Towards a sustainable social media archiving strategy

Country report: web and social media archiving in Belgium

Status-Version:	V2.0
Date:	8 April 2022

Authors	Friedel Geeraert and Fien Messens
Responsible partners	Cental (UCLouvain) CRIDS (UNamur) KBR UGhent (MICT, GhentCDH, IDLab)
Version	2.0

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
V1.0	2/12/2020	Add content related to KADOC, Amsab-ISG, ADVN, Liberas Letterenhuis, Universiteitsbibliotheek Gent	Friedel Geeraert
V2.0	8/12/2020	Update section Liberas and add content related to Archives of the University of Antwerp	Friedel Geeraert
V3.0	15/3/2021	Add structure (bibliography, conclusion, methodology, Walloon initiatives)	Fien Messens
V4.0	29/06/2021	Update section ADVN and Universiteitsbibliotheek Gent	Friedel Geeraert
V5.0	05/07/2021	Add content related to AMVB and IMS (KU Leuven)	Fien Messens
V6.0	06/07/2021	Add content related to Letterenhuis and Universiteitsbibliotheek Gent	Friedel Geeraert
V7.0	10/09/2021	Add content related to Vlaams Architectuurinitiatief	Friedel Geeraert
V8.0	26/11/2021	Add content related to CAVA, Archives de la ville de Bruxelles, and LIBIS	Fien Messens
V9.0	18/02/2022	Add content related to AAFB and Ville de Mons	Friedel Geeraert
V10.0	17/03/2022	Re-structure + add content la ville de Bruxelles, and Archives et Musée de la Littérature + update executive summary + update conclusion	Fien Messens
V11.0	28/03/2022	Transfer feedbacked versions	Fien Messens
V 12.0	31/03/2022	Proofread sections added for Letterenhuis, Archives de Quarantaine and the Archives of the City of Mons and Brussels	Friedel Geeraert

V 2.0	08/04/2022	Final proofreading	Friedel Geeraert & Fien Messens
V2.1	21/04/2022	Add content PhD student IMS/Libis	Fien Messens

Table of Contents

Introduction

The intention of this country report is to present a non-exhaustive overview of Belgian institutions that include archived social media material in their collections. This report is based on a previous report that was drafted in the context of the PROMISE project in 2019 in which the following Belgian web archiving initiatives were presented: 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. This report provides more information about the institutions who responded to the survey. In addition, other Belgian organisations were added to the overview 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, KADOC - Documentation and research center on religion, culture and society.

Given the discovery of other social media archives after the closing of the survey, further information was obtained through semi-structured interviews (Archives of the City of Mons) or was obtained by email (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 Instituut voor Media Studies (KU Leuven)). Additionally, information about the research projects ‘Catching the Digital Heritage’, ‘Best Practices voor sociale media archivering in Vlaanderen en Brussel’, and ‘Archives de Quarantaine Archief’ has also been included as a large number of Flemish, Brussels and Walloon institutions are involved in these projects.

In this report, the **first section** will focus on the methodological approach. In the **second section**, the three research projects will be briefly discussed. Then, in the **third section**, the institutions that provided information about their social media activities will be discussed from different perspectives: technical, selection, access, preservation, and legal. The final part of the report is the conclusion.

1. Methodology

This section discusses the research methodology, which is divided into three parts: 1) secondary research, 2) a survey and 3) additional information gathering through semi-structured interviews and email exchanges.

In the first phase, a secondary research approach (also known as desk research) was taken. This involved summarising, collating and/or synthesising documentation related to existing social media archiving projects.

With regards to the selection of our sample of web archiving initiatives, a number of characteristics were taken into account:

- Web archiving initiatives that were included in PROMISE project
- Convenience sampling (also known as grab sampling, accidental sampling, or opportunity sampling), a type of non-probability sampling that involves the sample being drawn from that part of the population that is close to hand. This type of sampling is most useful for pilot testing or exploratory research.; and

- Initiatives that are archiving or do not yet archive social media.

The main research question for this study was: how are institutions and initiatives in Belgium engaging in social media archiving? The archives were studied from an operational, legal and technical point of view. The aim was to fill in the gaps and enrich the information with regards to social media archiving in each of the institutions covering a) the selection, b) the social media archiving process itself, c) access to, and (re)use of the social media archive, d) preservation policy.

In the second research phase, a survey which ran from July 2020 to September 2020 was sent to representatives from the aforementioned institutions studied in the PROMISE project. The aim of this survey was to address the gaps that remained after the literature review. Each of the participants were sent a personalised spreadsheet with questions and were asked to provide written replies. Based on the desk research some questions were already answered beforehand and the respondents were asked to verify this information.

After the survey was closed, a number of new social media archiving initiatives were discovered. Additional information was gathered based on semi-structured interviews and email exchanges with representatives of these respective institutions.

2. Overarching research projects

2.1. Catching the Digital Heritage

Catching the Digital Heritage was a research project and a collaboration between Amsab-ISG and Liberas. The project started on March 18 2019 and ran until March 17 2020. The main aim of the project was to speed up the registration of born digital documents and more specifically of websites. The web archive collections of both institutions count about 1500 websites. The philosophy was that the registration of these collections is the basis of responsible collection management as it not only allows the resources to be managed internally, but also to be communicated to the public.

One of the aims of the project was to collect and analyse information about registration of websites (metadata) and the legal framework (GDPR, copyright, …). A metadata schema was created and pragmatic recommendations were made with regards to the legal framework. Another aim was to describe the websites in the collection management systems of both institutions (Adlib and Atlantis), including the rights statement, by means of ISAD(G). Archive producers have been contacted retroactively to discuss the rights statements. The third aim was to make the metadata publicly available via the online catalogues of both institutions and offer intra muros access to the collections. The project was described in a wiki report and a concluding seminar was organised in the spring of 2020.[1]

In parallel to this research, they also tested other tools than HTTrack in order to write guidelines so that other smaller institutions can set up their own web archiving initiatives without having to rely on IT experts. Archiving social media was another point the researchers reflected upon during the

[1] Amsab-ISG & meemoo. (2020). *Project CEST wiki*. Available online at: https://www.projectcest.be/wiki/Publicatie:Archiveren_van_sociale_media_in_Amsab-ISG.

project as the real 'action' usually takes place on social media and websites are usually more stable with regards to the information that is included.[2]

2.2. Best practices voor de archivering van sociale media in Vlaanderen en Brussel

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).

This research project is divided into three phases:

- Phase 1: research into the different ways and preconditions for capturing and archiving individual accounts.
- Phase 2: research into the added value of using APIs and tools for third parties to capture accounts, tags and hashtags. Together with phase 1, this will result in the development of a sustainable model for capturing and archiving social media.
- Phase 3: research how best to access and reuse archived social media.

The following institutions participate in the project as content providers and each of them participate in one or more pilot projects:

Table 1 - Institutions that participate in Best practices voor de archivering van sociale media in Vlaanderen en Brussel

Initiative	**Focus**
Liberas[3]	focuses on the Twitter, Facebook and Instagram account of Bart Tommelein and the Willemsfonds
Amsab-ISG[4]	focuses on the Twitter and Facebook account of Kunstencentrum Vooruit
Letterenhuis[5]	focuses on the Instagram and Facebook accounts of Gaea Schoeters and the Facebook account of Tom Naegels

[2] Buysse, J., Fernandez-Alonso, J. & Vekemans, T. (2019, April 8). *Interview with Jeroen Buysse, Jeroen Fernandez-Alonso & Tine Vekemans / Interviewers: Friedel Geeraert.*

[3] Surveyed for the BESOCIAL project. Will be discussed in more detail in section 3.

[4] Surveyed for the BESOCIAL project. Will be discussed in more detail in section 3.

[5] Surveyed for the BESOCIAL project. Will be discussed in more detail in section 3.

ADVN[6]	focuses on the Twitter, Facebook and Instagram accounts of the political party N-VA and the politician Bart De Wever
KADOC[7]	focuses on the public content on Twitter, Facebook, Instagram and if possible also the YouTube account of the Christian democratic political party CD&V and a number of individual politicians.
IMS[8]	focuses on (sports) gambling
AMVB[9]	focuses on Kaaitheater
CAVA[10]	focuses on deMens.nu and Humanistisch Verbond
VAi	focuses on the Twitter account of the construction company Willemen Groep, the Instagram account of the company Bovenbouw Architectuur and the Facebook account of Flanders Architecture Institute
Meemoo	focuses on covid-19 collecting
Centrum Kunstarchieven Vlaanderen	focuses on Philippe Van Snick[11]

2.3. Archives de Quarantaine Archief

The platform *Archives de Quarantaine Archief*[12] (#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.[13] An overview of projects that were archiving about the Covid-19 crisis is shown on their website .[14]

The goal of the *Archives de Quarantaine Archief* project was to centralise information about COVID-19 collecting in Belgium and display what has been done by Belgian archives and in different citizen initiatives. Web archiving was one of the ways in which AAFB and VVBAD members were collecting information. Flemish and Walloon partners shared knowledge about web and social media archiving and information on tools to collect websites and social media was also added to the project website to help institutions get started with web and social media archiving.[15]

[6] Surveyed for the BESOCIAL project. Will be discussed in more detail in section 3.
[7] Surveyed for the BESOCIAL project. Will be discussed in more detail in section 3.
[8] *Ibid*.
[9] *Ibid*.
[10] *Ibid*.
[11] KADOC. (2020). *BESOCIAL survey results.*
[12] Institutions that are mentioned on the platform of Archives quarantaine Archief and that harvest social media: Mundaneum, Archives de la ville de Bruxelles, Centre d'Archives et de Recherches pour l'Histoire des femmes (Carhif), AMSAB-Instituut voor Sociale Geschiedenis, Liberas, and Universiteit Antwerpen.
[13] Lessire, S. & Horge, V. (2022, January 5). *Interview with Sara Lessire and Virginien Horge / Interviewers: Fien Messens and Friedel Geeraert*.
[14] https://archivesquarantainearchief.be/nl/.
[15] Lessire, S. & Horge, V. (2022, January 5). *Interview with Sara Lessire and Virginien Horge / Interviewers: Fien Messens and Friedel Geeraert*. More information about the web archiving tools can be found on the website: https://archivesquarantainearchief.be/fr/2020/04/13/sites-web-blog-et-reseaux-sociaux-quelques-solutions-po

At the time of the interview (January 2022), collecting about COVID-19 had stopped in most institutions, but the plan is to move the COVID-19 content to a different platform and set up similar initiatives around other crises, for example the 2021 floods in Belgium and the war in Ukraine. The platform would become an aggregated search engine. AAFB created a virtual exhibition about the COVID-19 collections that went online in March 2022.[16] Each of the participants was asked to select three or four documents or artefacts and to explain their importance to the Belgian public.[17]

3. The studied initiatives

In this part, the institutions on which information was gathered through desk research, the survey, semi-structured interviews and through e-mail exchanges are discussed in detail. Table 3 shows an overview of these institutions, their overall approach to web and social media archiving and their involvement in research projects.

Table 3 - Overview of the surveyed institutions concerning the archiving of social media

Institution	Web archiving	Social media archiving	Social media platforms	Project
Liberas	Yes	Yes	Facebook, Twitter, YouTube and WhatsApp	
Amsab	Yes	Yes	Twitter, Instagram, Facebook, YouTube and a small amount on Flickr, Pinterest, Vimeo and Snapchat	
ADVN | archief voor nationale bewegingen	Yes	Yes	YouTube, Twitter, Facebook, Flickr, Instagram, TikTok, Tumblr, Vimeo, Bitchute, Pinterest and Slideshare	
Letterenhuis	Yes	Yes	Instagram and Facebook	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
KADOC	Yes	Yes	Twitter, Facebook, Instagram (and YouTube)	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
University Library Ghent	Yes	Yes	Twitter, Facebook, Wordpress and Flickr	
Archives of the UA	Yes	Yes	Facebook, Twitter, Instagram, LinkedIn and YouTube	
LIBIS (project ICANDID)		Yes	Twitter	
Instituut voor Media		Yes	Twitter and Facebook	Best practices voor de archivering

ur-archiver-le-web/; https://archivesquarantainearchief.be/fr/2020/04/22/tutoriel-enregistrer-une-page-web-avec-singlefile/.

[16] The exhibition can be consulted on the following web page: https://archivesquarantainearchief.be/expoaqa/s/expovirtuelle/page/expovirtuelle.

[17] Lessire, S. & Horge, V. (2022, January 5). *Interview with Sara Lessire and Virginien Horge / Interviewers: Fien Messens and Friedel Geeraert*.

Studies at KU Leuven				van sociale media in Vlaanderen en Brussel
AMVB	Yes	Yes	Facebook, Instagram and Twitter	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
CAVA	Yes	Yes	Facebook and Instagram	Best practices voor de archivering van sociale media in Vlaanderen en Brussel
City of Mons	Yes	Yes	Facebook	Archives de Quarantaine Archief
Archives et Musée de la Littérature	Yes	Yes	Facebook and YouTube	Archives de Quarantaine Archief
La ville de Bruxelles	Yes	No	Planned: Facebook, Twitter and Instagram	

3.1. Liberas

Liberas was initially conceived as a 'pillar' archive with a focus on the liberal philosophy in Belgium, but the scope of the archive has been broadened to free thinking and acting. Liberas has been archiving websites since 2003 and since 2018 they are also archiving social media. The web archive is focused on liberal persons and organisations. It contains mainly content related to liberal parties, their departments and the candidates that are partaking in the elections. Collecting efforts spike around the different European, federal, Flemish and local elections. Since 2009 websites of liberal organisations in Belgium and abroad are also regularly archived.[18] Three people are currently working on the web archive, but none full-time. No explicit permission of the right holders is asked before the websites or social media are crawled or scraped.[19] They also collect information about events; in 2020 they did three harvests related to the Covid-19 outbreak for example. Their collection comprised 367 GB (or 762 websites, 147 of which were added in 2020) of archived websites and 538 GB of social media in 2020.[20]

They used to work with HTTrack to capture websites and save the files in html, but they switched to Wget in 2019 because the tool outperforms HTTrack in capturing websites that make use of underlying content management systems and databases.[21] They also used Heritrix to capture 18 websites in 2020.[22] Since they work with Wget and Heritrix these websites are saved as WARC files.[23] Robots.txt is not respected as this leads to too much loss of information. They collect the following metadata elements based on ISAD(G): level of description, number, title, subtitle, archive producer,

[18] Buysse, J.. (2020). *Catching the digital heritage. De collectie websites van Liberas (2003-2019).* Available online https://www.liberas.eu/catching-the-digital-heritage-de-collectie-websites-van-liberas-2003-2019/.

[19] Buysse, J. (2020), Personal communication on 2 December 2020.

[20] Buysse, J. (2020), Personal communication on 2 December 2020.

[21] Buysse, J. (2020). *Catching the digital heritage. De collectie websites van Liberas (2003-2019).* Available online https://www.liberas.eu/catching-the-digital-heritage-de-collectie-websites-van-liberas-2003-2019/.

[22] Buysse, J. (2020), Personal communication on 2 December 2020.

[23] Buysse, J. (2020). *Catching the digital heritage. De collectie websites van Liberas (2003-2019).* Available online https://www.liberas.eu/catching-the-digital-heritage-de-collectie-websites-van-liberas-2003-2019/.

date, size, access rights, reproduction rights, physical description and technical requirements and language. [24]

Quality control is minimal. They verify automatically whether the files are corrupted or not. The harvested websites are also checked manually to see whether the archiving was successful but not exhaustively.[25] The archived websites are ingested in their e-deposit called Strongroom.[26] As is the case for all born-digital files, metadata is added after the ingest (upload via RM Tool and Seal Uploader). The metadata is harvested by Atlantis (developed by DEVENTit) afterwards so that the files can be consulted in the collection management system. The last step in this pipeline will be configured in 2021.[27] Access to the archived websites is possible onsite only. [28] The metadata of archived websites are made accessible via their catalogue and can be consulted as an inventory that is split between the websites of persons and the websites of organisations.[29] The archived websites have already been used by researchers, but mostly for textual and visual analyses.[30] Interested parties have to fill in a form indicating for what purpose the web archive will be used, but so far they have not received a lot of requests.[31]

Regarding social media, Liberas focuses mainly on Facebook, but they also capture some Twitter, YouTube and WhatsApp content. [32] They use Bino to scrape the social media content.[33] They collect specific profiles and results from specific search queries linked to events. They do not have a formal social media collection plan, but the list of social media content to harvest is revised and updated every year. Comments and other types of user-generated content are not captured. Normally the social media content is captured once a year, but in 2020 there were exceptionally three harvests due to the Covid-19 outbreak. Twitter data is archived in JSON, and Facebook content is scraped and saved in a csv-table. Embedded media content is preserved in png and mp4. Liberas focuses on the content and not on the look-and-feel of the social media platforms. Quality control is done based on sampling. Only staff members of Liberas have access to the social media archive that can be consulted as a CSV file that offers URL, keyword, temporal and full-text search.[34] Liberas only archives the metadata elements that are captured by the Bino scraping tool: ID, URL, Date of Creation, Likes, Comments, Shares, Description, Download Link and Page Name. Liberas will analyse how the metadata related to the archived social media can be added to the collection management system and how the content can be made accessible to users.[35]

[24] Buysse, J. (2020), Personal communication on 2 December 2020.
[25] Buysse, J. (2018, July 13). *Interview with Jeroen Buysse / Interviewers: Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot*.
[26] Buysse, J. (2020), Personal communication on 2 December 2020.
[27] *Ibid.*
[28] *Ibid.*
[29] Liberas. (2020). *Collectie websites*. Available online https://hdl.handle.net/21.12117/14379309.
[30] For one such example, see: Van Aelst, P. (2018). Politieke personalisering in het digitale tijdperk. In *Verkiezingskoorts. De strijd om de kiezer in de Belgische politiek (19de – 21ste eeuw)* (pp. 89-99). Gent, Liberas/Liberas.
[31] Buysse, J. (2018, July 13). *Interview with Jeroen Buysse / Interviewers: Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot*.
[32] Liberas. (2020). *BESOCIAL survey results.*
[33] Buysse, J. (2020), Personal communication on 2 December 2020.
[34] Liberas. (2020). *BESOCIAL survey results.*
[35] Buysse, J. (2020). *Catching the digital heritage. De collectie websites van Liberas (2003-2019).* Available online https://www.liberas.eu/catching-the-digital-heritage-de-collectie-websites-van-liberas-2003-2019/.

For the preservation of social media data they use Atlantis that is linked to their e-depot Strongroom that is developed by Data Matters. 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. [36]

3.2. Amsab – Instituut voor Sociale Geschiedenis

Amsab-ISG started web archiving in 2007 by doing research into which tools were available for doing web archiving. Until 2019 they used HTTrack and wget to capture websites. If these tools didn't permit to capture the lay-out of the content, they sometimes recorded a user session by means of a screen recorder. If the content couldn't be harvested either, the archive producer was asked to transfer the source files as well as a description of the software on which the website was hosted.[37] From 2019 they are using Heritrix to crawl websites.[38] Sometimes they also get requests to archive old websites. In some occasions, the old version of the website is put online again by the webmaster to enable Amsab-ISG to capture it. The websites are captured once a year and the collection of archived websites comprised about 150 GB in November 2020. [39] 0.25 FTE is currently involved in web archiving.[40]

They collect about 200-500 websites each year and ignore robots.txt.[41] Their collection profile is well described. The quality control is done by visually checking a sample of the archived websites: comparing lay-out etc. They consider the content to be the most important element of a website, so if they can capture the content more or less properly the archived website is integrated in the collection.[42] Incompleteness is seen as an inherent trait of web archives. [43]

When it comes to access, only one website is currently made available online: Indymedia.be. The articles of this magazine are made available as separate documents and are described. The website of Indymedia is considered as a publication. All other harvested websites have been described according to ISAD(G), where Amsab has created a sub-archive 'websites' for each creator of archives, with the various URLs under it at series level and then the snapshots taken periodically at component level. The additional advantage of considering websites as publications is that they can describe the website separately without having to describe the complete archive first. All websites except for the Indymedia.be are only accessible intra muros from one workstation in the reading room.[44] This is mainly due to the fact that copyright and GDPR regulations are not clear enough to make the archived web content available through the OPAC [45]

[36] Liberas. (2020). *BESOCIAL survey results.*

[37] Saevels, M. (2018, July 13). *Interview with Maarten Saevels / Interviewers: Sally Chambers, Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Eveline Vlassenroot and Tecle Zere*.

[38] Amsab-ISG. (2020). *BESOCIAL survey results;* Fernandez-Alonso, J., Personal communication on 30 November 2020.

[39] Saevels, M. (2018, July 13). *Interview with Maarten Saevels / Interviewers: Sally Chambers, Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Eveline Vlassenroot and Tecle Zere*.

[40] Amsab-ISG. (2020). *BESOCIAL survey results.*

[41] *Ibid.*

[42] Fernandez-Alonso, J., Personal communication on 30 November 2020.

[43] Saevels, M. (2018, July 13). *Interview with Maarten Saevels / Interviewers: Sally Chambers, Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Eveline Vlassenroot and Tecle Zere*.

[44] Saevels, M. (2018, July 13). *Interview with Maarten Saevels / Interviewers: Sally Chambers, Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Eveline Vlassenroot and Tecle Zere*.

[45] Fernandez-Alonso, J., Personal communication on 30 November 2020.

Amsab-ISG started capturing social media in March 2020. They capture content on Twitter, Instagram, Facebook and YouTube and a small amount of content on Flickr, Pinterest, Vimeo and Snapchat of institutions or persons whose archives they hold. They are in the process of drafting a formal social media collection policy. They focus on specific profiles and also proactively search for social media profiles of other interesting organisations or persons and either archive them in the context of event-driven harvesting or include them on the list for future periodic harvesting. They archive the profiles, pages and groups on social media once a year. They also check current events once a week and archive relevant hashtags and the results of specific search queries linked to current events and decide what else needs to be archived.[46] They do not use a specific tool for tracking hashtags and mainly archive hashtags based on specific search terms.[47] They also try to stay up to date about future events or campaigns within their community of archive producers and enter these events in a calendar to keep track of them. Social media content linked to events is archived more frequently than social media profiles. Twitter hashtags are captured every week. Instagram hashtags are captured less frequently since a methodology to limit this content to Belgium still needs to be developed. For social media profiles, pages or groups there are two different approaches depending on whether or not the organisations behind the content are archive producers or not. If the organisation in question is an archive producer, they will not do a separate harvest since the content is already included in Amsab-ISG's periodic harvest. If it does not concern an archive producer, they decide whether they wish to capture the content once or if the content is important enough to be captured periodically in the future. Some examples of thematic collections are Covid-19 and Black Lives Matter.[48]

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.[49] They published a Wiki on the subject of web and social media archiving that is regularly updated.[50] A digital strategy is currently being developed, but in the meantime they use the event calendar as the basis for keeping track of themes or events in which one or more of their archive producers are involved. Trending hashtags are also entered into the calendar and newspapers are also consulted. Once a week they check how much was published online about the event and then decide whether to archive it or not.[51]

To harvest the content they tested a number of tools: Nodexl, Twarc and Brozzler for Twitter, Youtube-dl for Youtube, Instaloader for Instagram and Webrecorder, Brozzler and Browsertrix for Facebook.[52] 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.[53] Hashtags are archived using Twarc and for some hashtags the media is archived separately. This is mainly done for content that has a high risk of disappearing very quickly such as protests or content linked to controversial subjects as they currently

[46] Amsab-ISG. (2020). *BESOCIAL survey results.*

[47] Fernandez-Alonso, J., Personal communication on 1 December 2020.

[48] Fernandez-Alonso, J., Personal communication on 30 November 2020. The Covid-19 hashtags they are focusing on are: #beternacorona, #ikbensolidair, #hetkananders, #zorgvoorelkaar, #degoedekantop, #samentegencorona, #zorgvoorlicht, #blijfthuis, #ditismijnzorg.

[49] Amsab-ISG. (2020). *BESOCIAL survey results.*

[50] Amsab-ISG & meemoo. (2020). *Project CEST wiki*. Available online at: https://www.projectcest.be/wiki/Publicatie:Archiveren_van_sociale_media_in_Amsab-ISG.

[51] Fernandez-Alonso, J., Personal communication on 30 November 2020.

[52] Amsab-ISG & meemoo. (2020). *Project CEST wiki*. Available online at: https://www.projectcest.be/wiki/Publicatie:Archiveren_van_sociale_media_in_Amsab-ISG.

[53] Fernandez-Alonso, J., Personal communication on 30 November 2020.

lack the necessary manpower to capture all embedded media.[54] Limiting content to Belgium is another challenge. They use geocodes to determine whether hashtags in English are linked to Belgium but this is far from perfect. For hashtags in Dutch that could also be linked to The Netherlands, they currently do not make a selection but they would like to develop this in the future. [55]

The social media archive comprised more than 200 GB in November 2020.[56] The content of the different social media platforms is stored in different formats: 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. Quality control on archived social media data is done based on manual sampling. Access to the social media archive is currently restricted to the staff members working on the archive and the existence of the social media archive has not yet been communicated to the research community. Access to the social media archive will be restricted to intra muros in the future.[57] They are interested in visualising a timeline showing the different hashtags and events that were collected by means of Timeline and in the long term they would like to develop an access platform similar to The Trump Archive (the code of which is open source) but for hashtags and including a timeline.[58]

They make use of SHA256 checksums and DROID to check file integrity and format identification. Only bit-level preservation is done so far. No specific preservation norms, standards or software are used, but the archived social media content is preserved in WARC, MP4, JSON and JPG. Copyright owners are not approached prior to capture since it is impossible to identify all the right holders, but the archive producers will be made aware of their social media archiving as they will add a clause about it in the contract they conclude with Amsab-ISG. The organisation also considers it to be the responsibility of the social media platform itself to take the appropriate action against illegal and harmful content: if they do not remove it from the social media content that falls within Amsab-ISG's collecting scope, they will archive it.[59]

3.3. ADVN | archief voor nationale bewegingen

ADVN | archief voor nationale bewegingen (literally 'archive for national movements') is an archive and scientific research centre . They currently manage about 7.750 m of paper archives, 188.000 images (photos, posters), audiovisual carriers and objects (paintings, flags, …), 56.000 books, 11.400 periodicals and a digital collection of more than 10 TB.[60] Their collections center around the Flemish movement in general and include archives of ministers, political parties, lobby groups such as the Vlaamse Volksbeweging etc. It is important to underline that acquiring archives for their institution is mostly based on trust as some of the material is highly coloured from a political point of view.[61]

ADVN first experimented with web archiving in 2005-2006 during which they made use of HTTrack and experimented with Heritrix. The last option was technically too difficult to implement and was

[54] Fernandez-Alonso, J., Personal communication on 1 December 2020.
[55] Fernandez-Alonso, J., Personal communication on 30 November 2020.
[56] *Ibid.*
[57] Amsab-ISG. (2020). *BESOCIAL survey results.*
[58] Fernandez-Alonso, J., Personal communication on 1 December 2020.
[59] Amsab-ISG. (2020). *BESOCIAL survey results.*
[60] Bossaert, S., Personal communication on 23 February 2021.
[61] Cobbaert, T. & Bossaert, S. (2018, July 25). *Interview with Tom Cobbaert & Sophie Bossaert / Interviewers: Sally Chambers, Friedel Geeraert, Gerald Haesendonck and Eveline Vlassenroot.*

therefore not further pursued. The project was then abandoned for 6 years and picked up again in 2012. Since 2016 ADVN harvests more than 1000 websites on a yearly basis with Wget, Heritrix, Browsertrix-crawler and ArchiveWeb.page.[62] The quality control is limited to manually checking in different browsers whether the capture is complete. In total 0.2 FTE work on the web archive.

With regards to selection, they focus on persons and organisations whose paper and digital archives are part of their collections: websites of ministers, political parties and their local branches, … They also do specific event harvests focusing e.g. on elections.[63] The entire web and social media collection comprises about 3TB.[64] In most cases the source files of the websites are not part of the transferred archives, which is why they ask for permission to harvest the websites. They respect robots.txt, but when robots.txt does not allow any harvesting at all, they contact the institution to see whether changes to their robots.txt can be made. It is important to note that ADVN considers websites to be publications.[65] They use EAD to describe archived social media data.[66]

ADVN also tested harvesting social media since 2019. They contacted the people in question and guided them in person through the procedure of harvesting their own Twitter, Facebook and other social media accounts. They focused on ministers and/or their chiefs of staff both on the federal and Flemish level. Currently they archive YouTube, Twitter, Facebook, Flickr, Instagram, TikTok, Tumblr, Vimeo, Bitchute and Slideshare. They collect specific social media profiles, hashtags and also results from specific search queries linked to their collection development plan. Between 500 GB and 1 TB of archived social media and website data is added every year. Comments and other types of interaction data of internet users are also collected if possible. Permission from the content creator is sought prior to capturing the social media content.[67]

As traditional web crawlers such as wget and Heritrix are not well-suited to harvest dynamic social media content, they use a combination of browser-based crawling and API archiving. For embedded multimedia, they use browser-based web crawlers to scrape the media content. The archived social media data is stored in the WARC, WACZ, CSV, JSON, MKV, MP4 and JPG formats. The quality of the archived social media is checked by replaying the WARC files in PyWB or ReplayWeb.page. [68]

ADVN is also 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. [69]

The archived websites and social media content cannot be accessed 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 of the organisation. However, if permission from the content creator can be obtained, access can already be

[62] ADVN. (2020). *BESOCIAL survey answers*.
[63] ADVN. (2020). *BESOCIAL survey answers*; Bossaert, S., Personal communication on 23 February 2021.
[64] Bossaert, S., Personal communication on 23 February 2021.
[65] Cobbaert, T. & Bossaert, S. (2018, July 25). *Interview with Tom Cobbaert & Sophie Bossaert / Interviewers: Sally Chambers, Friedel Geeraert, Gerald Haesendonck and Eveline Vlassenroot.*
[66] ADVN. (2020). *BESOCIAL survey answers*; Bossaert, S., Personal communication on 23 February 2021.
[67] ADVN. (2020). *BESOCIAL survey answers*.
[68] *Ibid.*
[69] ADVN. (2020). *BESOCIAL survey answers*.

provided beforehand.[70] They are in the process of developing a user interface. [71] ADVN received a request in 2007-2008 from a researcher who wished to do a discourse analysis for the political party Vlaams Belang, but as ADVN did not have a lot of digital material at this time, the researcher in question ended up scanning the paper versions of the newsletters. The fact that the archived web and social media material cannot yet be accessed explains why the organisation in question does not yet communicate about the existence of their web and social media archive. [72]

3.4. Letterenhuis

The Letterenhuis is the literary archive of Flanders and its mission is to take care of the Flemish literary heritage. They collect and preserve manuscripts, letters, documents, portraits and pictures of authors and open up these sources up for access and use.[73] They also have a large poster, snippet and documentation collection and a collection of objects such as death masks, furniture etc. as they used to have a much broader collection plan that focused on the cultural life in general. They manage 4.5 km of archives and their digital archive currently comprises 4.5 TB. On average, they manage the influx of about 20 personal archives and about 70 to 100 periodical transfers of archives every year.[74]

The Letterenhuis has undertaken a number of small initiatives with regards to web archiving. For instance, they contacted the authors of a well-known literary blog of two journalists: ‘De Papieren Man’ to see if they would like the Letterenarchief to archive their blog but they were not interested. Another initiative was linked with the organisation VLABIN (Vlaams Bibliografisch Informatiecentrum). The website of this organisation was archived by Felixarchief and integrated in their collections.[75]

They inquire about the online and digital presence of authors when paper archives are acquired or deposited in order to take into account the preferences of the archive producers in deciding which content should be preserved.they check what the online activities are of contemporary authors. The blogs of Henri-Floris Jespers and the web expo on Luuk Gruwez on the website of the Poëziecentrum were archived with HTTrack with permission of the right holders.[76] They plan to do more web archiving in the future. Letterenhuis is collaborating with the Erfgoedbibliotheek Hendrik Conscience on a plan to map and preserve online literary production for Dutch-language literature. They identified the different forms of literary production online that fit the collection development plans of both institutions and they will set up a dedicated project.[77] Regarding social media, they would like to archive social media of authors and other relevant content. Their participation in the project ‘Best practices voor de archivering van sociale media in Vlaanderen en Brussel’ is a first step in that direction.[78] In this project Letterenhuis focused on the social media accounts of the authors Tom

[70] Bossaert, S., Personal communication on 23 February 2021.

[71] ADVN. (2020). *BESOCIAL survey answers*.

[72] Cobbaert, T. & Bossaert, S. (2018, July 25). *Interview with Tom Cobbaert & Sophie Bossaert / Interviewers: Sally Chambers, Friedel Geeraert, Gerald Haesendonck and Eveline Vlassenroot.*

[73] Letterenhuis. (2018). *Over het Letterenhuis*. Retrieved from https://www.letterenhuis.be/nl/content/over-het-letterenhuis. Last accessed on 17/08/2018.

[74] Ferket, J. (2021). Personal communication, 6 July 2021.

[75] Van Ongeval, I. (2018, August 16). *Interview with Isabelle van Ongeval / Interviewers: Sally Chambers and Friedel Geeraert.*

[76] Ferket, J. (2021). Personal communication, 6 July 2021.

[77] Ferket, J. (2021). Personal communication, 6 July 2021.

[78] Letterenhuis. (2020). *BESOCIAL survey answers*.

Naegels and Gaea Schoeters. In the future, Letterenhuis will also do outreach to actively promote self archiving of social media accounts by authors themselves.[79]

Regarding preservation, 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 a digital preservation system that is E-ARK compliant. The firm Keep Solutions was awarded the contract and the development started in July 2021. The project partners plan to make the collaboration model sustainable and will submit a project proposal for the second phase of the project in October 2021.[80]

3.5. KADOC – Documentation and research center on religion, culture and society

KADOC is a documentation and research centre dedicated to the safeguarding, optimal management, dynamic retrieval and study of the historical heritage that has emerged from the interaction between religion, culture and society in Flanders, in its national and international context.[81]

KADOC started web archiving in 2015. The selection is done internally by the Consultant Digital Archives and the Head of the Archive. They focus on posts, blogs, events that fit within the acquisition plan and evaluate the 'uniqueness' of information, meaning that they prioritise capturing information that cannot be found elsewhere than on the web. They also archive websites when they are notified that these are at risk of disappearing, thanks to awareness-raising about the fact that websites are also part of the cultural heritage. They compiled an initial seed list of 445 websites and selected 160 that they considered to be priorities. They check the number of videos and the presence of calendars on certain pages in order to exclude content that contains too much audio-visual content or crawler traps. Social media is included in the collection scope (see Best practices voor de archivering van sociale media in Vlaanderen en Brussel). KADOC focuses in the project on the public content on Twitter, Facebook, Instagram and if possible also the YouTube account of the Christian democratic political party CD&V and a number of individual politicians, but builds a large seed list with relevant social media accounts and hashtags to collect in the future.

Regarding access, records of the captured websites aren't available yet. The description of metadata is done in ISAD(G) and comprises title, formal title (URL), creaton dates, collection period, archive producer, extent, content, appraisal and destruction, rights, type (website), technique (tools) and link to the e-deposit. They ask permission of the right holders of the content prior to capturing the content. The collection will be integrated in the central catalogue as a web archive collection and social media collection. No in-depth reflection was done in 2018 with regards to search options they wish to offer and the captured websites are not yet full-text indexed.[82] The archived websites are currently only available to staff members or on demand.[83]

[79] Ferket, J. (2022). Personal communication, 25 March 2022.

[80] Ferket, J. (2021). Personal communication, 6 July 2021. Meemoo. (2021). *Preserve assorted archive collections sustainably with SCALA*. Available online at https://meemoo.be/en/projects/preserve-assorted-archive-collections-sustainably-with-scala.

[81] KULeuven. (2020). *KADOC – KULeuven. Documentatie- en onderzoekscentrum voor religie, cultuur en samenleving*. Retrieved from https://kadoc.kuleuven.be/kadoc/index. Last accessed on 26/10/2020.

[82] Weyns, K. (2018, August 12). *Interview with Katrien Weyns / Interviewers: Friedel Geeraert and Eveline Vlassenroot*.

[83] KADOC. (2020). *BESOCIAL survey results*.

They investigated the possibility of using Archive It and tried to work with Heritrix and Web Curator Tool in 2015, which proved to be too complex to install and maintain. They opted for HTTrack to capture websites that will disappear as it is a more user-friendly tool. In general they only capture the websites about two links deep. They prioritise capturing the content and if possible they also capture the look and feel.[84] They are still searching for a good long-term solution to capture websites in a WARC format that will be preserved in LIAS, the repository of KADOC that uses Rosetta for long-term preservation.[85] Meanwhile they started a project to research sustainable solutions for capturing and archiving social media.

In general KADOC follows 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. They obtained authorisation of the right holders for the accounts included in the research project, but they did not seek authorisation from the social network platforms.[86] 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). The publication and dissemination of the captured social media data will be studied in the third phase of the project (2022-2023). During this phase, the management of copyright will also be investigated more in-depth.[87]

3.6. Ghent University Library

Ghent University Library has been archiving websites since 2007 by means of an Archive-It subscription.[88] In 2021, their web archive comprised 3 TB of data.[89] The selection is done in an organic manner, but they have a mission to archive material related to the Ghent University and for the city of Ghent.[90] They are looking into the possibility of working together with the city of Ghent although there are no concrete plans. The library used to collaborate with the ICT service of UGent to detect hosts that exist within Ghent University, but this practice has been discontinued because of GDPR regulations.[91] The crawl frequencies depend on how often the websites change. The UGent domain is crawled monthly for example. No quality control is done on the captured material.[92]

The collections of the Ghent University Library also include social media content from Twitter, Facebook, Wordpress and Flickr. The social media collecting is focused on specific social media profiles related to UGent and its faculties. Ghent University Library focuses exclusively on public social media data. The social media content is harvested via Archive-It and as is the case for archived websites, no quality assurance is done. The social media archive currently comprises about 300 GB in

[84] Weyns, K. (2018, August 12). *Interview with Katrien Weyns / Interviewers: Friedel Geeraert and Eveline Vlassenroot*.

[85] KADOC. (2020). *BESOCIAL survey results*.

[86] *Ibid.*

[87] *Ibid.*

[88] Archive-It. (n.d.). *Ghent University.* Retrieved from https://archive-it.org/organizations/146. Last accessed on 3/07/2018.

[89] Hochstenbach, P. Personal communication, 5 January 2021.

[90] Bastijns, P. & , Hochstenbach, P. (2018, June 28). *Interview with Paul Bastijns & Patrick Hochstenbach / Interviewers : Sally Chambers, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot.*

[91] Hochstenbach, P. Personal communication, 5 January 2021.

[92] Bastijns, P. & , Hochstenbach, P. (2018, June 28). *Interview with Paul Bastijns & Patrick Hochstenbach / Interviewers : Sally Chambers, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot.*

total. Only library administrators have access to the archived social media data as it is a dark archive. No specific legal restrictions are taken into account when archiving social media data and no authorisation is sought from the social media platforms or the copyright owners prior to archiving. Preservation procedures are not yet in place for social media content, nor are standards and norms for preservation used.[93]

It is important to note that the crawler does incremental archiving, meaning that only the differences between the different versions of the website. They only preserve the default metadata offered by Archive-It.[94] Archive-It offers the 15 Dublin Core elements (contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title, type) at the collection, seed and document level and possibilities to create custom metadata.[95]

Two people work on the web archive: Paul Bastijns who has a degree in library and information science and Patrick Hochstenbach who has a degree in exact science. Combined they work about 0.05 to 0.1 FTE on web archiving as web archiving is considered an ad hoc activity.[96] Not a lot of insight is available on the use of the crawled collections by the general public. Ghent University Library uses Archive-It for crawling targeted websites and making them accessible worldwide, but the platform does not offer user statistics. They also set the crawl frequency and depth of the targeted websites in Archive-It. These websites are not part of the library catalogue, but they plan to integrate them.[97]

Besides archiving web content via Archive-It, Ghent University Library is also working on archiving e-journals via the LOCKSS network. A Java application was developed by LOCKSS to undertake these specific crawls and agreements with the journal publishers are required. The archive currently comprises 9 TB of data. The LOCKSS network was set up within the University Library in 2018. They are also working on archiving the thesis databases via LOCKSS in cooperation with a number of universities including Université Libre de Bruxelles and the University of Bielefeld and there are also plans to archive the content in the Open Journals they publish via the LOCKSS network.[98]

3.7. Archives of the University of Antwerp

The Archives of the University of Antwerp started archiving websites in 2018. They started adding social media content in 2020. They focus on the university, its faculties or related organizations like student clubs but also collect content linked to specific events such as Covid-19.[99] 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.[100] The

[93] Ghent University Library. (2020). *BESOCIAL survey result*.

[94] *Ibid.*

[95] Archive-It. (n.d.). *Add, edit, and manage your metadata*. Retrieved from https://support.archive-it.org/hc/en-us/articles/208332603-Add-edit-and-manage-your-metadata. Last accessed on 3/07/2018.

[96] Hochstenbach, P. (2021). Personal communication, 5 January 2021.

[97] Bastijns, P. & , Hochstenbach, P. (2018, June 28). *Interview with Paul Bastijns & Patrick Hochstenbach / Interviewers : Sally Chambers, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot.*

[98] Hochstenbach, P. (2021). Personal communication, 6 July 2021; SAFE PLN project (2022). *SAFE Archive Federation,* Available online: https://www.safepln.org/.

[99] Archives of the University of Antwerp. (2020). *BESOCIAL survey results*.

[100] Vanneste, W. (2020). Personal communication on 7 December 2020.

content must have a direct link with the university in order to be scoped in. They focus on content on Facebook, Twitter, Instagram, LinkedIn and YouTube. They include profiles of university services, faculties, research groups, researchers, events and student clubs, but also hashtags linked to these specific social media profiles. Comments and other types of user-generated data are included.[101] A collection development plan for archived web content is currently being created. The web archive comprised about 100 GB and 260 snapshots in September 2020. Social media accounts for 10 GB. One staff member (+ two others are involved in setting policy) is currently working on the web archive. [102]

They work with a number of tools: HTTrack, Conifer (and Webrecorder is used sporadically if Conifer doesn't work) and the SingleFile Chrome Extension. HTTrack is not suitable to capture social media and Webrecorder also has its issues. They therefore use the SingleFile Chrome Extension to archive social media in combination with the 'self archiving' function that most social media platforms offer to their users. In this way they capture both the content of the accounts they manage themselves or for which they can make arrangements with the actual owners and the look and feel. The SingleFile output is a single HTML file. An added advantage is that such an html file can contain all messages linked to for example a Facebook account up until a year back.[103] Embedded media is either automatically included or captured manually if deemed necessary. The content is stored in WARC and html files. Not all embedded content is included in the collection as some material is not specific to the university. They still have to decide on the capture frequency and whether or not to archive incrementally or take a full snapshot each time. Quality control is done by means of visual checks. [104]

Access to the web and social media archive is only possible onsite although the archival descriptions in ISAD(G) are made available in their archive management system Brocade. 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 of the university. In practice, however, only the communications department and the people working on the web archive access the web archive via this way.[105] As the consultation of the archived copies is not tracked, it is difficult to say whether it has already been used or not. Access is provided to the WARC or html files or are shown as a folder structure containing the html and auxiliary files. 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[106]. URL-search is possible, but keyword or full-text search is not. Browsing through the hierarchical archival structure is possible. A proper digital archive is currently being developed as access via the file server is only a temporary solution.[107]

3.8. LIBIS - ICandid project[108]

Within LIBIS (in the iCANDID project), data from the social media platform Twitter is being captured -

[101] Archives of the University of Antwerp. (2020). *BESOCIAL survey results*.
[102] *Ibid.*
[103] Vanneste, W. (2020). Personal communication on 7 December 2020.
[104] Archives of the University of Antwerp. (2020). *BESOCIAL survey results*.
[105] Vanneste, W. (2020). Personal communication on 7 December 2020.
[106] *Ibid.*
[107] Archives of the University of Antwerp. (2020). *BESOCIAL survey results*.
[108]" iCANDID: Interdisciplinaire en Cross-Culturele Discoursanalyse vervolledigd met Data Mining Tools" http://intoinfo.blogspot.com/2019/05/icandid-interdisciplinaire-en-cross.html; "iCANDID: Interdisciplinaire en Cross-Culturele Discoursanalyse Vervolledigd met Data Mining Tools" https://www.kuleuven.be/onderzoek/portaal/#/projecten/3H180270.

but it is not intended to be kept for the long term. The project will enable efficient searching of relevant texts and fragments in large amounts of unstructured data. Due to the size of the data analyses, the new infrastructure will improve the quality of the analyses, enabling longitudinal, cross-media and country comparison studies. The focus of iCANDID is on data being searchable for researchers. Thus, it is not excluded that data will be deleted.

For now, iCANDID is limited to the social media platform Twitter. It is possible that data from websites will be added in the future, but currently this is not the scope of the project. The main selection criterion is "what is useful for researchers". The selection criteria depend on the use case that researchers submit to LIBIS. For now, selection was based on both profiles and hashtags. No particular content is excluded. URLs from tweets are shown embedded in iCANDID, but not stored separately.

The data is retrieved daily via the Twitter API, stored in an AgensGraph database and made searchable via an ElasticSearch index. The archived social media is stored in JSON format in an AgensGraph database. Currently (read: summer 2021) 3,063,075 tweets were archived in this format. No quality control is performed on this data. The data is stored in the Schema.org metadata model.

The data is accessible to a limited group of users. The content is searchable via the iCANDID user interface. It is currently only accessible with the personnel number of specific researchers.

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. They follow the Twitter terms of use to update or delete Tweets within 24h of change/deletion.

Obstacles: For obtaining a Twitter developer account, the use case must be specified. Thus, for new use cases, a new developer account must be requested. Deleted or updated Tweets are changed or deleted accordingly on the iCANDID side considering the Twitter terms of use. The number of likes and retweets are not recorded in iCANDID.

With the harvested data from iCANDID, a PhD student at the Institute of Media Studies (KU Leuven) has been working on the project: Social Media as Political Network: A Social Network Analysis Approach to Immigration Debates on Twitter since December 2021.[109] The focus is on the social media platform Twitter, websites are not archived. The goal of the PhD project is to work with network analysis of political communication on social media and expand this towards other platforms (such as Facebook), other national contexts (such as Turkey) and different subjects that generate political conflict and controversy. For this project iCandid archives tweets based on profiles selected by the OPPORTUNITIES team.[110]

The PhD is also interlinked with the OPPORTUNITIES project, which involves creating a comparative analysis of online political communication networks in the context of migration debates in four different European countries. The goal is to analyse social media communication and find ways to create a more level-telling field on the topic of migration. For example, they currently follow many political parties and actors' official accounts on Twitter in 4 different European countries (Germany, Italy, Austria and Hungary). As of 5th April 2022, they archived 17771 tweets from Hungary, 188937

[109] Kiyak, Sercan. (2022). Personal communication on 20 April 2022; Social Media as Political Network: A Social Network Analysis Approach to Immigration Debates on Twitter (3H220062). Retrieved from: https://www.kuleuven.be/onderzoek/portaal/?fbclid=IwAR3PTiMJx18wslr0eqXVw3pD82Tug5pfm_rLpdhF-aHnRJPVhdAoHLaDchA#/projecten/3H220062?hl=en&lang=en.

[110] OPPORTUNITIES is a Horizon 2020 project funded by the European Commission, strives to initiate a new, forward-looking debate, which is grounded on the principles of fair dialogue, multiperspectivity, and an ethics of listening.

tweets from Austria, 657898 tweets from Italy and 284199 tweets from Germany. The total of archived tweets amounts to 1148805. The Twitter data is transferred from iCandid to the PhD student via .csv files that contain attributes such as sender, date and content.

The archived content contains publicly available tweets. However, currently iCandid only shares the archived tweets and news media articles with authorised researchers.

3.9. IMS - Instituut voor Media Studies (KU Leuven) / Mintlab[111]

Like AMVB, the Institute for Media Studies is part of the Best Practices project on social media archiving in Flanders and Brussels. Besides this project, they are also working on 3 other projects with a link to SMA/WA, all three of which are interwoven.

1. In the context of a **PhD project** into the mediatisation of the Belgian (sports) gambling landscape, a number of websites were archived sporadically and non-systematically using the Wayback Machine (https://archive.org/web/). An overview of these websites can be found here.
2. In the context of a **qualitative research seminar** with students of the third bachelor of Communication Sciences at KU Leuven (for the course "Kwalitatief Seminarie 1" (B-KUL-S0E44A)), which is linked to the above-mentioned PhD research, social media posts of various gambling operators from the above-mentioned overview were captured and archived in the academic year 2019-2020. On the one hand, Facebook posts were collected using a paid AI-based software called Scrapestorm, and on the other hand, Twitter posts were captured by using the Twitter API and collaborating with a computer science PhD student.
3. In the context of a **master's thesis research**, which is also part of the above-mentioned PhD research, a list of social media apps from the Google Playstore was captured via Chrome plugins; via https://play.google.com/store.

The focus for selection for (1) and (2) is on gambling-related web content (gambling websites and social media pages and accounts of gambling operators) and for (3) on a webstore. In (1) and (2), the selection of websites/accounts depends on the availability/presence of a social media channel of a particular gambling operator (e.g. some operators have a website but no Facebook page or Twitter account). Within IMS, no particular content is excluded. Selection on the basis of hashtags has not yet been done for this PhD project. The objective in capturing social media was to obtain as many posts as possible.

The posts that have been extracted are stored in Excel sheets (or CSV) and JSON files. Images and movies are saved as JPEG or PNG and mp4. For the total number of archived social media, the IMS only has an overview of (2) the qualitative research seminar; the total number of Facebook and Twitter posts is several thousand entries. On average, between 1000-2000 entries are collected per account. The following tools have already been used: Scrapestorm (free and paid account), Table capture, Web scraper, and Wayback machine. The Wayback machine is used as part of the PhD project; Scrapestorm was used during the qualitative seminar; Table Scraper & Web scraper were used during the master thesis research.

Both urls and images were captured (with Scrapestorm and Web scraper). However, these tools do

[111] Institute for Media Studies. (2022). *Institute for Media Studies*. Retrieved from: https://soc.kuleuven.be/ims.

not allow saving the images directly, but they do allow saving the url of the images in question. To save the images themselves, the chrome-plugin 'Tab save' is used.

The capture is done in a focused manner and is usually completed within two weeks. There is no recurring frequency of crawls. In this process, there is no preservation and/or technical metadata capture.

In terms of quality control, the IMS only has an overview of (2) the qualitative research seminar. Here, the quality check consists of randomly checking whether the social media post has been encapsulated correctly, including all associated emojis, hashtags, hyperlinks and the likes. The check is done manually.

At preservation level, use is made of a personal KU Leuven OneDrive account. In case of collaboration with students, a password-protected folder in OneDrive is shared (or the KU Leuven J drive is used). There is no preservation policy as yet. On a legal level: for the capture of the data, a PRET-file (KU Leuven 'Privacy & Ethics' file) was filled in within IMS.

Until now, the data has only been used for student research. The objective is to conduct further doctoral research based on these data in the future.

Obstacles:

- The limited in-house knowledge; there is still a lot of research to be done independently before certain data can be properly captured and stored. There is a lot of trial and error involved. There is a need for a central point of contact.
- The limited processing capacity of KU Leuven laptops. When large amounts of data are being captured, there is a possibility that the laptop crashes. The same goes for processing the data into other file formats, once the data has been collected.
- The strict and unclear IT policy at KU Leuven concerning 'unknown' software, and the administration around it.

3.10. AMVB - Archief en Museum voor het Vlaamse leven in Brussel[112]

AMVB is a partner in the project project *Best practices voor sociale media archivering in Vlaanderen en Brussel* (started in September 2020), led by KADOC-KU Leuven and meemoo. Within this project AMVB, is currently testing tools, software and workflows for archiving social media posts and accounts. The archiving of social media is still in a test phase at AMVB.

Web archiving came on the AMVB radar in 2020, with some exploratory tests in July 2021. Following the project *Catching the Digital Heritage* by Amsab-ISG and Liberas, the archive itself has started testing a number of archiving tools. At the moment, no fixed workflows have been established for web archiving and/or social media archiving. The AMVB plans to start archiving websites more actively in the (long) term.

[112] AMVB | ARCHIVE AND MUSEUM FOR THE FLEMISH LIVING IN BRUSSELS. (2022) Retrieved from: https://www.amvb.be/.

In terms of selection policy, the knowledge in the archive is not yet advanced enough on the digital repository level to develop efficient workflows for a selection policy. The policy will probably be largely identical for their type of harvest for archiving social media and websites. The frequency of the crawls will depend on the character of the creator in question and the extent to which they are active on social media. The focus will be on profiles of creators. The social media capturing will be seen as an extension of the digital and the paper archives they already keep at AMVB.

The tools currently tested within AMVB are: ArchiveWeb.page (browser based), Snscrape stand-alone (CLI), Snscrape in combination with Browsertrix (CLI). All tools have been used in a Windows environment. The output of the harvested data are mainly WARC files. One tool (Snscrape) has the data stored in a JSON file. For the test-harvest, one archive creator was chosen to harvest accounts and posts from Facebook, Instagram and Twitter. A total of 807 MB of data has been captured so far. Embedded content and media are captured as much as possible to the extent that the tools allow this. With ArchiveWeb.page embedded media are fully included in the capture. URLs can also be captured with this tool, but must be manually clicked on during the capture. Snscrape does not include the embedded media, but does save an html link to the archived post. The media can be viewed via this link (if the post is still online).

The archived posts and accounts are currently not accessible to external users. The focus of the project at this stage is on capturing social media. Making the data accessible is planned for the next project phase. Apart from the project, captured social media accounts and websites at the AMVB will probably not be accessible in the coming years as they still need to develop a policy for this. However, in terms of the GDPR legislation, AMVB does plan to provide such an agreement in the short term.

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. In terms of metadata, most tools only give the title of the file, file size and file format. In the WARC and JSON files one can also find the title of the account being copied, the URL, the number of followers, likes, etc. Preservation and technical metadata are hardly or not at all provided by the tools used. However, the AMVB would like to record and keep track of data such as the date of capture, the tool used, etc.

Currently, only freeware tools are used for the capture of social media, each with their own advantages and disadvantages, which makes it difficult to formulate a policy. Tools like ArchiveWeb.page give good results but are very labour-intensive to use (among other things because the 'autopilot' function does not yet work properly). Tools like Snscrape and Browsertrix are already better at automating, but require a lot of IT knowledge to install and get up and running. For a small institution like AMVB, which does not have its own IT department, it is difficult to get started with these tools and to solve potential problems.

3.11. CAVA (Centrum voor Academische en Vrijzinnige Archieven)[113]

Within CAVA, both websites and social media are archived. The process of archiving websites is currently (read: December 2021) just over a year in joint. Their social media archiving is still in the

[113] CAVA, Centrum voor Academische en Vrijzinnige Archieven. (2022). Retrieved from: https://www.cavavub.be/.

starting phase. Their policy includes the platforms Facebook and Instagram, which for now are the main channels for their archives-creators. For these accounts, the organisation has permission for archiving. The steps taken within CAVA are mainly done within the framework of the project Best practices for social media archiving (Lead partners KADOC and meemoo). The goal for the future is to further fine-tune their archiving of both websites and social media.

In terms of selection policy for web and social media archiving, there is no difference in terms of themes. They select the profiles and the websites mainly on the basis of the archives-creator. So if a creator has a website and social media account, they acquire both. Currently no particular content is excluded for social media archiving, although images are not archived.

As archiving tools, ArchiveWeb.Page is used for social media archiving. Experiments were already conducted with the archiving functions of Facebook and Instagram, but these were less suitable because the look and feel were not archived. For preservation of the data, CAVA does not yet use specific software/tools. 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.

Metadata used within CAVA for archiving social media data:

- Archivist
- Account name
- Account URL
- Date of the crawl
- Medium (facebook/instagram/...)
- Size of the crawl
- File format
- Executor of crawl

The WARC files that are retrieved monthly through ArchiveWeb.Page are subject to a quality control. This consists of comparing the archived version with the online version. For now, the archived social media is stored on a hard drive and content is not accessible.

Obstacles:

- Time constraints: archiving social media is relatively time consuming. The retrieval is rather labour-intensive so CAVA is currently collecting limited metadata.
- Technical knowledge is required to deal with WARC files (and to work with the internal storage system).

3.12. Archives of the City of Mons

The Archives of the City of Mons were one of the contributing partners to the *Archives de Quarantaine Archief* initiative. At the beginning of the COVID-19 crisis and working from home, they captured about 30 Facebook pages until April 2020. These comprise the Facebook pages of the 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 such as the page of the Academy of Music or the Public Centre for Social Action (CPAS in French). Only the Facebook pages of the Mayor and the City were still captured at the time of the interview (January 2022) since activity had dropped too significantly on the others. At the beginning, all the pages were captured every two weeks, but this changed to monthly later on in the crisis. Whenever new measures were announced, active monitoring was done to discover other interesting pages. [114]

The choice to archive Facebook is motivated by the fact that it is a platform that is used by many different generations and that there is a lot of interaction in the comments section. Comments on the selected Facebook pages are archived. The guiding principle is that at least 100 comments per page are archived and when there are more than 100 comments, the Archive of the City of Mons aims to capture at least 10% of all comments. Other social media platforms were not included because of a number of reasons. TikTok for example is mainly used by a young audience and archiving videos results in larger amounts of data to be stored and preserved. Twitter also offers less interaction than Facebook and many Twitter profiles are private. Moreover, Twitter was more difficult to archive at the time of the events.[115]

At the beginning, the Archives of the City of Mons used SingleFile to archive Facebook, and HTTrack to archive websites. SingleFile produces a single html file. Afterwards, this service uses SingleFile which produces an html file which can be opened by Zip. This tool has the added advantage that the files remain easy to read as technology quickly evolves. Tools that produce WARC-files were considered less optimal as this file format requires technical knowledge. Accompanying metadata are stored in the title, for example 'Facebook_Ville_20211207_20210119_SingleFileZ'. Quality control is done by opening the html in Firefox. When blank spots show up in the archived version, the page is captured again. The readability of the files is also periodically checked. [116]

Rights management is an important part of the social media archiving at the City of Mons. Capturing and making available the pages of the Mayor and the City does not pose any issues but the same cannot be said for the other pages. The captured content can therefore only be consulted on site in the reading rooms, but an inventory describing the content they have captured will be published online. On one occasion a member of the public also approached the Archives to conclude an agreement to allow them to capture and make available the Facebook page of their initiative.[117]
The Archives of the City of Mons have communicated about their initiative on the website and in the podcast of Archives de Quarantaine.[118]

3.13. Archives et Musée de la Littérature

[114] Lessire, S. & Horge, V. (2022, January 5). *Interview with Sara Lessire and Virginien Horge / Interviewers: Fien Messens and Friedel Geeraert*.
[115] *Ibid*.
[116] *Ibid*.
[117] *Ibid*. The convention for donations can be found here: https://archivesquarantainearchief.be/fr/2020/08/04/conventions-de-don-archives-communales/.
[118] For more information, see: https://archivesquarantainearchief.be/fr/2021/05/06/memoire-de-confinement-a-mons-un-an-de-travail/; https://archivesquarantainearchief.be/fr/2021/12/06/fragment-2-archiver-facebook-pour-conserver-letat-desprit-des-gens/ and https://archivesquarantainearchief.be/fr/podcasts/ (Vis ma vie d'archiviste confiné.e, n°7).

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 organization participated collected texts and videos from social media such as Facebook and YouTube. These digital archives were collected during a literature review specifically related to the quarantine context, as a result of the Covid-19 pandemic, linked to themes of the French-speaking Belgian literary and theatrical heritage that constitutes the specificity of Archives et Musée de la Littérature. The focus was on collecting accounts. The collection continued after the lockdown period, but with lower intensity. In addition to this social media collected under the Quarantine Archives project, Archives et Musée de la Littérature also harvested media websites such as Le Soir, La Libre Belgique, RTBF, as well as websites on (Belgian) literature/theatre, such as Poète Nationale, Le carnet et les instants, Passa Porta, …

The organisation does not have a social media selection policy. However, they have set up a procedure for the permanent archiving of websites that originally belonged to documentation centres or non-profit organisations with similar fields of activity to theirs, but that no longer exist. In this case, they have archived their websites and recovered the domain name in order to prevail sustainable archiving activities.

The AML either archives the i) entire website and generates html files (even if it is a site with CMS for example) or it is ii) pages or parts of a website. For the latter it converts pdf or jpeg formats. Images are scoped out in the harvesting process because of their low quality. Harvested social media data (from Facebook and YouTube) is stored in mp4 or pdf formats.

The size of the collection is quite small since it was a harvest of a specifically short period of time. Due to this limited scope, no quality control is currently carried out. 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.

An obstacle in the harvesting process: moving from the social media page to a 'neutral' format that can be independently integrated into their digital collections.

A virtual exhibition "Quarantine archives", organised by the AAFB, is currently being created to visualize the output of this project. It will display the various data collected by different archive centres (including Archives et Musée de la Littérature) . In this exhibition, archivists who participated in these collection and preservation activities will also testify. The opening will take place on 17 March 2022.

3.14. Archives de La Ville de Bruxelles

Archives de La Ville de Bruxelles has already completed a test run to archive its own website for long-term preservation using WAIL. An obstacle here was that there was limited storage capacity. Their future goal is to harvest the websites and social media (Facebook, Twitter and Instagram) related to La Ville de Bruxelles and linked organisations (such as the ASBL Bruxelles-Musées-Expositions). For now, they can only aim to collect a limited collection until funding is released to provide additional storage for this born-digital data. At the moment there is no staff member who is responsible for web/social media archiving. No tools linked to social media archiving have been tested up till now.[119]

4. Conclusion

In this report we sought to provide a non-exhaustive overview of the state of art of social media archiving at Belgian memory institutions. Information was gathered through desk research, an extension of the survey that was done in the context of the PROMISE project that focused on archiving websites and in the case of the Archives de Quarantaine project and the Archives of the City of Mons through semi-structured interviews.

Some limitations in our research approach should be noted however. Firstly, 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 (e.g. 'preservation formats' or 'preservation standards') and that definitions of these terms should have been provided in the survey itself in order to improve clarity. Secondly, due to the convenience sample we gathered, we cannot claim to have used a representative sample of Belgian SMA initiatives. Thirdly, our study is based on self-reported data gathered using an online spreadsheet; although this proved to be a quick and easy method to collect data on current SMA initiatives. Other limitations such as reliability issues in online surveys and a possible self-selection bias need to be also taken into account (Gosling et al. 2004). Nonetheless, we hope that this Belgian study will be used as a point of departure for further research on SMA.

Our findings show that many institutions within Belgium are engaged in SMA, yet the stage and efforts vary in size and scope. Archiving social media happens through selective crawls that most often focus on specific events, manifestations or even emergencies and to a lesser extent through crawls on specific themes such as events. When creating a social media archive it is very difficult or near impossible to anticipate or plan for certain major events (e.g. covid19, and attacks of 22nd of March). Given that it is not feasible to archive the entire social web, selections must be made. These selections are often based on a specific topic; a hashtag (#) or keywords, a limited time period, or a crawl on one specific platform. Facebook and Twitter are the social media platforms most often archived by the institutions in our sample, followed by Instagram.

[119] Blanco, P., Personal communication on 22 February 22 2022.

5. Bibliography

ADVN. (2020). BESOCIAL survey answers.

Amsab-ISG. (2020). BESOCIAL survey results.

Amsab-ISG & meemoo. (2020). Project CEST wiki. Retrieved from https://www.projectcest.be/wiki/Publicatie:Archiveren_van_sociale_media_in_Amsab-ISG.

Archive-It. (n.d.). Add, edit, and manage your metadata. Retrieved from https://support.archive-it.org/hc/en-us/articles/208332603-Add-edit-and-manage-your-metadata. Last accessed on 3/07/2018.

Archives de Quarantaine Archief. Retrieved from https://archivesquarantainearchief.be/nl/.

Archives of the University of Antwerp. (2020). BESOCIAL survey results.

Bastijns, P. & , Hochstenbach, P. (2018, June 28). Interview with Paul Bastijns & Patrick Hochstenbach / Interviewers : Sally Chambers, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot.

Blanco, P., Personal communication on 22 February 22 2022.

Bossaert, S., Personal communication on 23 February 2021.

Behoud de Begeerte. (s.d.). Behoud de Begeerte. Retrieved from http://www.begeerte.be/. Last accessed on 17/08/2018.

Buysse, J. (2018, July 13). Interview with Jeroen Buysse / Interviewers: Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Peter Mechant and Eveline Vlassenroot.

Buysse, J., Fernandez-Alonso, J. & Vekemans, T. (2019, April 8). Interview with Jeroen Buysse, Jeroen Fernandez-Alonso & Tine Vekemans / Interviewers: Friedel Geeraert.

Buysse, J.. (2020). Catching the digital heritage. De collectie websites van Liberas (2003-2019). Retrieved from https://www.liberas.eu/catching-the-digital-heritage-de-collectie-websites-van-liberas-2003-2019/.

Buysse, J. (2020), Personal communication on 2 December 2020.

Cobbaert, T. & Bossaert, S. (2018, July 25). Interview with Tom Cobbaert & Sophie Bossaert / Interviewers: Sally Chambers, Friedel Geeraert, Gerald Haesendonck and Eveline Vlassenroot.

Database Archives de la Ville de Bruxelles relative au coronavirus retrieved from on Conifer, https://conifer.rhizome.org/AVB-ASB/coronarchivesief.

Ferket, J. (2021). Personal communication, 6 July 2021.

Fernandez-Alonso, J., Personal communication on 30 November 2020.

Fernandez-Alonso, J., Personal communication on 1 December 2020.

Ghent University Library. (2020). BESOCIAL survey result.

Hochstenbach, P. Personal communication, 5 January 2021.

Hochstenbach, P. (2021). Personal communication, 6 July 2021;

iCANDID: Interdisciplinaire en Cross-Culturele Discoursanalyse vervolledigd met Data Mining Tools" http://intoinfo.blogspot.com/2019/05/icandid-interdisciplinaire-en-cross.html; "iCANDID: Interdisciplinaire en Cross-Culturele Discoursanalyse Vervolledigd met Data Mining Tools" https://www.kuleuven.be/onderzoek/portaal/#/projecten/3H180270.

Institute for Media Studies. (2022). *Institute for Media Studies*. Retrieved from: https://soc.kuleuven.be/ims.

KADOC. (2020). BESOCIAL survey results.

KULeuven. (2020). KADOC – KULeuven. Documentatie- en onderzoekscentrum voor religie, cultuur en samenleving. Retrieved from https://kadoc.kuleuven.be/kadoc/index. Last accessed on 26/10/2020.

Lessire, S. & Horge, V. (2022, January 5). *Interview with Sara Lessire and Virginien Horge / Interviewers: Fien Messens and Friedel Geeraert*.

Letterenhuis. (2018). Over het Letterenhuis. Retrieved from https://www.letterenhuis.be/nl/content/over-het-letterenhuis. Last accessed on 17/08/2018.

Letterenhuis. (2020). BESOCIAL survey answers.

Liberas. (2020). Collectie websites. Retrieved from https://zoeken.liberas.eu/detail.php?id=14379309.

Meemoo. (2021). *Preserve assorted archive collections sustainably with SCALA*. Available online at https://meemoo.be/en/projects/preserve-assorted-archive-collections-sustainably-with-scala.

Messens, Fien; Vlassenroot, Eveline; Mechant, Peter; Watrin, Patrick; Rolin, Eva; Chambers, Sally; Birkholz, Julie M.; Geeraert, Friedel; Lieber, Sven; Michel, Alejandra, 2021, "BESOCIAL: Interview / Survey results WP1", https://doi.org/10.34934/DVN/RMKYKO, Social Sciences and Digital Humanities Archive – SODHA, V1.

Saevels, M. (2018, July 13). Interview with Maarten Saevels / Interviewers: Sally Chambers, Emmanuel Di Pretoro, Friedel Geeraert, Gerald Haesendonck, Eveline Vlassenroot and Tecle Zere.

SAFE PLN project (2022). *SAFE Archive Federation,* Available online: https://www.safepln.org/.

Stadsarchief Brussel (2020). Verwerven van documenten en getuigenissen van het leven in opsluiting-Pandemie. Retrieved from https://archivesquarantainearchief.be/nl/2020/04/07/collecte-de-documents-et-temoignages-de-la-vie-en-confinement-pandemie-coronavirus/.

Van Aelst, P. (2018). Politieke personalisering in het digitale tijdperk. In Verkiezingskoorts. De strijd om de kiezer in de Belgische politiek (19de – 21ste eeuw) (pp. 89-99). Gent, Liberas/Liberas.

Vanneste W. (2020). Personal communication on 7 December 2020.

Van Ongeval, I. (2018, August 16). Interview with Isabelle van Ongeval / Interviewers: Sally Chambers and Friedel Geeraert.

Vlassenroot, E., Chambers, S., Lieber, S., Michel, A., Geeraert, F., Pranger, J., Birkholz, J. & Mechant, P. (2021). Web-archiving and social media: an exploratory analysis. INTERNATIONAL JOURNAL OF DIGITAL HUMANITIES, in press.

Weyns, K. (2018, August 12). Interview with Katrien Weyns / Interviewers: Friedel Geeraert and Eveline Vlassenroot.