*Carolina de OLIVEIRA**

# SOCIAL MEDIA AS RECORDS IN THE CONTEXT OF THE OLYMPIC PUBLIC AUTHORITY (APO) AND THEIR PRESERVATION: A CASE STUDY

*Abstract:*

*In the management of social media records, the ideal scenario is to use archival policies plus an information system and digital repositories based on Archival Science requirements. Since the APO was unable to develop the ideal scenario, an alternative strategy was established. Using case study methodology, the purpose of this essay was to demonstrate the pros and cons of using Facebook, Flickr, Instagram, Twitter, and YouTube self-archiving or third-party tools to download their content to local hard drives. The tests provided results on the facility of using the tools for digital records archiving, and the file format for preservation. The practical implications of the results were to conciliate the producer's technical limitations, digital repository configurations, and Archival Science requirements.*

*Keywords:*

*social media, archival science, digital records, digital preservation, open government*

*Izvleček:*

*Družbeni mediji kot zapisi v kontekstu Olimpijskega javnega zavoda in njihova hramba: študija primera*

*Pri upravljanju z zapisi družbenih medijev bi bilo po idealnem scenariju potrebno uporabljati arhivsko politiko, skupaj z informacijskim sistemom in digitalnim repozitorijem na osnovi zahtev arhivske znanosti. Ker Olimpijski javni zavod ni uspel razviti idealnega scenarija, je bila vzpostavljena alternativna strategija. Z uporabo metodologije študije primera poskuša pričujoči prispevek prikazati prednosti in slabosti uporabe funkcije samoarhiviranja ali drugih orodij za shranjevanje vsebin na lokalne trde diske za naslednje platforme: Facebook, Flickr, Instagram, Twitter in YouTube. Poskusi so prinesli rezultate na področju uporabe orodij za arhiviranje digitalnih zapisov in formatov za hrambo. Praktične implikacije rezultatov so združile ustvarjalčeve tehnične omejitve, konfiguracije digitalnega repozitorija in zahteve arhivske znanosti.*

*Ključne besede:*

*družbeni mediji, arhivska znanost, digitalni zapisi, digitalna hramba, odprta vlada*

---

\* *Carolina de Oliveira, MA, records manager/archivist, National Archives of Brazil, Praça da República, 173 - Centro, Rio de Janeiro - RJ, 20211-350, contact: carolina.ufsm@gmail.com.*

## 1    INTRODUCTION

The use of social media by ordinary people is not a novelty, as we have always wanted to keep in touch with our relatives and friends; thus, information and communication technologies (ICT) have largely facilitated this in the past several decades. Nowadays, companies and governments have institutional accounts for one reason: to keep in touch with their clients/citizens.

However, as public bodies using private sites like Facebook, Instagram, YouTube, Flickr, etc., the main issue for governments is how to manage and preserve their records on those sites. The purpose of this article is to discuss and present the initiative of the Olympic Public Authority (APO). First, it will provide a revision of the meaning and purpose of social media. Second, it will provide a reflection on the identification of information recorded on social media platforms as archival documents. Finally, it will present the methodology and results of the case study developed at the APO to preserve the social media records produced during the first South America Olympic Games in Rio de Janeiro, Brazil.

## 2    THE MEANING AND PURPOSE OF SOCIAL MEDIA

Social media is comprehended as "the collective name given to Internet-based or mobile applications which allow users to form online networks or communities based on common interest, social or ideological orientations" (Hockx-Yu, 2014). This definition focuses on the Internet and the digital environment, but analogue television, newspapers, cinema, and radio should also be considered social media, as they comprise technologies developed to support communication among communities.

From Helen Hockx-Yu's definition, we may emphasize "networks." We understand that the main purpose of social media platforms is to connect people for many different reasons, but particularly to express an opinion about trending topics or to make fun of such topics on the Internet. On the other hand, the use of social networking sites by governments represents transparency, social participation, and democracy, but also control.

The adoption of these types of digital media, via the Internet, supports the initiative of Open Government. The main organization promoting this initiative is the Open Government Partnership (OGP),[1] whose vision "is that more governments become sustainably more transparent, more accountable, and more responsive to their own citizens, with the ultimate goal of improving the quality of governance, as well as the quality of services that citizens receive. This will require a shift in norms and culture to ensure genuine dialogue and collaboration between governments and civil society."

A shift in "norms and culture" must be applied to the use of social media by public bodies. In Brazil, there is a norm to regulate the use of social media by institutions, particularly in relation to information security and responsibilities (Brazil, 2012).

---

[1]    *Brazil has been a member of the Open Government Partnership (OGP) since 2011 (http://www.opengovpartnership.org/country/brazil).*

However, since governments have adopted the social media provided by private companies, how do they maintain information security? Are they assigned a different set of "Terms of Use"? Moreover, aside from information security and other vulnerabilities, another relevant issue is the long-term preservation of social media records.

In this century, we can assume that an important source of information about the behavior of members of a society is social media and "This is perhaps the most demanded content by researchers [...] to understand our time" (Hockx-Yu, 2014).

Memory institutions must assume responsibility for defining procedures and best practices to preserve this other type of primary source. According to Hockx-Yu (2014, p. 1), "a combination of legal, curatorial and technical issues has made archiving of social media content a non-trivial task. To date there are no scalable solutions to preserving such content. There is thus a need to address the problem collectively, starting with discussing common issues and practices."

## 3    IS SOCIAL MEDIA A RECORD?

In Archival Science theory, a record is the byproduct of an activity developed by an organization, person, or family. If the producer of the information recorded in a social media platform is a public body giving notice about its activities and actions, that comprises a record, and archival procedures should be applied to its preservation.

New media present a challenge for archivists to recognize as records: What activities do they represent? Are the activities unique? Is it a simple or complex document? What about the archival bond, persons involved, reliability, and authenticity?

According to Duranti and Thibodeau (2008, p. 348), "what distinguishes a record from a document that is not a record is the nature of its relationship with the activity of the creator rather than its formal or content characteristics." The use of social media platforms by governments is associated with their activity of "Social Communication/Publicity."

In 2001, the Brazilian National Archives Council (Conarq) published a basic disposition schedule for activities that do not comprise the final business of Federal Public Administration bodies[2]. Records referring to the publicity of public bodies were defined as permanent and documents from social media platforms are included in this purpose.

An Archival Institution in the public sphere, unless exempted by legislative frameworks, will probably face problems archiving private social media platforms. The intention of archiving is to capture and replay the content, look, feel, and context as much as possible, but how can this be done without facing legislative problems?

---

[2]    *Available at http://www.conarq.arquivonacional.gov.br/images/publicacoes_textos/Codigo_de_ classificacao.pdf.*

One reasonable measure would be the inclusion of social media as records in archival policies and programs. Further, technological tools (Thomson, 2016)[3] should be developed to facilitate the capture of such content to archival information systems and its transfer to a trustworthy archival repository. To date, this has not been an easy task, as can be seen in the case of the UK: "Due to the difficulty in identifying automatically the UK content on non .uk domains, the coverage of social media content will remain limited until scalable solutions are developed for UK territoriality" (Hockx-Yu, 2014, p. 8).

Once the documents from social media platforms are recognized as records, an alternative for their long-term preservation is to maintain stable content and fixed form, and then to make use of a digital preservation system based on archival premises.

## 4    METHODOLOGY AND RESULTS

### 4.1    Olympic Public Authority (APO)

The Olympic Public Authority (APO), a Brazilian public consortium created by Federal Law n. 12.396, of March 21, 2011, guaranteed the International Olympic Committee (IOC) the delivery of the Rio 2016 Olympic and Paralympic Games. The APO´s mission was to coordinate the actions of the Federal Government, Rio de Janeiro State Government, and Rio de Janeiro City Hall for the preparation and delivery of the first Olympic Games in South America.

Institutional accounts were created on social media, such as Facebook, Twitter, Flickr, Instagram, and YouTube to keep in touch and interact with the Brazilian and international communities, above all complying with the public bodies' transparency initiatives in a commitment to the Freedom of Information Law, n. 12.527, from November 18, 2011.

Based on the law binding its creation, the APO should be disbanded in 2018 – plus or minus two years. The Olympic Public Council (CPO), responsible for the top decisions, defined the end date of the APO as March 31, 2017. Therefore, the Social Communication Department (ASCOM), Records Management Department (SUGED), and Information Technology Department (STI) had to define a methodology for the long-term preservation of its social media records.

### 4.2    Methodology

This case study was developed at the APO with the aim of defining the best method to capture and archive social media recognized as records. The team responsible for so doing was composed of professionals from the Social Communication Department (ASCOM), Records Management Department (SUGED), and Information Technology Department (STI).

---

[3]    *See item "9.2 Tools and Resources" at p. 37.*

According to Hockx-Yu (2014), "Web crawlers are commonly used to capture snapshots of websites." Based on this, we defined three steps for the case study:

1. Review the guidelines emanating from the national archives of Brazil, the United Kingdom, the United States of America, and Australia regarding the capture of social media as records and best practice in this regard[4];

2. Research open source tools for the capture of social media records to a trusted archival repository;

3. Test whether the tools selected are appropriate for social media record archiving.

## 4.3    Results

On the websites of the national archives of the United Kingdom, the United States of America, and Australia, guidelines are given to recognize social media as records and reports of experiences, but no recommendations are provided about specific tools that are appropriate to capture the content therein. In the case of Brazil, guidelines are provided for the preservation of digital records in general, but not social media.

According to Hockx-Yu (2014, p. 3), "The current crawling technology is inadequate in dealing with such dynamic content, which one would encounter extensively on social media platforms." On this basis, we researched open source tools, using Google research tool and the sentence "How to archive social media," to capture documents on social media platforms. The results were derived from the blogs of IT experts, academic papers, and companies selling such services.

One of the academic papers discussed the preservation of Facebook content and proposed the use of an add-on: "We have developed a framework for capturing Facebook accounts for a personal archive via a browser extension and by third party web archivers" (McCown and Nelson, 2009, p. 3). Despite APO not being "personal," we also tested this tool.

For each of the social media platforms (figure 1), a tool was used to capture the records.

---

[4]    *Brazil http://www.siga.arquivonacional.gov.br/images/an_digital_completo_2010.pdf; USA http://www.archives.gov/records-mgmt/bulletins/201 4/2014-02.html; Australia http://www.naa.gov.au/records-management/agency/digital/socialmedia/social-media-policy.aspx; UK http://www.nationalarchives.gov.uk/documents/information-management/web-archive-tweets-and-videos.pdf.*

*Picture 1: APO on social media*

### 4.3.1 Facebook (https://www.facebook.com/AutoridadePublicaOlimpica)

*A- ArchiveFacebook – Add-on for Mozilla (https://addons.mozilla.org/pt-br/firefox/addon/archivefacebook/).*

To install the add-on is an easy task for anyone who understands how this kind of tool works. Once installed, it captured the content of the profile account (photos, posts, etc.) using the format Resource Description Framework file (.rdf).

The positive aspect of ArchiveFacebook is that it is a free tool that allows the producer to save content from institutional Facebook accounts directly to hard drive. The negative aspect is that a program is required to visualize the site downloaded in .rdf format.

*B- MIME HTML by Internet Explorer (IE)*

MIME HTML (.mht or .mhtml) is a web page archive format used to combine HTML code and its companion resources that are otherwise represented by external links in a single document (Wikipedia, 2017). For this case study, this comprised saving the APO Facebook page in .mht format.

The benefit was that user experiences were maintained during navigation with the same appearance of the on-line page and functional external links. A negative aspect was that its functionality worked best in the Internet Explorer environment.

*C- PDF extension*

To save a Facebook page in .pdf format is another way to maintain stable content and fixed form, corresponding to one of the characteristics of a record; in so doing, the posts and external links remained available for navigation. The negative aspect was that it did not have the dynamic environment of a social media platform.

*D- Print screen*

The print screen of APO's Facebook account in .png format maintained stable content and a fixed form for the daily post. A negative aspect was that it did not have the dynamic environment of a social media platform, i.e., the information was incomplete due to the inoperative external links.

### 4.3.2 Twitter ([https://twitter.com/apogovbr](https://twitter.com/apogovbr))

*A- Twitter archive (self-archiving)*

The Twitter platform includes a tool to archive the content of a user's account. This section is called *Twitter archive: Download your entire Tweet history* (*[https://blog.twitter.com/2012/your-twitter-archive](https://blog.twitter.com/2012/your-twitter-archive)*). To request the archives, the user must go to *Settings*, scroll down to the bottom of the page, and check the option to request a Twitter archive. The user will then receive an email with instructions on how to access the archive when it is ready for download. During processing, if the volume of data is too large, a notice will appear regarding how long it will take for the link to be delivered to the user's email: the time taken may range from a few minutes up to 24 hours. The positive aspect is that the tool is easy to use and includes a .csv format sheet with metadata about the tweets; however, it may take a day to receive the email.

In tests with *MIME HTML by Internet Explorer (IE)*, *PDF extension*, and *Print screen*, we obtained the same results as described in section 4.3.1 Facebook.

### 4.3.3 Instagram ([https://www.instagram.com/apogovbr/](https://www.instagram.com/apogovbr/))

*A- Instaport.me ([https://vibbi.com/instaport/](https://vibbi.com/instaport/))*

This is a free tool offered by the company Vibbi to back up Instagram photos and posts to a local hard drive.

The user must insert the username or URL account into a textbook to begin the download process. In this instance, the file format chosen for photos was .jpg. The positive aspect of this tool was its ease of use; the negative aspect was that it did not include user experience navigation and no descriptive metadata were embedded in photos; instead, each photo was identified by a general number.

In the tests with *MIME HTML by Internet Explorer (IE)*, *PDF extension*, and *Print screen*, we obtained the same results as those described in section 4.3.1 Facebook.

### 4.3.4 Flickr ([https://www.flickr.com/photos/apogov](https://www.flickr.com/photos/apogov))

*A- Bulkr ([http://getbulkr.com/](http://getbulkr.com/))*

This is a desktop application to back up, browse, and download photos and videos from Flickr. It was developed by an individual computer programmer and is available for free. In this study, we downloaded photos in .jpg format from APO's Flickr account.

To start the download, it is necessary to configure settings such as size, folder, and metadata (embed to photo or save as text file; in our study, the former was chosen).

A positive aspect is the short time for download and that secure access via user authorization is mandatory using the Yahoo! login account. Before authorizing user consent, a notice pops up asking for authorization to allow third parties to access the user's Flickr data. However, this authorization can be revoked. The negative aspect is

that there is no user experience navigation and no descriptive metadata embedded in photos.

In the tests with *MIME HTML by Internet Explorer (IE)*, *PDF extension*, and *Print screen*, we obtained the same results as described in section 4.3.1 Facebook.

### 4.3.5    YouTube (https://www.youtube.com/apogovbr)

*A- Download your data (self-archiving)*

On Google platforms, a tool is available to download data from all of the products offered by that company. On the *Download your data* page (https://takeout.google.com/settings/takeout), the user selects which product to archive – in this case, YouTube – and configures the settings. On the next page, options for the file type, archive size, and delivery method are required. In this study, we chose .zip format, 4MB, and *send download link via email*.

If the archive size is too large, the delivery could take from a few minutes to 24 hours. Once they have received the link, the user can download the archive to a local hard drive within a certain period. A positive aspect of this tool is the ease with which it can be managed, but there is no user experience navigation.

In the tests with *MIME HTML by Internet Explorer (IE)*, *PDF extension*, and *Print screen*, we obtained the same results as described in section 4.3.1 Facebook.

According to the researcher Sara Day Thomson: "Each platform presents its own obstacles to selection and curation, but all platforms with a networking element, from Twitter to Instagram to YouTube, pose difficulties for deciding the content that comprises a coherent collection and, more pragmatically, the content that can be effectively preserved within the legal, ethical, and technical framework in which it is collected" (Thomson, 2016).

For this case study, we used the same formats to download APO's Facebook, Instagram, Flickr, Twitter, and YouTube accounts as an option to minimize the difficulties and maximize the time required to capture those digital objects for long-term preservation. In further studies, we will analyze the pros and cons of this decision.

### 5    FINAL CONSIDERATIONS

This paper aimed to provoke ideas regarding the preservation of social media as records from the perspective of Archival Science. To recognize social media as a record demands a diplomatic analysis and classification scheme.

It is not feasible to collect and retain all content from social media platforms, which has become a dilemma in archiving this type of record. For this reason, public bodies must debate and include social media records in archival policies.

The ideal scenario for the treatment of social media as records would be archival policies plus an information system for record management and digital repositories based on archival requirements. The Olympic Public Authority was unable to develop

the ideal scenario, so an alternative strategy was established to avoid loss of records. Two of the four social media test-beds in this case study offered their own self-archiving tools for user´s download content – Twitter and YouTube. The others were developed by third parties. However, platform developers and individual computer programmers do not view social media as records, and therefore do not anticipate the need for an archival methodology and metadata.

As far as we know, IT development teams are not concerned about the preservation of long-term digital records, probably because this is not within their remit; thus, archivists must deal with this problem. But how? Herein resides the archivist's challenge and we do not yet have a clear answer. This case study thus comprises an attempt to find some feasible alternatives.

Based on the results of our experiences, the worst choice was to create a print screen from each social media platform as doing so did not capture the dynamic environment of a social media platform and the information was incomplete due to inoperative external links. For dynamic digital records, the strongest recommendation is to maintain the user´s navigation experience.

To date, the Application Platform Interface (API) is the one made available by Facebook/Instagram/Twitter/YouTube/Flickr platforms for developers to build non-automated apps and services, corresponding to an actual scenario of reusing and monetizing data. Thus, an effort to program plug-ins for open-source capture and preservation platforms could help archivists in their challenge to preserve social media records. Furthermore, information technology experts and technological equipment/tools are essential for new media tests.

For this essay, the APO team studied the manner of capturing and preserving documents from a social media platform as records, which means maintaining the archival characteristics of a digital record. The practical implications of the results were to conciliate the producer's theoretical and technical limitations, with digital repository configurations and Archival Science requirements. In further studies, issues related to privacy, copyright, and third-party tools are some of the topics that could be investigated.


## 6    ACKNOWLEDGEMENT

## SOURCES AND LITERATURE

- Brazil. National Defense Council. *Portaria nº 38, de 11 de junho de 2012. Homologa a Norma Complementar nº 15/IN01/DSIC/GSIPR* [Estabelecer diretrizes de Segurança da Informação e Comunicações para o uso das redes sociais, nos órgãos e entidades da Administração Pública Federal (APF), direta e indireta.]. Diário Oficial da União, nº 119, seção 1, p. 3, de 21/6/2012. Retrieved 08. 12. 2016 from: http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=21/06/2012&jornal=1&pagina=3&totalArquivos=68.

- Duranti, Luciana & Thibodeau, Ken. The concept of record in interactive, experiential and dynamic environments: The view of InterPARES (Appendix 2). In Luciana Duranti & R. Preston (Eds.), *International research on permanent authentic records in electronic systems (InterPARES) 2: experiential, interactive and dynamic records*. Pádova: Associazione Nacionale Archivistica Italiana (Anai), 2008. Retrieved 18. 12. 2016 from: http://www.interpares.org/display_file.cfm?doc=ip2_book_complete.pdf.

- Hockx-Yu, H. (2014). *Archiving social media in the context of non-print legal deposit*, IFLA WLIC Libraries, Citizens, Societies: Confluence for Knowledge in Lyon. Retrieved 10. 12. 2016 from: http://library.ifla.org/999/1/107-hockxyu-en.pdf.

- McCown, F. & Nelson, M. L. (2009). *What happens when Facebook is gone?* JCDL'09, June 15–19, 2009, Austin, Texas, USA. Retrieved 14. 12. 2016 from: http://www.harding.edu/fmccown/pubs/archiving-facebook-jcdl2009.pdf.

- Thomson, S. D. (2016). *Preserving social media*. Digital Preservation Coalition (DPC) Technology Watch Report 16-01 February 2016. ISSN: 2048-7916. Retrieved 17. 12. 2016 from: http://dx.doi.org/10.7207/twr16-01.

- Wikipedia. *MHTML: MIME HTML*. Retrieved 17. 12. 2016 from: https://en.wikipedia.org/wiki/MHTML.

## POVZETEK

## DRUŽBENI MEDIJI KOT ZAPISI V KONTEKSTU OLIMPIJSKEGA JAVNEGA ZAVODA IN NJIHOVA HRAMBA: ŠTUDIJA PRIMERA

Ljudje po svetu uporabljajo družbene medije kot način ohranjanja stikov s prijatelji in družino, razpravljanja o aktualnih temah, širjenja idej, sledenja poslovanju podjetij itd.

Za telesa javne uprave pomeni uporaba družbenih medijev možnost ostajati blizu populaciji, ki ji služijo; vendar, kako ta telesa upravljajo z družbenimi mediji kot dokumenti, ki odražajo njihove aktivnosti? Kako lahko družbene medije ohranimo za daljši čas? Kaj narediti, če organizacija nima idealnih orodij za upravljanje z zapisi?

Idealen scenarij za upravljanje z zapisi družbenih medijev bi bil, da bi se ustvaril informacijski sistem za upravljanje z zapisi in digitalnimi repozitoriji, ki bi slonel na arhivskih zahtevah. Če organizacija nima možnosti, da bi razvila tak sistem, bi morala, v izogib izgubi zapisov, najti alternativno strategijo – to pa je tudi namen te študije.

Študija primera je bila opravljena v Olimpijskem javnem zavodu (APO), brazilskem javnem konzorciju, ki je bil ustanovljen na podlagi Zveznega zakona št. 12.396 z dne 21. marca 2011 kot zagotovilo Mednarodnemu olimpijskemu komiteju za uspešno izvedbo olimpijskih in paraolimpijskih iger leta 2016 v Riu de Janeiru. Naloge APO so zajemale koordinacijo med Zvezno vlado, vlado zvezne države Rio de Janeiro in mestne občine Rio de Janeiro v zvezi s pripravo in izvedbo prvih olimpijskih iger v Južni Ameriki.

Za ta prispevek so bili dokumenti družbenih medijev, kot so Facebook, Instagram, Flickr, Twitter in YouTube, prepoznani kot zapisi, ki odsevajo javne aktivnosti APO in bi tako morali biti ohranjeni. Novi mediji so arhivistom in upravljalcem z dokumenti predstavljali velik izziv pri iskanju najboljše metodologije in tehničnih orodij za zajem in hrambo digitalnih zapisov.