
Belgian Social Media Archiving Initiatives Mapped

Cartographie des initiatives belges d'archivage des médias sociaux

Belgische initiatieven voor de archivering van sociale media in kaart gebracht

Fien Messens



Electronic version

URL: <https://journals.openedition.org/ima/322>

ISSN: 2507-0312

Publisher

Bibliothèque royale de Belgique | Koninklijke Bibliotheek van België (KBR)

Electronic reference

Fien Messens, "Belgian Social Media Archiving Initiatives Mapped", *In Monte Artium* [Online], 15 | 2023, Online since 01 December 2023, connection on 11 December 2023. URL: <http://journals.openedition.org/ima/322>

This text was automatically generated on December 11, 2023.



The text only may be used under licence CC BY-NC-ND 4.0. All other elements (illustrations, imported files) are "All rights reserved", unless otherwise stated.

Belgian Social Media Archiving Initiatives Mapped

Cartographie des initiatives belges d'archivage des médias sociaux

Belgische initiatieven voor de archivering van sociale media in kaart gebracht

Fien Messens

EDITOR'S NOTE

This article was written in the summer of 2022, at the end of the BESOCIAL project. Fien Messens was at that time researcher on the BESOCIAL project, affiliated to KBR (department Innovation and Research) and UGhent.

I would like to express my very great appreciation to Friedel Geeraert and the other members of the BESOCIAL research team for their valuable and constructive suggestions during the planning and development of this article. I would also like to thank the representatives of all the organisations who have provided information for this study for enabling me to gain insight into their social media archiving practices.

Introduction

- 1 Social media archiving is neither new nor unknown. Over the past decade, the output of social media, being big data, has become one of the major sources for quantitative (and qualitative) research.¹ It has been crucial to reveal fascinating insights into a variety of questions of human behaviour.² In addition to collecting social media data solely for research purposes, cultural institutions such as national libraries are also finding their way to archiving and preserving this rapidly changing data type, by creating a framework for web and/or social media archiving. Research of KBR (under the BESOCIAL project) showed that a number of (national) institutions in various countries are starting up pilot projects for social media archiving (SMA).³ While efforts are being

made, and more and more institutions are getting engaged and interested in this swift media type, the stage and efforts vary in scope and size.

- 2 Despite the efforts and the ongoing interest in capturing these born-digital collections, an overview of the Belgian SMA landscape remains to be thoroughly investigated. In the framework of the BESOCIAL research project a descriptive approach was chosen to map the existing Belgian SMA landscape. The main contribution of this paper is to provide insights into how Belgian cultural organisations are archiving social media collections.
- 3 The data for this analysis was collected in the context of the BESOCIAL research project, which was initiated in July 2020 and ran until September 2022. The project was funded by the Belgian Science Policy Office (BELSPO) as part of its BRAIN.be programme. It aimed to develop a sustainable strategy for the preservation and archiving of social media in Belgium. The Royal Library of Belgium (KBR) was the coordinator of this project that was managed in close collaboration with CRIDS (Research Centre in Information, Law and Society) at the University of Namur, CENTAL (Centre de traitement automatique du langage), at the UCLouvain, IDLab (Internet Technology & Data Science Lab), GhentCDH (Ghent Centre for Digital Humanities), and MICT (Research Group for Media, Innovation and Communication Technologies), at Ghent University.
- 4 The article starts by providing the methodological context, after which a brief overview of the importance of having a legal framework for SMA is outlined. Then, the different steps in the SMA process are discussed: selection, harvesting, preservation and access. The paper ends with a conclusion and a reflection on future prospects.

Methodology

- 5 To investigate how the Belgian landscape of SMA is represented, we took a **descriptive research approach**. This entails a number of steps: finding initiatives that harvest social media data via secondary research, collecting data of these initiatives, and analysing and interpreting the data, using a qualitative thematic analysis approach.
- 6 Prior to creating an overview of SMA in Belgium in the context of the BESOCIAL research project, SMA on the international level was studied.⁴ We first observed that policies and access conditions with regard to social media widely differ between institutions. However, some common trends were identified: i) Twitter is the social media platform most often archived; ii) collections of accounts and/or hashtags focus on important people, organisations, and events; iii) priority is given to selective harvests; iv) open source tools are mostly used; v) the WARC data format is most often used, and finally, vi) a clear lack of a common understanding of preservation concepts (e.g. ‘preservation formats’ or ‘preservation standards’) was observed.⁵ In this article, we examine the same themes with a magnifying glass view and focus on the Belgian initiatives.
- 7 A number of different approaches were used to compile a **list of SMA initiatives**:
 - In the context of the PROMISE research project a study of web archiving (collections of websites) initiatives in Belgium was undertaken. The network of web archiving institutions was approached again in the context of the BESOCIAL project. The organisations that also archive social media content were asked to reply to a survey.⁶

- Asking identified institutions if they knew about other SMA initiatives in Belgium proved to be another fruitful approach as a number of additional initiatives were uncovered this way.
 - Desk research was also used to identify more organisations.
- 8 Different **data collection methods** were used in this study. Through desk research additional information about social media archiving at the identified institutions was found. Furthermore, the institutions identified during the PROMISE project were asked to fill out a survey. In addition, semi-structured interviews were organised with representatives of new SMA initiatives that were discovered or, alternatively, the questions were answered over email exchanges. The combination of the different data collection methods allowed us to answer: **how are institutions and initiatives in Belgium engaging in social media archiving?** This research question addresses a) the selection, b) the social media harvesting, c) access to, and (re)use of the social media archive, and d) the preservation policy.
- 9 The survey which ran from July 2020 to September 2020 was sent to representatives from the institutions studied in the PROMISE project.⁷ These institutions were asked to participate in the BESOCIAL survey to update the information related to archiving websites and provide information about social media archiving at their institution.⁸ In addition, other Belgian organisations were contacted to report on their approach to social media archiving. The following initiatives answered the survey: Archives of the University of Antwerp, Ghent University Library, Liberas, Amsab - Instituut voor Sociale Geschiedenis, ADVN | archief voor nationale bewegingen, Letterenhuis and, KADOC-KU Leuven - Documentation and research centre on religion, culture and society. Each of the participants were sent a personalised spreadsheet with questions and were asked to provide written feedback. Based on the desk research, some questions were already answered beforehand and the respondents were asked to verify this information.
- 10 After the survey was closed, several new social media archiving initiatives were discovered. Additional information was gathered based on semi-structured interviews (Archives of the City of Mons) and email exchanges with representatives of these respective institutions: Archives of the City of Brussels, CAVA, AMVB | archief en museum voor het Vlaamse leven in Brussel, Archives et Musée de la Littérature, LIBIS, and Institute for Media Studies (KU Leuven). Information about the BESOCIAL project was also added to the table in order to make the overview of Belgian social media initiatives comprehensive.

Table 1

Institution	Social media archiving	Social media platforms	Linked to project
BESOCIAL ⁹	Yes	Twitter and Instagram	
Liberas	Yes	Facebook, Twitter, YouTube and WhatsApp	
Amsab	Yes	Twitter, Instagram, Facebook, YouTube and a small amount on	

		Flickr, Pinterest, Vimeo and Snapchat	
ADVN	Yes	YouTube, Twitter, Facebook, Flickr, Instagram, TikTok, Tumblr, Vimeo, Bitchute, Pinterest and Slideshare	
Letterenhuis	Yes	Instagram and Facebook	Best practices voor de archivering van sociale media in Vlaanderen en Brussel ¹⁰
KADOC	Yes	Twitter, Facebook, Instagram (and YouTube)	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
University Library Ghent	Yes	Twitter, Facebook, Wordpress and Flickr	
Archives of the University of Antwerp	Yes	Facebook, Twitter, Instagram, LinkedIn and YouTube	
LIBIS (project ICANDID)	Yes	Twitter	
Instituut voor Media Studies at KU Leuven	Yes	Twitter and Facebook	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
AMVB	Yes	Facebook, Instagram and Twitter	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
CAVA	Yes	Facebook and Instagram	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
City of Mons	Yes	Facebook	Archives de Quarantaine Archief ¹¹
Archives et Musée de la Littérature	Yes	Facebook and YouTube	Archives de Quarantaine Archief
La ville de Bruxelles	No	Planned: Facebook, Twitter and Instagram	

Overview of the final list of institutions concerning the archiving of social media¹²

- 11 This collected corpus of data underwent a qualitative thematic analysis. This technique uses labels to identify, categorise and summarise important concepts within a set of data. These concepts or themes (e.g. selection and harvesting) will be used to highlight certain patterns in the text. This is driven by the research question, so it is not necessary to identify every possible theme in the data, but rather focus on the key aspects that relate to our research question.
- 12 We have approached this by using a deductive point of view. This means that we went over the data and analysed them with a set of themes that we already expected to find in our corpus since we already had prior knowledge of the different steps in the process of social media archiving. In our case, focus went to finding topics linked to the legal aspects of social media archiving, selection, harvesting, preservation and access to and (re)use of the social media corpus.
- 13 The findings of this study have to be seen in light of some limitations. The list of organisations studied was compiled by convenience sampling, where data is collected from a conveniently available pool of respondents. This can be seen as a limitation as the sampling leads to a non-exhaustive list of Belgian initiatives. Secondly, communication plays a role. There is a chance that some initiatives that do harvest social media were not included in this study. On the one hand, this may be because they were not currently promoting their SMA activities, and we simply missed them. On the other hand, because we identified and contacted certain institutions but did not receive a response (e.g. Felixarchief and M HKA). A third limitation was the timing. There was a time gap between the starting point of the surveys and the analysis. Social media, harvesting tools, interfaces, etc. are subject to rapid changes. It is therefore strongly recommended to initiate a follow-up study on social media archiving initiatives in Belgium.¹³

Importance of having a legal framework

- 14 In the context of archiving social media, the law has the potential to either strike a balance between safeguarding citizens' rights and allowing for the growth of the digital society. On one hand, laws such as GDPR and the right to privacy are crucial in light of the rapidly changing digital landscape and must be upheld to ensure the protection of citizens' fundamental rights. On the other hand, overly restrictive regulations can hinder the evolution of the digital society by limiting the ability of individuals and organisations to share and access information. As nice as digital evolution sounds, it also jeopardises the rights that need to be protected, which complicates the process of social media archiving. Therefore, it is essential to strike a balance that protects citizens' rights while also promoting innovation and progress in the digital world.¹⁴
- 15 Many of the choices made in establishing social media archiving practices are a result of the legal framework within which the respective institutions operate.¹⁵ For the initiatives studied, only a few emphasised the importance to comply to the legal boundaries, emphasising the advantages and especially the disadvantages.
- 16 If we look at the BESOCIAL research project, the role of KBR as a national library comes into the picture. The Royal Library of Belgium (KBR) was established by Royal Decree in 1837.¹⁶ The decree has been revised multiple times since then. Social media archiving is not specifically mentioned, but it does fit with the missions of "acquiring and keeping

for legal deposit, whatever the medium used, all publications published in Belgium and those of authors domiciled in Belgium published abroad, as well as publications on Belgium published abroad” and “managing, preserving and enhancing cultural heritage”. Legal deposit legislation exists in Belgium, but it has not been extended to web and social media content yet. KBR is currently working on this extension of the law.

- 17 During the BESOCIAL project, legal recommendations were made about topics such as selection, preservation, and access. First, the right to privacy must not be overlooked for the benefit of the right to information of the public. Social media content is highly prone to contain private information and personal data which should be treated with the necessary caution. Regarding the respect of privacy as such, a balance must be found on a case by case basis between a person’s right to privacy (art. 8 of the European Convention on Human Rights) and the right of information of the general public (art. 10 of the European Convention on Human Rights) in order to determine whether private information should be archived or not. In addition to balancing these two rights, the BESOCIAL project team also kept the GDPR (General Data Protection Regulation) in mind. The right to personal data protection stems directly from the more general right to privacy, but is particularly crucial when it comes to the digital world. The GDPR provides a framework to protect living people’s rights when their personal data is processed, which is often the case when archiving social media content.¹⁷
- 18 Most other Belgian institutions that archive or plan to archive social media don’t have an extensive research team with specialists focussing on legal issues such as giving access to born-digital collections. The Best Practices project (led by KADOC-KU Leuven and meemoo) is, next to BESOCIAL, currently the only research project to conduct extensive legal analyses regarding social media archiving in the cultural heritage context. For the qualification of archiving tools, they analysed among others the user terms and platform conditions. In addition, they already stressed the importance of complying with (sometimes complex) copyright, image rights and data protection rules.¹⁸
- 19 When it comes to legal considerations that constitute obstacles to the evolution of the digital society, we can say that this evolution happened at such a fast pace that it has been difficult for the law to keep up. The limitations regarding available legal expertise and complexity of having different rights involved, are a contributing factor to smaller institutions deciding to only make certain data publicly available (or after request, intra muros).

Selecting the corpus

- 20 Manually curating collections of archived social media data is expensive and time-consuming, but they are a critical source of information for researchers studying historical and sociological events.¹⁹ With human-generated selection criteria, the coverage and outlining of these archives are undoubtedly influenced by the curators of the collections. Archivists define strategies to populate web and social media archives according to the scope of their missions and the resources available.²⁰
- 21 Selection is therefore a critical first step in shaping a social media corpus. This can incorporate the following criteria: capturing the look and feel of social media platforms

or focusing on raw data only, the choice of social media platforms from which to archive content, topics on which the collection should focus, the choice of language(s) to include, the time frame, the geographic region to focus on (e.g. only Flanders), the type of content to select (accounts and/or hashtags).

Table 2

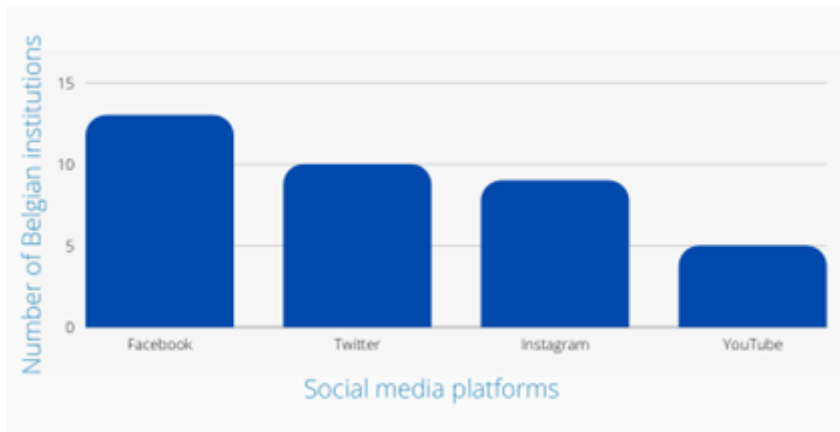
Initiative	Social Media Platform(s)	Topic(s)	Type of content
BESOCIAL	Twitter and Instagram	Cultural heritage in Flanders, Brussels and Wallonia	Profiles, hashtags and comments
Liberas	Facebook, Twitter, YouTube and WhatsApp	Bart Tommelein and the Willemsfonds, COVID-19	Profiles (no comments)
Amsab	Twitter, Instagram, Facebook, YouTube and a small amount on Flickr, Pinterest, Vimeo and Snapchat	Kunstencentrum Vooruit/404, (COVID-19 and BlackLivesMatter)	Profiles, hashtags, pages, groups, (and comments) ²¹
ADVN	YouTube, Twitter, Facebook, Flickr, Instagram, TikTok, Tumblr, Vimeo, Bitchute, Pinterest and Slideshare	Ministers and/or their chiefs of staff both on the federal and Flemish level	Profiles, hashtags and also results from specific search queries linked to their collection development plan. Comments and other types of interaction data of internet users are also collected if possible.
Letterenhuis	Instagram and Facebook	Gaea Schoeters and Tom Naegels	Profiles
KADOC	Twitter, Facebook, Instagram (and YouTube)	Christian democratic political party CD&V and a number of individual politicians.	Profiles
University Library Ghent	Twitter, Facebook, Wordpress and Flickr	Content related to UGent and its faculties	Profiles
Archives of the University of Antwerp	Facebook, Twitter, Instagram, LinkedIn and YouTube	The university, its faculties or related organisations like student clubs but also collect content linked to specific events such as COVID-19. ²²	Profiles, hashtags and comments
LIBIS (project ICANDID)	Twitter	Content that is useful for researchers	Profiles and hashtags

Instituut voor Media Studies at KU Leuven	Twitter and Facebook	Gambling-related web content	Profiles
AMVB	Facebook, Instagram and Twitter	Kaaitheater, and AMVB's creators ²³	Profiles
CAVA	Facebook and Instagram	Creators of archives (e.g. deMens.nu and Humanistisch Verbond)	Profiles
City of Mons	Facebook	COVID-19 in Mons ²⁴	Profiles and comments
Archives et Musée de la Littérature ²⁵	Facebook and YouTube	COVID-19, linked to themes of the French-speaking Belgian literary and theatrical heritage that constitutes the specificity of Archives et Musée de la Littérature	Profiles
La ville de Bruxelles	Planned: Facebook, Twitter and Instagram	La Ville de Bruxelles and linked organisations (such as the ASBL Bruxelles-Musées-Expositions)	NA

Overview of the most important characteristics of corpus selection (choice of platform, topic(s) and type(s) of content)

- 22 When analysing the data from Table 2 we can see some interesting trends linked to the choice of **social media platforms**. The majority of the initiatives archive data from Facebook, while this is one of the most difficult platforms to archive (e.g. has no Application Programming Interface).²⁶ We notice that the preference for this is driven by the thematic aspect. Organisations find that the accounts they want to capture are more likely to be found on Facebook (see Figure 1). The city of Mons can serve as an example. They motivate their choice of Facebook because of the advantage that the platform is used by many different generations, with much interaction in the comment section. In comparison, the TikTok platform appeals to a much younger audience and, due to the many videos, results in larger amounts of data that must be stored and saved.²⁷

Figure 1



The top 4 of chosen social media platforms in Belgium. Facebook 13, Twitter 10, Instagram 9 and YouTube 5

- 23 We can say that the number of platforms selected is not proportional to the size and resources of particular organisations. It could be reasoned that smaller organisations, with fewer resources tend to archive 1 or 2 social media platforms. This is not the case: the data collected shows that organisations that opt for a strongly focused corpus (e.g. archiving around one institution such as Amsab with the Vooruit/viernulvier) have more room to experiment with different platforms and thus also archive platforms such as Snapchat. One can therefore conclude that the choice of the number of accounts and/or hashtags included in the collection, balances the number of social media platforms in the sense that if one wants to archive a lot of accounts and hashtags, one will opt for fewer platforms.
- 24 Furthermore, it can also be concluded that institutions that opt to archive a focused corpus with only a few accounts, often also have the permission of the account holder to archive their data. This gives them more options to, for example, reproducing archived content without violating copyright rules.
- 25 The chosen **topics** that the organisation wishes to archive, are mostly linked to the focus of the collection development of their institution. KADOC-KU Leuven, the Documentation and Research Centre for Religion, Culture and Society, for example, will focus on archiving profiles of Christian democratic political party CD&V members and a number of individual politicians. Another popular approach for selecting social media content is to focus on specific events. When creating a social media archive, it is very difficult or nearly impossible to anticipate or plan for certain major events (e.g. COVID-19, and attacks of 22nd of March 2016 in Brussels). Although, we see that some organisations want to archive the content of such kinds of themes (e.g. City of Mons).
- 26 At the level of types of **content**, we do not see any striking trends. The choice is usually made to archive only profiles, based on institutional thematic collections. Some Belgian institutions indicated that archiving hashtags is not always easy. Delineating a “Belgian” hashtag is quite difficult. Content-wise, it often results in overfishing. If an organisation does choose to capture hashtags, they focus on hashtags based on specific search terms (e.g. for the COVID-19 collection: #zorgvoorelkaar, #degoedekantop, #samentegencorona, and #blijfthuis). Not all institutions aim to capture comments and/or other user-generated content. The institutions that aim to capture this content

agree that it is very difficult to do so, due to software restrictions. We observed the same trend on a global level.²⁸

Harvesting the corpus

- 27 The next level in shaping a social media corpus depends on the choice of harvesting tools and their settings. The decisions that are made, from frequency of harvesting, to the final format of the data, will help determine the completeness as well as how the social media data will be presented to the end user.

Table 3

Initiative	Tools	Format	Frequency	Data size
BESOCIAL	Instaloader and Social Feed Manager	WARC and JSON	Weekly	14 GB
Liberas	Bino	csv, JSON	Normally once a year (during COVID more frequently)	538 GB in December 2020
Amsab	Twitter: Nodexl, Twarc ²⁹ and Brozzler YouTube: Youtube-dl Instagram: Instaloader Facebook ³⁰ : Webrecorder, Brozzler and Browsertrix	MP4 for Youtube, WARC for regular archiving of Twitter and a combination of JSON, JPG and MP4 for event-driven archiving, WARC or a combination of JSON, JPG, MP4 and TXT for Instagram and WARC for Facebook.	yearly	200 GB in November 2020
ADVN	combination of browser-based crawling and archiving and API	WARC, WACZ, CSV, JSON, MKV, MP4 and JPG formats	/	3TB February 2021 ³¹
Letterenhuis	/	/		NA
KADOC	/	/	/	/
University Library Ghent	/	/	/	300 GB
Archives of the University of Antwerp	Conifer (and Webrecorder is used sporadically if Conifer doesn't work) and the SingleFile Chrome Extension	WARC and html	Still have to decide on the capture frequency	10 GB

LIBIS (project ICANDID)	Twitter API	JSON	Daily	3,063,075 tweets
Instituut voor Media Studies at KU Leuven	Scrapestorm (free and paid account), Table capture, Web scraper, and Wayback machine	Excel sheets (or CSV), JSON, JPEG or PNG and mp4.	There is no recurring frequency of crawls. The capture is usually completed within two weeks.	/
AMVB	ArchiveWeb.page (browser based), Snsrape stand-alone (CLI), Snsrape in combination with Browsertrix (CLI).	WARC and JSON	depends on the character of the creator in question and the extent to which they are active on social media	807 MB
CAVA	ArchiveWeb.Page	WARC	Monthly	/
City of Mons	SingleFile	html (opened by zip)	/	/
Archives et Musée de la Littérature	/	mp4 or pdf	/	/
La ville de Bruxelles	/	/	/	/

Overview of the most important characteristics of harvesting social media (tools, format, frequency and data size)

- 28 Almost all of the archives mentioned above are already actively capturing websites (also called web archiving) in addition to social media archiving. On an international level we observed that this often has an impact on the choice of tool for social media archiving.³² It is possible to archive web and social media with the same tools (e.g. Heritrix). Archiving web content usually starts with crawling websites, which will generate files that preserve the original website in the Web ARChive file format (WARC). However, it is recommended to move more towards the choice of specialised tools due to the dynamic nature of social media.³³ The Belgian initiatives, in contrast to the international community, do not often choose such web archiving (website collections) tools for archiving social media data.
- 29 We can identify three types of tools: 1) simulated browsers such as Browsertrix, Webrecorder, and Conifer, 2) API-based tools such as Social Feed Manager and Twarc and 3) regular web harvesting tools such as Brozzler, Snsrape or Heritrix.³⁴ The choice in Belgium goes more towards simulated browsers, which capture the look and feel of a social media account. Although, we have to note that these tools also have their

disadvantages. Using simulated browsers as a tool can be very labour-intensive to use because among other things the 'autopilot' function does not yet work properly. Tools like Snsrape and Browsertrix are already better at scaling up, but require a lot of IT knowledge to install and to get them up and running. For small institutions that do not have their own IT department or in absence of dedicated in-house knowledge, it is difficult to get going with these tools, to test them, and to solve potential problems.

- 30 We see that some institutions with a more technical background (e.g. BESOCIAL and LIBIS) tend to use specialised tools such as Social Feed Manager (SFM) (type 2) which crawl data from the social media providers API's. They can then intervene more quickly in case of harvester hiccups. Although harvesting with an (official) API also has its limits. BESOCIAL expressed their concerns regarding fragile API harvesting in which you "can be kicked out of the API" if certain rate limits are reached. This was especially the case for harvesting data from Instagram with Instaloder. However, the computational mimicry of human scrolling limits this.
- 31 From a helicopter view on the tools chosen by Belgian cultural heritage institutions, we see that most of the institutions use open source tools, only a single institution uses a commercial archiving service. IMS uses both the free and paid account of the tool scrapestorm to archive Facebook. Despite having commercial tools that could support the smaller institutions without the necessary technical in-house knowledge, the high prices asked by the providers do not outweigh the advantages of free open-source options.
- 32 The choice of output format is mostly defined by the choice of tools. Type 1 tools, simulated browsers, harvest social media content via its HTML and CSS representation which allows for archiving the look and feel. This approach is chosen by the city of Mons (HTML format) and the Archives of the University of Antwerp (HTML and for embedded media, WARC format). This is in contrast to data harvested from an API by means of a tool such as Social Feed Manager, type 2 tools, which usually consists of raw data in a structured format (WARC and JSON format). Both harvesting methods have their merits, i.e. preserved look and feel allows browsing of archived content by users in its original context, whereas archived structured data facilitates linguistic analysis or general post-processing (e.g. basic graphs) of archived data.³⁵
- 33 Looking at the data size, we observed that it relates mainly to the start date of SMA and the number of seeds for the thematic corpus. Most initiatives started their social media archiving practices around 2020, with some outliers (e.g. University Library of Ghent started around 2007). A standout is ADVN | archive for national movements who started archiving ministers and/or their chiefs of staff both on the federal and Flemish level in 2019. So far, they are on top in terms of size with 3TB of data by February 2021. The social media archives of most other initiatives comprise a number of GBs.³⁶
- 34 The size of the corpus is also influenced by the frequency of harvests. Approaches differ between institutions. In some cases, collecting is done on a daily basis (e.g. iCANDID project from LIBIS), weekly (e.g. BESOCIAL), some yearly (e.g. Amsab) and in other institutions it is a mix. At Liberas for example, normally the social media content is captured once a year, but in 2020 there were exceptionally three harvests due to the COVID-19 outbreak.³⁷ We can surmise that institutions that use API harvesting tools (which are (semi) automated via code) harvest somewhat more frequently, than institutions that use simulated browser tools.

Preserving the corpus

- 35 Any heritage format can qualify for preservation: in this case, it is social media data. Despite the (research) opportunities that social media creates, the data remains fragile. Social media platforms are complicated and fluid, and are changing every day. “Their [digital data] longevity and utility is threatened where contents or context are lost: engagement and exploitation are enabled when digital materials endure. The greater the importance of digital materials, the greater the need for their preservation: digital preservation protects investment, captures potential, and transmits opportunities to future generations and our own”.³⁸
- 36 Several archives and libraries have already developed and implemented plans to avert the digital dark age. Yet the emerging formats remain a shared challenge. To guarantee the quality of digital data in the long term, we must also set our sights on access, which in turn means we need to understand and mitigate rapid changes in technology and organisations. The research done within the BESOCIAL project shows that there is lack of harmonisation, and that there is currently no real consensus (in Belgium and abroad) on the meaning of preservation concepts used, which could lead to misunderstandings that may be detrimental to the quality of the work and could also hinder cooperation between organisations.³⁹
- 37 At present, there is no commonly accepted standard data format for archived social media.⁴⁰ Furthermore, social media platforms have not published internal preservation plans.⁴¹ To date, most of the institutions decide on an ad hoc basis, on the format used for the preservation of social media content. At best, institutions (especially international ones) publish lists of recommended formats for digital preservation in their institution. However, it is important to note that they do not directly refer to social media.⁴²
- 38 Preservation of social media data actually means monitoring the archive and alerting archivists and librarians, if needed. The procedure has an umbrella function; it keeps track of technical innovations, preservation formats and standards. If an error occurs indicating that a file format is a risk for long term preservation, a migration plan is put into action to prevent losing data.

Table 4

Initiative	Preservation procedure	Tools
BESOCIAL	The data will be stored at KBR and linked to the Long Term Preservation Platform of BELSPO. The aim of KBR is to implement the OAIS model in their data storage structure in the near future. The first steps will be taken in the context of the Digital Data Strategy Working Group.	/
Liberas	The social media content is currently stored on a server in csv, png and mp4 format but they plan to transfer the csv to the collection management system and the multimedia to the e-deposit.	Atlantis that is linked to their e-depot Strongroom (developed by Data Matters)

Amsab	Bit-level preservation. No specific preservation norms, standards or software are used.	SHA256 checksums and DROID to check file integrity and format identification
ADVN	Looking into using specific preservation systems or software, but that currently still is a work in progress.	They use DROID and Brunnhilde for file identification, checksum generation and verification and JHOVE for file validation. No format migration is done and formal preservation procedures are currently being developed.
Letterenhuis	Letterenhuis is part of the SCALA project together with Amsab, ADVN, AMVB, AVG-Carhif, CAVA, VAI, Archiefbank Vlaanderen en meemoo to set up a shared test infrastructure for digital preservation system that is E-ARK compliant.	/
KADOC	OAIS principles and ISAD(G), but specific standards for social media are being investigated during the 'Best practices voor de archivering van sociale media in Vlaanderen en Brussel' project. The preservation workflow for social media will be the same as for other material: harvest, pre-processing and pre-ingest (virus scan, control checksums, metadata collection, MIME-type control, creation of derivatives), ingest in preservation system (creating and linking PID and adding access rights).	/
University Library Ghent	Preservation procedures are not yet in place for social media content, nor are standards and norms for preservation used.	/
Archives of the University of Antwerp	/	/
LIBIS (project ICANDID)	iCANDID has a preservation policy, but the harvested Twitter data is intended to be used by researchers for big-data analysis, not kept for long-term storage.	/
Instituut voor Media Studies at KU Leuven	Use is made of a personal KU Leuven OneDrive account. There is no preservation policy as yet.	/
AMVB	In the long run, the harvested social media data will be included in the digital repository of AMVB. At the moment, it is not known whether this will require any	/

	specific software or tools. They are still actively working on the development of a digital repository (among others with projects such as AIDA/SCALA). This includes the development of a preservation policy.	
CAVA	CAVA does have a preservation plan for digital collections in which it focuses primarily on bit preservation. Currently this plan is being revised to include websites and social media in the long term.	/
City of Mons	/	/
Archives et Musée de la Littérature	For preservation the AML rents a professional digital repository from Dropbox (n+3 replication), and an in-house developed WebApp allows remote management of this repository by AML archivists. The digital-born data are made accessible in OPAC of AML and are available for consultation on request. The URL links are notified in the “source” zone of the records. Digital documents are described in the same way as other documents, but their format is specified. Regarding the reuse of data: the format of the data is institution-specific, but APIs allow reuse by others. AML has full control over the databases and therefore all export formats are theoretically possible.	/
La ville de Bruxelles	/	/

Overview of the most important characteristics of long-term preservation of social media data (procedures and tools)

- 39 A common trend in the institutions surveyed is that there are occasionally plans for the long-term preservation of social media data, but very few initiatives have a working policy to prevent obsolescence of this swift media type. Some institutions are using bit-level preservation as a first step, such as Amsab. They use SHA256 checksums and DROID to check file integrity and format identification. Bit-level preservation is the building block for the more complete set of practices and processes to ensure the survival of digital content over time. This minimum level of preservation alone, does not ensure that content is fully usable in the many ways that users may want to interact with it and doesn't ensure readability on a long term basis. That being said, setting up and executing such a minimal level preservation plan will minimise the risks of losing important data due to media obsolescence and media failure. Bit-level preservation is also an area of practice that is maturing because of a growing community sharing expertise.⁴³
- 40 When following this first step of preservation the following two criteria needs to be taken into account:
- The more independent copies, the safer. As the size of the data increases, the per-copy cost increases, reducing the number of backup copies that can be afforded.

- The more frequently the copies are audited, the safer. As the size of the data increases, the time and cost needed for each audit - again - increases, reducing their audit frequency. These copies are distinct from backups that are part of the routine operation of data centres. It is expected that these copies will serve to replace damaged or lost files. The auditing is most often done using tools that create a unique hash of the file, or a checksum. If any part of the file changes, the next checksum will not match, therefore flagging an error. This approach is especially appropriate to use when moving content from one device to another or even to another organisation. The inventory is usually managed with a database or locally-developed tool or set of repository services.
- 41 Some organisations outsource some parts of the (bit-level) preservation process, such as Liberas; they use Atlantis, a web-based collection management system. Overall we can say that preservation is key (for institutions) when creating a social media archive. It is normal that not many institutions have developed operational preservation plans yet for this swift media type. There is little concrete documentation both internationally and nationally on how to specifically preserve social media in the long term, and which preservation standards and norms to use. Through social media archiving projects, we are gradually getting closer to a long-term solution. One way to bring uniformity to the preservation of social media is to apply the OAIS (Open Archival Information System) model (e.g. BESOCIAL and KADOC-KU Leuven).⁴⁴

Access to and (re)use of the corpus

- 42 From an international point of view, we can say that giving access to national social media (and web) archives is in most cases very restricted due to copyright legislation. This has led to these archives being underused by scholars although national libraries are undertaking active efforts to facilitate access by for example setting up dedicated research labs (cf. BnF Data Lab, National Library of Sweden, British Library Labs) and to engage with the research community to raise awareness about the existence and potential of born-digital collections.⁴⁵ In Belgium, such labs are created less frequently. Some of the studied initiatives (e.g. Amsab) have decided not to communicate the existence of their social media archive to the research community yet. Only large Belgian projects where researchers work full time on social media archiving have been noticed by the public. The BESOCIAL research project for example has put social media archiving practices in the spotlight by getting a lot of press coverage through its crowdsourcing campaign⁴⁶ and the Quarantine archive also got a lot of media attention.⁴⁷ But whether this has paved the way for future researchers or users to find their way to the social media archives, remains to be seen. One way to help with increasing discoverability and (re)use of the archived data could be by setting up a shared social media metadata archive available in Belgium (e.g. Nationaal Register Webarchieven in the Netherlands).⁴⁸
- 43 Taking a closer look at the different initiatives in Belgium, we can conclude that setting up an access platform for social media archiving practices is far from easy. Legal, financial, and technical restrictions hold back institutions.

Table 5

Initiative	Output	Access
------------	--------	--------

BESOCIAL	Pilot interface (with full-text search, filters, NER, etc) with visualisations and statistics	Only staff members of the BESOCIAL project.
Liberas	Social media archive with CSV file that offers URL, keyword, temporal and full-text search.	Only staff members of Liberas
Amsab	Planned: timeline showing the different hashtags and events that were collected.	Only staff members of Amsab working on social media. Access to the social media archive will be restricted to intra muros in the future
ADVN	Planned: social media user interface	No access by users as all datasets are closed for 30 years. Researchers wishing to gain access to archives older than 30 years will need to sign a research statement and will only have access to the content in the reading room. However, if permission from the content creator can be obtained, access can already be provided beforehand.
Letterenhuis	/	/
KADOC	/	/
University Library Ghent	/	Only library administrators have access to the archived social media data.
Archives of the University of Antwerp	Social media archive where WARC or html files are shown as a folder structure containing the html and auxiliary files. URL-search is possible, and browsing through the hierarchical archival structure is possible.	Onsite access to the web and social media archive. Users need to submit a request to be able to consult an archive copy in the reading room. Alternatively, the requested file(s) can be made available via OneDrive. Members of staff of the university can also access the open part of the temporary digital archive since it is a shared folder on the central file server (temporary solution) of the university. In practice, however, only the communications department and the people working on the website archive can access the archive via this way. ⁴⁹
LIBIS (project ICANDID)	iCANDID user interface.	The data is accessible to a limited group of users. It is currently only accessible using the personnel number of specific researchers
Instituut voor Media Studies at KU Leuven	/	The data has only been used for student research. The objective is to conduct further doctoral research based on these data in the future.
AMVB	Planned: inclusion of social media data in the digital repository of AMVB.	Only staff members of AMVB

CAVA	Stored on a hard drive	Not accessible
City of Mons	A virtual exhibition “Quarantine archives”, organised by the AAFB, was created to visualise the output of this project.	Onsite access in the reading rooms, but an inventory describing the content they have captured will be published online.
Archives et Musée de la Littérature	Planned: implementation of social media data in their catalogue. Participation in the virtual exhibition “Quarantine archives”, organised by the AAFB.	The digital-born data are made accessible in the OPAC of AML and are available for consultation on request. Onsite access to the exhibition “Quarantine archives” in the reading rooms, but an inventory describing the content they have captured will be published online.
La ville de Bruxelles	Planned: creation of a (web and) social media platform in the future	/

Overview of the most important characteristics of providing access to (web and) social media archive(s) (output and access)

- 44 Overall, it is noticeable that the knowledge of user needs and research practices of social media and collections of archived websites is quite scarce, and as such, hardly documented.⁵⁰ However, there are several publications that have identified the needs for archiving websites (and its importance) through semi-structured in-depth interviews⁵¹ or by means of online surveys.⁵² According to Vlassenroot *et al.* (2019) and Littman *et al.* (2018), end-users expressed the need of provenance important to scholars, more concretely the context of archiving why, by whom, and on whose behalf data were collected. In a more recent study (Vlassenroot *et al.* (2022)), four general criteria were notified that should be taken into account (from the end-user perspective) when creating an access platform: 1) sufficient communication about and promotion of the social media archive; 2) clear and exhaustive descriptions of the collection scope 3) need to engage with (non-)professional researchers working with the archived social media and to involve them in the creation of methodology and identification of sources within the archives, and 4) development of specific social media data interfaces allowing simple and advanced search.⁵³
- 45 Within the Belgian studied initiatives, we see that only a minority has done research on the requirements of their users.⁵⁴ BESOCIAL is one of the exceptions; through the creation of several fictional personas, their corpus was tested. A mock-up was also made of an ideal interface based on these findings.⁵⁵
- 46 In recent years the focus has also shifted from facilitating close reading of archived web content to distant reading by facilitating big-data analyses of data held in web and social media archives.⁵⁶ Functionalities of playback engines are increasing (such as SolrWayback⁵⁷), which offers, among others, the following options: creation of word clouds, image search, interactive and large-scale link graphs. When we look at the **interface formats**, we can identify two approaches 1) institutions that (will) make their social media data accessible within their existing catalogue (e.g. AML, and

KADOC), and 2) institutions that (will) set up a specific social media interface, possibly in combination with their collections of archived websites (e.g. BESOCIAL).⁵⁸ The majority of the heritage institutions surveyed fall into the second category. Some see potential in already existing social media platforms, such as ADVN, which in the long term wants to create an interactive timeline just like the Trump Archive.⁵⁹

- 47 One other link between the different institutions is that most intend to develop **output** for the public, but currently have no publicly accessible platform - although it may already exist internally. Some of them do share (parts of) their collections of archived websites with the public (e.g. Amsab and Ghent University Library), but currently restrict the access to their social media archive to only staff members. These constitute dark archives as they are a repository for archival resources that is closed to researchers and only accessible to its custodian, in this case the employees of the specific heritage institution.⁶⁰ Legal implications lie mainly behind these choices. Some institutes can avoid the legal implications, for example, by obtaining permission from the account holder to capture and share the archived content. The University of Antwerp, which archives its own accounts, can make its data available to the public - but only intra muros and after researchers submit a request.

Future prospects and conclusion

- 48 In recent years in Belgium, several social media archiving (SMA) initiatives have been started, but so far no detailed overview of this development was provided. Archiving and preserving the rich and rapidly changing data type of social media will play a significant role in keeping our heritage collections up-to-date through the use of the latest digital technologies. Mapping the Belgian SMA landscape is a first step towards gaining insights in and identifying overlapping characteristics of the different approaches of social media archiving applied in Belgium. Similar to the study of SMA practices in other countries, our findings showed that many Belgian institutions are engaged in SMA, but that the stage and efforts vary both in size and scope.
- 49 It is important to highlight three limitations of our research. First, we for sure do not claim to have studied all Belgian institutions that are archiving social media. During our study, efforts were made to find institutions in Flanders, Brussels, Wallonia, and the German-speaking region. Several institutions, however, are engaged in SMA internally, but without communicating about it. It is therefore likely that we failed to notice some SMA projects. Second, there is a general tendency to first introduce archiving websites into the existing institutional structures before also including social media data in archives. We concluded that further research is needed to determine the effects of integrating the archiving of both data types in an institution and adapting the existing infrastructure to expose the data in a user-friendly manner to a broad audience. After the BESOCIAL project, KBR (as the national library) luckily can fall back on a new BELSPO-financed project, BelgicaWeb. This project aims to bring together social media and web content in an interactive access platform. Third, similar to the study of social media practices in other countries, there is little knowledge on good preservation practices of social media data. During the analysis of the results it became clear that there is a lack of common understanding of certain concepts amongst the participants who filled out the survey and that definitions of the used terminology

should have been provided in the survey itself to improve clarity. Due to the lack of homogeneity in preservation practices, we were not able to delve deeply into this topic.

- 50 Overall, we can conclude that the choices and different steps taken in the social media archiving process are resource-dependent. There are three stumbling blocks: i) time to explore SMA next to the ongoing tasks within cultural institutions, ii) (technical) in-house knowledge and iii) limited budget to, for example, switch to commercial tools and thereby gaining time. In addition, the legal framework also plays an important role in SMA processes. When it comes to legal considerations that constitute obstacles to the evolution of the digital society, we can say that this evolution has happened at such a fast pace that it has been difficult for the law to keep up with the latest developments. Smaller institutions without much legal baggage decide to scope a focused corpus that allows them to ask for permission of the account holder. This approach can make the access to data less restrictive, although most institutions still decided on making certain data only internally available (or publicly but only *intra muros* and upon request).
- 51 Regarding the choice of social media platform(s), we noticed that Facebook is most often archived by the Belgium institutions, followed by Twitter and Instagram. This preference is driven by the thematic aspect; organisations find that the accounts they want to capture are more likely to be found on Facebook. We also concluded that the number of accounts and/or hashtags is well balanced between the number of social media platforms. In case that one wishes to archive a lot of accounts and hashtags, organisations opt for fewer platforms. Our findings showed that most of the harvesting is executed by means of open source tools with a focus on capturing the look and feel of an account.
- 52 Over time, the concept of social media and the way society uses these platforms will without doubt considerably change. Whether small institutions will be able to monitor and follow-up on these changes has to be seen. Based on these conclusions, practitioners should consider follow-up research by updating this first Belgian overview to better understand the implications of certain decisions for the future.

BIBLIOGRAPHY

ARCHIVES DE LA QUARANTAINE ARCHIEF. On parle de nous – In de pers. *Archives de la Quarantaine Archief* [online]. 2020-2022. [consulted on 23/08/2022]. Available on: <<https://archivesquarantainearchief.be/nl/category/presse/>>

BAILEY, Jefferson, GROTKE, Abigail, MCCAIN Edward, MOFFATT, Christie, TAYLOR, Nicholas. Web Archiving in the United States: A 2016 survey. *National Digital Stewardship Alliance*. February 2017.

BOTÉ VERICAD, Juanjo, MINGUILLÓN, Julià. Preservation of Learning Objects in Digital Repositories. *RUSC. Universities and Knowledge Society Journal* [online]. 2012, 9, n 1, p. 22. Available on: <<http://dx.doi.org/10.7238/rusc.v9i1.1036>>

- BOYD, Danah, CRAWDFORD, Kate. Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, & Society* [online]. 2019, 15, n 5, p. 662-79. Available on: <<https://www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878>>
- BROWN, Brendan. Trump Twitter Archive. *Trump Twitter Archive* [online]. 2016. [consulted on 2016]. Available on: <<https://www.thetrumparchive.com>>
- CORBEEL, Adrien, PEIGNOIS, Aurore. La KBR vous invite à sauver le patrimoine belge en ligne. *RTBF* [online]. 17 février 2022. Available on: <<https://www.rtb.be/article/la-kbr-vous-invite-a-sauver-le-patrimoine-belge-en-ligne-10937416>>
- DE BAST, Anne-Catherine. Réseaux sociaux : un miroir de la société à préserver. *Athena* [online]. 2022, 359, nov-déc. [consulted on 25/08/2022]. Available on: <<http://athena-magazine.be/magazine/le-magazine-n355/reseaux-sociaux-un-miroir-de-la-societe-a-preserver/>>
- [D.O.D.]. KBR vraagt hulp voor archiveren sociale media. *De Standaard* [online] 29 october 2021. Available on: <https://www.standaard.be/cnt/dmf20211028_97696140>
- DOUGHERTY, Meghan, MEYER, Eric T. Community, Tools, and Practices in Web Archiving: The State-of-The-Art in Relation to Social Science and Humanities Research Needs. *Journal of the Association for Information Science and Technology* [online]. 2014, 65, n 11, p. 2195-2209. Available on: <<https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.23099>>
- DPC. Home. *Digital Preservation Handbook* [online]. 2021. [consulted on 25/08/2022]. Available on: <<https://www.dpconline.org/handbook>>
- FERNANDO, Zeon Trevor, MARENZI, Ivana, NEJDL, Wolfgang. ArchiveWeb: Collaboratively Extending and Exploring Web Archive Collections—How Would You like to Work with Your Collections?. *International Journal on Digital Libraries* [online]. 2017, 19, n 1, p. 39-55. Available on: <<https://link.springer.com/article/10.1007/s00799-016-0206-2>>
- GOMES, Daniel, FREITAS, Sérgio, SILVA, Mário J. Design and Selection Criteria for a National Web Archive. *Research and Advanced Technology for Digital Libraries* [online]. 2003, 4172. Available on: <https://link.springer.com/chapter/10.1007/11863878_17>
- HAI, Liang, ZHU, Jonathan. Big Data, Collection of (Social Media, Harvesting). *The International Encyclopedia of Communication Research Methods*. Hoboken: Wiley-Blackwell, 2017. p. 1-18.
- HOCKX-YU, Helen. Access and Scholarly Use of Web Archives. *Alexandria: The Journal of National and International Library and Information Issues* [online]. 2014, 25, 1-2, p. 113-27. Available on: <<https://journals.sagepub.com/doi/10.7227/ALX.0023>>
- HOLZMANN, Helge, RISSE, THOMAS. Accessing Web Archives from Different Perspectives with Potential Synergies - SAS-Space. *Researchers, practitioners and their use of the archived web* [online]. School of the Advanced Study, 14-16 June 2017. Available on: <<https://sas-space.sas.ac.uk/9678/>>
- JAUMOTTE, Marion. Les Médias Sociaux Belges Archivés et Conservés Par La Bibliothèque Nationale. *RTBF* [online]. 28 octobre 2021. [consulted on 25/08/2022]. Available on: <<https://www.rtb.be/article/les-medias-sociaux-belges-archives-et-conserves-par-la-bibliotheque-nationale-marion-jaumotte-10869305?id=10869305>>
- KBR. PROMISE Project. *KBR* [online]. [consulted on 23/08/2022]. Available on: <<https://www.kbr.be/en/projects/promise-project/>>

- LAURIDSEN, Jesper. SolrWayback 4.0 release! What's it all about?. *International Internet Preservation Consortium* [online]. 25 February 2021. [consulted on 23/08/2022]. Available on: <<https://netpreserveblog.wordpress.com/2021/02/25/solrwayback-4-0-release-whats-it-all-about/>>
- LAZER, David, PENTLAND, Alex, ADAMIC, Laura, ARAL, Sinan, BARABÁSI, Albert-László, BREWER, Devon, CHRISTAKIS, Nicholas, CONTRACTOR, Noshir, FOWLER, James, VAN ALSTYNE, James, *et al.* Review of Computational Social Science. *Science* [online]. 2009, 323, 5915, p. 721-23. Available on: <<https://www.science.org/doi/10.1126/science.1167742>>
- LEE, So-Yeon. Standardization of Digital Archiving and OAIS Reference Model. *Journal of Information Management* [online]. 2002, 33, n 3, p. 45-68. Available on: <<https://doi.org/10.1633/jim.2002.33.3.045>>
- LITTMAN, Justin, CHUDNOV, Daniel, KERCHNER, Daniel, PETERSON, Christie, TAN, Yecheng, TRENT, Rachel, VIJ, Rajat, WRUBEL, Laura. API-Based Social Media Collecting as a Form of Web Archiving. *International Journal on Digital Libraries* [online]. 2016, 19, n°1, p. 21-38. Available on: <<https://link.springer.com/article/10.1007/s00799-016-0201-7>>
- MECHANT, Peter, VLASSENROOT, Eveline. BESOCIAL: Analysis of User Requirements (Task 2.2). *Orféo* [report, online], July 2021. Available on: <<https://orfeo.belnet.be/handle/internal/10011>>
- MECHANT, Peter, VLASSENROOT, Eveline, MESSENS, Fien, GEERAERT, Friedel, CHAMBERS, Sally. BESOCIAL: Quality Control of Harvested Content (Task 3.2). *Orféo* [report, online]. March 2022. Available on: <<https://orfeo.belnet.be/handle/internal/10012>>
- MECHANT, Peter, THEYS, Tim, VLASSENROOT, Eveline. BESOCIAL: Evaluation of the Belgian Pilot Social Media Archive (Task 4.3). *Orféo* [report, online]. June 2022. Available on: <<https://orfeo.belnet.be/handle/internal/10035>>
- MESSENS, Fien. Hoe Gaan We Binnen 200 Jaar Posts Op Sociale Media Terugvinden? KBR Wil Belgische Content Uit Sociale Media Archiveren. *Radio 1* [online]. 29/10/2021. [consulted on 25/08/2022] Available on: <<https://radio1.be/hoe-gaan-we-binnen-200-jaar-posts-op-sociale-media-terugvinden-kbr-wil-belgische-content-uit-sociale>>
- MESSENS, Fien, DENIS, Lise-Anne. To harvest or not to harvest? The importance of legal advice in BESOCIAL. *Zenodo* [poster, online]. 23/05/2022. Available on: <<https://zenodo.org/record/6572703>>
- MESSENS, Fien, GEERAERT, Friedel. Country Report: Web and Social Media Archiving in Belgium. *Orféo* [report, online]. May 2022. Available on: <<https://orfeo.belnet.be/handle/internal/9980>>
- MICHEL, Alejandra, PRANGER, Jessica, GEERAERT, Friedel, LIEBER, Sven, MECHANT, Peter, VLASSENROOT, Eveline, CHAMBERS, Sally, BIRKHOLZ, Julie. M., MESSENS, Fien. Towards a Sustainable Social Media Archiving Strategy for Belgium: BESOCIAL: WP1 Report: An International Review of Social Media Archiving Initiatives. *Orféo* [report, online]. 2021. Available on: <<https://orfeo.belnet.be/handle/internal/7741>>
- MILLIGAN, Ian. Exploring Web Archives in the Age of Abundance: A Social History Case Study of GeoCities. *The SAGE Handbook of Web History* [online]. 2019, p. 344-58. Available on: <<https://sk.sagepub.com/reference/the-sage-handbook-of-web-history>>
- NETWERK DIGITAAL ERFGOED. Register. *National Register Webarchieven* [online]. 2018. Available on: <<https://www.registerwebarchieven.nl/register>>

- NIX, Adam, DECKER, Stephanie. Using Digital Sources: The Future of Business History?. *Business History* [online]. 22/04/2021, p. 1-23. Available on: <<https://www.tandfonline.com/doi/full/10.1080/00076791.2021.1909572>>
- NWALA, Alexander C., WEIGLE, Michele C., NELSON, Michael L. Bootstrapping Web Archive Collections from Social Media. *Proceedings of the 29th on Hypertext and Social Media* [online]. 2019, 29, July 2019, p. 64-72. Available on: <<https://doi.org/10.1145/3209542.3209560>>
- OLMÜTZ ZIERAU, Eld Maj-Britt. A Holistic Approach to Bit Preservation. *Library Hi Tech* [online]. 2012, 30, n°3, p. 472-89. Available on: <<https://www.emerald.com/insight/content/doi/10.1108/07378831211266618/full/html>>
- PAELINCK, Gianni. #Stoofvlees, Rock Werchter of Een Tweet van Rik Torfs? Koninklijke Bibliotheek Wil Online Erfgoed Bewaren. *VRT NWS* [online]. 28 oktober 2021. Available on: <<https://www.vrt.be/vrtnws/nl/2021/10/28/online-erfgoed>>
- PEETERS, Thomas, VAN KEER, Ellen, WEYNS, Katrien, VANDERPERREN, Nastasia, VISSERS, Rony. Best Practices voor de Archivering van Sociale Media in Vlaanderen en Brussel: Dase 2 – D6: Eindrapport – Een duurzaam model voor de captatie en archivering van sociale media. *KADOC* [online]. 2020. Available on: <https://kadoc.kuleuven.be/8_projecten/2020/sm-20220831-d6-eindrapport-v2.pdf>
- ROEKENS, Anne, FRANÇOIS, Aurore, FILLIEUX, Véronique. *Pérenniser l'Éphémère : Archiving et Médias Sociaux*. Louvain : Archives de l'Université, Presses Universitaires de Louvain, 2018.
- ROSENTHAL, David S. H. Bit Preservation: A Solved Problem?. *International Journal of Digital Curation* [online]. 2010, 5, n°1, p. 134-48. Available on: <<http://www.ijdc.net/article/view/151>>
- RUEST, Nick, LIN, Jimmy, MILLIGAN, Ian, FRITZ, Samantha. The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* [online]. August 2020, p. 157-166. Available on: <<https://dl.acm.org/doi/10.1145/3383583.3398513>>
- RUEST, Nick, FRITZ, Samantha, DESCHAMPS, Ryan, LIN, Jimmy, MILLIGAN, Ian. From Archive to Analysis: Accessing Web Archives at Scale through a Cloud-Based Interface. *International Journal of Digital Humanities* [online]. 2021, 2, p. 5-24. Available on: <<https://link.springer.com/article/10.1007/s42803-020-00029-6>>
- SIERMAN, Barbara. Het OAIS-model, een leidraad voor duurzame toegankelijkheid. *Zenodo* [online]. 01/12/2012. Available on: <<https://zenodo.org/record/1188260>>
- STASSART, Camille. Tweets et Publications Instagram, Un Patrimoine Numérique à Part Entière. *Daily Science* [online]. 11 janvier 2022. Available on: <<https://dailyscience.be/11/01/2022/tweets-et-publications-instagram-un-patrimoine-numerique-a-part-entiere/>>
- STIRLING, Peter, CHEVALLIER, Philippe, ILLIEN, Gildas. Web Archives for Researchers: Representations, Expectations and Potential Uses. *D-Lib Magazine* [online]. 20102, 18, n°3/4. Available on: <<http://www.dlib.org/dlib/march12/stirling/03stirling.html>>
- THOMSON, Sara Day, KILBRIDE, William. Preserving Social Media: The Problem of Access. *New Review of Information Networking* [online]. 2015, 20, n°1-2, p. 261-75. Available on: <<https://www.tandfonline.com/doi/full/10.1080/13614576.2015.1114842>>
- TUFEKCI, Zeynep. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *ICWNS '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*. 2014.

VAN HUFFEL, Jozefien. Archief van Vluchtigheid. *Kerk & Leven* [online]. 28 december 2021. Available on: <<https://www.kerkenleven.be/uitgave/2152/artikel/archief-van-vluchtigheid>>

VLASSENROOT, Eveline, MESSENS, Fien, MECHANT, Peter, ROLIN, Eva, CHAMBERS, Sally, BIRKHOLZ, Julie. M., GEERAERT, Friedel, LIEBER, Sven, WATRIN, Patrick. BESOCIAL: Interview / Survey Results WP1. SODHA [online]. 2021. Available on: <<https://www.sodha.be/dataset.xhtml?persistentId=doi:10.34934/DVN/RMKYKO>>

VLASSENROOT, Eveline, CHAMBERS, Sally, LIEBER, Sven, MICHEL, Alejandra, GEERAERT, Friedel, PRANGER, Jessica, BIRKHOLZ, Julie, MECHANT, Peter. Web-Archiving and Social Media: An Exploratory Analysis. *International Journal of Digital Humanities* [online]. 2021, 2, p. 107-128. Available on: <<https://link.springer.com/article/10.1007/s42803-021-00036-1>>

[s.n.]. Robots.txt. MOZ [online]. 2019. [consulted on 219]. Available on: <<https://moz.com/learn/seo/robotstxt>>

NOTES

1. HAI, Liang, ZHU, Jonathan. Big Data, Collection of (Social Media, Harvesting), *The International Encyclopedia of Communication Research Methods*. Hoboken: Wiley-Blackwell, 2017. p. 1-18;

TUFEKCI, Zeynep. n.d. Review of Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the 8th International Conference on Weblogs and Social Media*. 2014. AAAI. 2014, 8, n°1.

2. BOYD, D., CRAWDFORD, K.. Critical Questions for Big Data. *Crawford. Information, Communication & Society*. 2019, 15, n° 5, p. 662-79;

LAZER, D., PENTLAND, A., ADAMIC, L. Review of Computational Social Science. *Science* 323. 2009, n°5915, p. 721-23.

3. A tendency can be seen for initiatives to archive websites first before social media.

VLASSENROOT, Eveline, CHAMBERS, Sally, LIEBER, Sven, MICHEL, Alejandra, GEERAERT, Friedel, PRANGER, Jessica, BIRKHOLZ, Julie, MECHANT, Peter, Web-Archiving and Social Media: An Exploratory Analysis. *International Journal of Digital Humanities*. 2021. Available on: <<https://doi.org/10.1007/s42803-021-00036-1>>

4. MICHEL, Alejandra, PRANGER, Jessica, GEERAERT, Friedel, LIEBER, Sven, MECHANT, Peter, VLASSENROOT, Eveline, CHAMBERS, Sally, BIRKHOLZ, Julie M., MESSENS, Fien. Towards a Sustainable Social Media Archiving Strategy for Belgium: BESOCIAL: WP1 Report: An International Review of Social Media Archiving Initiatives [online]. Commissioned report. Namur Digital Institute, Centre de Recherche Information, 2021. Available on: <<https://orfeo.belnet.be/handle/internal/7741>>;

VLASSENROOT, Eveline, MESSENS, Fien, MECHANT, Peter, ROLIN, Eva, CHAMBERS, Sally, BIRKHOLZ, Julie M., GEERAERT, Friedel, LIEBER, Sven, WATRIN, Patrick. BESOCIAL: Interview / Survey Results WP1 [dataset, online]. SODHA, 2021. Available on: <<https://doi.org/10.34934/DVN/RMKYKO>>

5. PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium.

6. The project Preserving Online Multiple Information: towards a Belgian strategy (PROMISE) started on 1 June 2017 and aimed to develop a federal strategy for the preservation of the Belgian web. The project is the predecessor of the BESOCIAL project. octopix. n.d. PROMISE Project • KBR. KBR. [consulted on 23/08/2022]. Available on: <<https://www.kbr.be/en/projects/promise-project/>>

7. Felixarchief, meemoo (formerly VIAA), University Library Ghent, Liberaal Archief, Amsab-ISG, ADVN, Letterenhuis, KADOC, Vlaams Architectuurarchief Vlaanderen, Archief Gent and Université Catholique de Louvain. These institutions were asked to participate in the BESOCIAL

survey to update the information related to web archiving and provide information about social media archiving at their institution.

8. This article provides more information about the institutions who responded to the survey. It should be noted that a large portion of our survey respondents are affiliated with the "Best Practices for Archiving Social Media in Flanders and Brussels" project, which started in 2020. This affiliation may have influenced our participation rates in the survey.

9. Information about the BESOCIAL project was also added to the table in order to make the overview of Belgian social media initiatives more comprehensive.

10. The project 'Best practices voor de archivering van sociale media in Vlaanderen en Brussel' kicked off in September 2020. The aims of the project are to: 1) archive social media, 2) develop guidelines for data capture, metadata, preservation and opening up the collections for access and reuse, 3) work towards a sustainable model for harvest and 4) work within a network of cultural heritage institutions and Digital Humanities researchers. KADOC initiated the project with meemoo. Other partners involved in the research project are: Centrum Kunstarchieven Vlaanderen / MHKA, ADVN, AMVB (Archief en Museum voor het Vlaams Leven te Brussel), CAVA (Centrum voor Academische en Vrijzinnige Archieven), Liberas, Letterenhuis, VAI (Vlaams Architectuurinitiatief) en het Instituut voor Mediastudies KU Leuven (IMS). MESSENS, Fien, GEERAERT, Friedel.. Country Report: Web and Social Media Archiving in Belgium [report, online]. *Orfeo*, 2022. Available on: <<https://orfeo.belnet.be/handle/internal/9980>>

11. The platform Archives de Quarantaine Archief (#AQA for social media) wants to contribute to archiving and documenting the COVID-19 crisis in Belgium. The project kicked off in April 2020 under the supervision of the Vlaamse Vereniging voor Bibliotheek, Archief & Documentatie vzw (VVBAD) and the Association des Archivistes Francophones de Belgique (AAFB). Working groups were created based on specific themes, one of which focused on web and social media archiving.

12. GEERAERT, MESSENS. Country report.

13. With the constantly evolving landscape of social media, it is crucial to understand what initiatives are being undertaken in the country, and to identify potential gaps or areas for improvement. By conducting a follow-up study, researchers can gain up-to-date valuable insights into the strategies being used to archive social media data, evaluate effectiveness, and identify any challenges faced by organisations. Ultimately, this information can help organisations make informed decisions about archiving social media data.

14. MESSENS, Fien, DENIS, Lise-Anne, To Harvest or Not to Harvest? The Importance of Legal Advice in BESOCIAL. *DH Benelux 2022 - ReMIX: Creation and Alteration 2022* [poster, online]. DH, 2022. Available on: <<https://doi.org/10.5281/zenodo.6572702>>

15. Some (international) institutions also choose to disregard the legal framework within which they operate and act or harvest data without regard for legal compliance.

16. Royal Decree of 19 June 1837. 1837. Arrêté royal du 19 juin 1837 portant constitution en établissement scientifique de la Bibliothèque royale de Belgique. *Moniteur belge*. 8 July 1837.

17. MESSENS, DENIS. To harvest or not to harvest?.

18. VAN KEER, Ellen, WEYNS, Katrien. Klankbordgroep - Best Practices Voor de Archivering van Sociale Media in Vlaanderen En Brussel, Presentation [email to Fien MESSENS]. August 2022;

PEETERS, Thomas, VAN KEER, Ellen, WEYNS, Katrien, VANDERPERREN, Nastasia, VISSERS, Rony. Fase 2 - D6 Eindrapport - Een duurzaam model voor de captatie en archivering van sociale media. KU Leuven, 2022. Available on: <https://kadoc.kuleuven.be/8_projecten/2020/sm-20220831-d6-eindrapport-v2.pdf>

19. NWALA, Alexander C., WEIGLE, Michele C., NELSON, Michael L. Bootstrapping Web Archive Collections from Social Media. *Proceedings of the 29th on Hypertext and Social Media*. 2018. Available on: <<https://doi.org/10.1145/3209542.3209560>>;

LIANG, ZHU, Big data, collection of (social media, harvesting), p. 1-18.

20. GOMES, Daniel, FREITAS, Sérgio, SILVA, Mário J. Design and Selection Criteria for a National Web Archive. *Research and Advanced Technology for Digital Libraries*. 2006, p. 196–207. Available on: <https://doi.org/10.1007/11863878_17>
21. Comments or other types of user-generated content are of interest to the organisation and they would like to include this type of content in the collection, but it is not always possible to do so because of software restrictions.
22. The Covid-19 collection is focused on the university and its community and more specifically on the measures taken within the university and their impact, research initiatives and special events and initiatives.
23. The social media capturing will be seen as an extension of the digital and the paper archives they already keep at AMVB.
24. City Council and College, nine pages of non-for profit organisations, two pages of hospitals, eight pages of citizen initiatives, two pages of schools, one page that grouped the press articles linked to the situation in Mons and ten more pages that had a direct link with the activities of the City of Mons. Currently only the Facebook pages of the Mayor and the City are still captured since the activity had dropped too significantly on the others.
25. Archives et Musée de la Littérature does not monitor social media or the web on a frequent basis. In the past, the Quarantine Archives project in which the organisation participated collected texts and videos from social media such as Facebook and YouTube.
26. It is worth noting that most of the organisations mentioned in this context were involved in the Best Practices project, which could potentially introduce bias to this statement.
27. GEERAERT, MESSENS. Country report, p. 26.
28. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium.
29. This tool is used for capturing hashtags.
30. On a technical level they find harvesting Facebook to be problematic. For the Black Lives Matter collection, they used Webrecorder but the automated version still leaves them wanting and tests with Brozzler were not satisfactory either.
31. This data size is linked to their web and social media archive.
32. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium.
33. VLASSENROOT, Eveline, CHAMBERS, Sally, DI PRETORO, Emmanuel, GEERAERT, Friedel, HAESSENDONCK, Gerald, MICHEL, Alejandra, MECHANT, Peter. Web archives as a data resource for digital scholars. *International Journal of Digital Humanities* [online]. 2019, p. 85–111. Available on: <<https://link.springer.com/article/10.1007/s42803-019-00007-7>>
34. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium. p. 37.
35. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium. p. 37.
36. Note that at the time of writing this article the size of certain corpora may have seriously increased.
37. GEERAERT, MESSENS. Country report. p. 11.
38. DIGITAL PRESERVATION COALITION. *Home - Digital Preservation Handbook*. s.d. Available on: <<https://www.dpconline.org/handbook>>
39. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium.
40. ROEKENS, Anne, FRANÇOIS, Aurore, FILLIEUX, Véronique. *Review of Pérenniser l'Éphémère : Archivage et Médias Sociaux*. Academia Bruylant/Presses Universitaires de Louvain, 2018.

41. THOMSON, Sara Day, KILBRIDE, William. Preserving Social Media: The Problem of Access. *New Review of Information Networking* [online]. 2015, 20, n 1-2, p. 261-75. Available on: <<https://doi.org/10.1080/13614576.2015.1114842>>
42. MICHEL, PRANGER, GEERAERT, LIEBER, MECHANT, VLASSENROOT, CHAMBERS, BIRKHOLZ, MESSENS. Towards a sustainable social media archiving strategy for Belgium.
43. OLMÜTZ ZIERAU, Eld Maj-Britt. A Holistic Approach to Bit Preservation. *Library Hi Tech* [online]. 2012, 30, n 3, p. 472-89. Available on: <<https://doi.org/10.1108/07378831211266618>>;
BOTÉ VERICAD, Juanjo, MINGUILLÓN, Julià. Preservation of Learning Objects in Digital Repositories. *RUSC. Universities and Knowledge Society Journal* [online]. 2012, 9, n°1, p. 22. Available on: <<https://doi.org/10.7238/rusc.v9i1.1036>>;
ROSENTHAL, David S. H. Bit Preservation: A Solved Problem?. *International Journal of Digital Curation* [online]. 2010, 5, n°1, p. 134-48. Available on: <<https://doi.org/10.2218/ijdc.v5i1.148>>
44. LEE, So-Yeon. Standardization of Digital Archiving and OAIS Reference Model. *Journal of Information Management* [online]. 2002, 33, n 3, p. 45-68. Available on: <<https://doi.org/10.1633/jim.2002.33.3.045>>;
SIERMAN, Barbara. *Het OAIS-model, een leidraad voor duurzame toegankelijkheid* [word document]. 2012. Available on: <<https://doi.org/10.5281/zenodo.1188260>>
45. HOCKX-YU, Helen. Access and Scholarly Use of Web Archives. *Alexandria: The Journal of National and International Library and Information Issues* [online]. 2014, 25, n°1-2, p. 113-27. Available on: <<https://doi.org/10.7227/alx.0023>>;
VLASSENROOT, E. et al., Web archives as a data resource for digital scholars.
46. STASSART, Camille. Tweets et Publications Instagram, Un Patrimoine Numérique à Part Entière. *DAILY SCIENCE* [online]. 11/01/2022. Available on: <<https://dailyscience.be/11/01/2022/tweets-et-publications-instagram-un-patrimoine-numerique-a-part-entiere>>;
D.O.D. KBR vraagt hulp voor archiveren sociale media. *De Standaard*. 29/10/2021.
VAN HUFFEL, J. Archief van Vluchtigheid. *Kerk & Leven* [online]. 28/12/2021. Available on: <<https://www.kerkenleven.be/uitgave/2152/artikel/archief-van-vluchtigheid>>;
[s.n.]. Hoe Gaan We Binnen 200 Jaar Posts Op Sociale Media Terugvinden? KBR Wil Belgische Content Uit Sociale Media Archiveren. *Radio 1*. [consulted on 25/08/2022]. Available on: <<https://radio1.be/hoe-gaan-we-binnen-200-jaar-posts-op-sociale-media-terugvinden-kbr-wil-belgische-content-uit-sociale>>;
PAELINCK, Gianni. #Stoofvlees, Rock Werchter of Een Tweet van Rik Torfs? Koninklijke Bibliotheek Wil Online Erfgoed Bewaren. *NWS, VRT* [online]. 28/10/2021. Available on: <<https://www.vrt.be/vrtnws/nl/2021/10/28/online-erfgoed>>;
JAU MOTTE, M. Les Médias Sociaux Belges Archivés et Conservés Par La Bibliothèque Nationale. *RTBF* [online]. 2021. [consulted on 25/08/2022]. Available on: <<https://www.rtbf.be/article/les-medias-sociaux-belges-archives-et-conserves-par-la-bibliotheque-nationale-marion-jaumotte-10869305?id=10869305>>;
[s.n.]. Réseaux Sociaux: Un Miroir de La Société à Préserver. *Athena* [online]. s.d. [consulted on 25/08/2022]. Available on: <<http://athena-magazine.be/magazine/le-magazine-n355/reseaux-sociaux-un-miroir-de-la-societe-a-preserver/>>;
CORBEEL, A. A. P. La KBR vous invite à sauver le patrimoine belge en ligne. *RTBF* [online]. 17/02/2022. Available on: <<https://www.rtbf.be/article/la-kbr-vous-invite-a-sauver-le-patrimoine-belge-en-ligne-10937416>>
47. [s.n.] On Parle de Nous !. *ArchivesQuarantaineArchief* [online]. [consulted on 23/08/2022]. Available on: <<https://archivesquarantainearchief.be/nl/category/presse/>>
48. NETWERK DIGITAAL ERFGOED. Register | Nationaal Register Webarchieven. *Registerwebarchieven* [online]. 2018. Available on: <<https://www.registerwebarchieven.nl/register>>

49. The necessary software to consult the archival copies will be made available in the reading room and they also provide users with the necessary information about software to consult the copies on their own device. GEERAERT, MESSENS. Country report.
50. BAILEY, Jefferson, GROTKE, Abigail, MCCAIN, Reynolds, MOFFATT, Christie, TAYLOR, Nicholas. *Review of Web Archiving in the United States: A 2016 Survey*. National Digital Stewardship Alliance. 2017.
51. STIRLING, Peter, CHEVALLIER, Philippe, ILLIEN, Gildas. Web Archives for Researchers: Representations, Expectations and Potential Uses. *D-Lib Magazine* [online]. 2012, 18, n°3/4. Available on: <<https://doi.org/10.1045/march2012-stirling>>; DOUGHERTY, Meghan, MEYER, Eric T. Community, Tools, and Practices in Web Archiving: The State-of-The-Art in Relation to Social Science and Humanities Research Needs. *Journal of the Association for Information Science and Technology* [online]. 2014, 65, n°11, p. 2195-2209. Available on: <<https://doi.org/10.1002/asi.23099>>; FERNANDO, Zeon Trevor, MARENZI, Ivana, NEJDL, Wolfgang. ArchiveWeb: Collaboratively Extending and Exploring Web Archive Collections—How Would You like to Work with Your Collections?. *International Journal on Digital Libraries* [online]. 2017, 19, n°1, p. 39–55. Available on: <<https://doi.org/10.1007/s00799-016-0206-2>>
52. HOCKX-YU, Helen. Access and Scholarly Use of Web Archives. *Alexandria* [online]. 2014, 25, n°1–2, p. 113–27. Available on: <<https://doi.org/10.7227/ALX.0023>>; VLASSENROOT, et al. 2021.
53. MECHANT, P., THEYS, T., VLASSENROOT, E. BESOCIAL: Evaluation of the Belgian pilot social media archive (Task 4.3) [report, online]. *Orfeo*, 2022. Available on: <<https://orfeo.belnet.be/handle/internal/10035>>
54. However, phase 3 of the Best Practices Project, starting in 2023, will explore this extensively.
55. MECHANT, Peter, THEYS, Tim, VLASSENROOT, Eveline. BESOCIAL: Evaluation of the Belgian Pilot Social Media Archive (Task 4.3) [report, online]. *Orfeo*, 2022. Available on: <<https://orfeo.belnet.be/handle/internal/10035>>; MECHANT, Peter, et al. BESOCIAL: Quality Control of Harvested Content (Task 3.2) [report, online]. *Orfeo*, 2022. Available on: <<https://orfeo.belnet.be/handle/internal/10012>>; MECHANT, Peter, VLASSENROOT, Eveline. Review of BESOCIAL: Analysis of User Requirements (Task 2.2) [report, online]. *Orfeo*, 2021. Available on: <<https://orfeo.belnet.be/handle/internal/10011>>
56. RUEST, Nick, FRITZ, Samantha, DESCHAMPS, Ryan, LIN, Jimmy, MILLIGAN, Ian. From Archive to Analysis: Accessing Web Archives at Scale through a Cloud-Based Interface. *International Journal of Digital Humanities* [online]. 2021. Available on: <<https://doi.org/10.1007/s42803-020-00029-6>>; RUEST, Nick, LIN, Jimmy, MILLIGAN, Ian, FRITZ, Samantha. The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* [online]. 2020. Available from: <<https://doi.org/10.1145/3383583.3398513>>; HOLZMANN, Helge, RISSEM, Thomas. Accessing Web Archives from Different Perspectives with Potential Synergies - SAS-Spacem. *Doi.org* [online]. 2017. Available on: <https://doi.org/https://sas-space.sas.ac.uk/9678/1/RESAW2017-HolzmannRisse-Accessing_web_archives_from_different_perspectives_with_potential_synergies.pdf>; MILLIGAN, I. Review of Exploring Web Archives in the Age of Abundance: A Social History Case Study of GeoCities. *The SAGE Handbook of Web History*. 2019, p. 344–58.
57. “As the name implies, SolrWayback is a fusion of discovery (Solr) and playback (Wayback) functionality. Besides full-text search, Solr provides multiple ways of aggregating data, moving common net archive statistics tasks from slow batch processing to interactive requests. Based on input from researchers the feature set is continuously expanding with aggregation, visualisation and extraction of data.” LAURIDSEN, J. SolrWayback. *Netpreserveblog.wordpress.com* [online]. 2021.

[consulted on 23/08/2022]. Available on: <<https://netpreserveblog.wordpress.com/tag/solrwayback/>>

58. Most of the surveyed institutions have a webarchive. GEERAERT, MESSENS. Country report.

59. BROWN, Brendan. Trump Twitter Archive. www.thetrumparchive.com [online]. 2016. Available on: <<https://www.thetrumparchive.com>>

60. NIX, Adam, DECKER, Stephanie. Using Digital Sources: The Future of Business History?. *Business History* [online]. 2021, april, p. 1-23. Available on: <<https://doi.org/10.1080/00076791.2021.1909572>>

ABSTRACTS

Research has demonstrated that born-digital collections, in addition to the known physical objects, are increasingly becoming more present in cultural institutions. Still, a clear overview of Belgian social media archiving (SMA) initiatives is currently lacking. In this study, we map the Belgian landscape of social media archiving initiatives. The BESOCIAL research project used surveys and semi-structured interviews to find out how different initiatives in Belgium are archiving these volatile types of data, such as tweets. Our findings demonstrate that the scope and size of the archiving activities depend on 3 major thresholds: i) time to explore the field of SMA, ii) (technical) in-house knowledge and iii) the limited budget. Results are discussed in terms of selection, harvesting, preservation, and access. Based on our conclusions, we advise that practitioners should consider follow-up research to better understand the implications of certain SMA decisions for the future.

La recherche a démontré que les collections *born-digital*, en plus des objets physiques connus, sont de plus en plus présentes dans les institutions culturelles. Pourtant, une vue d'ensemble claire des initiatives belges d'archivage des médias sociaux (SMA) fait actuellement défaut. Dans cette étude, nous cartographions le paysage belge des initiatives d'archivage des médias sociaux. Le projet de recherche BESOCIAL a utilisé des enquêtes et des entretiens semi-structurés pour découvrir comment différentes initiatives en Belgique archivent ces types de données volatiles, tels que les tweets. Nos résultats démontrent que la portée et l'ampleur des activités d'archivage dépendent de trois seuils principaux : i) le temps nécessaire pour explorer le domaine de la SMA, ii) les connaissances (techniques) internes et iii) le budget limité. Les résultats sont discutés en termes de sélection, de récolte, de préservation et d'accès. Sur la base de nos conclusions, nous conseillons aux praticiens d'envisager une recherche de suivi afin de mieux comprendre les implications de certaines décisions SMA pour l'avenir.

Onderzoek heeft aangetoond dat er behalve de gekende fysieke objecten ook steeds meer born-digital collecties aanwezig zijn in culturele instellingen. Toch ontbreekt momenteel een duidelijk overzicht van Belgische initiatieven op het vlak van sociale media-archivering (SMA). In deze studie brengen we het Belgische landschap van sociale media archivering in kaart. Het BESOCIAL onderzoeksproject gebruikte enquêtes en semigestructureerde interviews om na te gaan hoe verschillende initiatieven in België deze vluchtige datatypes, zoals tweets, archiveren. Onze bevindingen tonen aan dat de omvang van de archiveringsactiviteiten afhangt van drie belangrijke drempels: i) tijd om het SMA-veld te verkennen, ii) (technische) interne kennis en iii) het beperkte budget. De resultaten worden besproken in termen van selectie, oogst, bewaring en

toegang. Op basis van onze conclusies adviseren we dat praktijkmensen vervolgonderzoek zouden moeten overwegen om de implicaties van bepaalde SMA-beslissingen voor de toekomst beter te begrijpen.

INDEX

Keywords: social media archiving, cultural heritage institutions, digital preservation, born digital data

Trefwoorden archivering van sociale media, instellingen voor cultureel erfgoed, digitale bewaring, geboren digitale gegevens

Mots-clés: archivage des médias sociaux, institutions du patrimoine culturel, préservation numérique, données numériques nées

AUTHOR

FIEN MESSENS

Fien Messens is a passionate professional specializing in the realm of digital humanities and the field of art history. She has a strong focus on the preservation and archiving of born-digital collections, including web and social media content. Throughout her career, Messens has held various positions where she made significant contributions to research projects, such as the BESOCIAL project. Her work in these projects aims to develop sustainable strategies for safeguarding and archiving social media content.

Currently, Messens holds the position of New Media Expert at KADOC-KU Leuven. In this role, she not only manages new media heritage but also continues to make valuable contributions to the field of web and social media archiving. Her expertise extends beyond social media archiving, as she is set to embark on a digital humanities-project related to Belgian 19th Century handwritten letters in the near future.