bande dessinée et les archives numérisées. *Comicalités* [online]. 2022. [consulted on 20/01/2025]. Available on: <<https://doi.org/10.4000/comicalites.7723>>.

367. GROENSTEEN, Thierry. *La Bande dessinée en France à la belle Époque: 1880-1914*. Brussels (Belgium): Les Impressions Nouvelles, 2022; SAUSVERD, Antoine. *Dans l’Ombre du Professeur Nimbus. Aventures et méasaventures d’un héros de bandes dessinées et de son créateur André Daix*. Montrouge (France): PLG, 2023; <https://gallica.bnf.fr/>.

digitised press corpora, before presenting a number of collaborative research initiatives which have been developed within the framework of the research project BeComicsLab (KBR and ULB).

Comics in the Periodical Press

Contemporary researchers, particularly in Belgium, have hitherto been focused primarily on the material available in albums, limiting the attention paid to periodical publication. Some of this may be blamed on the structure of the comics market itself, which today revolves around the album format. Although comics albums have existed since at least the middle of the nineteenth century,³⁶⁸ the history of comics cannot be separated from the history of the press. Belgian most popular series – *Tintin* (1929), *Spirou* (1938), *Suske en Wiske* (1945), *Blake et Mortimer* (1946), *Lucky Luke* (1946), *Nero* (1947), *Jommeke* (1951), *Gaston Lagaffe* (1957), *Les Schtroumpfs* (1958) – were all found in a periodical publication before being reprinted in albums. It was not until the end of the 1980s that most comics were published directly in the album format. Failure to take dailies and magazines into account would mean writing a truncated history.

The reasons why *Spirou* (1938) and *Tintin* (1946)³⁶⁹ are seen as the two flagship magazines in Belgian comics history is undoubtedly related to the intrinsic qualities of some of the creations they have launched. Nevertheless, the reasons for their success should be considered from a broader perspective, and that includes their publishing policies: Dupuis and Le Lombard, the publishers that issue *Spirou* and *Tintin*, have both piggybacked a collection of albums onto their famous weeklies. These regularly reprinted volumes with a firmer cover are less fragile than the magazines and thus can better withstand the test of time.

368. Often described as the first comics artist in history, Swiss drawer Rodolphe TÖPFFER published *Histoire de Mr. Jabot* (History of Mr. Jabot, Geneva, Imprimerie Caillet) in 1833. This hardcover book in landscape format was then used as model for numerous stories in pictures, including *Le Déluge à Bruxelles* (The Brussels Flood, Brussels, Jules Gérard, 1843) by French-born artist Richard DE QUERELLES. In the light of the present state of knowledge, it is the first comics volume published in Belgium.

369. It is necessary to distinguish between the *Tintin* and *Spirou* series - created in 1929 and 1938 respectively – and the eponymous weeklies – launched in 1946 and 1938.



Fig. 1. Kleine Zondagsvriend, n° 52, 1950. Source: KBR.

The leading role and the great success of *Spirou* and *Tintin* tends to overshadow the importance of a multitude of similar publications, like *Ons Volkske*, *Petits Belges*, *Bravo!*, *Bimbo*, *Kleine Zondagsvriend* (Fig. 1), *Heroïc-Album*, *Paddy or Wrill*. Other youth magazines, such as *Annette*, *Arim Djim Djim*, *Blondine* or *Grand Cœur*, are rarely if ever mentioned in books devoted to the history of

Belgian comics. Whereas youth weeklies are the best-known press support for comics in Belgium, periodicals for adult readers also have a long history in the country. During the second half of the nineteenth century and the very beginning of the twentieth century, comics flourished in satirical magazines such as *L'Uylenspiegel*, *Caprice Revue*, *La Chaudière* or *Les Corbeaux* (Fig. 2).³⁷⁰ Monthly, bimonthly or quarterly publications such as *Curiosity Magazine*, *Ice Crim's*, *Le Journal illustré le plus grand du monde* or *Spetters* testify to the thematic evolution of such works in the 1970s and the 1980s as well the increasing age of the audience.



Fig. 2. *La Chaudière*, n° 1, 1894. Source: KBR.

370. PAQUES, Frédéric. *Avant Hergé. Étude des premières apparitions de bande dessinée en Belgique francophone (1830–1914)*. Liège (Belgium): Presses universitaires de Liège, 2025.

From the 1990s onwards, alternative publishers entered the scene. These new productions were characterised by a willingness to consider comics as the unique expression of a single author.³⁷¹ They enlarged the field of possibilities, questioned the limits of the medium and continue to do so. During their first years of their existence, these publishers generally printed anthologies with refined layouts and at various intervals, with examples including *La 5e couche*, *Bill or Frigobox*. Alternative comics were preceded by fanzines (or zines) with whom they share a strong desire of independence from the publishing industry. These self-made print works are often produced by a collective of authors who generally eschewed official trappings like ISBN or ISSN numbers and did not benefit from a professional distribution network. *Courant d'Air*, *Cuistax*, *Mycose*, *Spon*, *Superstructure*, *Tieten met haar*, *Forgeries* are some of the numerous zines that have flourished in Belgium during the last decades. Lastly, the picture would be incomplete without taking into consideration a lot of press publications where comics only appear in specific sections, without forming the main content: dailies and newsmagazines, journals for specific audiences (growers, doctors, scouts...) or publications specialising in a particular subject (fashion, music, video games...).

The Royal Library of Belgium (KBR) has begun to digitise some parts of this huge but heterogenous corpus. Like most of the other national libraries, they have undertaken digitisation policies by offering a large set of more than 450 digitised periodicals (newspapers and magazines). Thanks to the research projects Pop Heritage Lab – devoted to the Belgian mass-market print heritage – created in 2022, and BeComicsLab in 2023, the number of digitised comics from within press periodicals being made available has significantly increased.³⁷² Digitising media in this way facilitates access to the comics without damaging

371. DEJASSE, Erwin. Le regard cosmopolite et rétrospectif de la bande dessinée alternative. In: DONY, Christophe, HABRAND, Tanguy & MEESTERS, Gert (eds.). *La bande dessinée en dissidence – Alternative, indépendance, auto-édition / Comics in Dissent – Alternative, Independence, Self-Publishing*. Liège (Belgium): Presses universitaires de Liège, 2014.

372. Pop Heritage Lab (www.kbr.be/en/projets/pop-heritage-lab) and BeComicsLab (www.kbr.be/en/projects/be-comics-lab) are both FED-TWIN research projects co-hosted by KBR and a university, respectively KULeuven and ULB, which are exploring mass-market print heritage and comics in Belgium.

the original objects, which are often fragile, valuable and hard to find. Moreover, these initiatives are underlined by the conviction that comics are a part of our cultural heritage that needs to be saved, perpetuated and made available for study and enjoyment.

The induction of comics into the protective arc of 'heritage' has come relatively late, considering that they have existed since at least the middle of the 19th century, and when compared to the cultural recognition of other art forms – in particular, movies. It began in the mid-60s with albums that reprint works that had, in large part, originally been published in the press. This phenomenon has continued its acceleration unabated, albeit with some serious caveats. Works were often chosen because of their nostalgic power or their market potential rather than their artistic value. Furthermore, they have often suffered from poor print quality, inelegant lettering and rather conventional forewords unfitting for creations supposed to belong to the heritage of the medium.

Yet, there is no doubt that, since the turn of this century, both mainstream and independent publishers have taken great pains to publish carefully designed and well printed omnibus volumes, often enriched with critical apparatus. Despite these worthy endeavours, the rich past of comics is still being rediscovered. As the Heritage Selection of the yearly Angoulême International Comics Festival Prize eloquently illustrates, works that have hitherto flown under the radar are regularly brought back into the light, encouraging what may appear to be a perpetual rewriting of the history of comics in the process. Nevertheless, the manufacturing costs of these books are often high while their sales potential is generally limited. Many series cover thousands of pages as they existed over more than half a century. Therefore, publishers often limit themselves to reprinting a sample, encouraging readers to find digitised versions to complete them or to get a comprehensive body of works for research purposes. Moreover, not all comics that have been published in the press are masterpieces. If a large part of the production does not necessarily deserve a complete hardcover collection, reprinting them could be very precious nevertheless to understand the *Zeitgeist*, the publishing policies or their ideological background. Comparing 'minor works' with canonical contemporary creations is also a way to better understand the idiosyncrasies of the latter.

Studying Digitised Comics

By engaging in routine digitisation campaigns, KBR, like other libraries, has made comics that once seemed consigned to obscurity more easily accessible. Examples include 19th century magazines or comics made by women and aimed mostly at a young female readership, such as the post-war weeklies *Annette* and *Grand Cœur* (see above). The latter are a ready-to-investigate set, the study of which will contribute to rewriting the history of comics beyond the male canon, a process which is under way.³⁷³ Even the study of the most popular series will benefit from this process and be reinvigorated. Reading these in albums and finding them back in their original periodical are two very different experiences, however. This is not only because the two versions may differ, but because the latter allows for reading them from their original context - i.e. the ‘environment’ in which they first found life and dialogue with other comics and sections such as editorials, puzzles, news, readers’ letters or advertisements.

As far as possible, researchers should always have at least a look at the original physical publications they study. The size of the copy, the choice of paper, the type of ink, the binding or the format are key pieces of data that digitised versions can hardly provide – this is perhaps especially so in the case of comics which are a visually-dominant medium. The proportion of the computer screen is rarely homothetic with the printed page. To see a comic full-size, the viewer must view it in fragments. “A standard laptop or desktop screen cannot display a typical 6.625” by 10.25” comic book page at both its proper size and shape simultaneously. Standard [...] offer a variety of viewing options, but the two basic options are to view the entire page at once, which makes it too small to read, or to fit it to the size of the viewing window, which causes half the page to be cut off”.³⁷⁴ Therefore, the researcher must always be aware that his reading experience is most likely radically different from that of someone reading the

373. HERTIMAN, Marys Renné & DE SINGLY, Camille (ed.). *Construire un Matrimoine de la BD – Créations, mobilisations et transmissions des femmes dans le neuvième art, en Europe et en Amérique*. Dijon (France): Les Presses du Réel, 2024.

374. KASHTAN, Aaron. *Between Pen and Pixel: Comics, Materiality, and the Book of the Future* [online]. Columbus (United States): Ohio State University Press. 2018. [consulted on 20/01/2025]. Available on <<https://doi.org/10.2307/j.ctv1khdqnk>>.

original version.

Nevertheless, digitisation opens several possibilities, such as optical character recognition, which can be most advantageous to the researcher even on a very bulky set of pages. Anyone who has had to find specific contents from a set of periodicals that covers several years knows how painstaking and time-consuming such a search can be. By allowing full-text search, the development and democratisation of OCR software has completely transformed the methods to the point that it is today possible for an individual to process corpora of many hundreds of thousands of pages. A short time ago, this was impossible unless you had a full team of several researchers. OCR could be also used to find comics simply by typing “*bande dessinée*” for a French-speaking title or “*strips*” or “*stripverhaal*” for a Dutch-speaking one. Nevertheless, it is likely that such searches will lead to incomplete results. For many years there was no term that identified comics as a specific medium. For example, in the French-speaking areas, the phrase *bande dessinée* does not seem to have appeared until the 1940s and it was only at the end of the 1950s that it supplanted other designations.³⁷⁵ In the press, the comics section may have gone under the title of “*Feuilletons dessinés*” or “*Humour*” or, as was most often the case, been graced with no title at all. The authors were not always mentioned, and the translators often freely frenchified the name of the foreign characters – for example, Superman was rechristened *Marc Hercule moderne*, *Yordi* or *François l’Imbattable*, among others.

Research Projects

Because of the visual nature of comics, researchers cannot rely solely on devices that were forged for the literary model. Simple reliance on OCR will lead to mixed results if it is not associated with tools that are able to detect graphics. It is with this in mind that we have decided – along with a group of researchers from KBR (including Pop Heritage Lab, BeComicsLab and Digital Research Lab)³⁷⁶ and computer scientists from the University of Ghent – to bring our

375. GLASSER, Jean-Claude. *Courrier de Jean-Claude Glasser. Les Cahiers de la bande dessinée*. 1988, n° 80, 8.

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

complementary competences together. By exploring the application of computer vision, we have developed a workflow for the automated extraction of comics. This had been tested on an 80,000-page corpus of Belgian illustrated periodicals from the Interwar period that had already been digitised in the frame of Artpresse project (see the contribution by Sébastien Hermans and Morgane Ott in this volume).³⁷⁷



Fig. 3. A sample of how an object detector model can detect comics, cartoons and others with VGG Image Annotator.

377. Intermedial research program on Belgian Art through the analysis of mass media magazines: <https://www.kbr.be/en/projects/artpresse>.

The choice of the Interwar period allowed the model to be trained on a very heterogenous set of comics. Indeed, the corpus is characterised by a huge variety of plastic and narrative devices. Here, sequences of images with balloons, inherited from the US, cohabit with speechless strips or strips accompanied by captions in the tradition of the 19th century. The semi-automatic workflow first consists of taking some 4% of the whole corpus, identifying comics and labelling them with the opensource tool VGG Image Annotator (Fig. 3). This labelling allows for the model to be trained to recognize comics' plastic and narrative devices. Comics could be described as suites of images articulated together to produce a story or a global meaning. One of the main challenges here is to not confuse them with other visual layouts such as amalgams of independent images with no semantic links between them. To meet this challenge, further research should engage a model with semantic capabilities, for example by using Multimodal Large Language Models.

Digitised documents also offer enormous opportunities in the field of stylometry. By computing the recurrences of words, groups of words, syntactic structures, punctuation, etc, it can be used to identify authors or to generate an in-depth understanding of their distinctive stylistic features. Here again, the analysis of comics through the techniques and methodologies of Digital Humanities needs to consider the visual nature of the medium and therefore design suitable tools. The recent creation of QuaDiHum (Quantitative and Digital Humanities Centre) at ULB offers a great opportunity to develop them.³⁷⁸ I will point to a few avenues of possibility, as we have just begun to scratch the surface of this vast field of research through the StyloBD project with my colleague Sébastien de Valeriola (ULB). In the case of comics, the concept of style is multifaceted. The codes and devices through which cartoonists may make their mark are numerous: the style of writing in the literary sense (dialogues and captions) but also the style of the lines (including the way the character and backgrounds are drawn but also the letters, the balloons or the frames), the nature of the represented motives, the viewing angles or the categories of shots, the colour schemes, the type of layout (Fig. 4) or even the way ellipses between the panels are managed (what has been chosen to show and not to show in a given sequence). Each criterium requires

378. <https://quadihum.ulb.be/>.

a specific approach, and all criteria must be combined to create a model. Because of the multifaceted nature of comics, developing such a model will necessarily be a complex task. It will, however, be a tool of great potential, due to its ability to perform analyses by taking in account the variability of a very large set of factors. Among the possible applications of such a model: studying the aesthetic impact of masters on their assistants and through which idiosyncratic features the latter frees themselves from the former's influence; helping the attribution of a specific work or identifying all the persons involved in a comic where several hands are involved; examining, on the base of distinct corpora, the singularities of comics from different cultural areas by comparing their similarities and dissimilarities. I assume that because of the multimodal nature and the plurality of semantic registers available to comics the development of such a model will not only be fruitful for their study but may also become a useful toolbox for many other plastic and/or narrative art forms.

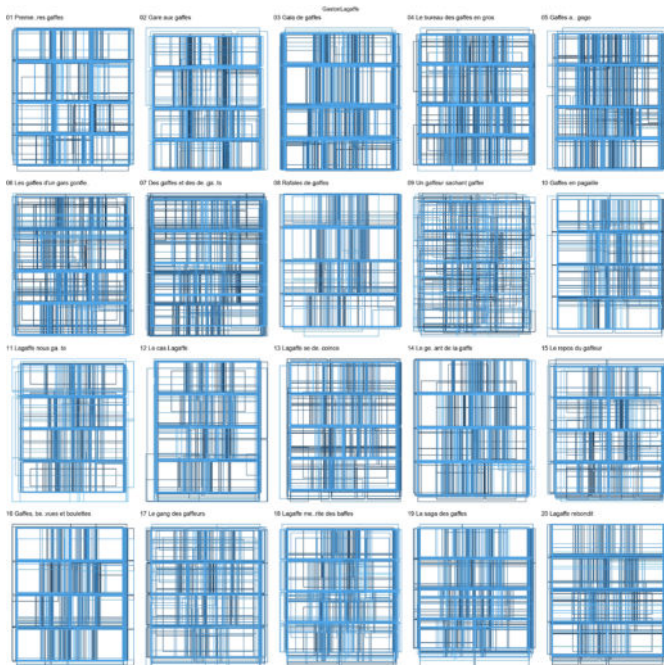


Fig. 4. Sébastien de Valeriola & Ervin Dejasse. Modelling of the layout variation in the series Gil Jourdan by Maurice Tillieux.

13. CAMille. A Research Centre on the History of Belgian Journalism

Florence Le Cam, Brecht Deseure & Max De Wilde

Abstract

CAMille is research centre created in 2020 as a joint venture between ULB and KBR with a view to further the history of journalism in Belgium via a focus on critical discourse about journalism and the development of professional identities among journalists. To reach this goal, CAMille stimulates the development of new methodologies within digital humanities in order to facilitate the analysis of digitised press sources on a massive scale. The project brings together and digitises a wide range of documents relating to journalism in Belgium, which are made accessible via a platform which serves as a testing ground for research methodologies, including automatic signature detection and the reconstruction of media careers. The prosopographic aspect of the project is pursued via the creation of a database of Belgian journalists developed within the RDF data framework. While these infrastructures and research techniques are in full development, they have already led to significant contributions to the historiography of Belgian journalism.

Authors

Brecht Deseure obtained his doctorate in history at the University of Antwerp in 2011. He publishes on various aspects of Belgian history in the revolutionary and contemporary eras, including the history of political and constitutional culture, iconography, and the press. He is the lead researcher of the research centre *CAMille, Centre d'archives sur les médias et l'information*, hosted by KBR and the Université libre de Bruxelles. At KBR he is responsible for the historical newspaper collections. His publications include *Onboudbaar verleden. Geschiedenis als politiek instrument tijdens de Franse periode in België* (2014), *Constitution belge. Drafts and Proposals for the Belgian Constitution of 1831* (2020)

and *Revolutie in Antwerpen. De aquarellen van Pierre Goetsbloets, 1794-1797* (2021).
E-mail: brecht.deseure@kbr.be.

Florence Le Cam is a professor in the department of Information and Communication Sciences at the Université libre de Bruxelles and at the University of Montreal. She obtained her PhD at the University of Laval and the University of Rennes with a thesis entitled *L'identité du groupe des journalistes du Québec au défi d'Internet*. Her research focuses on journalists and their emotions, communication, and experiences. Her works include *Le Journalisme en ligne mondialisé* (2022, PUR) with Pereira, *Les Journalistes et leurs médias en Afrique. Pensées mêlées en souvenir de Marie-Soleil Frère* (2022, VUB Press), *Etre femme et journaliste. Enquête sociologique dans un monde au masculin* (2021, VUB Press) with Libert et Ménalque, and *Emotions de journalistes: Sel et sens du métier* (2017, Grenoble University Press) with Ruellan. Currently, she is working on publishing her book *L'usine du journalisme. Espaces, matérialités et relations de pouvoir*.
E-mail: florence.le.cam@ulb.be.

Max De Wilde holds a PhD in Information and Communication Sciences. He has been teaching natural language processing at the Université libre de Bruxelles and the University of Geneva since 2016. Alongside his academic responsibilities, he works as an independent consultant in artificial intelligence, notably for the European Commission. E-mail: max.de.wilde@ulb.be.

Keywords

Newspapers; history; journalism; prosopography; database; signatures; RDF

Introduction

The history of journalism in Belgium is an insufficiently explored, inadequately documented, and fragmented subject of research. Historiography of the Belgian press is consequently meagre, certainly when compared to the situation in neighbouring countries. Among the reasons for the limited development of

such research are the wide dispersal of sources and the difficulty of accessing and exploiting them. It is our view that, to be able to gainfully pursue research in this area, researchers need to gain access to archives, historical press collections, and the contents of journalists' personal collections.

In 2017, the Belgian government introduced legislation that put in place a framework for sustainable cooperation between the ten Federal Scientific Institutions (FSI) which fall under the Belgian Science Policy (BELSPO), and the eleven Belgian universities.³⁷⁹ As the programme was being launched, the researcher Antoine Jacquet was completing his doctoral thesis at the Université libre de Bruxelles under supervision of Florence Le Cam, chair of journalism.³⁸⁰ From the convergence of these circumstances emerged the plan to set up a large-scale research project aimed at writing the history of Belgian journalism by studying the critical discourse on journalism. Such a history would consider the growing public criticism directed at the media and Belgian journalists in recent times. Furthermore, we wanted to understand how a collective professional identity was defined, both by the members of the professional group and by the broader public, which projects its expectations and representations onto this professional world.

Given the potential of this research, we saw an ideal partner in the Royal Library of Belgium (KBR), which holds the largest collection of Belgian newspapers anywhere. The connections we had established with Marc D'Hoore, curator of newspapers at the institution, encouraged us to apply for funding under the FED-tWIN call in 2018-19, which laid the foundation for what would become CAMille.

The Original Project

We therefore worked on a proposal to create the Digital Archives Laboratory for the History of Belgian Journalism (*Digi-Lab Journalisme*). The aim of the

379. https://www.belspo.be/belspo/research/FEDtWIN_en.stm.

380. JACQUET, Antoine. *Le français des journalistes en ligne. Régulation de la langue par les représentations et les pratiques*. Unpublished doctoral thesis, supervised by Florence Le Cam and Laurence Rosier. Université libre de Bruxelles, 2018.

project was and remains to enable KBR and researchers to benefit from a comprehensive and innovative laboratory of digital archives of the history of Belgian journalism and to produce knowledge about the media and journalism in Belgium from 1886 — a date marking the professionalisation of journalism in Belgium — until the present day. The project aims to bring together and digitise a wide range of documents relating to journalism in Belgium (archives and publications of professional associations, media companies, articles from intellectual journals, publications, reports, etc.) and to include, enhance and complete the corpus of press that has already been digitised.

On the research side, two central projects were proposed. The first was to produce a history of Belgian journalism since its professionalisation from a specific angle: analysis of the critical discourse to which journalism has been subjected by both journalists and by all the social players who have taken an interest in it. The second project aimed to produce a biographical dictionary of Belgian journalists. The Digi-Lab journalism project stands out for its work on the use of digital humanities in these research projects. Working on huge corpora of old press requires innovative research methods, and we wanted to follow in the footsteps of international projects such as NewsEye,³⁸¹ Numapresse³⁸² and Impresso.³⁸³

First Steps

The project was approved, funded for ten years and implemented in March 2020. Antoine Jacquet became the project lead, taking up a post-doctoral position half-time at KBR and half-time at ULB. Our first goal was to create a closed platform enabling researchers to access a selection of Belgian dailies from the 19th and early 20th centuries. The idea was to deploy dedicated search tools for these dailies, enabling them to be searched using a wide range of filters (title of newspapers, dates, editions, etc.) and allowing searches by exact expression, proximity search or weighting of terms.

381. <https://www.newseye.eu/>.

382. <http://www.numapresse.org/>.

383. <https://www.impresso-project.ch/>.



Fig. 1. CAMille homepage for researchers.

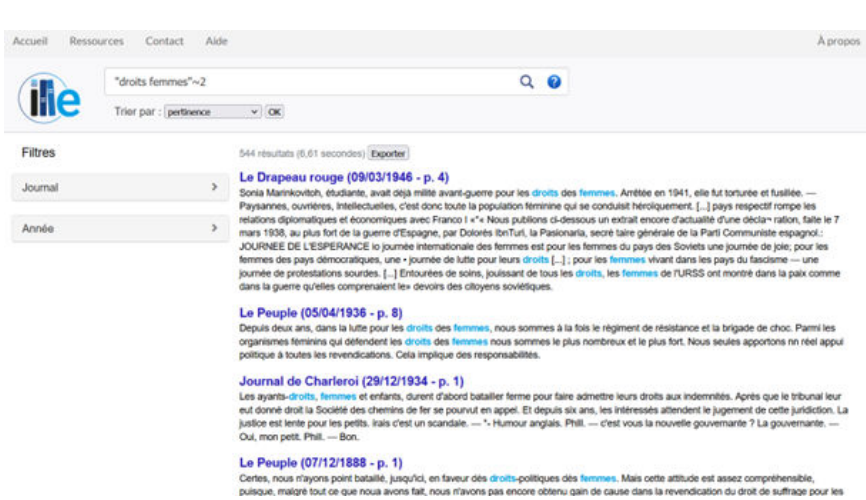


Fig. 2. Capturing search results.

The platform was created in June 2020. The viewer provides access to OCR-d PDFs, which are exportable in .xlsx and .txt format to facilitate text and data analysis. At this stage, the platform is intended to operate first and foremost as a testing ground, allowing the users to test the methods of archiving and interrogating the corpus before submitting funding applications for its further development.

As of November 2024, more than 1.4 million pages have been uploaded to the CAMille platform, representing 50GB of raw data. These texts have been extracted from the ALTO-XML files and are stored in an Elasticsearch index to allow for optimised full-text search in real time, along with filtering based on metadata (newspaper title, date, day of the week, edition, page number, language, etc.).

The application is currently deployed on ULB infrastructure. The Elasticsearch index is being queried from a web interface written in Python (Flask) and hosted on a virtual machine. PDF files are stored on a separate filesystem and can be retrieved on demand (i.e. when a user clicks on a given page).

The source code for the project has been published as open source (MIT license) and can be found online.³⁸⁴ This includes the Flask interface, but also various scripts to convert the data, and the contents of the Lambda functions used to automatically index new files on the fly when uploaded.

CAMille currently contains 13 digitised Belgian newspapers, which have been chosen to represent the francophone Belgian press across geographical original and ideological orientations.³⁸⁵ It also contains one professional periodical (*La Presse*). The platform will be progressively enriched with additional digitised sources. A central goal of our approach is indeed to reconstitute dispersed series of periodical publications and to enable them to be interrogated through application of various filters. Thanks to additional funding from ULB, we have been able to finance the reconstruction and digitisation of two sector-specific journals: the annual directories of the Association of the Belgian Press (created in 1886)³⁸⁶ and *Le Journaliste*, the journal for professional journalists (created in 1920). No complete series of either of these journals survives, with individual issues scattered across the collections of numerous university and non-university libraries. The integration of these documents into the platform is scheduled for 2026.

384. <https://github.com/madewild/CAMille>.

385. *L'Avenir du Luxembourg, La Libre Belgique, L'Indépendance belge, L'Indépendance belge* (édité en Angleterre), *Journal de Bruxelles, Journal de Charleroi, La Meuse, Le Petit bleu, Le Vingtième siècle, Vers l'avenir, Le Peuple, Le Soir, Le Drapeau rouge*.

386. *Association de la presse belge*, later called *Annuaire officiel de la presse belge/Officieel jaarboek van de Belgische pers*.

Alongside this focus on infrastructure, collaborative research initiatives were being initiated with the participation of a small team of associated researchers (Antoine Jacquet, Alexia Vidalenche, Manon Libert and Sébastien de Valeriola), in the context of the French research project Numapresse.³⁸⁷ Inspired by the prosopographical approach taken within Numapresse, we decided on two lines of research:

- Work on the automatic search for signatures in the press in order to reconstruct the organisational lives of the dailies studied
- Work on research into the careers of journalists, based on the dictionary of Belgian journalists left to us by Pierre Van den Dungen, a historian at the ULB who did a remarkable job of writing biographical notes with his colleagues and students.

These initial initiatives were accompanied by regular meetings with the project's scientific committee.³⁸⁸ In 2021 it was decided to change the project's initial name *Digi-Lab Journalisme* into *CAMille - Centre d'Archives sur les Médias et l'Information*, a reference to the emblematic Belgian writer and journalist Camille Lemonnier (1844-1913). In September 2021, Antoine Jacquet left the project due to a career change and was succeeded in February 2022 by Brecht Deseure. In order to make the research centre known, gather expertise and establish links with colleagues in the field, a conference dedicated to research on Belgian print media was organised for September 2023 (of which this volume is the result).³⁸⁹

387. See the Numapresse website: <http://www.numapresse.org/presentation-du-projet/>.

388. The committee consists of Paul Aron (literature researcher, F.R.S.-FNRS/ULB), Marc D'Hoore (curator of Contemporary Newspapers and Media Section, KBR), Juliette De Maeyer (professor in journalism, Université de Montréal), Antoine Jacquet (researcher in journalism ULB-KBR), Irene Di Jorio (professor in media history, ULB), Florence Le Cam (professor in journalism, ULB), Frédéric Lemmers (head of digitisation service, KBR), Manon Libert (professor in journalism, UMONS-LaPij), Thuy-An Pham (ICT service, KBR), Pierre Van den Dungen (historian of the Belgian press, ULB), Sophie Vandepontseele (director of Contemporary Collections, KBR).

389. *Belgian Print Media in the Age of Digitization. Current Challenges of Research into the History of Media and Journalism in Belgium* (Université libre de Bruxelles, 4-5 September 2023).

In 2024 we successfully applied for funding under the PDR (*projet de recherche*) program at ULB with a project called *A History of Belgian Journalism through Press Releases*. This joint project between ULB and KBR, co-authored by Florence Le Cam and Frédéric Lemmers, aims to digitise, make accessible, and study KBR's collection of press releases from the Belgian press agency Belga. This unique collection consists of thousands of paper volumes (1954-1995) and digital files under different formats and is the only known series of such press releases in existence. Constituting a precious and vast source for the history of Belgian journalism, it will be exploited for research via the development of digital research techniques. The project has been launched in August 2025.

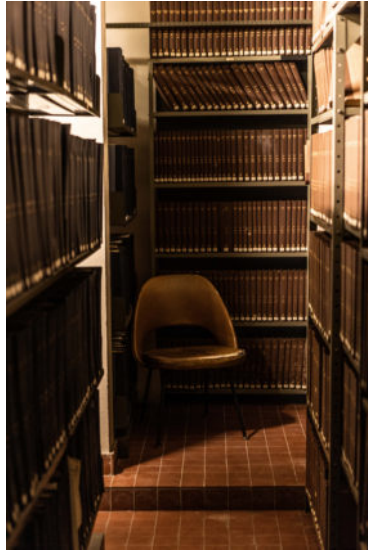


Fig. 3. A view of the collection of Belga press releaser (photo by Johanna de Tessières)

Current Projects and Lines of Research

When the CAMille project had been running for two years, it was presented in this way:

CAMille is working to enrich studies of Belgian journalism from three standpoints: the prosopographical approach; the reconstruction of newsrooms; and research into the careers of journalists, particularly women journalists. These three approaches very often overlap.

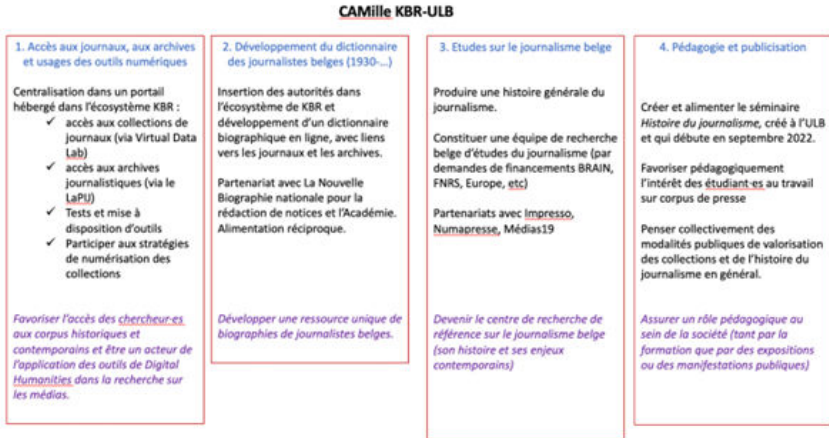


Fig. 4. Presentation of CAMille in April 2022.

Prosopography of Belgian Journalists and the Reconstruction of Newsrooms

The dictionary bequeathed by Pierre Van den Dungen contains 3,990 entries on journalists from the nineteenth and twentieth centuries, each encompassing a range of information: name, pseudonym, press outlets they worked for, functions, titles of works, parallel activities, etc. This dictionary is a gateway to understanding the transformations of the Belgian press and will help to make up for the lack of research devoted to this history, while also contributing to the development of digital tools in the history of journalism. Transforming the initial manuscript, which took the form of a typewritten printout, into a digital format required considerable manual labour. An Excel spreadsheet was created that enabled us to encode each item of biographical data in the dictionary indi-

vidually. This major encoding work was carried out, using CAMille's own funds, in 2020 and 2021. The dictionary was subsequently expanded thanks to the work of trainees at KBR who digitally harvested additional names of journalists from authority lists of European national libraries, as well as from a number of digitised sources. This brought the total number of records to over 6,000.

Based on this dictionary, we then launched two research projects:

- 1) How can the press-wide search for signatures in the press, using digital humanities tools, contribute to reconstructing the editorial staff of a newspaper? This project aims to understand the journalistic activities and professional mobility of journalists, and to identify possible organisational crises within the editorial staffs of newspapers.³⁹⁰
- 2) How can the search for a signature and, more generally, the mention of a journalist in the press, enable us to reconstruct professional careers? More specifically, this project primarily concerns the case of Alice Bron, a socialist woman journalist working in Brussels between 1895 and 1904.

The dictionary enables us to locate the names of journalists, identify them according to their press outlet and then carry out a systematic search for these names in the newspaper itself. A case study of the newspaper *Le Vingtième siècle* was chosen in order to develop a methodology, using Named Entity Recognition and other extraction techniques for recovering a maximum number of contributor names. This resulted in a list of journalists, many of whom were previously unknown to us. We then created charts (*présençogrammes*) indicating the prevalence of each journalist (by signature or mention of their name) over time, which enable us to reconstruct the newspaper's successive editorial teams and further analyse professional mobility. In addition, tracing the mentions of individual journalists' names from the dictionary opens the way to analysing the

390. JACQUET, Antione, LE CAM, Florence, LIBERT, Manon, VIDALENCHE, Alexia, & DE VALERIOLA, Sébastien. *Prosopography and digital humanities through the Belgian daily Le XXe siècle (1895-1940)*. Conference paper in *Atelier-séminaire Numapresse "Individus, auteurs, signature"* (Paris, June 2021).

uses and usefulness of the signature in their careers. The functions and uses of signatures by and for journalists throughout their lives and careers contribute to developing an enhanced understanding of the development of professional identities in Belgian journalism.³⁹¹

SPARQuLb est basé sur une instance [Wikibase](#) (un logiciel libre qui offre un espace collaboratif pour mettre en commun, éditer et conserver des informations de manière structurée) déployée sur les serveurs de l'université. Wikibase est particulièrement efficace pour ouvrir l'accès à vos données, non seulement pour les ordinateurs, mais aussi pour votre public humain dans sa propre langue, afin qu'il puisse faire partie du réseau de données ouvertes liées (Linked Open Data).

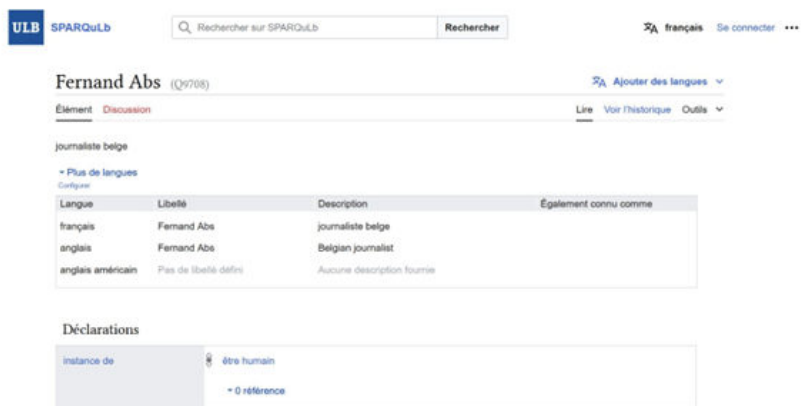


Fig. 5. Presentation of the SPARQuLb project led by Sébastien de Valeriola and Anthony Leroy.

The work on the reorganisation of newsrooms is continuing, thanks to advances in automatic signature detection deployed by Sébastien de Valeriola (see his chapter in this volume).³⁹² The first automatic signature detection list result-

391. DESEURE, Brecht, DE VALERIOLA, Sébastien, LE CAM, Florence, LIBERT, Manon & VIDALENCHÉ, Alexia. *Héritier d'un dictionnaire biographique: enjeux méthodologiques autour d'une base de données belges*. Conference paper at *Congrès international Médias 19-Numapresse, Presses anciennes et modernes à l'ère du numérique* (Paris, May 2022).

392. LIBERT, Manon, JACQUET, Antonie, DE VALERIOLA, Sébastien, LE CAM, Florence, VIDALENCHÉ, Alexia & DESEURE, Brecht. Repeupler les rédactions du passé à l'aide des humanités numériques. *Le cas du quotidien belge Le XXe Siècle*. *Questions de communication* 48 (2025), <https://doi.org/10.4000/15b19>.

ing from searches carried out in the daily *Le Vingtième siècle* is currently being analysed. This will not only provide a better understanding of the newspaper's organisational history, but also enable us to add new names to the dictionary of journalists, and perhaps make new connections to this particular newspaper.

The dictionary was also the subject of a funding application to the FNRS (Research Credit - CDR obtained in 2022) for the transfer of dictionary data from an Excel spreadsheet to SPARQuLb,³⁹³ an RDF data ecosystem linked to an instance of the Wikibase software, thanks to collaborative work with Anthony Leroy, Sébastien de Valeriola and Max De Wilde (Fig. 5).

Reconstructing Careers

Since 2022, our research into the history of journalism has focused on one particular area: Belgian women journalists. Hitherto, the history of Belgian journalism has almost exclusively been understood as a history of male journalists. As in many other countries, historians in Belgium need to work on increasing women's visibility in history, including the history of the media. Only 205 of the Dictionary's 6000+ entries refer to women journalists. Not only have women historically been a minority among Belgian journalists,³⁹⁴ but they have also been the subject of little historical research.³⁹⁵ Our starting point was one particular journalist, Alice Bron, who rarely appears in the pantheon of women journalists in Belgium.³⁹⁶ She seemed an interesting case study for examining the visibility

393. <https://quadihum.ulb.be/sparqulb/>.

394. LE CAM, Florence & TANT, Cédric. Premiers pas de la carte de presse en Belgique francophone (1885-1966). *Le Temps des médias*. 2018, vol. 30, n° 1, p., 200-226. DOI: <https://doi.org/10.3917/tm.030.0200>; LE CAM, Florence, LIBERT, Manon & MÉNALQUE, Lise. *Être femme et journaliste. Enquête sociologique dans un monde au masculin*. Brussels: Éditions de l'Université de Bruxelles, 2021.

395. Apart from: GEMIS, Vanessa. Femmes écrivains-journalistes (1880-1940): questions de genre(s). Pistes de recherche et réflexions autour de Marguerite Van de Wiele. *Textyles. Revue des lettres belges de langue française*. 2010, n° 39, p. 39-50, DOI: <https://doi.org/10.4000/textyles.107>; GEMIS, Vanessa. Occupation: Flying Parcel. Portrait of Marc Augis, Woman, Journalist, and Writer. *Literary Journalism Studies*. 2016, vol. 8, n° 2, p. 38-51, DOI: <https://doi.org/10.4000/textyles.107>; VANDERPELEN-DIAGRE, Cécile. Magasins d'amour. Les romans et chroniques de France Adine, journaliste de Femmes d'aujourd'hui. *Belpégor. Littérature populaire et culture médiatique*. 2021, vol. 19, n° 2. DOI: <https://doi.org/10.4000/belphegor.4219>.

396. VIDALENCHÉ, Alexia, LE CAM, Florence, DE VALERIOLA, Sébastien, LIBERT, Manon & DESEURE, Brecht. Retracer (massivement) et analyser une carrière médiatique. Désinvisibiliser Alice Bron, jour-

and media career of a Belgian journalist whose identity as a social activist cannot be summed up in a single word.

Our approach has been informed by research that relies on the interactionist perspective³⁹⁷ as a way of reconstructing media careers.³⁹⁸ This approach takes into account the different stages and bifurcations that an individual's trajectory can take and, above all, the way in which career moments circulate in the newspapers, are commented on, picked up on, etc. This reconstruction of a career examines the various ways in which a name is represented within newspaper columns.

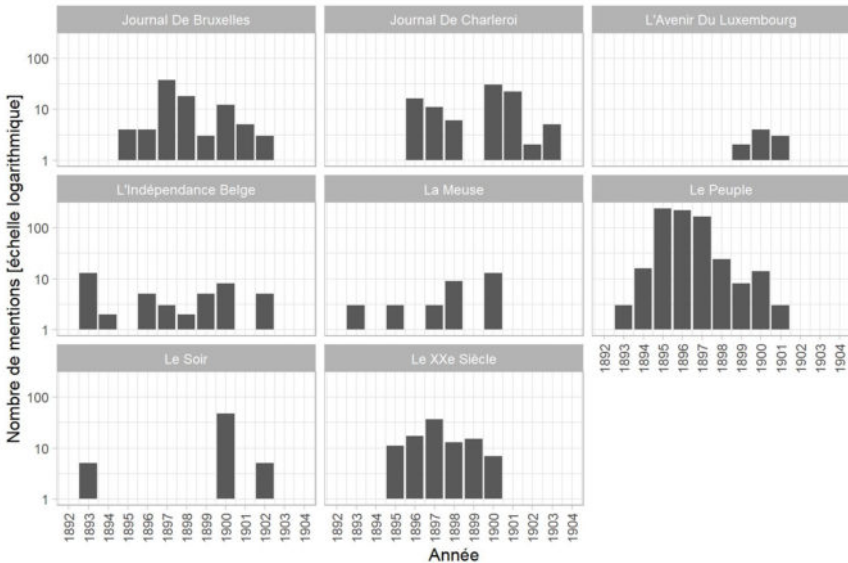


Fig. 6. Mentions of Alice Bron in the corpus of newspapers, particularly in 8 titles (VIDALENCHE et al., 2024).

naliste, socialiste et féministe belge (1850-1904). *Le Temps des médias*. 2024, vol. 42, n° 1, p. 169-189.

397. BECKER, Howard S. *Les mondes de l'art*. Paris: Flammarion, 2010; DARMON, Muriel. La notion de carrière: un instrument interactionniste d'objectivation. *Politix*. 2008, vol. 82, n° 8, p. 149-167.

398. TASSET, Cyprien. Entre sciences sociales, journalisme et manifestes. La représentation de groupes sociaux réputés émergents dans la France des années 2000. *Les enjeux de l'information et de la communication*. 2011, vol. 12, n° 3, p. 139-157; KACIAF, Nicolas & LE DREF, Paul. Du devenir des enquêtes de Mediapart et du Canard enchaîné: contribution à l'analyse de la «carrière médiatique» des révélations journalistiques. *Les cahiers du journalisme - Recherches*. 2018, vol 2, n° 1, R119-R129.

In the case of Alice Bron, she is mentioned in the newspapers as the author of articles, as the subject of ideologically informed controversies between newspapers, and in a variety of other roles and functions, professional and otherwise.

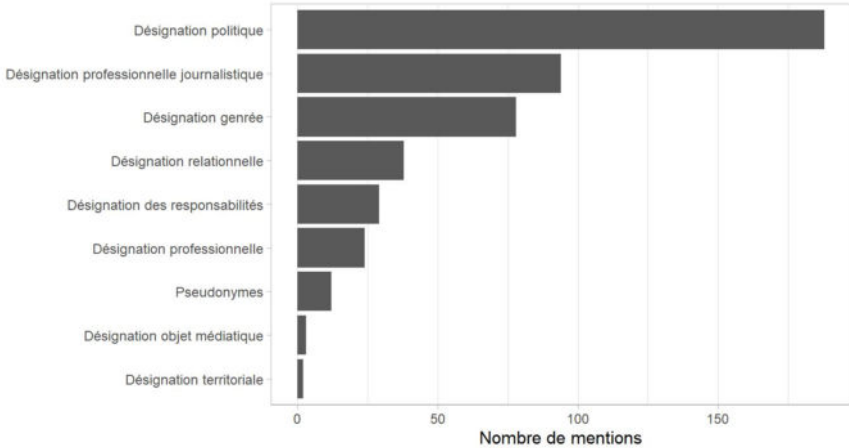


Fig. 7. Typology distribution of designations of Alice Bron in the corpus of newspapers (VIDALENCHE et al., 2024).

Σαπλ



Fig. 8. Types of designation over time (VIDALENCHE et al., 2024).

Interroger la domination masculine dans l'histoire de la presse



Fig. 9. Opening slide from a paper presented at the *Journée de l'histoire contemporaine* devoted to *The Margins of History* on 13 May 2024 in Brussels (DESEURE, Brecht, GRIBOMONT, Isabelle, & LE CAM, Florence).

This one name thus carries with it multiple possible significations. Studying the figure of Alice Bron through the newspapers in our corpus sheds significant new light on her life and career in the media.³⁹⁹

Thanks to Sébastien de Valeriola's expertise in digital humanities, this research enabled us to create charts displaying the presence of her name in the media across time and across newspapers (Fig. 10)

The approach also allowed us to chart the terms used to refer to this individual in newspaper discussions about her. These approaches were conducive to identifying periods that included many mentions of Bron and subsequently to performing a qualitative analysis of the discourses identified per period and per media.

This methodology has therefore proven to be a useful tool to broaden our study of the careers of Belgian women journalists. In the future we will replicate the approach with other profiles. This perspective is also pursued in our history of journalism seminar, which is part of the master's program in Journalism at ULB. The focus on women journalists has led students to work on profiles such as Lily Pousset, Marguerite Inghels and Louise Corbusier.

399. VIDALENCHE et al., Retracer (massivement) et analyser.

Critical Discourse about (Women) Journalists

In parallel with this research focused on careers, we wanted to get a view of the discourse about women journalists in the Belgian press and its evolution through time. We composed a corpus by extracting all of the text fragments that used terms referring to women journalists from our 11 dailies.⁴⁰⁰ This enabled us to visualise the distribution frequencies of the terms (figure 10), their frequency over time (figure 11), and their distribution according to media (figure 12). We next performed a qualitative analysis which showed, for the first time, the discursive tendencies relating to the representation of women journalists in the Belgian newspaper press and their chronological and ideological variations (figure 13). Many Belgian dailies and archives talked about the feminisation of the profession during his time periode. This internal critical discourse reveals collective professional identities and the way in which the professional group constructed itself.⁴⁰¹ We also hypothesised a convergence between the discourse of the dailies on women journalists and that of feminism and feminists. A combination of quantitative text mining methods and qualitative analysis allowed us to qualify that assumption, while uncovering linguistic patterns related to the use of the terms feminists and feminism.⁴⁰²

400. 'cette journaliste', 'collaboratrice, journal', 'collaboratrice, journaux', 'collaboratrice, revue', 'collaboratrice, revues', 'collaboratrices, journal', 'collaboratrices journaux', 'collaboratrices revue', 'dame journaliste', 'dames journaliste', 'dames journalistes', 'femme, correspondante', 'femme, journalisme', 'femme journaliste', 'femme rédacteur', 'femme, reporter', 'femmes, correspondantes', 'femmes, journalisme', 'femmes journalistes', 'femmes reporters', 'journalisme féminin', 'journaliste féminin', 'journaliste femme', 'journalistes femmes', 'la journaliste', 'mme journaliste', 'reporter femme', 'reporters femmes', 'reportrice', 'une journaliste'. Queries including commas are those which detect the occurrence of terms within spans of words. 'this journalist', 'contributor, newspaper', 'contributor, newspapers', 'contributor, review', 'contributor, reviews', 'contributors, newspaper', 'contributors newspapers', 'contributors review', 'lady journalist', 'ladies journalist', 'ladies journalists', 'woman, correspondent', 'woman, journalism', 'woman journalist', 'female editor', 'female reporter', 'female correspondent', 'female journalism', 'female journalists', 'female reporters', 'female journalism', 'female journalist', 'female journalist', 'female journalists', 'la journaliste', 'mme journaliste', 'female reporter', 'female reporters', 'reportrice', 'une journaliste'.

401. DESEURE, Brecht, GRIBOMONT, Isabelle & LE CAM, Florence. Journalistic Gender Trouble. Women Journalists as Seen by and in the Belgian Press (1850-1950), *Journal of Belgian History*. In press.

402. GRIBOMONT, Isabelle, DESEURE, Brecht & LE CAM, Florence. The Perception of Women Journalists and Feminism in Belgian Newspapers (1880-1950). *Digest. Journal of Diversity and Gender Studies* 12 (2025/2), p. 45-67, <https://www.digest.ugent.be/article/id/91568/>.

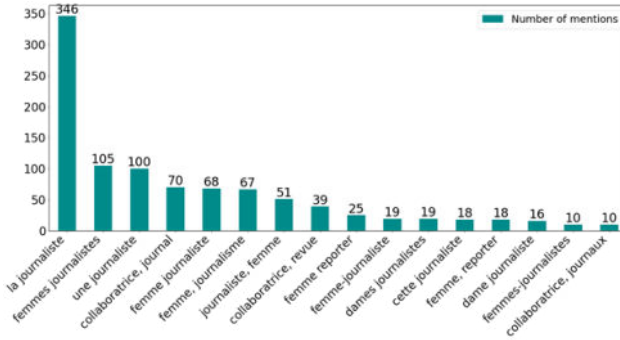


Fig. 10. Distribution frequency (DESEURE et al., Journalistic Gender Trouble). Graph by Isabelle Gribomont.

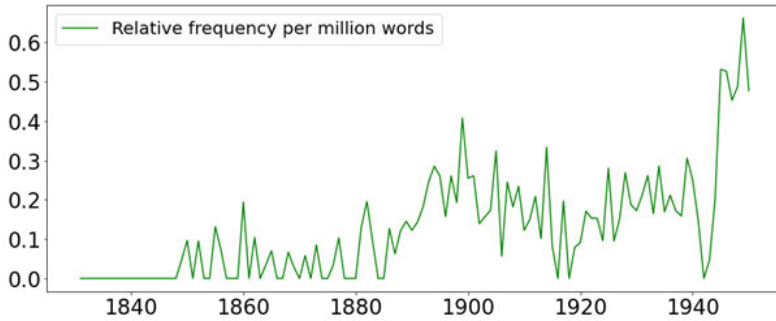


Fig. 11. Diachronic view of distribution frequency (DESEURE et al., Journalistic Gender Trouble). Graph by Isabelle Gribomont.

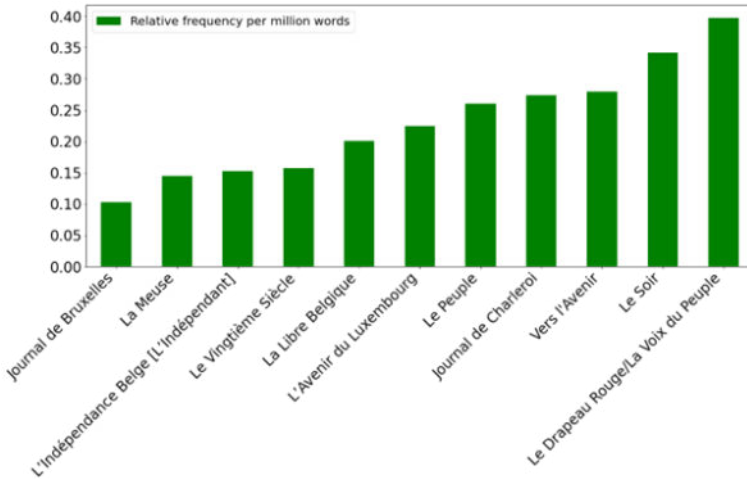


Fig. 12. Distribution of terms in daily newspapers (DESEURE et al., Journalistic Gender Trouble). Graph by Isabelle Gribomont.

LA FEMME-JOURNALISTE

**Il est vrai qu'il y a des femmes
journalistes de talent, mais le
reportage n'est pas une carrière
féminine.**

Le journalisme est-il un métier pour
la femme?

Le Conseil National des Femmes françaises, au cours d'une récente soirée, tenue au Collège féminin et présidée par M. Georges Lecomte, de l'Académie Française, vient de répondre par l'affir-

Fig. 13. Example of discourse found by qualitative analysis (close reading) based on quantitative results. Article by Robert Dubard in *La Meuse*, 05/02/1925, p. 1. Source: *BelgicaPress*.

Conclusion and Future Perspectives

CAMille celebrated its fifth anniversary in March 2025. Since its inception, the research centre has taken important steps towards a more nuanced history of Belgian newspaper journalism, thanks to the use of digital humanities techniques for the study of a corpus of digitised press sources. Collaboration with other researchers has been central to the project's success. The main challenge encountered in the development of CAMille has, without doubt, been finding adequate digital expertise. The leads of the research centre are qualitatively trained scholars without a background in digital humanities. We consequently depend on collaboration with associated researchers with adequate digital skills, none of whom is permanently attached to the project. In such an environment, progress can only ever be piecemeal, and even so, it relies heavily on the availability and goodwill of our colleagues.

Developing and maintaining digital infrastructures is, moreover, an expensive affair; we can only achieve it thanks to obtaining additional funding. Our research is ongoing and often exploratory. The digital infrastructure is continually evolving in response to our research needs, financial capacity and available data.

Several additional funding applications, which are aimed specifically at strengthening and expanding the digital component of the project, have either been submitted or are being prepared for submission.

An additional challenge of working with archives and analysing millions of pages of old press is the sheer scale of the task. In order to find a manageable point of entry, we have chosen to focus initially on the history of women journalists. This has already proven to be a fruitful and very exciting topic, which has allowed us to gather expertise, develop research methods, and make meaningful contributions to historiography. It will also help to prepare work on the prosopography and discourse analysis of the entire group of journalists.

Overall, the application of digital humanities to digitised press and archival sources has proven to be an excellent way of shedding light onto forgotten stories, obscured careers and other little-known parts of the history of Belgian journalism. We continue to explore these forgotten histories, seeking to connect with the research community of historians, journalism researchers, and digital humanities specialists. We do this in part through organising conferences and study days like the international conference *Forgotten Journalists: Lived Experiences and Professional Identities in the Past*, which took place in June 2025 in Ghent and was organised in collaboration with the archive and research centre Liberas and Ghent University.

14. News Nuggets: Mining Imperfect Historical Press Data

Sébastien de Valeriola (Ratio DH, Université libre de Bruxelles)

Abstract

In this short note, we present the main problems involved in applying Optical Character Recognition models to old press. We give an overview of the solutions we have put in place to overcome these problems, as part of a collaborative research project with the CAMille laboratory dedicated to the history of Belgian journalism.

Author

Mathematician, actuary and historian, Sébastien de Valeriola is associate professor at ULB, where he holds the Chair of Digital Humanities. He is coordinator of the Ratio DH research centre. His research focuses on the application of quantitative methods to the human sciences in general, and the epistemological issues this process raises.

Keywords

Digital humanities; newspapers; signatures; OCR; segmentation

The emergence, development and democratisation of Optical Character Recognition (OCR) models, which enable text to be extracted from an image (e.g. a scan of a printed document), has opened extraordinary possibilities in humanities research. The associated technical problem has been considered solved by specialists for the last ten years or so:⁴⁰³ the methods implemented (generally derived from machine learning) can produce output of impressive quality.

403. DOERMANN, David & TOMBRE, Karl. *Handbook of Document Image Processing and Recognition*. London: Springer, 2014. Available on: <https://doi.org/10.1007/978-0-85729-859-1>.

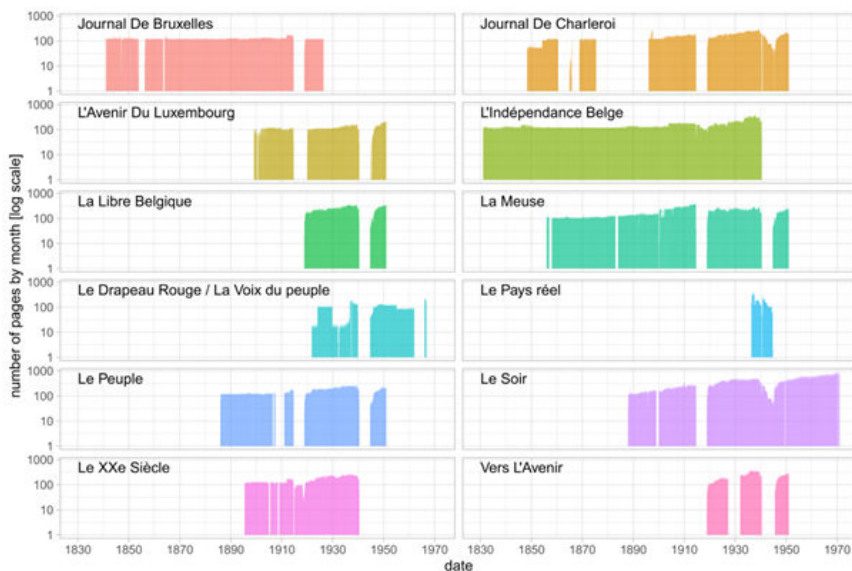


Fig. 1. Chronological and newspaper distribution of data in the corpus studied.

This quality can be assessed in various ways,⁴⁰⁴ for example using the Character Error Rate (CER), which gives the frequency of errors in individual characters. Nowadays, it is possible to get below a CER of 0.2%,⁴⁰⁵ which means that over 99.8% of characters are recognised correctly.

However, these enthusiastic results correspond to an ideal situation, in which the basic material (images containing the text to be extracted) is of excellent quality. In reality, however, this is often far from the case, especially when

404. NEUDECKER, Clemens, BAIERER, Konstantin, GERBER, Mike, CLAUSNER, Christian, ANTONACOPOULOS, Apostolos & PLETSCHACHER, Stefan. A Survey of OCR Evaluation Tools and Metrics. In: *HIP '21: Proceedings of the 6th International Workshop on Historical Document Imaging and Processing*. New York: Association for Computing Machinery. 2021, p. 13–18. Available on: <https://doi.org/10.1145/3476887.3476888>.

405. WICK, Christoph, REUL, Christian & PUPPE, Frank. Calamari – A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition. *Digital Humanities Quarterly*. 2020, vol. 14, n° 2. Available on: <https://www.digitalhumanities.org/dhq/vol/14/2/000451/000451.html>; IGNAT, Oana, MAILLARD, Jean, CHAUDHARY, Vishrav & GUZMAN, Francisco. OCR Improves Machine Translation for Low-Resource Languages. In: *Findings of the Association for Computational Linguistics: ACL 2022*. New York: Association for Computational Linguistics. 2022, p. 1164-1174. Available on: <https://doi.org/10.18653/v1/2022.findings-acl.92>.

working with historical documents. The result of the OCR process can, in such situations, be of only mediocre quality, which can lead to serious problems.

This is the situation we are facing as part of a collaborative research project with the CAMILLE laboratory, dedicated to the history of Belgian journalism. This involves, amongst other things, trying to recover the identities of forgotten journalists, or reconstructing the shape of the newsrooms of the country's leading newspapers as they changed over time. To do this, we draw on a substantial press corpus. It comprises the daily editions of thirteen newspapers from the mid-19th century to the mid-20th — a total of 1,391,802 pages, covering a period from the mid-19th to the mid-20th century (Fig. 1). This corpus has been digitised and OCR-d (that is, submitted to OCR) by the Royal Library of Belgium (KBR), as part of the BelgicaPress project.

When a digital humanist starts to examine such questions on the basis of a corpus of such richness and size, he or she immediately brings to mind those modern texts mining tools that enable highly detailed and deep analyses. For example, we might consider applying topic modelling tools to the various articles that appear in newspaper issues in order to extract the topics they cover, study their chronological distribution, compare them, etc. With projects such as ours, however, matters are somewhat less clear-cut. The sheer number of textual errors in the digitised texts renders this kind of study far more difficult, if not impossible, without the mobilisation of special strategies.

Two main factors explain the presence of errors in our OCR-d corpus. On the one hand, the quality of the sources that have been scanned and used as input for the OCR model (whether paper or, as in this case, microfilm) is not always excellent. On the other hand, the OCR process was applied by KBR many years ago, when the process was less efficient than it is now. Of course, an institution of this type is not going to reapply the OCR process to all its digitised collections every 5 or 10 years, to take advantage of the evolution of relevant technologies.

The errors that appear in the data we manipulate fall into two categories. First, there are the character recognition errors themselves, i.e. the mistakes made by the classification model trained to recognise the shape of letters, numbers and so on. Within our corpus, there is considerable variance in CERs. For

example, for the same medium (here *Le XXe Siècle*), we find an issue from 1932 with a CER of 1%, and another from 1898 with a CER of 7%. This variance can also be observed within the same issue, from one paragraph to the next. Some text blocks are almost perfectly clear, while others are completely illegible, even to a human, as in the example shown in Figure 2.



Fig. 2. *Le XXe Siècle*, issue of December 1, 1898, p. 1: comparison of the scan and the output of the OCR model for the subscription prices paragraph (the title and the two elements to the left of the braces were ignored by the OCR model).

The range of CER values observed in our corpus assigns it an intermediate level of quality according to the system established by the literature.⁴⁰⁶ However, this categorisation needs to be qualified: considering that a French word has an average length of 5 characters,⁴⁰⁷ a CER of 7% means that more than one word in three contains an erroneous character.

Second, there are segmentation errors. Segmentation is a step of the OCR

406. Good OCR accuracy for CER between 1% and 2%; Average OCR accuracy for CER between 2% and 10%; Poor OCR accuracy for CER higher than 10%. HOLLEY, ROSE. How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. *D Lib Magazine*. 2009, vol. 15, n° 3-4. Available on: <https://www.dlib.org/dlib/march09/holley/03holley.html>.

407. This calculation was carried out on a manually corrected corpus page of 4622 words.

process that takes place before character recognition itself, and involves detecting where text is located on the scanned page, identifying coherent blocks and lines of text, and then breaking them down into sequences of words and characters. If this stage goes well, the model correctly identifies columns, sets of lines forming articles, etc. Unfortunately, there are many errors of this type in our corpus. On many pages, the text blocks do not follow the newspaper's division of lines into editorial units. This means that some articles are split into several blocks, or on the contrary, some blocks cover several articles. By manipulating data with this type of error, you lose the editorial textual unit, which can lead to multiple problems. It is also not uncommon for the segmentation model to fail to distinguish between columns. The result is that the model reads two columns as if it were a single one. Each line of the block then consists of a concatenation of a line from the left-hand column with a line from the right-hand column, leading to text that makes no sense. Figure 3 shows an example of this problem.

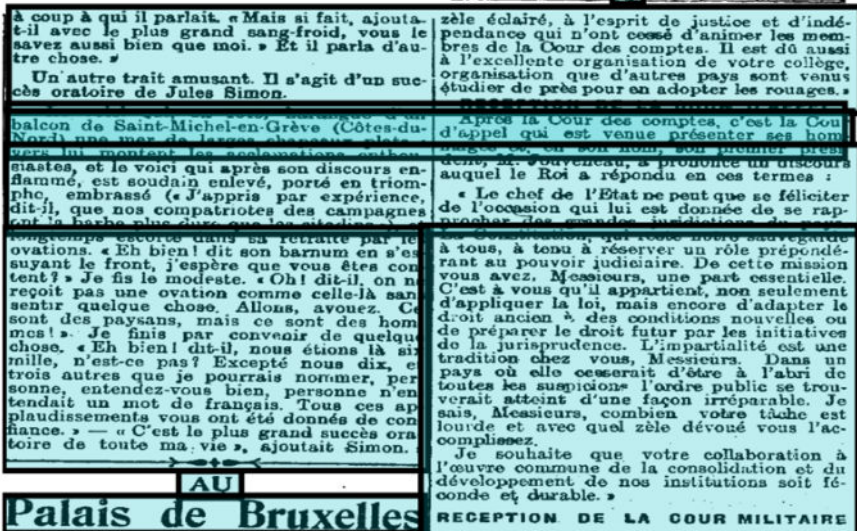


Fig. 3. Le XXe Siècle, issue of September 1, 1910, p. 1 (black frames correspond to text blocks, the result of the page segmentation process).

It is, of course, crucial to take account of these quality problems when posing research questions to the corpus. One might ask what sense there is in applying lexical analytical tools that work through fine manipulation of the words within a text when we know that almost a third of them contain at least one erroneous character? One solution to this problem is to test the robustness of the methods implemented in relation to these errors, in order to establish whether they are sufficiently 'stable' when applied to imperfect corpora. Several teams of researchers have undertaken this task, with variable but rather optimistic conclusions for the NLP tools most often applied in digital humanities.⁴⁰⁸ While very convincing, these methodological results do not ensure that a particular set of research questions on a particular corpus will yield relevant results, and caution and nuance remain the order of the day, of course.

The problem associated with segmentation errors is more difficult to deal with, because it is, so to speak, structural: it deprives us of the ability to divide the text of a newspaper issue into articles. All tools based on the comparison of well-defined documents, such as topic modelling, stylometry or sentiment analysis, are therefore *a priori* inapplicable. Does this mean that we should give up working on the corpus, and leave it to punctual, local queries that are not directly linked to research projects that look at things from a global perspective? No, we should not lose hope: it is possible to achieve good results, provided we work with these quality issues in mind and implement flexible methods that are adapted to the situation.

This is what we have started to do as part of our work to reconstruct news-

408. FRANZINI, Greta, KESTEMONT, Mike, ROTARI, Gabriela, JANDER, Melina, OCHAB, Jeremi K., FRANZINI, Emily, BYSZUK, Joanna & RYBICKI, Jan. Attributing Authorship in the Noisy Digitised Correspondence of Jacob and Wilhelm Grimm. *Frontiers in Digital Humanities*. 2018, vol. 5, n° 4. Available on: <https://doi.org/10.3389/fdigh.2018.00004>; MUTUVI, Stephen, DOUCET, Antoine, Odeo, Moses & JATOWT, Adam. Evaluating the Impact of OCR Errors on Topic Modelling. In: *International Conference on Asian Digital Libraries*. Cham: Springer. 2018, p. 3-14. Available on: https://doi.org/10.1007/978-3-030-04257-8_1; HAMDI, Ahmed, JEAN-CAURANT, Axel, SIDERE, Nicolas, COUSTATY, Mickaël & DOUCET, Antoine. An Analysis of the Performance of Named Entity Recognition over OCRed Documents. In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. New York: Institute of Electrical and Electronics Engineers (IEEE), 2019, p. 333-334. Available on: <https://doi.org/10.1109/JCDL.2019.00057>; HILL, Mark J. & HENGCHEN, Simon. Quantifying the Impact of Dirty OCR on Historical Text Analysis: Eighteenth Century Collections Online as a Case Study. *Digital Scholarship in the Humanities*. 2019, vol. 34, n° 4, p. 825-843. Available on: <https://doi.org/10.1093/lc/fqz024>.

rooms.⁴⁰⁹ To achieve this goal, we need to extract the names of the journalists who signed the articles in the issues of the corpus. Unfortunately, this task is heavily impacted by the two categories of error mentioned above. Firstly, the words that make up the signatures, most often anthroponyms, are more prone to character recognition errors. Most OCR models use language dictionaries, which enable them to correct reading errors, or at least to bend certain decisions made during the process in one direction or another, depending on the words that exist in the language being processed (and their frequencies). For example, if a model sees a word made up of the three letters ‘uno’ and knows that it is dealing with French, it could correct the last letter to an ‘e’ by consulting a dictionary (or a frequency list) and realising that ‘une’ is a much more common word in French than ‘uno’. Of course, this does not work for all strings. In the case of words that are not in the dictionary, such as names of people, places or institutions, it is not possible to use a list of existing forms, so the error rate is even greater. Based on a set of hand-scanned data (see below), we estimate that only 57.8% of signatures are free of OCR errors.

Secondly, segmentation errors have a major impact on signature identification. Since we cannot rely on the article delimitations given by the segmentation model, we cannot simply check whether the last line of each text block is a signature or not. We have to go through all the strings, even those that appear at the beginning or in the middle of a text block, as this block could correspond to two chunks of articles artificially merged by the segmentation model.

To get around this problem, we set up an automatic classification model designed to distinguish between those lines on a page that are signatures and those which are not. This is a supervised task: to work, such a model needs to be fed with a set of labelled data. In our case, this means supplying the computer with lines labelled ‘signature’ or ‘non-signature’. To obtain these, we manually extracted the signatures present in a sample of 756 pages from the corpus. We

409. LIBERT, Manon, JACQUET, Antoine, DE VALERIOLA, Sébastien, LE CAM, Florence, VIDALENCHÉ, Alexia & DESEURE, Brecht. Repeupler les rédactions du passé à l’aide des humanités numériques. Le cas du quotidien belge *Le XXe Siècle*. *Questions de communication* 48 (2025), <https://doi.org/10.4000/15b19>; VIDALENCHÉ, Alexia, LE CAM, Florence, DE VALERIOLA, Sébastien, LIBERT, Manon & DESEURE, Brecht. Retracer (massivement) et analyser une carrière médiatique. Désinvisibiliser Alice Bron, journaliste, socialiste et féministe belge (1850-1904). *Le Temps des médias*. 2024, vol. 42, n° 1, p. 169-189. Available on: <https://doi.org/10.3917/tdm.042.0169>.

then provided the computer with a set of 20 pieces of information for each line (size, position, format, etc.), which are the features on which the model will rely to distinguish signatures from non-signatures. Four classification models were tested: Generalised Linear Model, Random Forest, Extreme Boosting and Neural Networks. The latter produced the best results, with a precision score (positive predictive value) of around 80%.

In the next stage of our research process, we will apply the model trained on this sample to all the pages of the newspaper *Le XXe siècle* and then, after possible re-training (perhaps made necessary by other editorial choices, etc.), to those of the other newspapers in the corpus. Before we have a list of the journalists who took part in each newspaper's writing process, two further steps will be necessary. First, we will need to clean up the model's output, which will contain a relatively large amount of noise, given the 80% precision score and the fact that some article signatures are not from journalists (but, for example, from characters in a feuilleton whose text includes a letter). Second, we will have to disambiguate the anthroponyms extracted, as they will sometimes appear with variations in spelling, due to OCR errors and possible different signature formats (with or without initial for the first name, etc.). Once these two steps have been completed, we will have all the pieces of the puzzle we need to reconstruct newsrooms from the past.



Bibliography

Pressing Matters. Introduction

- Abraham. Catalogue of Belgian Newspapers [Online]. Available on: <http://www.krantencatalogus.be>.
- Abraham 2020. Nieuw leven voor oud nieuws [Online]. Available on: <https://www.bibliothecairerfgoed.be/projecten/abraham-2020>.
- Het Archief [Online]. Available on: <https://hetarchief.be/>.
- Artpresse [Online]. Available on: <https://www.kbr.be/en/projects/artpresse/>.
- BelgicaPeriodicals [Online]. Available on: <https://www.belgicaperiodicals.be/>.
- BelgicaPress [Online]. Available on: <https://www.belgicapress.be>.
- Bruges Public Library [Online]. Available on: <https://brugge.bibliotheek.be/erfgoed>.
- CAMille. Centre d'archives sur les médias et l'information [Online]. Available on: <https://www.kbr.be/en/projects/camille/>.
- Chronicling America. Historic American Newspapers [Online]. Available on: <https://chroniclingamerica.loc.gov/>.
- CultuurOverleg Zeven [Online]. Available on: <https://www.co7.be/>.
- Delpher [Online]. Available on: <https://www.delpher.nl/>.
- Digitaal Krantenarchief – Stadsarchief Aalst [Online]. Available on: <https://aalst.courant.nu/>.
- Erfgoed Brugge [Online]. Available on: <https://erfgoedbrugge.be/>.
- Flanders Heritage Libraries [Online]. Available on: <https://bibliothecairerfgoed.be>.
- Gallica BNF [Online]. Available on: <https://gallica.bnf.fr/>.
- Ghent Centre for Digital Humanities [Online]. Available on: <https://www.ghentcdh.ugent.be/>.
- Hendrik Conscience Heritage Library [Online]. Available on: <https://consciencebibliotheek.be/>.
- Historische Kranten [Online]. Available on: <https://historischekranten.be/>.
- iCandid [Online]. Available on: <https://icandid.t.libis.be/>.
- Images et histoires des patrimoines numérisés [Online]. Available on: <http://www.numeriques.be/>.
- Impresso. Media Monitoring of the Past [Online]. Available on: <https://www.impresso-project.ch/>.
- KADOC-KU Leuven [Online]. Available on: <https://kadoc.kuleuven.be/>.

- Kempens Erfgoed [Online]. Available on: <https://www.kempenserfgoed.be/>.
- Krantencollectie.be. Historische kranten uit Leuven en Halle [Online]. Available on: <https://krantencollectie.be/>.
- Laboratoire des pratiques et identités journalistiques [Online]. Available on: <https://lapij.ulb.ac.be/rubrique/publications/carnets-lapij/>.
- Liberas [Online]. Available on: <https://www.liberas.eu/en/>.
- Meemoo. Flemish institute for archives [Online]. Available on: <https://meemoo.be/>.
- Mundaneum [Online]. Available on: <https://mundaneum.org/>.
- News of the Great War [Online]. Available on: <https://nieuwsvandegrooteoorlog.hetarchief.be>.
- NewsEye. A Digital Investigator for Historical Newspapers [Online]. Available on: <https://www.newseye.eu/>.
- Nieuwe Tijdingen [Online]. Available on: <https://www.bibliothecaierfgoed.be/projecten/nieuwe-tijdingen>.
- Nieuws van onder de kerktoeren [Online]. Available on: https://kadoc.kuleuven.be/8_projecten/2021/2021_06_parchiebladen.
- Numapresse. Du papier à l'écran [Online]. Available on: <http://www.numapresse.org/>.
- ODIS [Online]. Available on: <https://www.odis.be/>.
- QuaDiHum Centre [Online]. Available on: <https://quadihum.ulb.be/>.
- Ratio DH [Online]. Available on: <https://quadihum.ulb.be/ratio-dh/>.
- Tijdschriften in the picture [Online]. Available on: https://kadoc.kuleuven.be/8_projecten/2022/tijdschriften_picture/2022_03_tijdschriften-in-the-picture.
- Weekbladpers te Sint-Truiden [Online]. Available on: <https://www.wudesk-sast.eu/ucat/sast>.
- BINGHAM, Adrian. The Digitization of Newspaper Archives: Opportunities and Challenges for Historians. *Twentieth Century British History*. 2010, vol. 21, n° 2. DOI: <https://doi.org/10.1093/tcbh/hwq007>.
- BUNOUT, Estelle, EHRMANN, Maud & CLAVERT, Frédéric (eds.). *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. Berlin (Germany): De Gruyter Oldenbourg, 2023. DOI: <https://doi.org/10.1515/9783110729214>.
- EHRMANN, Maud, BUNOUT, Estelle & CLAVERT, Frédéric. Introduction. In: BUNOUT, Estelle, EHRMANN, Maud & CLAVERT, Frédéric (eds.). *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*.

- Berlin (Germany): De Gruyter Oldenbourg, 2023, p. 1-24. DOI: <https://doi.org/10.1515/9783110729214>.
- FICKERS, Andreas. Towards a New Digital Historicism? Doing History in the Age of Abundance. *VIEW Journal of European Television History and Culture*. 2012, vol. 1, n° 1, pp. 19–26.
 - KOOLEN, Marijn, VAN GORP, Jasmijn & VAN OSSENBRUGGEN, Jacco. Toward a Model for Digital Tool Criticism: Reflection as Integrative Practice. *Digital Scholarship in the Humanities*. 2019, vol. 34, n° 2, pp. 368–385, DOI: <https://doi.org/10.1093/llc/fqy048>.
 - KRIEG, Alice. Analyser le discours de presse, *Communication*. 2000, vol. 20, n° 1. Available online: <http://journals.openedition.org/communication/6432>. DOI: <https://doi.org/10.4000/communication.6432>.
 - LALLICH-BOIDIN, Geneviève & BRUN-PICARD, Céline. Collections de presse, numérisation et patrimoine des bibliothèques. *Culture & musées*. 2013, vol. 21, n° 1, pp. 89-110.
 - MOIRAND, Sophie. *Les discours de la presse quotidienne. Observer, analyser, comprendre*. Paris (France): Presse universitaires de France, 2017.
 - PRÉMEL, Benjamin. La numérisation de la presse rétrospective à la BnF. *Le Temps des médias*. 2007, vol. 8, n° 1, 267-278. DOI: <https://doi.org/10.3917/tdm.008.0267>.
 - WILKINSON, Mark D., DUMONTIER, Michel, AALBERSBERG, Ijsbrand Jan, and others. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data*. 2026, vol. 15, n° 3. DOI: <https://doi.org/10.1038/sdata.2016.18>.

1. Reflections on a Workshop: Research Challenges and Opportunities for Belgian Print Media in the Age of Digitisation

- Abraham. Catalogue of Belgian Newspapers [Online]. Available on: <http://www.krantencatalogus.be>.
- Always Already Computational: Collections as Data [Online]. Available on: <https://collectionsasdata.github.io/>.
- Impresso. Media Monitoring of the Past [Online]. Available on: <https://www.impresso-project.ch/>.
- ONB LABS [Online]. Available on: <https://labs.onb.ac.at/en/>.
- ARENAS, Diego, ATKINS, John, AUSTIN, Claire and others. *Design Choices for*

Productive, Secure, Data-Intensive Research at Scale in the Cloud. 2019 [Online]. Available on: <https://arxiv.org/abs/1908.08737>.

- BEELEN, Kasper. *Digitising Newspapers Press Directories to Understand the Landscape of Historical Newspapers*. 2023 [Online]. Available at: <https://livingwithmachines.ac.uk/digitising-newspapers-press-directories-to-understand-the-landscape-of-historical-newspapers/>.

- BIRKHOLZ, Julie M. *Digital Workflows for Historical Newspapers & Periodicals: Why Both the Start and the End Matters*. Conference lecture. 2023.

- DE VALERIOIA, Sébastien. *Automatic Recognition of Signatures in a Corpus of Belgian Newspapers*. Conference lecture. 2023.

- DÜRING, Martin & EHRMANN, Maud. *Impresso - Media Monitoring of the Past*. Conference lecture. 2023.

- HERMANS, Sébastien. *Cataloguing Individual Issues of Digitised Historical Magazines*. Conference lecture. 2023.

- KEMMAN, Max & CLAEYSENS, Steven. User Demand for Supporting Advanced Analysis of Historical Text Collections. *DH Benelux 2022 - ReMIX: Creation and alteration in DH* (hybrid), Belval Campus, Esch-sur-Alzette, Luxembourg and online. Zenodo. DOI: <https://doi.org/10.5281/zenodo.6595769>.

- WEYNS, Katrien. *Automation of journal selection for digitisation in a fragmented library landscape*. Conference lecture. 2023.

2. Digital Access to Historical Newspapers in Belgium and Abroad. From Endangered Heritage to a Gold Mine for Machine Learning Technology

- BelgicaPress [Online]. Available on: <https://www.belgicapress.be>.

- Chronicling America. Historic American Newspaper [Online]. Available on: <https://chroniclingamerica.loc.gov/>.

- DATA-KBR-BE [Online]. Available on: <https://www.kbr.be/en/projects/data-kbr-be/>.

- Delpher [Online]. Available on: <https://www.delpher.nl/>.

- DSPH [Online]. Available on: <https://ai.vub.ac.be/dsph/>.

- Europeana Newspapers. Available on: <http://www.europeana-newspapers.eu/>.

- Faro. Vlaams steunpunt voor cultureel erfgoed [Online]. Available on: <https://faro.be/>.
- Flanders Heritage Libraries [Online]. Available on: <https://bibliothecaierfgoed.be>.
- Het Archief [Online]. Available on: <https://hetarchief.be/>.
- Impresso. Media Monitoring of the Past [Online]. Available on: <https://www.impresso-project.ch/>.
- *Increased Contextualisation of Humanist Intangible Cultural Heritage (ICHICH)*. CAVA-VUB [online]. Elsene (Belgium): Universiteitsarchief van de VUB en het Centrum voor Vrijzinnig Humanistisch Erfgoed vzw. [consulted on 09/02/2024]. Available on: <https://cavavub.be/nl/over-ons/projecten/increased-contextualisation-of-humanist-intangible-cultural-heritage-ichich-2022-2024>.
- *KBR lanceert een nieuw en groots opgezet project voor de digitalisering van de Belgische dagbladders*. Kbr.be [online]. Brussels (Belgium): KBR. 25/04/2024. [consulted on 28/10/2024]. Available on: <https://www.kbr.be/nl/kbr-lanceert-een-nieuw-en-groots-opgezet-project-voor-de-digitalisering-van-de-belgische-dagbladders/>.
- Liberas [Online] Available on: <https://www.liberas.eu/en/>.
- Meemoo. Flemish institute for archives [Online]. Available on: <https://meemoo.be/>.
- NewsEye. A Digital Investigator for Historical Newspapers [Online]. Available on: <https://www.newseye.eu/>.
- Nieuwe Tijden [Online]. Available on: <https://www.bibliothecaierfgoed.be/projecten/nieuwe-tijden>.
- News of the Great War [Online]. Available on: <https://nieuwsvandegrooteoorlog.hetarchief.be>.
- RetroNews [Online]. Available on: <https://www.retronews.fr/>.
- *Strategische Adviesraad voor Cultuur, Jeugd, Sport en Media, Advies over het ontwerpbesluit bij het decreet digitale transformatie van de cultuursector*, 6 July 2023,
- VR 2023 0809 DOC.1161/5. [consulted on 10/12/2023]. Available on: <https://beslissingenvlaamseregering.vlaanderen.be/document-view/64F6D7383605E1AC863BE1D3>.
- Trove. [consulted on 19/04/2025]. Available on: <https://trove.nla.gov.au/>.
- *Vlaamse Erfgoedbibliotheken vzw, Rapport Communicatie & Valorisatie Gedigitaliseerde Kranten*, January 2024, D/2024/12377/1. [consulted on 08/02/2024]. Available on: <https://vlaamse-erfgoedbibliotheken.be/bron/5088>.
- ALI, Dilawar, MILLEVILLE, Kenzo, VERSTOCKT, Steven, VAN DE WEGHE,

Nico, CHAMBERS, Sally & BIRKHOLZ, Julie M. Computer Vision and Machine Learning Approaches for Metadata Enrichment to Improve Searchability of Historical Newspaper Collections. *Journal of Documentation*. 2024, vol. 80, n° 5, p. 1031-1056. DOI: <https://doi.org/10.1108/JD-01-2022-0029>.

- BEALS, Melodee & BELL, Emily, with additional contributions by CORDELL, Ryan, FYFE, Paul, GALINA RUSSELL, Isabel, HAUSWEDELL, Tessa, NEUDECKER, Clemens, NYHAN, Julianne, PADÓ, Sebastian, PEÑA PIMENTEL, Miriam, OIVA, Mila, ROSE, Lara, SALMI, Hannu, TERRAS, Melissa & VIOLA, Lorella. *The Atlas of Digitised Newspapers and Metadata: Reports from Oceanic Exchanges* [online]. Loughborough (United Kingdom), 2020. [consulted on 15/11/2023]. Available on: <https://doi.org/10.6084/m9.figshare.11560059>.

- BUNOUT, Estelle, EHRMANN, Maud & CLAVERT, Frédéric (eds.). *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. Berlin (Germany): De Gruyter Oldenbourg, 2023. Available on: <https://doi.org/10.1515/9783110729214>.

- BUYSSE, Jeroen. Sterke aangroei aanbod gedigitaliseerde kranten. *Liberas.eu* [online]. Ghent (Belgium): Liberaz vzw. 01/04/2022. [consulted on 14/02/2024]. Available on: <https://www.liberas.eu/sterke-aangroei-aanbod-gedigitaliseerde-periodieken/>.

- BUYSSE, Jeroen. Journal de Gand (1815-1923): Een nieuwe digitale bron. *Liberas.eu* [online]. Ghent (Belgium): Liberaz vzw. 07/06/2024. [consulted on 28/10/2024]. Available on: <https://www.liberas.eu/journal-de-gand-1815-1923-eeen-nieuwe-digitale-bron/>.

- DE SPIEGELEER, Christoph. De digitale toegang tot historische kranten. Van bedreigd erfgoed tot een goudmijn voor machinelearning. *META. Tijdschrift voor Bibliotheek en Archief*. 2023, n° 8, p. 10-15. Available on: <https://www.vvbad.be/meta/meta-nummer-20238/de-digitale-toegang-tot-historische-kranten>.

- DE VALERIOLO, Sébastien. Fouille au sein d'un corpus massif de presse numérisée pour étudier l'histoire du journalisme belge. *QuaDiHum Lab* [online]. Ixelles (Belgium): Université Libre de Bruxelles. 30/08/2023. [consulted on 07/11/2023]. Available on: <https://quadihum.ulb.be/fouille-au-sein-dun-corpor-massif-de-presse-numerisee-pour-etudier-lhistoire-du-journalisme-belge/>.

- DENYS, Montaine. Een terugblik op de vierde Collegagroep 'Digitalisering van Periodieken'. *Vlaamse Erfgoedbibliotheken* [online]. Berchem (Belgium): Vlaamse Erfgoedbibliotheken vzw. 11/10/2023. [consulted on

- 08/02/2024]. Available on: <https://www.bibliothecairerfgoed.be/nieuws/een-terugblik-op-de-vierde-collegagroep-digitalisering-van-periodieken>.
- EHRMANN, Maud, Explorer la presse numérisée: le projet Impresso. *Revue historique vandoise*, 129 (2021), p. 159-173.
 - EHRMANN, Maud, DÜRING, Marten, NEUDECKER, Clemens & DOUCET, Antoine. Computational Approaches to Digitised Historical Newspapers. *Dagstuhl Reports* [online]. 2023, vol. 12, n° 7, p. 112-179. [consulted on 05/12/2023]. Available on: 10.4230/DagRep.12.7.112.
 - ROCHMES, Sophia, DENYS, Montaine & COPPOOLSE, David. Nieuwe OCR voor bestaande digitale krantencollecties: de moeite waard? *META. Tijdschrift voor Bibliotheek en Archief*. 2023, n° 5, p. 10-18. Available on: <https://www.vvbad.be/meta/meta-nummer-20235/nieuwe-ocr-voor-bestaande-digitale-krantencollecties-de-moeite-waard>.
 - VAN DORMOLEN, Hans. *Metamorfoze Preservation Imaging Guidelines*. The Hague: Koninklijke Bibliotheek, 2012. Available online: https://www.metamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf.

3. Belgian Press Heritage Online. Digitisation of Printed Media Collections at KBR

- ACDC [Online]. Available on: <https://www.uantwerpen.be/en/research-groups/digitalhumanities/>.
- *Always Already Computational: Collections as Data* [Online]. Available on: <https://collectionsasdata.github.io/>.
- AMSAB [Online]. Available on: <https://www.amsab.be/>.
- Artpresse [Online]. Available on: <https://www.kbr.be/en/projects/artpresse/>.
- The Belgian War Press [Online]. Available on: <https://warpress.cegesoma.be/>.
- Belgica [Online]. Available on: <https://belgica.kbr.be/>.
- BelgicaPeriodicals [Online]. Available on: <https://www.belgicaperiodicals.be/>.
- BelgicaPress [Online]. Available on: <https://www.belgicapress.be>.
- Bibliothèque Ulysse Capitaine [Online]. Available on: <https://lesmuseesdeliege.be/les-fonds-patrimoniaux/>.
- BRAIN-be 2.0 [Online]. Available on: <https://www.belspo.be/BRAIN-be/>.
- British Library's Collection Datasets [Online]. Available on: <https://bl.iro.bl.uk/>

- collections/64e3804a-788a-4c4b-962c-ae180d955455?locale=en.
- British Library's Research Repository [Online]. Available on: <https://bl.iro.bl.uk/>.
 - Brussels City Archives [Online]. Available on: <https://archives.brussels.be/>.
 - CAMille. Centre d'archives sur les médias et l'information [Online]. Available on: <https://www.kbr.be/en/projects/camille/>.
 - CARCoB [Online]. Available on: <https://carcob.eu/>.
 - CegeSoma [Online]. Available on: <https://www.cegesoma.be/>.
 - *Collections as Data: Part to Whole* [Online]. Available on: <https://collectionsasdata.github.io/part2whole/>.
 - *Common European data space for cultural heritage* [Online]. Available on: <https://pro.europeana.eu/>.
 - DARIAH-EU [Online]. Available on: <https://www.dariah.eu/>.
 - DATA-KBR-BE [Online]. Available on: <https://www.kbr.be/en/projects/data-kbr-be/>.
 - Data Science Lab [Online]. Available on: <https://www.kbr.be/en/projects/data-science-lab/>.
 - Digital Library of the Caribbean's DLoc as a Data repository [Online]. Available on: <https://dataverse.fiu.edu/dataverse/dloc-as-data>.
 - Digital Research Lab [Online]. Available on: <https://www.kbr.be/en/projects/digital-research-lab/>.
 - European Open Science Cloud [Online]. Available on: <https://open-science-cloud.ec.europa.eu/>.
 - ESFRI-FED [Online]. Available on: https://www.belspo.be/belspo/organisation/call/ESFRI-FED_2021_en.stm.
 - Ghent Centre for Digital Humanities [Online]. Available on: <https://www.ghentcdh.ugent.be/>.
 - Impresso. Media Monitoring of the Past [Online]. Available on: <https://www.impresso-project.ch/>.
 - International GLAM Labs Community [Online]. Available on: <https://glamlabs.io/>.
 - KBR General Catalogue [Online]. Available on: <https://opac.kbr.be/>.
 - KBR Virtual Lab [Online]. Available on: https://www.belspo.be/belspo/ESFRI-FED/project_en.stm#KBRVirtualLab.
 - National Library of Luxembourg's Open Data Platform [Online]. Available on: <https://data.bnl.lu/data/>.

- National Library of Scotland's Data Foundry [Online]. Available on: <https://data.nls.uk/>.
- National Library of the Netherlands' Data Services [Online]. Available on: <https://www.kb.nl/en/research-find/for-researchers/data-services-apis-and-downloads>.
- Pop Heritage Lab [Online]. Available on: <https://www.kbr.be/en/projects/pop-heritage-lab/>.
- Presse18 [Online]. Available on: <https://www.gazettes18e.fr/>.
- ReConFort [Online]. Available on: <https://www.reconfort.uni-passau.de/en/>.
- Réseau namurois de lecture publique [Online]. Available on: <https://www.namur.be/fr/loisirs/culture/lectures/bibliotheques/bibliotheques-du-reseau>.
- Zentrum für Ostbelgische Geschichte [Online]. Available on: <https://www.geschichte.be/>.
- ALKEMADE, Henk, CLAEYSSSENS, Steven, COLAVIZZA, Giovanni, FREIRE, Nuno, LEHMANN, Jörg, NEUDECKER, Clemens, OSTI, Giulia & VAN STRIEN, Daniel. Datasheets for Digital Cultural Heritage Datasets. *Journal of Open Humanities Data*. 2023, vol. 9, n° 1. DOI: <https://doi.org/10.5334/johd.124>.
- ALI, Dilawar, MILLEVILLE, Kenzo, VERSTOCKT, Steven, VAN DE WEGHE, Nico, CHAMBERS, Sally & BIRKHOLZ, Julie M. Computer Vision and Machine Learning Approaches for Metadata Enrichment to Improve Searchability of Historical Newspaper Collections. *Journal of Documentation*. 2024, vol. 80, n° 5, p. 1031-1056. DOI: <https://doi.org/10.1108/JD-01-2022-0029>.
- AMES, Sarah & LEWIS, Stuart. Disrupting the Library: Digital Scholarship and Big Data at the National Library of Scotland. *Big Data & Society*. 2020, vol. 7, n° 2. DOI: <https://doi.org/10.1177/2053951720970576>.
- ARBLASTER, Paul. *From Ghent to Aix: How They Brought the News in the Habsburg Netherlands*. Leiden and Boston: Brill, 2014.
- BRUCHER, Roger. *Les collections de périodiques et de presse de la bibliothèque royale Albert Ier. Historique, structure et gestion*. Brussels: Royal Library of Belgium, 1984.
- BRUCHER, Roger. *Les fonds de journaux: de la Bibliothèque royale Albert Ier. Évolution, aménagement et gestion récente d'une structure bibliothéconomique spécifique*. Brussels: Royal Library of Belgium, 1990.
- CANDELA, Gustavo, SÁEZ, María Dolores, ESCOBAR ESTEBAN, M Pilar & MARCO-SUCH, Manuel. Reusing Digital Collections from GLAM Institutions. *Journal of Information Science*. 2020, vol. 48, n° 2, p. 251-267. DOI: <https://doi.org/10.1177/0165551520950246>.

- CANDELA, Gustavo, GABRIËLS, Nele, CHAMBERS, Sally, PHAM, Thuy-An, AMES, Sarah, FITZGERALD, Neil, HOFMANN, Katrine, HARBO, Victor, POTTER, Abigail, FERRITER, Meghan, MANCHESTER, Eileen, IROLLO, Alba, VAN KEER, Ellen, MAHEY, Mahendra, HOLOWNIA, Olga & DOBREVA, Milena. *A Checklist to Publish Collections as Data in GLAM Institutions*. 2023. Available at: <https://arxiv.org/abs/2304.02603v1>.
- CANDELA, Gustavo, CHAMBERS, Sally & IROLLO, Alba. *A Workflow to Publish Collections as Data: the Case of Cultural Heritage Data Spaces*. 2023. Available at: <https://marketplace.sshopencloud.eu/workflow/I3jvP6>.
- COPPOOLSE, David. De krachten bundelen. De Vlaamse Erfgoedbibliotheek als expertisenetwerk digitalisering. *Archief- en bibliotheekwezen in België*. 2011, vol. 83, n° 1-4, p. 31-46.
- D'HOORE, Marc. Un exemple de projet de numérisation de masse. Les journaux de la Bibliothèque royale de Belgique (KBR). *Archief- en bibliotheekwezen in België*, 2011, vol. 83, n° 1-4, p. 117-134.
- D'HOORE, Marc, LEMMERS, Frédéric & LEYDER, Dirk. Digitising the Written Cultural Heritage: from Virtual to Real. *Archief- en bibliotheekwezen in België*. 2011, vol. 83, n° 1-4, p. 11-14.
- DER WEDUWEN, Arthur. *Dutch and Flemish Newspapers of the Seventeenth Century, 1618-1700*. Leiden: Brill, 2017.
- GOOSSENS, Alain. Le projet fédéral belge de numérisation et coupures de presse. Une préoccupation quotidienne! *Archief- en bibliotheekwezen in België*, 2007, vol. 78, n° 1-4, p. 185-196.
- LENTZEN, Evelyne. La politique de numérisation des patrimoines culturels en Fédération Wallonie-Bruxelles. *Archief- en bibliotheekwezen in België*. 2011, vol. 83, n° 1-4, p. 47-58.
- LUYTEN, Dirk & D'HOORE, Marc. Nationaal project voor de bewaring van Belgische kranten "Presse-Papier". *Bladen voor documentatie*. 2006, n° 2, p. 4-8.
- METTENS, Philippe. Het Belgisch digitaliseringsplan voor de Federale Wetenschappelijke Instellingen en het Koninklijk Belgisch Filmarchief. *Archief- en bibliotheekwezen in België*. 2011, vol. 83, n° 1-4, p. 15-30.
- OBERBICHLER, Sarah, BOROŞ, Emanuela, DOUCET, Antoine, MARJANEN, Jani, PFANZELTER, Eva, RAUTIAINEN, Juha, TOIVONEN, Hannu &

TOLONEN, Mikko. Integrated Interdisciplinary Workflows for Research on Historical Newspapers: Perspectives from Humanities Scholars, Computer Scientists, and Librarians. *Journal of the Association for Information Science and Technology*. 2022, vol. 73, n° 2, p. 225-239. DOI: <https://doi.org/10.1002/asi.24565>.

- PADILLA, Thomas, ALLEN, Laurie, FROST, Hannah, POTVIN, Sarah, RUSSEY Roke, Elizabeth & VARNER, Stewart. *Final Report: Always Already Computational: Collections as Data*. 2019. DOI: <https://osf.io/mx6uk/wiki/home/>.

- TASOVAC, Toma, CHAMBERS, Sally & TÓTH-CZIFRA, Erzsébet. 2020. *Cultural Heritage Data from a Humanities Research Perspective: a DARLAH Position Paper*. Available at: <https://hal.science/hal-02961317/>.

- VAN DEN DUNGEN, Pierre. *Milieux de presse et journalistes en Belgique (1828-1914)*. Brussels: Académie royale de Belgique, 2005.

- VAN DE HEUVEL, Marcel. Achter de schermen van een bewaarplaats voor oude en nieuwe tijdingen. Kranten in de Stadsbibliotheek Antwerpen. *Archief- en bibliotheekwezen in België*. 2007, vol. 78, n° 1-4, p. 133-146.

- VAN DORMOLEN, Hans. *Metamorfoze Preservation Imaging Guidelines*. The Hague: Koninklijke Bibliotheek, 2012. Available online: https://www.metamorfoze.nl/sites/default/files/documents/Metamorfoze_Preservation_Imaging_Guidelines_1.0.pdf.

- WITTMANN, Rachel, NEATROUR, Anna, CUMMINGS, Rebekah & MYN'TTI, Jeremy. From Digital Library to Open Datasets. *Information Technology and Libraries*. 2019, vol. 38, n° 4, p. 49-61. DOI: <https://doi.org/10.6017/ital.v38i4.11101>.

4. Nieuwe Tijdingen: Towards a Newspaper Preservation Programme for Flanders

- Abraham — Catalogus van Belgische kranten [Online]. Available on: <https://krantencatalogus.be/>.

- Abraham — Participating institutions [Online]. Available on: <https://manual.krantencatalogus.be/p/BNC:Lists/Collections>.

- ADVN [Online]. Available on: <https://advn.be>.

- AMSAB [Online]. Available on: <https://www.amsab.be/>.

- Het Archief [Online]. Available on: <https://hetarchief.be/>.

- BelgicaPress [Online]. Available on: <https://www.belgicapress.be>.

- Bibliotheek Hasselt Limburg [Online]. Available on: <https://hasselt.bibliotheek.be/>.
- Bruges public library [Online]. Available on: <https://brugge.bibliotheek.be/erfgoed>.
- Courant [Online]. Available on: <https://courant.krantencatalogus.be/>.
- Erfgoed Brugge [Online]. Available on: <https://erfgoedbrugge.be/>.
- GIVE-krantenproject Primeur [Online]. Available on: <https://meemoo.be/nl/projecten/give-krantenproject-primeur>.
- Hendrik Conscience Heritage Library [Online]. Available on: <https://consciencebibliotheek.be/>.
- Historische Kranten [Online]. Available on: <https://historischekranten.be/>.
- KADOC [Online]. Available on: <https://kadoc.kuleuven.be/>.
- Kempens Erfgoed [Online]. Available on: <https://www.kempenserfgoed.be/>.
- KU Leuven Libraries [Online]. Available on: <https://bib.kuleuven.be/>.
- Meemoo. Vlaams Instituut voor het Archief [Online]. Available on: <https://meemoo.be/>.
- Nieuwe Tijdingen [Online]. Available on: <https://meemoo.be/nl/projecten/nieuwe-tijdingen-redt-historische-kranten-van-het-verval>.
- Nieuws van de Grote Oorlog [Online]. Available on: <https://nieuwsvandegrooteoorlog.hetarchief.be>.
- University Library Ghent [Online]. Available on: <https://lib.ugent.be/>.
- Vlaamse Erfgoedbibliotheken [Online]. Since 2026: bibliothecairerfgoed.be. Available on: <https://www.bibliothecairerfgoed.be/>.

5. Transforming ULB Libraries Digitisation Service: Going Beyond Scanning Our Own Collections to Also Become a Data Provider for Our Researchers.

- BelgicaPress [Online]. Available on: <https://www.belgicapress.be>.
- Collectif interuniversitaire d'étude du littéraire [Online]. Available at <http://ciel.philo.uliege.be/cielcms/>.
- La Digithèque des Bibliothèques de l'ULB [Online]. Available at: <https://digitheque.ulb.ac.be/>.
- OrdoNum [Online]. Available on: <https://quadihum.ulb.be/ordonum/>.
- Praesto DH [Online]. Available on: <https://quadihum.ulb.be/praesto-dh/>.
- ARON, Paul. & SOUCY, Pierre-Yves. *Les Revues littéraires belges de langue française de 1830 à nos jours*. Brussels: Labor, 1998.

- HOLLEY, Rose. How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitization Programs, *D-Lib Magazine*. 2009, vol. 15, n° 3-4.
- KIESSLING, Benjamin. Kraken—A Universal Text Recognizer for the Humanities, *DataverseNL* [Online]. 2019, V2, <https://doi.org/10.34894/Z9G2EX>.
- KIESSLING, Benjamin, TISSOT, Robin, STOKES, Peter & STÖKL BEN EZRA, Daniel (2019), EScriptorium: An Open Source Platform for Historical Document Analysis, *International Conference on Document Analysis and Recognition Workshops* (online). Sydney, NSW, Australia, 2019, <https://doi.org/10.1109/ICDARW.2019.10032>.
- LEROY, Anthony. LOCKSS Distributed Digital Preservation Networks, *Proceedings of the 11th Conference on Grey Literature and Repositories* (online). Prague: National Library of Technology, 2018, ISSN 2336-5021. Available on: <https://nusl.techlib.cz/en/conference/conference-proceedings>.
- LEROY, Anthony, & PAUWELS, Benoit. Ph.D. *Theses Mass Digitization at ULB*, unpublished poster session presented at *ETD 2016 Data and Dissertations 19th International Symposium on Electronic Theses and Dissertations* (11-13 July 2016): Lille (France), <http://hdl.handle.net/2013/ULB-DIPOT:oai:dipot.ulb.ac.be:2013/231956>.
- MAYEUR, Ingrid. *ARC 'Presse et littérature en Belgique francophone' (1920-1960) : Note générale au sujet des numérisations*, 3 October 2013 [internal report].
- NGUYEN, Thi Tuyet Hai, JATOWT, Adam, COUSTATY, Mickael. & DOUCET, Antoine. Survey of Post-OCR Processing Approaches, *ACM Comput. Surv.* 2022, vol. 54, n° 6, article 124, <https://doi.org/10.1145/3453476>.
- PADILLA, Thomas, ALLEN Laurie, FROST, Hannah, POTVIN, Sarah, RUSSEY ROKE, Elisabeth & VARNER, Stewart. *Santa Barbara Statement on Collections as Data. Always Already Computational: Collections as Data*. Zenodo [Online], 20 May 2019, <https://doi.org/10.5281/zenodo.3066209>.
- PADILLA, Thomas, SKATES KETTLER, Hannah, VARNER, Stewart, & SHORISH, Yasmeen. *Vancouver Statement on Collections as Data*. Zenodo [Online], 13 September 2023, <https://doi.org/10.5281/zenodo.8342171>.
- ROSENTHAL, David S.H., ROBERTSON, Thomas, LIPKIS, Tom, REICH, Vicky & MORABITO, Seth. (2005). Requirements for Digital Preservation Systems: A Bottom-Up Approach. *D-Lib Magazine*, n° 11. Available on: <https://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.
- SKINNER Katherine & SCHULTZ Matt (eds.), *A Guide to Distributed Digital*

Preservation. Atlanta, GA (USA): Educopia Institute, 2010. Available on: https://metaarchive.org/wp-content/uploads/2017/03/A_Guide_to_Distributed_Digital_Preservation_0.pdf.

6. From Paper to ‘Paper’. Facilitating Structural Digitisation and Reuse of Print Media Preserved in a Heritage Institution

- Abraham. Catalogus van Belgische Kranten [online]. Antwerp (Belgium): Vlaamse Erfgoedbibliotheken. [consulted on 25/03/2026]. Available on: <https://www.bibliothecairerfgoed.be/>.
- Anet projecten [online]. Antwerp (Belgium): UAntwerpen. Since 2026: Wander. [consulted on 25/03/2026]. Available on: <https://www.wander.be>.
- *Commission opens formal proceedings against X under the Digital Services Act* [online]. Brussels (Belgium): European Commission. 18 December 2023. [consulted on 28/03/2024]. Available on: https://ec.europa.eu/commission/presscorner/detail/en/IP_23_6709.
- EduGAIN Entities [online]. GÉANT. 2024. [consulted on 28/03/2024]. Available on: <https://technical.edugain.org/entities>.
- *Europese richtlijn over auteursrechten en naburige rechten in de digitale vengemaakte markt – omzetting in Belgisch recht* [online]. Brussels (Belgium): FOD Economie. 27 September 2024. [consulted on 28/03/2024]. Available on: <https://economie.fgov.be/nl/themas/intellectuele-eigendom/intellectuele-eigendomsrechten/auteursrecht-en-naburige/auteursrechten/europese-richtlijn-over>.
- Fair by Design. Blog [online]. (Netherlands): UNLOCK. 2024. [consulted on 28/03/2024]. Available on: <https://fairbydesign.nl/>.
- FAIR principles [online]. Hamburg (Germany): GO FAIR. 2024. [consulted on 28/03/2024]. Available on: <https://www.go-fair.org/fair-principles/>.
- iCandid [Online]. Available on: <https://icandid.t.libis.be/>.
- IISN [online]. Brussels (Belgium): KBR. 2024. [consulted on 28/03/2024]. Available on: <https://www.kbr.be/en/issn/>.
- *Journals in the picture. Digital disclosure of periodicals via Wikimedia platforms and preparation for future digitisation* [online]. Leuven (Belgium): KADOC-KU Leuven. [consulted on 28/03/2024]. Available on: https://kadoc.kuleuven.be/8_projecten/2022/tijdschriften_picture/2022_03_tijdschriften-in-the-picture.

- KADOC-KU Leuven [online]. Available on: <https://kadoc.kuleuven.be>.
- Libis [online]. Leuven (Belgium): Libis. [consulted on 28/03/2024]. Available on: <https://libis.be/en/home>.
- Nieuwe Tijdingen [Online]. Available on: <https://www.bibliothecairerfgoed.be/projecten/nieuwe-tijdingen>.
- Publishing GLAM data as FAIR data [online]. Den Haag (Netherlands): Europeana Foundation. 2024. [consulted on 28/03/2024]. Available on: <https://pro.europeana.eu/event/publishing-glam-data-as-fair-data>; <https://arxiv.org/abs/2304.02603>.
- RDF AND JSON-LD UseCases [online]. W3C. 15 September 2024. [consulted on 28/03/2024]. Available on: https://www.w3.org/2013/dwbp/wiki/RDF_AND_JSON-LD_UseCases.
- Research API [online]. TikTok. 2024. [consulted on 25/03/2026]. Available on: <https://developers.tiktok.com/doc/research-api-get-started>.
- The International Centre for the registration of serial publications – CIEPS [online]. Paris (France): ISSN. 2024. [consulted on 28/03/2024]. Available on: <https://www.issn.org/the-centre-and-the-network/our-mission/the-international-centre-for-the-registration-of-serial-publications-cieps/>.
- TikTok Research API Terms of Service [online]. TikTok. 2024. [consulted on 28/03/2024]. Available on: <https://www.tiktok.com/legal/page/global/terms-of-service-research-api/en>.
- Twitter moderation research consortium application [online]. (US): X. 2024. [consulted on 25/03/2026]. Available on: https://blog.x.com/en_us/topics/company/2022/twitter-moderation-research-consortium-open-researchers.
- BERNERS-LEE, Tim. *Linked Data*. [online] 27 July 2006 [consulted on 28/03/2024]. Available on: <https://www.w3.org/DesignIssues/LinkedData>.
- CLEGG, Nick. *New Tools to Support Independent Research* [online]. (US): Facebook. 21 November 2023. [consulted on 28/03/2024]. Available on: <https://about.fb.com/news/2023/11/new-tools-to-support-independent-research/>.
- ROCHMES, Sophia, DENYS, Montaine & COPPOOLSE, David. Nieuwe OCR voor bestaande digitale krantencollecties: de moeite waard? *META Tijdschrift voor bibliotheek & archief*. 2023, n° 3, p. 10-16.
- TRIELOFF, Lars. Three Principles of API First Design. *Adobe Tech Blog* [online]. (US): Adobe. 2 June 2017. [consulted on 28/03/2024].

Available on: <https://blog.developer.adobe.com/three-principles-of-api-first-design-fa666d9f694>.
 - WEYNS, Katrien. Kom dat zien! Tijdschriften nog even in the picture. *KADOC-e-Nieuwsbrief* [online]. 2024, n° 1. [consulted on 28/03/2024]. Available on: https://kadoc.kuleuven.be/6_nieuwsbrieven/enbr/enbr_2024_01.

7. News from Around the Church. The Participatory Digitisation and Contextual Disclosure of Local Parish Journals

- KADOC-KU Leuven [Online]. Available on: <https://kadoc.kuleuven.be/>.
- De Katholieke Kerk in België 2023 [online]. S.l.: 2023. [consulted on 26/03/2024]. Available on: <https://www.kerknet.be/sites/default/files/Jaarrapport%20Katholieke%20Kerk%202023.pdf>.
- *De Kerk in het midden. Een halve eeuw parochieblad*. Antwerp (Belgium): Uitgeverij Halewijn, 1993.
- Libis [Online]. Available on: <https://libis.be/en/home>.
- Limo (KADOC) [Online]. Available on: <https://kadoc.limo.libis.be>.
- ODIS [Online]. Available on: <https://www.odis.be>.
- Parochiebladen/Kerk en Leven. [Online]. Available on: https://kadoc.limo.libis.be/discovery/collectionDiscovery?vid=32KUL_KADOC:KADOC&collectionId=815834000001480.
- COLLA, Joris. Voor de eeuwigheid. Zorg dragen voor parochiearchieven. *META. Tijdschrift voor bibliotheek & archief*. 2018, n° 2, pp. 26-30.
- COLLA, Joris. Het Sint-Mauritiusklokje online. *Erfgoedkrant Haspengouw*. 2023, vol. 10, n° 1, p. 7.
- COLLA, Joris & HEYRMAN, Peter. Het middenveld en zijn erfgoed in context: het veelzijdig gebruik van de onlinedatabank ODIS (2000-2017). In: LEYDER, Dirk & ISTASSE, Nathaël (eds.). *Ten dienste van de gebruiker. Over de zoekinstrumenten ontwikkeld door archivariissen en bibliothecarissen. Akten van het colloquium van 8 december 2017* (Archief- en Bibliotheekwezen in België, extra nummer 105). Brussels (Belgium): Archief- en Bibliotheekwezen in België, 2019. pp. 99-111.
- HERMANS, Roeland & VINTS, Luc (eds.). *Hemel en aarde. Religie in Vlaanderen sinds 1750*. Leuven (Belgium): Leuven University Press, 2017.
- HEYRMAN, Peter. Online tools and instruments supporting the study of transnational

élites: the functionalities of the contextual database ODIS. In: CIAMPANI, Andrea and KROLL, Thomas (eds.). *Transnational Encounters? European Elites, International Associations and National States (1882-1914)*. Berlin (Germany)/Boston (United States of America): De Gruyter/Oldenbourg, 2025, p. 117-128.

- MARYL, Maciej & GRABORSKA, Klaudia (eds.). *How to Facilitate the Cooperation between Humanities Researchers and Cultural Heritage Institutions. Guidelines* [online]. Warsaw (Poland): Digital Humanities Centre at the Institute of Literary Research of the Polish Academy of Sciences, 2019. [consulted on 28/03/2024]. Available on: <https://zenodo.org/records/2587481>.

- NEIRYNCK, Pieterjan. Kerk & Leven-uitgever wil 21 mensen ontslaan. *De Tijd*. 2024, vol. 54, n° 176 (8 September), p. 17.

- WEYNS, Katrien. Nieuws vanonder de kerktoren. META. Tijdschrift voor bibliotheek & archief. 2021, n° 9, pp. 10-15.

8. Newspapers in the Mundaneum's collection

- *Bibliothèque Collective des Sociétés Savantes*. Brussels: Institut International de Bibliographie, 1908, n° 89.

- *Le Musée international de la presse. Section de l'Institut international de bibliographie. Notice-catalogue*. Brussels: Institut International de Bibliographie, 1908, n° 108.

- *Le Musée international de la presse. Recueil publié par l'Association auxiliaire du Musée international de la presse*. 1912, vol. 1, n° 1-2, 6-8.

- *Union de la presse périodique belge. Bulletin officiel mensuel*. 1909, vol. 13, n° 7-8.

- GILLEN, Jacques, MANFROID, Stéphanie, MONTIGNIE, Aurélie & STRAGIER, Justine. *Le Mundaneum, machine à penser le monde*. Mons: Mundaneum, 2022.

- MAUS, Octave & MERTENS, Gaston. Création d'un musée de la presse à Bruxelles. Rapport présenté à l'Administration communale de Bruxelles, par le Comité de l'Union de la presse périodique belge. *La Presse universelle*, n° 2, 1896.

- MERTENS, Gaston. Projet de création à Bruxelles d'un Musée international de la presse. Rapport présenté par M. Gaston Mertens, au nom du bureau de l'Union de la presse périodique belge. *Union de la presse périodique belge. Bulletin officiel mensuel*. 1907, vol. 16, n° 10-12, p. 259.

- O'TLET, Paul. Conférence internationale de bibliographie et de documentation.

Mouvement sociologique international. 1908, vol. 9, n° 4, p. 5.

- WARZÉE, André. *Essai historique et critique sur les journaux belges*. Ghent: Hebbelynck, 1945.

9. *Historische Kranten: 20 Years of Digitising and Unlocking Newspapers in the Westhoek Region*

- Het Archief [Online]. Available on: <https://hetarchief.be/>.

- CultuurOverleg Zeven (CO₇) [Online]. Available on: <https://www.co7.be/>.

- ErgoedInzicht [Online]. Available on: <https://erfgoedinzicht.be/>.

- Meemoo. Flemish institute for archives [Online]. Available on: <https://meemoo.be/>.

- Geheugen van Nederland [Online]. Available on: <https://geheugenvannederland.nl/>.

- Nieuwe Tijdingen [Online]. Available on: <https://meemoo.be/nl/projecten/nieuwe-tijdingen-redt-historische-kranten-van-het-verval>.

- Westhoek Verbeeldt [Online]. Available on: <https://westhoekverbeeldt.be/>.

- SIMOENS, Tine & THIERS, Liesbeth. ‘Westhoek verbeeldt’: een buitenbeentje onder de beeldbanken. *FARO: Tijdschrift over cultureel erfgoed*. 2009, vol. 2, n° 1, pp. 28-33.

- WUYTS, Eva. Geheugen van Ieper: de digitale ontsluiting van lokale krantenarchieven. *Bibliotheek & Archiefgids*. 2007, vol. 83, n° 3, pp. 23-28.

10. *The Digitisation of Historical Newspapers in Bruges Public Library: Experiences and Perspectives*

- ALTO. Technical Metadata for Layout and Text Objects [online]. Washington DC: The Library of Congress. 08/06/2022 [consulted on 21/03/2024]. Available on: <https://www.loc.gov/standards/alto/>.

- ANNO Historische Zeitungen und Zeitschriften [online]. Available on: <https://anno.onb.ac.at/>.

- Bruges City Archives [online]. Available on: <https://www.brugge.be/archief>.

- Bruges Public Library [online]. Available on: <https://brugge.bibliotheek.be/erfgoed>.

- Collectie Musea Brugge [online]. Brugge: Musea Brugge. S.d. [consulted on 21/03/2024]. Available on: <https://www.museabrugge.be/collecties/ontdek>.

- Collectie Raakvlak [online]. Brugge: Raakvlak. S.d. [consulted on 21/03/2024].

- Available on: <https://collectie.raakvlak.be/>.
- Delpher [online]. Available on: <https://www.delpher.nl/>.
 - Deutsches Zeitungportal [online]. Available on: <https://www.deutsche-digitale-bibliothek.de/newspaper>.
 - Erfgoed Brugge [online]. Available on: <https://erfgoedbrugge.be/>.
 - ICCH [online]. Available on: <http://commissionroyalehistoire.be/en/reference-works/publications-of-the-icch/>.
 - Meemoo. Vlaams Instituut voor het Archief [Online]. Available on: <https://meemoo.be/>.
 - *Het meemoo-archiefsysteem* [online]. Gent: meemoo. S.d. [consulted on 21/03/2024]. Available on: <https://support.meemoo.be/hc/nl/categories/6039894771857-Het-meemoo-archiefsysteem>.
 - *Metadatauitwisseling met OAI-PMH. Protocol en implementatie* [online]. Hilversum: Beeld & Geluid. s.d. [consulted on 21/03/2024]. Available on: https://publications.beeldengeluid.nl/pub/451/BenG_Metadatauitwisseling_met_OAI-PMH.pdf.
 - Historische kranten [online]. Poperinge: CO₇ [consulted on 21/03/2024]. Available on: <https://historischekranten.be/>.
 - Nieuwe tijdingen: landelijk programma voor het behoud van het Vlaamse krantenerfgoed [online]. Antwerpen: Vlaamse Erfgoedbibliotheken vzw. 08/02/2021 [consulted on 25/03/2026]. Available on: <https://www.bibliotheccairerfgoed.be/projecten/nieuwe-tijdingen>.
 - Open Archives Initiative Protocol for Metadata Harvesting [online]. S.l. s.d. [consulted on 21/03/2024]. Available on: <https://www.openarchives.org/pmh/>.
 - Vlaamse Erfgoedbibliotheken [Online]. Available on: <https://vlaamse-erfgoedbibliotheken.be/>.
 - HUYSENTRUYT, Bart. Zo maken ze geen gazetten meer. *Het Laatste Nieuws*, 27/11/2016, p. 20.
 - PEVERELLI, Tymen. *De stad als vaderland. Brugge, Leuwarden en Maastricht in de eeuwen van de natiestaat, 1815–1914*. Nijmegen: Vantilt, 2019.
 - SOETAERT, Alexander & VIAENE, Dieter. Beeldbank Brugge: twenty years of making historic images accessible online [online]. Peccioli (Italy): *Photoconsortium*. 25/03/2021 [consulted on 21/03/2024]. Available on: <https://www.photoconsortium.net/beeldbank-brugge/>.
 - VAN EENOO, Romain. *De pers te Brugge 1792-1914: bouwstoffen*. Leuven: Nauwelaerts,

1961.

11. Constructing a Digital Corpus of Interwar Belgian Magazines to Investigate the Representations of Fine Arts. An Artpresse Case Study

- Actualités A-Z. (‘A l’exposition de Bruxelles, un monument vient d’être élevé à la gloire de Pieter Breughel’). *A-Z* 4:22 (18-08-1935), p. 14. Online: <<https://uurl.kbr.be/1901667/p14>>.
- *Annuaire publicitaire et commercial de Belgique*. Liège: Pim, 1932; 1933.
- Artpresse [online]. Available on: <<https://www.kbr.be/en/projects/artpresse>>.
- Bij het werk van Pieter Breugel. *Ons Volk* 10:37 (14-09-1924), p. 466-469. Online: <<https://uurl.kbr.be/1936064>>.
- Bij James Ensor. *Wereldrevue* 3:29 (17-08-1933), p. 18. Online: <<https://uurl.kbr.be/1903987/p18>>.
- BelgicaPeriodicals [online]. Available on: <https://www.belgicaperiodicals.be/>.
- BelgicaPress [online]. Available on: <https://www.belgicapress.be>.
- Ca représente quoi? *Le Centaure* 1:4 (01-01-1927), p. 75-77. Online: <<https://uurl.kbr.be/2202289/p7>>.
- De tentoonstelling van de oude Vlaamsche kunst. Het levensgevoel bij de Vlaamsche meesters. *De Stad* 3:23 (21-08-1930), p. 529-533; 540-541. Online: <<https://uurl.kbr.be/1921563>>.
- De persmacht der Katholieke Arbeidersjeugd. *KAJ*, 1934, 2:1, p. 8-9. Online: <<https://uurl.kbr.be/1970041/p8>>.
- James Ensor. *ABC* 3:44 (25-11-1934), p. 6-7 (ABC-0071). Online: <<https://uurl.kbr.be/1900653/p6>>.
- Kerstmis bij Breughel. *Ons Volk* XXI:51 (22-12-1935), p. 810-813. Online: <<https://uurl.kbr.be/1948132/p4>>.
- L’inauguration du monument Brueghel l’Ancien à l’exposition. *Le Patriote illustré* 51:33 (18-08-1935), p. 1029. Online: <<https://uurl.kbr.be/1953826/p5>>.
- Le monument national aux victimes civiles de la guerre. *Le Soir illustré* 9:445 (29-08-1936), p. 2. Online: <<https://uurl.kbr.be/1955283/p2>>.
- *Livre d’or de l’Exposition universelle et internationale de Bruxelles 1935*. Brussels: Comité exécutif de l’Exposition, 1935.

- Onthulling van een Brueghel-Gedenkteken op de tentoonstelling te Brussel. 1935 *Bulletin officiel de l'exposition universelle* 3:41 (17-08-1935), p. 1042. Online: <<https://uurl.kbr.be/2125480/p36>>.
- Onthulling van het borstbeeld James Ensor te Oostende. *Ons Land* 13:4 (26-04-1930), p. 59. Online: <<https://uurl.kbr.be/1942145/p11>>.
- SODHA [online]. Available on: <<https://www.sodha.be/>>.
- VERMETTE, Roger. Beaux-Arts. James Ensor à Paris. *Tout* 1:26 (08-07-1932), p. 27. Online: <<https://uurl.kbr.be/1903230/p27>>.
- BACHA, Eugène & DUPIERREUX, Richard. *Périodiques belges. Répertoire par titres et par sujets*. Brussels: Albert Dewit, 1928. Supplément, 1929.
- BIERME, Maria. Le quatrième centenaire de Pierre Bruegel. *Le Patriote illustré* 40:21 (25-05-1924), p. 328-329. Online: <<https://uurl.kbr.be/1938356/p8>>.
- BRUCHER, Roger. *Les collections de périodiques et de presse de la Bibliothèque royale Albert Ier, historique, structure et gestion*. Brussels: Bibliothèque royale Albert Ier, 1984.
- CASIER, Micheline. Monument au peintre James Ensor (Oostende). *Be Monumen. Répertoire participatif des bronzes et fontes belges*, March 2024. Online: <<https://be-monumen.be/patrimoine-belge/monument-au-peintre-james-ensor-oostende-ostende/>>.
- CURRIE, Christina & ALLART, Dominique. *The Brueg(h)el Phenomenon*. Brussels: Royal Institute for cultural Heritage (KIK-IRPA), 2012. Online: <<https://www.kikirpa.be/fr/publications/the-brueghel-phenomenon>>.
- D'HOORE, Marc. Un exemple de projet de numérisation patrimoniale de masse: les journaux de la Bibliothèque royale de Belgique (KBR). *Archives et Bibliothèques de Belgique. La numérisation du patrimoine écrit: du virtuel à la réalité*. 2011, vol. 82, n° 3, p. 117-134.
- DE BENS, Els and RAEYMAECKERS, Karin. *De pers in België*. Tiel: Lannoo, 2010.
- DE KEYSER, Léon. *La presse périodique. Annuaire de la presse périodique belge 1939-1940*. Brussels: Association des journalistes périodiques belges et étrangers, 1940.
- DE SOETE, Pierre. *Les incidents de Louvain. Furor Teutonico... Ses origines et ses polémiques. Le droit artistique et spirituel. 1914-1929*. Brussels: Imprimerie scientifique et littéraire, 1929.
- DENIS, Jean. *Bilan. Soirées* 109 (1933). Online: <<https://uurl.kbr.be/1927412/p3>>.
- DEVILLEZ, Virginie. Les peintres belges dans la tourmente. Du krach économique à la seconde guerre mondiale. *CHTP-BEG*. 1997, vol. 2, p. 36-38. Online: https://www.journalbelgianhistory.be/fr/system/files/article_pdf/cht2_002_Devillez.pdf.
- DEVILLEZ, Virginie. *Le retour à l'ordre. Art et politique en Belgique 1918-1945*. Brussels:

Labor, 2002.

- DEVILLEZ, Virginie. The Avant-Garde on the Reworking of Tradition. In: AUBERT, Nathalie, FRAITURE, Pierre-Philippe & MCGUINNESS, Patrick (eds.). *From Art Nouveau to Surrealism, Belgian Modernity in the Making*. London: Legenda, 2007, 142-151.

- EECKELS, Guido. Vlaanderen leeft. James Ensor. *Ons Land* 15:14 (02-07-1932), p. 300-301. Online: <<https://uurl.kbr.be/1942462/p28>>

- FIERENS, Paul. L'exposition James Ensor à Paris. *Hebdo* 1:31 (08-07-1932), p. 17. Online: <<https://uurl.kbr.be/1915261/p17>>.

- FLORIZOONE, Patrick & COPPEJANS, Willem. James Ensor tussen avant-garde en traditie. *James Ensor een online Museum*. Online: <<https://jamesensor.vlaamsekunstcollectie.be/nl/bronnen/webpublicaties/james-ensor-tussen-avant-garde-en-traditie/>>.

- FLOUQUET, P.L. Cinq siècles d'art bruxellois à l'exposition de 1935. *1935 Bulletin officiel de l'exposition universelle* 1:4 (15-11-1933), p. 139-142. Online: <<https://uurl.kbr.be/2125406/p21>>.

- FRIZOT, Michel. Photo/graphismes de magazines: les possibles de la rotogravure, 1926-1935. In: *Photo/graphisme. Actes du colloque « Photo / Graphisme » tenu au Jeu de Paume le 20 octobre 2007*. Paris: Editions de Jeu de Paume, 2008, p. 5-12.

- GEUKENS, Eric. *Het Dagblad «De Dag» (1934-1944)*. Unpublished dissertation, Ghent University, 1975.

- GHESQUIÈRE, Rita & QUAGHEBEUR, Patricia. *Averbode, een uitgever apart 1877-2002*. Leuven: Universitaire pers, 2002.

- GHIJS, Inge. *100 jaar Gazet van Antwerpen*. Antwerp: De Vlijt, 1991.

- HOOZEE, Robert. *Vlaams expressionisme in Europese context 1900-1930*. Ghent: Snoeck-Ducaju, 1990.

- HOTON, Edmond. Furore Teutonico Diruta. *Libra illustré* 1:23 (30-09-1933), p. 4-5. Online: <<https://uurl.kbr.be/1902425/p4>>.

- JOURDAIN, Virginie. Promouvoir l'Exposition: le règne de l'image et de la presse. In: JAUMAIN, Serge & BALCERS, Wanda (eds.). *Bruxelles 1910: de l'exposition universelle à l'université*. Brussels: Dexia-Racine, 2010, p. 94-98.

- L.R. L'art ancien. *1935 Bulletin officiel de l'exposition universelle* 3:30 (01-06-1935), p. 519-521. Online: <<https://uurl.kbr.be/2125458/p9>>.

- LAMBOTTE, Paul. L'exposition d'art italien à Londres. *Le Soir illustré* 3:101 (25-01-

- 1930), p. 6-7. Online: <<https://uurl.kbr.be/1952352/p6>>.
- LAUDE, Jean. Retour et/ou rappel à l'ordre. In: *Le Retour à l'ordre dans les arts plastiques et l'architecture, 1919-1925*. Saint-Étienne: CIEREC, Université de Saint-Étienne, 1975, p. 7-45.
 - LAUDE, Jean. La crise de l'humanisme et la fin des utopies (Sur quelques problèmes de la peinture et de la pensée européennes 1929-1939). In: *L'Art face à la crise. L'art en Occident, 1929-1939*. Saint-Étienne: CIEREC, 1980, p. 295-391.
 - LAUWAERT, Helke. Néo-réalisme. In: HOOZEE, Robert (ed.). *L'art moderne en Belgique 1900-1945*. Brussels: Mercatorfonds, 1992.
 - LEFÈVRE, G., James Ensor. *Ons Volk* XVI:17 (27-04-1930), p. 256-257. Online: <<https://uurl.kbr.be/1947183/p12>>.
 - LEMMERS, Frédéric, OTT, Morgane & HERMANS, Sébastien. Printed mass media and automatic digitisation: the case of Belgian illustrated magazines from the interbellum. In *Monte Artium* [online]. 2023, 15. Available on: <<http://journals.openedition.org/ima/427>>.
 - LIEBRECHT, Henri, Cinq siècles d'art. Maîtres de jadis. *Le Soir illustré* 8:392 (24-08-1935), p. 16-18. Online: <<https://uurl.kbr.be/1954193/p16>>.
 - LIESEN, Bruno. Sur les traces des imprimeurs bruxellois dans l'entre-deux-guerres: l'imprimerie J. Felix et fils. *Histoire et civilisation du livre*. 2018, vol. 14, p. 81-93.
 - LINEPHTY, Maurice Guillaume. A propos de la presse périodique belge. *Le musée du livre*, 1934, 29:7. Online: <<https://uurl.kbr.be/2186283>>.
 - LUYKX, Theo. *Evolutie van de communicatiewetenschap*. Brussels: Elsevier, 1978.
 - René Lyr [VANDERHAEGHE, René]. La noble aventure de James Ensor. *A-Z* 2:25 (10-09-1933), p. 3; 6-7. Online: <<https://uurl.kbr.be/1901465/p3>>.
 - MERTENS, Gaston (ed.). *Union de la presse périodique belge. Bulletin officiel*. Brussels: Union de la presse périodique belge, 1930, vol. 40, n° 3, p. 62-63.
 - MULS, Jozef, De bethlehemsche kindermoord. *Zondagsvriend* 14:52 (25-12-1946), p. 27. Online: <<https://uurl.kbr.be/2136714/p27>>.
 - NAHON, Valérie. Profil d'une critique moderne. Charles Bernard et la défense de l'art vivant dans l'entre-deux-guerres. *Textyles*. 2010, Vol. 39, p. 63-84. Online: <<https://journals.openedition.org/textyles/101>>.
 - OULET, Paul. *La presse périodique. Annuaire de la presse périodique belge 1923-1924*. Brussels: Maison du livre, 1924.

- OVINK, Gerrit Willem. *Het aanzien van een eeuw: de periode 1856-1956 weerspiegeld in 950 illustraties uit de voornaamste Nederlandse familiebladen*. Haarlem: De Spaarnestad, 1959.
- PAUMEN, Vanessa. *The Flemish primitives: What's in a name?* Online: <<https://vlaamseprimitieven.vlaamsekunstcollectie.be/en/research/webpublications/the-flemish-primitives-whats-in-a-name/>>.
- SAROLEA, R. [Leo Arras], Vlaamsche koppen. Evarist De Buck. *De Stad* 2:36 (22-11-1929), p. 779. Online: <<https://uurl.kbr.be/1921485/p15>>.
- SEYL, Antoine. *La rotogravure. Hélogravure en creux rotative*. Brussels: Imprimerie scientifique et littéraire, 1928.
- STRUYE, Johan et al. (eds.). *150 ans à la une: un siècle et demi d'information illustrée en Belgique. Exposition organisée par le Crédit Communal de Belgique du 19-9 au 2-11 1980*. Brussels: Crédit Communal de Belgique, 1980.
- STULIK, Dusan & KAPLAN, Art. *Half-tone*. Los Angeles: J. Paul Getty Trust, 2013.
- STULIK, Dusan & KAPLAN, Art. *Photogravure*. Los Angeles: J. Paul Getty Trust, 2013.
- TIMMERMANS, Felix. Vlaamsch volksleven gezien door Peter Brueghel. *Ons Land*, p. 1358-1359. Online: <<https://uurl.kbr.be/1937085/p16>>.
- VAN DELFT, Marieke et al. (eds.). *Magazine! 150 jaar Nederlandse publiektijdschriften*. The Hague: Koninklijke Bibliotheek, 2006.
- VAN HOVE, Julien. De algemene verzamelingen van gedrukte werken. In: *Koninklijke Bibliotheek. Liber Memorialis 1559-1969*. Brussels: Koninklijke Bibliotheek van België, 1969, p. 138-156
- VAN TRIEST, Jaap. Vormgeving. In: VAN DELFT, Marieke et al. (eds.). *Magazine! 150 jaar Nederlandse publiektijdschriften*. The Hague: Koninklijke Bibliotheek, 2006, p. 50-51.
- Z.L., Vlaamsche koppen. De prins der Vlaamsche schilders James Ensor. *De Stad* 3:35 (14-11-1930), p. 831. Online: <<https://uurl.kbr.be/1921587/p15>>.
- Z.L. [Lode Zielens], De Vlaamsche schilder Albert Saverijs. *De Stad* 4:33 (30-10-1931), p. 795. Online: <<https://uurl.kbr.be/1921688/p19>>.
- ZERVIGÓN, Andrés Mario. Rotogravure and the Modern Aesthetic of News Reporting. In: HILL, Jason & SCHWARTZ, Vanessa (eds). *Getting the Picture. The Visual Culture of the News*. London: Routledge, 2015, p. 197-205.

12. Studying Comics in the Belgian Press through Digitised Corpora: Challenges and Opportunities

- Artpresse [online]. Available on: <https://www.kbr.be/en/projects/artpresse/>.
- BeComicsLab [online]. Available on: www.kbr.be/en/projects/be-comics-lab.
- Digital Research Lab [online]. Available on: <https://www.kbr.be/en/projects/digital-research-lab/>.
- Gallica BNF [online]. Available on: <https://gallica.bnf.fr/>.
- Pop Heritage Lab [online]. Available on: www.kbr.be/en/projets/pop-heritage-lab.
- QuaDiHum Centre [Online]. Available on: <https://quadihum.ulb.be/>.
- BAUDRY, Julien & LABARRE, Nicolas. De l'imprimé au numérique et retour: lecture et perception d'une bande dessinée d'un support à l'autre. *Alternative Francophone* [online]. 2020, vol. 2, n° 7, 32-54. [consulted on 20/01/2025]. DOI: <<https://doi.org/10.29173/af29402>>.
- CAPART, Philippe & DEJASSE, Erwin. In Search of the Lost Serial. In: BEATY, Bart & MILLER, Ann (eds.). *The French Comics Theory Reader*. Leuven (Belgium): Leuven University Press, 2014.
- CRUCIFIX, Benoît. Sortie des encombrants. Fabrique du patrimoine de la bande dessinée de journal aux États-Unis. *Sociétés & Représentations*. 2022, vol. 53, 167-186.
- DEJASSE, Erwin. Le regard cosmopolite et rétrospectif de la bande dessinée alternative. In: DONY, Christophe, HABRAND Tanguy & MEEESTERS, Gert (eds.). *La bande dessinée en dissidence – Alternative, indépendance, auto-édition / Comics in Dissent – Alternative, Independence, Self-Publishing*. Liège (Belgium): Presses universitaires de Liège, 2014.
- GLASSER, Jean-Claude. Courrier de Jean-Claude Glasser. *Les Cahiers de la bande dessinée*. 1988, n° 80, 8.
- GROENSTEEN, Thierry. *La Bande dessinée en France à la belle Époque: 1880-1914*. Brussels (Belgium): Les Impressions Nouvelles, 2022.
- HERTIMAN, Marys Renné & DE SINGLY, Camille (eds.). *Construire un Matrimoine de la BD – Créations, mobilisations et transmissions des femmes dans le neuvième art, en Europe et en Amérique*. Dijon (France): Les Presses du Réel, 2024.
- IYYER, Mohit, MANJUNATHA, Varun, GUHA, Anupam, VYAS, Yogarshi, BOYD-GRABER, Jordan, DAUMÉ, Hal III, & DAVIS, Larry S. The Amazing Mysteries of the Gutter: Drawing Inferences between Panels in Comic Book Narratives. *ArXiv* [online]. 2017. [consulted on 20/01/2025]. Available on: <<https://arxiv.org/abs/1611.05118>>.
- KASHTAN, Aaron. *Between Pen and Pixel: Comics, Materiality, and the Book of the Future* [online]. Columbus (United States): Ohio State University Press. 2018. [consulted on

- 20/01/2025]. Available on: <<https://doi.org/10.2307/j.ctv1khdqnk>>.
- LESAGE, Sylvain. Le canon et l'écran. Le patrimoine de la bande dessinée et les archives numérisées. *Comicalités* [online]. 2022. [consulted on 20/01/2025]. Available on: <<https://doi.org/10.4000/comicalites.7723>>.
 - LESAGE, Sylvain & SUVILAY, Bounthavy. Introduction thématique: pour un tournant matériel des études sur la bande dessinée, *Comicalités* [online]. 2019. [consulted on 20/01/2025]. Available on: <<https://doi.org/10.4000/comicalites.3692>>.
 - MITCHELL, W.J.T. & SPIEGELMAN, Art (2014). Public Conversation: What the %\$#! Happened to Comics? May 19, 2012. *Critical Inquiry* [online]. 2014, vol. 40, n° 3, p. 20–35. [consulted on 20/01/2025]. Available on <<https://doi.org/10.1086/677327>>.
 - PAQUES, Frédéric. *Avant Hergé. Étude des premières apparitions de bande dessinée en Belgique francophone (1830–1914)*. Liège (Belgium): Presses universitaires de Liège, 2025.
 - PIFFAULT, Olivier. Destinées et aléas! La bande dessinée dans les bibliothèques françaises. *Bibliothèque(s)*. 2010, n° 51, p. 11–15.
 - DE QUERELLES, Richard. *Le Déluge à Bruxelles*. Brussels (Belgium): Jules Gérardet, 1843.
 - SAUSVERD, Antoine. *Dans l'Ombre du Professeur Nimbus. Aventures et méasaventures d'un héros de bandes dessinées et de son créateur André Daix*. Montrouge (France): PLG, 2023.
 - TÖPFFER, Rodolphe. *Histoire de Mr. Jabot*. Geneva (Switzerland): Imprimerie Caillet, 1833.
 - VERCRUYSE, Bas. Discovering Hidden Comics Using AI. *Ghent Centre for Digital Humanities* [online]. 2024 [consulted on 20/01/2025]. Available on: <www.ghentcdh.ugent.be/discovering-hidden-comics-using-ai>.

13. CAMille. A Research Centre on the History of Belgian Journalism

- CAMille platform source code [Online]. Available on: <https://github.com/madewild/CAMille>.
- Impreso. Media Monitoring of the Past [Online]. Available on: <https://www.impresso-project.ch/>. <https://www.newseye.eu/>.
- NewsEye. A Digital Investigator for Historical Newspapers [Online]. Available on: <https://www.newseye.eu/>.
- Numapresse. Du papier à l'écran [Online]. Available on: <http://www.numapresse.org/>.
- SPARQuLb [Online]. Available on: <https://quadihum.ulb.be/sparqulb/>.
- BECKER, Howard S. *Les mondes de l'art*. Paris: Flammarion, 2010.

- BIRKNER, Thomas, KOENEN, Erik & SCHWARZENEGGER, Christian. A Century of Journalism History as Challenge: Digital archives, sources, and methods. *Digital Journalism*. 2020, vol. 6, n° 9, p. DOI: <https://doi.org/10.1080/21670811.2018.1514271>.
- DARMON, Muriel. La notion de carrière: un instrument interactionniste d'objectivation. *Politix*. 2008, vol. 82, n° 8, p. 149-167.
- DESEURE, Brecht, GRIBOMONT, Isabelle & LE CAM, Florence. Journalistic Gender Trouble. Women Journalists as Seen by and in the Belgian Press (1850-1950), *Journal of Belgian History*. In press.
- DESEURE, Brecht, DE VALERIOLA, Sébastien, LE CAM, Florence, LIBERT, Manon & VIDALENCHE, Alexia. *Hériter d'un dictionnaire biographique: enjeux méthodologiques autour d'une base de données belges*. Conference paper at *Congrès international Médias 19-Numapresse, Presses anciennes et modernes à l'ère du numérique* (Paris, May 2022).
- DOZO, Björn-Olav. Portrait statistique de l'écrivain journaliste en Belgique francophone entre 1918 et 1960. *Textyles. Revue des lettres belges de langue française*. 2010, vol. 39, p. 123-143.
- GEMIS, Vanessa. Femmes écrivains-journalistes (1880-1940): questions de genre (s). Pistes de recherche et réflexions autour de Marguerite Van de Wiele. *Textyles. Revue des lettres belges de langue française*. 2010, n° 39, p. 39-50. <https://journals.openedition.org/textyles/107>.
- GEMIS, Vanessa. Occupation: Flying Parcel. Portrait of Marc Augis, Woman, Journalist, and Writer. *Literary Journalism Studies*. 2016, vol. 8, n° 2, p. 38-51. DOI: <https://doi.org/10.4000/textyles.107>.
- GRIBOMONT, Isabelle, DESEURE, Brecht, & LE CAM, Florence. The Perception of Women Journalists and Feminism in Belgian Newspapers (1880-1950). *Digest. Journal of Diversity and Gender Studies* 12 (2025/2), p. 45-67, <https://www.digest.ugent.be/article/id/91568/>.
- JACQUET, Antoine. *Le français des journalistes en ligne. Régulation de la langue par les représentations et les pratiques*. Unpublished PhD thesis, under the supervision of Florence Le Cam and Laurence Rosier. Université libre de Bruxelles, 2018.
- JACQUET, Antoine, LE CAM, Florence, LIBERT, Manon, VIDALENCHE, Alexia, & DE VALERIOLA, Sébastien. *Prosopography and digital humanities through the Belgian daily Le XXe siècle (1895-1940)*. Conference paper in *Atelier-séminaire Numapresse "Individus, auteurs, signature"* (Paris, June 2021).

- KACIAF, Nicolas & LE DREF, Paul. Du devenir des enquêtes de Mediapart et du Canard enchaîné: contribution à l'analyse de la «carrière médiatique» des révélations journalistiques. *Les cahiers du journalisme - Recherches*. 2018, vol 2, n° 1.
- LE CAM, Florence, LIBERT, Manon & MÉNALQUE, Lise. *Être femme et journaliste. Enquête sociologique dans un monde au masculin*. Brussels: Éditions de l'Université de Bruxelles, 2021.
- LE CAM, Florence & TANT, Cédric. Premiers pas de la carte de presse en Belgique francophone (1885-1966). *Le Temps des médias*. 2018, vol. 30, n° 1, p. 200-226. DOI: <https://doi.org/10.3917/tdm.030.0200>.
- LIBERT, Manon, JACQUET, Antonie, DE VALERIOla, Sebastien, LE CAM, Florence, VIDALENCHE, Alexia & DESEURE, Brecht. Repeupler les rédactions du passé à l'aide des humanités numériques. Le cas du quotidien belge Le XXe Siècle. *Questions de communication* 48 (2025), <https://doi.org/10.4000/15b19>.
- TASSET, Cyprien. Entre sciences sociales, journalisme et manifestes. La représentation de groupes sociaux réputés émergents dans la France des années 2000. *Les enjeux de l'information et de la communication*. 2011, vol. 12, n° 3, p. 139-157.
- VANDERPELEN-DIAGRE, Cécile. Magasins d'amour. Les romans et chroniques de France Adine, journaliste de Femmes d'aujourd'hui. *Belpégor. Littérature populaire et culture médiatique*. 2021, vol. 19, n° 2. DOI: <https://doi.org/10.4000/belpégor.4219>.
- VIDALENCHE, Alexia, LE CAM, Florence, DE VALERIOla, Sébastien, LIBERT, Manon & DESEURE, Brecht. Retracer (massivement) et analyser une carrière médiatique. Désinvisibiliser Alice Bron, journaliste, socialiste et féministe belge (1850-1904). *Le Temps des médias*. 2024, vol. 42, n° 1, p. 169-189. Available on: <https://doi.org/10.3917/tdm.042.0169>.

14. News Nuggets: Mining Imperfect Historical Press Data

- DOERMANN, David & TOMBRE, Karl. *Handbook of Document Image Processing and Recognition*. London: Springer, 2014. Available on: <https://doi.org/10.1007/978-0-85729-859-1>.
- FRANZINI, Greta, KESTEMONT, Mike, ROTARI, Gabriela, JANDER, Melina, OCHAB, Jeremi K., FRANZINI, Emily, BYSZUK & Joanna, RYBICKI, Jan. Attributing Authorship in the Noisy Digitized Correspondence of Jacob and Wilhelm

- Grimm. *Frontiers in Digital Humanities*. 2018, vol. 5, n° 4. Available on: <https://doi.org/10.3389/fdigh.2018.00004>.
- HAMDI, Ahmed, JEAN-CAURANT, Axel, SIDERE, Nicolas, COUSTATY, Mickaël & DOUCET, Antoine. An Analysis of the Performance of Named Entity Recognition over OCREd Documents. In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. New York: Institute of Electrical and Electronics Engineers (IEEE). 2019, p. 333–334. Available on: <https://doi.org/10.1109/JCDL.2019.00057>.
 - HILL, Mark J. & HENGCHEN, Simon. Quantifying the Impact of Dirty OCR on Historical Text Analysis: Eighteenth Century Collections Online as a Case Study. *Digital Scholarship in the Humanities*. 2019, vol. 34, n° 4, p. 825-843. Available on: <https://doi.org/10.1093/llc/fqz024>.
 - HOLLEY, Rose. How Good Can It Get? Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs. *D Lib Magazine*. 2009, vol. 15, n° 3-4. Available on: <https://www.dlib.org/dlib/march09/holley/03holley.html>.
 - IGNAT, Oana, MAILLARD, Jean, CHAUDHARY, Vishrav & GUZMÁN, Francisco. OCR Improves Machine Translation for Low-Resource Languages. In: *Findings of the Association for Computational Linguistics: ACL 2022*. New York: Association for Computational Linguistics. 2022, p. 1164-1174. Available on: <https://doi.org/10.18653/v1/2022.findings-acl.92>.
 - LIBERT, Manon, JACQUET, Antoine, DE VALERIOLO, Sébastien, LE CAM, Florence, VIDALENCHE, Alexia & DESEURE, Brecht. Repeupler les rédactions du passé à l'aide des humanités numériques. Le cas du quotidien belge Le XXe Siècle. *Questions de communication*. 48 (2025), <https://doi.org/10.4000/15b19>.
 - MUTUVI, Stephen, DOUCET, Antoine, ODEO, Moses & JATOWT, Adam. Evaluating the Impact of OCR Errors on Topic Modelling. In: *International Conference on Asian Digital Libraries*. Cham: Springer. 2018, p. 3-14. Available on: https://doi.org/10.1007/978-3-030-04257-8_1.
 - NEUDECKER, Clemens, BAIERER, Konstantin, GERBER, Mike, CLAUSNER, Christian, ANTONACOPOULOS, Apostolos & PLETSCHACHER, Stefan. A Survey of OCR Evaluation Tools and Metrics. In: *HIP '21: Proceedings of the 6th International Workshop on Historical Document Imaging and Processing*. New York: Association for Computing Machinery. 2021, p. 13–18. Available on: <https://doi.org/10.1145/3476887.3476888>.

- VIDALENCHE, Alexia, LE CAM, Florence, DE VALERIOLA, Sébastien, LIBERT, Manon & DESEURE, Brecht. Retracer (massivement) et analyser une carrière médiatique. Désinvisibiliser Alice Bron, journaliste, socialiste et féministe belge (1850-1904). *Le Temps des médias*. 2024, vol. 42, n° 1, p. 169-189. Available on: <https://doi.org/10.3917/tdm.042.0169>.
- WICK, Christoph, REUL, Christian & PUPPE, Frank. Calamari – A High-Performance Tensorflow-based Deep Learning Package for Optical Character Recognition. *Digital Humanities Quarterly*. 2020, vol. 14, n° 2. Available on: <https://www.digitalhumanities.org/dhq/vol/14/2/000451/000451.html>.

Σαράη

300



Σαράη

302



