

ESSACHESS - Journal for Communication Studies

The Technical Infrastructure of Cultural Initiatives on Wikimedia: Three Case Studies From Brazil

Éder Porto ALVES Wiki Movimento Brasil BRAZIL e-mail: eder.porto@wmnobrasil.org

Paul R. BURLEY Northwestern University Libraries UNITED STATES e-mail: p-burley@northwestern.edu

Solange Ferraz de LIMA Universidade de São Paulo BRAZIL e-mail: sflima@usp.br

Michael William PEEL Instituto de Astrofísica de Canarias, SPAIN SPAIN e-mail: email@mikepeel.net

João Alexandre PESCHANSKI Faculdade Cásper Líbero, BRAZIL BRAZIL

Article received on the. Article accepted on the. Conflict of Interest: The author(s) declare(s) no conflict of interest.

ESSACHESS – Journal for Communication Studies Volume X Issue X(XX), p. 00-00 © The Author(s) XXXX Reprints and Permission: © ESSACHESS https://www.essachess.com/ DOI: 10.21409/essachess.1775-352x

e-mail: japeschanski@casperlibero.edu.br

Abstract: We use the concept of *cultural movements* as a theoretical basis to reflect on the sociotechnical infrastructure of three Wikimedia initiatives in Brazil: the photography contest Wiki Loves Monuments, the National Museum Cross-Wiki Project and the Museu Paulista GLAM-Wiki partnership. We analyse how they were organised and how their workflow is coordinated in a collaborative environment. We also present their strategies for the preservation of cultural heritage collections and for the innovation in content dissemination. The evidence points to the relevance of Wikidata—a collaborative, open database—for structuring information and optimising coordination and crowdsourcing. It also points to the importance of establishing mechanisms for shared curatorship between the Wikimedia community and cultural institutions.

Keywords: Wikimedia; cultural initiatives; technical infrastructure; digital preservation; digital dissemination

L'infrastructure technique des initiatives culturelles sur Wikimédia: trois études de cas au Brésil

Résumé : À partir de la base théorique des *mouvements culturels*, nous présentons l'infrastructure technique de trois initiatives de Wikimédia au Brésil: la compétition de photographie Wiki Loves Monuments, le Projet Inter-Wiki du Musée national et la coopération GLAM-Wiki avec le Musée Paulista. Nous analysons comment ces initiatives sont organisées et comment leurs activités ont été coordonnées dans un cadre collaboratif. Nous présentons aussi leurs stratégies de préservation de collections culturelles et d'innovation pour la diffusion des informations. Les résultats indiquent l'importance de Wikidata, une base de données ouverte collaborative, pour structurer les informations et perfectionner la coordination et le travail collectif. Ils indiquent aussi l'importance de développer des mécanismes curatoriaux partagés entre la communauté de Wikimédiens et les institutions culturelles.

Mots-clés: Wikimédia; initiatives culturelles; infrastructure technique; préservation digitale; diffusion digitale

Introduction

Cyberspace, networked society, and collective intelligence are sociological categories that are of particular relevance at the start of the 21st century (Castells, 2001, 2013; Lévy, 2007). These categories help explain how the internet has evolved as a global communication system that is capable of substantially changing

the systems of social relationships and, especially, contemporary social and cultural movements. According to Castells (2001), current cultural movements are "movements in defence of specific ways of life" that are not only supported by communications systems but are fundamentally structured by these systems, the most important of which is the internet.

The global reach of the internet has relied on a shared cultural ground as well as the dissemination of social mobilisation strategies. It has weakened national boundaries and shaken rigid and vertical organisations that hitherto prevailed in culture (Castells, 2001). Pierre Lévy (2007) conceptualised this as a "collective intelligence" when documenting this unprecedented form of global sharing of values, worldviews, and knowledge. However, no phenomenon is exclusively dependent on cyberspace. Castells claims that the strength of contemporary social and cultural movements comes from the intertwining of network communication activities with social movements in tangible places (Castells, 2001). In other words, influential cultural and social movements will emerge from the intertwining of local contexts and digital strategies, which will eventually lead to global impact.

Movements for the preservation of material and immaterial heritage are included in the above perspective. For instance, Meneses et al. (2006) claim that the "world heritage" concept created by the United Nations Educational, Scientific and Cultural Organization (UNESCO) is only able to serve as an effective means of cultural preservation for monuments and immaterial practices when it also has meaning for local communities. In other words, cultural heritage must not limit itself to formal recognition, such as an identifier allocated by international agencies with a preservation mission. It must be effective locally, and relevant to the actual individuals and groups that frame their identities in relation to this heritage.

In this sense, the Wikimedia movement has become a cultural collective agent with a global reach that voluntarily and independently values, fosters and provides capacity for sharing knowledge. The mission of this movement is to empower and engage people around the world to collect knowledge and disseminate it under a free copyright licence. Knowledge in this movement is defined from the perspective of the potential contributors. Based on Castells' definitions discussed above, it is possible to consider, therefore, that the growth capacity of the Wikimedia movement is directly related to its capacity of sharing common values across stakeholders and its embeddedness in local contexts. This is because contributors rely on digital tools for producing and disseminating information that matters to their identity, system of social relationships and material cultural background. The socio-technical infrastructure that is needed to make a cultural movement of this sort of work is of key interest, as what we have here is to some extent a cybernetic community (Lévy, 2007) that coordinates, organises and finally acts to share cultural information.

Two of the three case studies in this article focus on material cultural heritage protected by national legislation: Museu Paulista and Museu Nacional. Both share trajectories linked to the development of local memories; to some extent they are "places of memory" (Nora, 1992, 1993; Brefe, 2003) firmly grounded in the Brazilian national imaginaire. They additionally act as centres that produce knowledge and higher education, since they are linked to prominent Brazilian public universities (Schwarcz, 1989; Knauss, 2018; V. C. de Carvalho et al., 2021). Both also share, unfortunately, the consequences of instability in public policies for the protection of material and immaterial heritage in Brazil. From the perspective of digital preservation, the Wikimedia initiatives developed in the context of these museums emerge and evolve as a web of stakeholders and bear an activist strategy —in the sense of an actor in a cultural movement that is working to protect and disseminate digital records of cultural items in their collections—combining various local actors, including Wikimedians, researchers, professors, students and the public who frequent these spaces of memory and knowledge with digital recovery strategies and valuing the preserved collections of both institutions.

The other case that is presented here—Wiki Loves Monuments (WLM)—brings an even more challenging action, as it is run without a connection with an institution as in the previous cases. Here, the project depends on the engagement and coordination of local agents to achieve results that are shared on a global scale.

What we investigate from these cases is specifically how the sociotechnical infrastructure of the cultural movement they manifest has emerged and evolved. This line of investigation is especially applicable in the context of the Wikimedia movement as the whole initiative is collaborative, thus technical decisions are crowdsourced. This investigation also provides an opportunity to move from abstract to practical discussions on digital cultural movements and their work around preservation and dissemination of cultural heritage. Questions we have asked ourselves as we have investigated these cases include:

- What is the optimised infrastructure for cultural movements on Wikimedia?
- How does the development of processes and resources for cultural preservation and dissemination in a collaborative digital environment happen?

This article focuses on presenting the technical infrastructure behind three cultural initiatives on the Wikimedia projects—a growing ecosystem of free and collaborative knowledge on the Web. All three cases presented are directed towards the discussion and investigation of what is the optimal infrastructure of cultural initiatives on this ecosystem and how their processes and tools have been collaboratively developed.

1. National Museum

1.1 Brief history

The National Museum is one of the oldest cultural institutions in Brazil. It was established in 1818 with educational and scientific dissemination objectives. Initially called the Royal Museum, it is located in Rio de Janeiro, then the capital of the country. After Brazil became a Republic in 1889, the museum adopted its current name and was moved to the Paço de São Cristóvão, a historical building that served as the residence of the Portuguese Royal Family. In 1964, the National Museum was integrated into the Federal University of Rio de Janeiro (Pires, 2017).

The National Museum held one of the largest collections of natural history and anthropology in the world, with around 20 million artefacts, as well as one of the most comprehensive scientific libraries in Brazil. It also played a major role in advancing Brazilian science as it also functioned as a research institution. The museum notably organised scientific expeditions beginning in the 19th century; this resulted in discoveries across a range of scientific fields that were subsequently documented in research literature and sometimes exhibited at the museum (Sá et al., 2018).

Disaster struck the National Museum of Brazil on September 2, 2018: a massive fire devastated the building and its collections. The catastrophe occurred amidst an ongoing crisis in the cultural sector in Brazil, with shrinking investment for public institutions and disorganisation of the national museum agency (Sá et al., 2018). It was estimated that the fire damaged around three-quarters of the museum collection (Motta & Silva, 2020). Digital projects emerged in this context to continue the legacy of the collection, notably the National Museum Cross-Wiki Project, hosted on the Wikipedia in Portuguese.1

1.2 Context and goals

Digital preservation and dissemination have emerged as increasingly significant topics in the literature on disaster preparedness and management for cultural institutions (Dearborn & Meister, 2017; Idiegbeyan-Ose et al., 2018; Rachman & Afidhan, 2018; Rinehart et al., 2014). The Wikimedia response to the National Museum fire in 2018 has been mentioned as an exemplary model of digital crowdsourcing in response to a cultural disaster (Kumar, 2019, 2020; Motta & Silva, 2020; Tosun & Bostan, 2021). No prior research has focused on the National Museum Cross-Wiki Project collaborative technical infrastructure.

¹ https://pt.wikipedia.org/wiki/Wikipédia:Projetos/Museu_Nacional

The National Museum Cross-Wiki Project emerged spontaneously from within the Wikimedia community after the fire. No institutional partnership between the Brazilian Wikimedia groups and this cultural institution existed in 2018. Edits on Wikipedia in Portuguese about the National Museum fire started minutes after the first reports on TV; a Wikipedia article on the disaster was created within two hours; articles on the fire existed in over 20 languages within ten days (Peschanski, 2018). The cross-wiki project was established on September 22, 2018 to better coordinate contributions across languages and Wikimedia projects.

The goals of the National Museum Cross-Wiki Project were twofold. Firstly, there was a multilingual call for media uploads on the National Museum's collection to Wikimedia Commons. This call was disseminated on banners on the top of Wikimedia projects and through various social media channels. It sparked "tremendous attention from the public" (Kumar, 2019). Secondly, the project aimed to coordinate Wikimedians to curate and create content on the National Museum. Curation involved triaging and using uploaded media files. Data curation associated with media uploads was referred to as "data archeology" (Evenstein, 2019), as editors gathered to match metadata and images. Content creation was undertaken on: Wikipedia in several languages, chiefly Portuguese, Catalan, English, German, Spanish, and French; Wikimedia Commons; Wikisource in Portuguese and French; and Wikidata (Motta & Silva, 2020). A notable action was scraping and uploading metadata and media files available on the National Museum websites. Rare books that were previously digitised by the institution were given special attention.

1.3 Technical infrastructure

To realise the outreach and coordination objectives of the Wikipedia project, a landing page was created. Its main goal was to provide a reference point to Wikimedians and the general public and to coordinate activities related to the cross-wiki project. The landing page was created as a content block structure similar to the National Archives Galleries, Libraries, Archives, and Museums (GLAM)-Wiki landing page and shortly after became TGLAM (meaning "template GLAM"), a semi-automated landing page generator used by the majority of the Brazilian GLAM-Wiki and Wikiprojects initiatives (Araújo & Knipel, 2018; Peschanski et al., 2018). The landing page presented the context of the project, along with links to activities pages focused on different Wikimedia Projects. These included Wikipedia, Wikidata, Wikimedia Commons and Wikisource.

Wikipedia activities focused on creating articles in multiple languages. A module called BRAtable (a contraction of "Brazilian table") was created by Turíbio Branco to build a wikitable and list the articles in multiple Wikipedias according to their Wikidata items.2 Listeria was used in articles about the exhibition rooms or

² https://pt.wikipedia.org/wiki/Módulo:BRAtable

collections to automatically list items that had been located in those rooms or were part of the collections before the fire. As the images were being uploaded and items were being created, those lists continued to be updated dynamically3.

The Wikimedia Commons activities consisted of categorising media files uploaded through the campaign set for the project. Users were asked to try to identify the objects photographed and where in the museum they had been located (Motta & Silva, 2020). The mapping of the museum collections and rooms and the correct categorization and identification of its contents allowed the articles being created in multiple Wikipedia to be enriched with images. Another project that gained from this work was Wikisource, a sister Wikimedia project responsible for holding free-content textual sources. With the correct curation of the uploaded files, activities on this project could be organised, and revolved around transcribing the documents and rare books in the public domain that were scraped from the National Museum websites.

Scraping metadata and media files were a relevant action throughout the whole project. All the metadata about the objects, rooms, collections, etc. were harvested from various websites and files using and adapting *ad hoc* scripts in Python language and uploaded to Wikidata. This upload was made using QuickStatements, a tool created by Magnus Manske for batch editing and creating Wikidata items4. The inherently multilingual aspect of Wikidata and its structured content (Kaffee et al., 2017) enabled the meaningful engagement of contributors who do not speak Portuguese, but also allowed the coordination of the content being created and improved in several languages.

One of the tools used for improvement of the content uploaded to Wikidata itself was also developed by Magnus Manske and is called TABernacle5. This tool creates a dynamic table of contents with different metadata columns so users can help improve statements, labels and descriptions of Wikidata items. Another tool taking advantage of the Wikidata structure is the aforementioned BRAtable, which relies on Wikidata information to build its tables.

Although it is desired that another event like this never occur again, cultural institutions, particularly public ones neglected or underfunded by the government, are open to such catastrophes. In the case of such an occurrence, though, it is important that a well-documented and flexible plan be developed and improved from past experiences, and emergency protocols be set in place to call volunteers to action and the general public to contribute and salvage as much as possible the sum of human knowledge safeguarded by cultural institutions.

³ https://pt.wikipedia.org/wiki/Sala_Kumbukumbu

⁴ https://quickstatements.toolforge.org

⁵ https://tabernacle.toolforge.org

1.4 Numerical impact and results

The number of files in the category of the Museu Nacional at Wikimedia Commons before the fire was 1822, uploaded by 47 users. After the tragedy, 2240 files were uploaded by 129 users, totaling 4062 files in the category. 84 (or 65%) of the 129 uploaders made their first contribution to Wikimedia Commons after the fire.

Figure 1 shows the number of views of the files in the category before and after the fire. It is clear in the graph that the number of views of files related to the museum peaked in September 2018 with more than 156 million views, and rapidly decreased, as focus on the event faded (although the views stayed higher after the event than they were before it). This reinforces that action needs to be taken quickly.

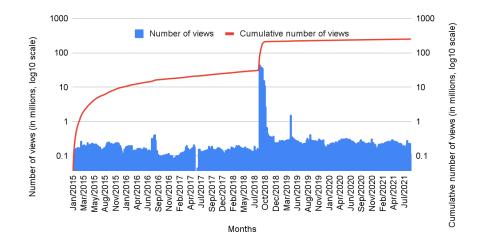


Figure 1. Number of views of files in the category Museu Nacional, Rio de Janeiro, from January 2015 to July 2021.

1.5 Digital dissemination circuit

Bringing an object to Wikimedia is not simply uploading content to a drive; the object is brought into an open media ecosystem that in theory contributes to crowdsourced curatorship and content dissemination. The Wikimedia semantic and convergent environment is illustrated on Figure 2, having as an example a scientific

report on the transportation of the Bendegó Meteorite to the Museu Nacional (J. C. de Carvalho, 1888). It is not publicly known yet if this report was destroyed during the 2018 National Museum fire and in any case, it has been digitally reborn as a central element in the example web.

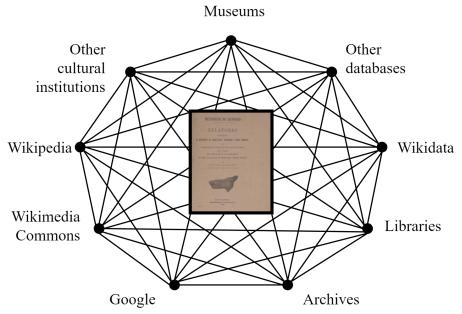


Figure 2. *Model of digital convergence illustrating the Meteorito de Bendegó report case.*

The process to bring the Bendegó report to Wikimedia involved two steps. Firstly, a Wikidata item was created, which included descriptive statements on the publication and became a central hub to connect content related to the report across Wikimedia projects. The item alone now suffices for easily using the publication as a reference on Wikipedia, relying on a template called Cite Q6. In parallel, a PDF file that was scraped from the National Museum digital library was uploaded to Wikimedia Commons. This file was matched to the Wikidata item identifier, and the Optical Character Recognition (OCR)-friendly format made it easily convertible to Wikisource.

⁶ https://pt.wikipedia.org/wiki/Predefinição:Citar_Q

Wikidata and Wikimedia Commons provided the basis upon which dissemination on other Wikimedia projects happened. Wikimedians created articles on the report in four Wikipedias. Images that were extracted from the report were used in almost 100 pages. Contributors fully transcribed the report on Wikisource in both Portuguese and French. This framework ultimately provided an optimal structure for community contributions and dissemination.

2. Museu Paulista

2.1 Context and goals

The Museu Paulista da USP is the oldest public museum of São Paulo state. It was created in 1893 by the state government and was integrated into the State University of São Paulo (USP), a public institution in 1963. As a university museum, it is specialised in History and Material Culture and preserved collections of arts and also collections originated from Brazilian domestic and public working spaces (tools of crafts such as carpentry, tailoring, shoemaking, typography). The museum is built in the eclectic style of the late 19th century, inspired by Italian Renaissance palatial architecture. In 2013, the building was closed to the general public for a large-scale restoration project; teams and collections were temporarily transferred to other buildings.

The GLAM-Wiki initiative with Museu Paulista is the result of an official partnership between the museum direction and the Wikimedia affiliate in Brazil. This partnership was ratified on July 25, 2017. In February 2017, an internal document of the museum stated the goal of the partnership, stating that Wikimedia "works as an important tool for producing and disseminating content, which will lead to greater visibility of the Museu Paulista collection" (Lima, 2017).

The objective was to expand the quantity and quality of content related to the collection and research of the Museu Paulista on the Wikimedia projects. The consideration was that, at the beginning of the partnership, content on the institution's central themes was poorly covered on Wikimedia and were critically absent in the Portuguese language Wikipedia (Peschanski, 2021). Unlike other GLAM-Wiki initiatives, which are centred on disseminating images from the collections, the case of the Museu Paulista was aimed at improving the entire ecosystem of knowledge in a broader way.

As the museum was closed during the time in which the partnership was established, the GLAM-Wiki initiative became an opportunity for "opening up the collection of a closed museum" (Azzellini & Peschanski, 2020).

2.2 Technical infrastructure

The technical infrastructure of the Museu Paulista GLAM-Wiki initiative may be divided into two phases. Firstly, from 2017 and 2019, there was mostly a focus on bringing media files and corresponding metadata to Wikimedia projects. As described in Peschanski (2021), this process involved a sequence of steps: data refinement and reconciliation, Wikidata mass editing, and Wikimedia Commons batch uploads. Then, some work was done to improve content on Wikipedia based on semi-automated technologies for producing artwork lists and entries.

A second phase was launched in 2020, with a specific emphasis on fostering a culture of digital collaboration between the museum and the Wikimedia Movement. This phase was organised as two formal projects, respectively called "The University of São Paulo Museu Paulista Wikipedia Initiative" (February to November, 2020) and "The Museu Paulista Wiki-Dissemination Initiative and Collaborative Digital Training" (February to July, 2021). These projects comprised several types of activities, including Wikimedia contests, edit-a-thons, tool development, and technical training (Alves, Burley, et al., 2021). The idea was that the GLAM-Wiki technical infrastructure needed to be embedded in a renewed strategic direction and processes and the museum staff.

A "Guide to the creation of a digital dissemination strategy for the Museu Paulista" was released in the context of the 2020 Wikimedia project in which a vision for the museum technical infrastructure was laid out (Wiki Movimento Brasil, 2020). This document stated the need "to consider infrastructure as an ongoing process, not as a mere acquisition and use of technological resources", as "Infrastructure is understood here as a process of reflection, investment and engagement with digital technologies and audiences". Engaging with technical infrastructure decisions, including Wikimedia processes, was described as a key element in achieving the museum mission.

The 2020 Digital Dissemination Strategy Guide led in the following year to the organisation of a course for the museum staff: "The Museu Paulista in the Digital Culture", hosted on Wikiversity in Portuguese (Wiki Movimento Brasil, 2021). The course goals included "To provide a theoretical and practical framework on new ways of social relationships on the Internet, helping to build a digital culture in the institution and with its audiences" and had a special focus on open-knowledge and collaborative infrastructure, especially Wikimedia projects.

The Museu Paulista GLAM-Wiki evolved in the sense of organically connecting Wikimedia and museum practices. This involved, in both 2020 and 2021, the participation of researchers from the Museu Paulista in conducting Wikimedia

dissemination activities such as webinars and edit-a-thons. The symbiotic strategy of Museu Paulista with Wikimedia also guided the development of technical infrastructure, seeking ways to establish technologies in which crowdsourcing was thought to act directly in niches relevant to the museum and think of ways to roundtrip Wikimedia contributions, principally on Wikidata, to the local database of the museum, after undergoing curation and validation processes.

2.3 Numerical impact and results of the initiative

Media and metadata from around 33,000 collection items of the Museu Paulista have been shared on Wikimedia projects since 2017. Image uploads have mostly occurred in batches, with peaks in late 2018, mid 2019, early 2020 and early 20211. Figure 3 shows the number of views of the files in the Museu Paulista GLAM-Wiki category on Wikimedia Commons across the years.

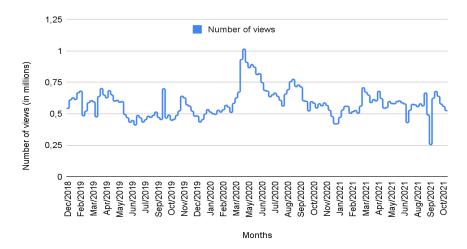


Figure 3. Number of views of files in the Museu Paulista GLAM-Wiki, from December 2018 to October 2021. Source: GLAM-Wiki Dashboard.

Content on Wikipedia on topics of interest related to the Museu Paulista has increased markedly since the GLAM-Wiki ratification, as shown in Figure 4. Peaks are associated with the deployment of tools that semi-automated artwork lists and contests to improve Wikipedia entries.

¹ https://glamwikidashboard.org/MPUSP/user-contributions

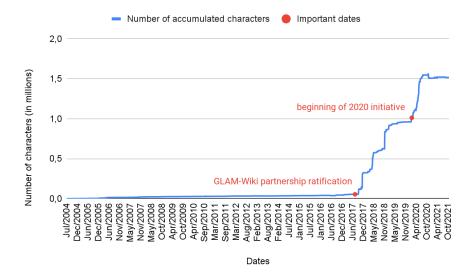


Figure 4. Evolution of the size of the articles related to the Museu Paulista on Wikipedia in Portuguese.

There is no global tool to assess how much metadata that was contributed by the museum has been improved on Wikidata. A test case was items in the Alberto Santos Dumont collection, a 1,700-item set in the Museu Paulista connected to the Brazilian inventor. From a manual check, there were 2957 statements added and 770 edited from December 2018 until June 2021.

2.4 Shared curatorship, a future challenge

The case reported here is part of an increasing investment in collaborative digital actions in the Museu Paulista during the COVID-19 pandemic. These actions included a plan for the interoperability between the museum local database and Wikidata, and a set of tools for the crowdsourced curation of metadata from the museum collection. The former has led to a partnership involving the Museu Paulista, two public universities—Universidade Federal de Brasília (UnB) and Universidade de São Paulo (USP)—and the national Wikimedia affiliate. The interoperability framework remains undeveloped and is currently being worked on in the context of a project funded by FAPESP, a state research agency, and results are expected to be available in 2024. Its concept, however, was presented publicly in two panels in which cultural institutions from North America and Europe have also presented their strategies to roundtrip metadata to and from Wikidata (Alves, Lih, et al., 2021; Martins, 2021).

The second action, the crowdsourced metadata curation tools, has been aimed at creating interactive and playful ways to engage Wikimedian volunteers in the process of describing and identifying objects and images content from the Museu Paulista collection. A set of online apps, called Wiki Museu do Ipiranga, have been developed and include a tool for identifying brands of historical toys and other objects2, describing outfits in visual arts3 and improving the description of heraldic elements on traditional dishware4. Contributions lead to edits to Wikidata items, which may subsequently be brought to the local museum database. Metadata curation is normally a task assigned to internal teams, and the metadata apps have been an infrastructure of shared curatorship.

The vision embedded in these actions rests upon the understanding of culturalheritage preservation in the digital environment as both active and engaging. Involving volunteers in digital tasks of description and identification of collection items could result in two desirable situations: the awareness about the complex research works curation requires and the sense of belonging. It is a practical and activist strategy of realising the ICOM definition of a museum: they must serve society.

3. Wiki Loves Monuments

3.1. Overview and context

WLM began as a pilot project in the Netherlands in 2010. A major aim is to significantly increase the number of freely licensed photos of built cultural heritage that are available on Wikimedia Commons and are also used to illustrate Wikipedia articles. It self-describes itself as an "international photo contest for monuments" or "(re)discover the cultural, historical, or scientific significance of their neighbourhood"1. The 2012 edition of the competition, coordinated by Vereniging Wikimedia Nederland, was recognized by a Guinness World Record as the "largest photography competition", with 353,768 entries2. As of 2021, the competition has gathered over 1.7 million images of 1.5 million monuments, submitted by over 60,000 participants.

The competition has a federated model and consists of two layers: national competitions and an international competition. National competitions are organised by national level organising committees, and follow the rules of the international

1 https://www.wikilovesmonuments.org/contest

2https://www.guinnessworldrecords.com/world-records/largest-photography-competition

² https://wikimarcas.toolforge.org

³ https://wikiroupas.toolforge.org

⁴ https://wikibrasoes.toolforge.org

competition. Lists of monuments are gathered from national and regional heritage institutions. National competitions follow the same general rules: images must be self-taken; self-uploaded; uploaded within a specified time period, typically the month of September or October; freely licensed; and containing an identified monument. The international competition gathers up to ten submissions from each country, and an international jury selects international winners from that pool. The prizes awarded at the national and international level vary; first prize at the international competition in 2021 was €1500.

WLM can be seen as part of the history of selection and documentation of monuments of historic or cultural interest. In the Brazilian context, WLM is rooted in the history of both architectural preservation and its later formal documentation in the country. The registry of historic monuments in Brazil began in the early 20th century and culminated with the establishment of the Serviço do Patrimônio Histórico e Artístico Nacional (SPHAN, now IPHAN). The work of IPHAN was amplified by the passage of the International Charter for the Conservation and Restoration of Monuments and Sites, commonly known as the Venice Charter, in 1964, and subsequently codified by the International Council on Monuments and Sites (ICOMOS) in the late 1960s. Beyond its role in asserting norms of conservation, restoration, and archaeological excavation, the charter states that "In all works of preservation, restoration or excavation, there should always be precise documentation in the form of analytical and critical reports, illustrated with drawings and photographs." (International Council on Monuments and Sites, 1964; Azevedo, 1987).

Detailed inventories of monuments, inspired by the Venice Charter, included a set of elements to describe a monument: its name, a brief summary of its history and the structure, and photographs of both the exterior, interior, and at times, the monument in situ. A full inventory of historic sites was first carried out in Bahia beginning in 1973 and published in seven volumes as the *Inventário de Proteção do Acervo Cultural* (Azevedo & Lima, 1975); subsequent inventories were only carried out in Minas Gerais and Pernambuco. (Guedes, 1987). This reflects the stark disparity between regions of Brazil: wealthier states in the country produce lists and related information freely on the internet, while less wealthier states, notably in the Northeast and North, lack state-level lists. Lists of designated monuments at the municipal level in Brazil are rarely available on the internet.

WLM can broadly be seen as part of a global movement to catalogue heritage sites but expanding it to have wider general public contributions through the internet. It has been investigated as a case study of participatory movement for cultural preservation and dissemination (Posada et al., 2012; Magrini, 2018). WLM started in Brazil in 2015 and is part of a general drive to improve content about Brazil in the Wikimedia projects. Significant work has been invested in collecting

data and structuring lists of monuments in Brazil, since these are not easily available either offline or online, and also to improve the infrastructure used on the Wikimedia projects.

3.2 Technical infrastructure

Since the first edition of the contest in Brazil in 2015, contestants have used a step-by-step tool to upload photographs to Wikimedia Commons called UploadWizard3. An internationalised UploadCampaign overlay is implemented to pre-fill some fields for the uploader and simplify the upload process. Other tools can be used as well, like Commonist4 or Pattypan5, but they require more knowledge of wiki markup and thus are less useful for new editors. This structure comes from the international effort to facilitate the work of local organisers.

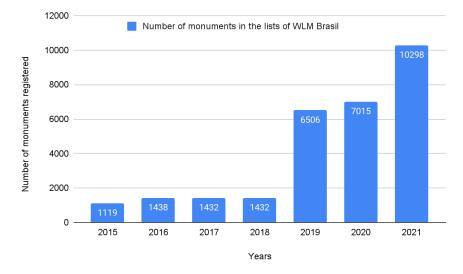


Figure 5. The number of monuments listed within the WLM Brasil competition from 2015 to 2021.

The upload process starts with users accessing pages filled with lists of monuments, which were manually created and curated until 2019. From 2015 to 2018, the landing page was constructed based on tabs with different information and the monuments lists were manually maintained with little improvement of the quantity and illustration of the monuments. Since 2019, this structure has been simplified into a single landing page and the lists are generated based on Wikidata

³ https://commons.wikimedia.org/wiki/Commons:Upload_Wizard

⁴ https://commons.wikimedia.org/wiki/Commons:Commonist

⁵ https://commons.wikimedia.org/wiki/Commons:Pattypan

metadata through the tool Listeria by Magnus Manske6. This tool, based on a query written in Simple Protocol and RDF Query Language (SPARQL), fetches every monument in a particular state or municipality and writes a table with its metadata. This allows the information about monuments to be updated in a structured format directly in Wikidata, and also to add new monuments as new entries in Wikidata, with the lists of monuments automatically updating. This significantly reduces the maintenance burden and makes them easier to expand. Figure 5 shows the number of monuments in Brazil present in the lists for each year from 2015 to 2021.

It is difficult to find official lists of Brazilian monuments outside of those of well-funded cultural heritage and governmental institutions. In 2019, the organisers of Wiki Loves Monuments began to import metadata on monuments by jurisdiction: international (e.g., UNESCO, or Heritage of Portuguese Influence, HPIP), national (i.e., IPHAN), state (e.g., Espírito Santo State Council of Culture - SECULT-ES), and municipality (e.g., Conselho Municipal de Preservação do Patrimônio Histórico, Cultural e Ambiental da Cidade de São Paulo - CONPRESP). Most of the content of the corresponding websites was contained within Portable Document Format (PDF) files or HyperText Markup Language (HTML) pages with a minimal set of metadata. To retrieve this information, scripts were written in Python language ad hoc for each website, as there is no standardised format in which the information is presented. The most-used website for this step was iPatrimônio7, a website which federates metadata from thousands of Brazilian monuments at the federal, state, and municipal levels. The most common Python package used to extract content from the websites was called BeautifulSoup, which makes it easier for the programmer to navigate in the HTML elements of the page. (Thota & Elmasri, 2021) The metadata harvested from the websites was then cleaned of any obvious inconsistencies and formatted to be uploaded to Wikidata using QuickStatements.

Images are tracked on Commons using templates for individual authority controls to start with, and later using the Wikidata Q-identifier (QID) with {{MonumentID}}8 included in the file pages in Brazil. This new template is used to link the photograph of the monument with the Wikidata item using the QID code; they can then be linked with different institutional identifiers via the Wikidata item. This generalising of the identifier code to use Wikidata both makes it easier to handle cases where monuments have multiple IDs from different organisations, and for cases where the monument does not have a specific official ID but is nevertheless included in lists of monuments or bulletins recording newly-listed monuments These, as well as the main WLM template, include the files in the competition categories, as well as displaying a main photograph of the monument

⁶ http://magnusmanske.de/wordpress/?p=301

⁷ https://www.ipatrimonio.org

⁸ https://commons.wikimedia.org/wiki/Template:MonumentID

(where present on Wikidata—it prompts for a photo to be added to Wikidata if not9, as well as listing the monument's various IDs and including the corresponding tracking categories for them.

The national jury for the contest is selected by local organisers and usually includes a diverse set of volunteers, such as photographers, journalists and wikimedia editors. The process of voting an image occurs in a tool called Montage10, in at least three rounds of yes/no selections to include or exclude the images, followed by rating and ranking rounds where the scores are averaged over multiple jury members' votes to decide which images go on to the next stage, or are ultimately selected as winning images (best quality image and best contributors).

The increase of listed monuments and the expansion of the properties used in the lists has helped to fully register Brazilian cultural heritage sites and minimise the existing bias in terms of the photographic coverage of monuments across different regions. That, combined with a solid outreach plan for both on wiki and social media platforms has the potential to improve the quality of information of the monuments as well as help illustrate the thousands of listed monuments without images.

By adding the image to the Wikidata item, the image is automatically used in the monument lists, and in Wikipedia articles that use Wikidata-enabled infoboxes. Where a Wikimedia Commons category for the monument exists, the image is also shown in the Wikidata infobox there. Additionally, the colour of the monument's item link is updated in geographically based tools such as Wikishootme11 to indicate that this monument already has photos, enabling better discovery of unphotographed monuments in the area. Finally, it makes it easier for search engine tools to find the picture for relevant search terms, given that Wikimedia projects typically have high google rankings.

3.3 Numerical impact and results

⁹ https://commons.wikimedia.org/wiki/Category:Uses_of_MonumentID_with_no_pi cture_on_Wikidata

¹⁰ https://commons.wikimedia.org/wiki/Commons:Montage

¹¹ https://wikishootme.toolforge.org

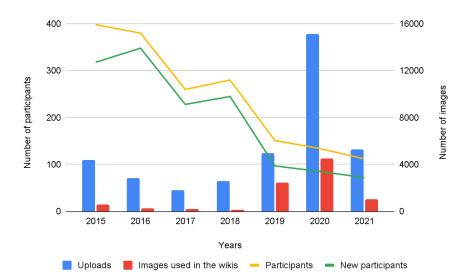


Figure 6. Number of uploads and participants in WLM Brasil, from 2015 to 2021. Source: https://wikiloves.toolforge.org/country/Brazil

As seen in Figure 6, the number of images contributed to the WLM Brasil from 2015 to 2017 followed a relatively consistent pattern of reduced uploads and image usage, in 2018 this recovered partially in terms of images uploaded. In 2019, with the diversification and extension of the lists, there was a record of around 5,000 images and in 2020, there were more than 15,000 images—despite the Covid-19 pandemic. The usage of the images also dramatically changed: in 2019 more than 2,400 images (49%) and in 2020 a little more than 4,500 images (29%) illustrated a monument on at least one Wikimedia project (including Wikidata). Figure 6 also shows the decreasing number of new and existing participants and participants in the contest over the years, particularly due to the COVID-19 pandemic.

Figure 7 shows the number of monuments depicted by the participants of the contest. With the expansion of the list in 2019, it is possible to see an increase in the number of monuments depicted. In 2020, there were 417 monuments depicted for the first time on Wikimedia Commons, and in 2021 this number went up to 636 monuments depicted for the first time, out of 925 photographed in total.

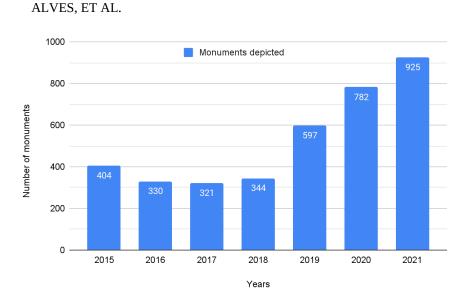


Figure 7. *The number of monuments depicted over time. 'Monuments depicted' means the number of monuments with new photographs uploaded in that year.*

3.4 Future of WLM in Brazil

00

WLM in Brazil has worked in a challenging environment, with complicated access to lists of monuments and other historical sites. Despite that, its organisers learned from the competition year-by-year, moving from static analog print materials to a technical infrastructure based on Wikidata. This enables the collection, organisation, and automation of the processes to identify monuments; it additionally allows for the addition and enhancement of data related to them.

While the organisers of Wikimedia projects often say that there is no deadline by which articles or items have to be completely written by, the architectural literature and case studies in Brazil demonstrate that they can unexpectedly be destroyed, robbed, or lost at any time. This is particularly true of the Global South, but also elsewhere in the world. WLM ensures that listed monuments are documented at least for the first time, but if necessary for the last time.

As the project grows to have more lists of monuments; more photographs; and more participants, the technical structures and workflow to support WLM will be stress-tested and require improvements. This will include expanding the completeness of the lists and the photographic coverage of a full range of monuments; developing more user-friendly interfaces, such as maps; and adjusting to new technologies such as mobile-only interfaces. The enhancement of both Wikidata properties and maturation of the competition itself has led to additional types of images and other media like with architectural descriptive practices. WLM increasingly includes more interior and aerial views; while comprehensive floor plans, site plans, and 3D visualisations are aspirational and might be added in future years. The enhancement of technology to support WLM and description of sites is balanced against significant gaps: the competition and organisers must address the regional imbalances of the competition in Brazil, coverage gaps such as the period between copyright-expired materials and modern photography, and the documentation of monuments in Brazil that are not yet covered by official lists.

Conclusions

What is the optimised infrastructure for cultural movements on Wikimedia? The development of the infrastructure has to be open and collaborative, two core values of Wikimedia projects. For full optimisation, it is best if content is first added to Wikidata as a central data repository hub. From there, structured data can be used across Wikimedia projects and beyond, using Wikidata-enabled templates to display the data according to the project needs. Coupled with this, having a variety of input toolkits is vital to semi-automate the process of feeding data and media into Wikidata in a user-friendly way, which enables both complex data structures and simplicity so that the general public can use them.

How does the development of processes and resources for cultural preservation and dissemination in a collaborative digital environment happen? It can occur organically, as a result of specific significant events, but also needs to be driven by the organisations, who engage with Wikimedians to make specific activities possible. It can also optimise engagement—ensuring that pre-existing media is available for the public to reuse, and round-tripping publicly contributed data back into the museum database. It is always highly collaborative work, where individual players can have big roles, but fundamentally relies on the engagement of many editors and knowledge experts acting together to improve content.

We have discussed three case studies demonstrating the link between Wikimedia projects and the cultural sector. In the first case study, we covered how WLM exists within a wider background of monument curation and preservation. Increasing the number of listed monuments, their photographic coverage, and quality of photography and other imaging as they are now and across time will help digitally preserve these monuments for future generations. For Museu Nacional we described how a cross-wiki project naturally came into being after the 2018 disaster, and the impact that the on-wiki work has had with the context of preserving knowledge about objects and collections lost in the fire. For Museu Paulista, we have described an ongoing multi-year project to fully integrate the museum's content within Wikimedia projects, including round-tripping to and from those projects to improve

the museum's local database. All three of these cases studies have involved significant tooling, both on- and off-wiki.

We have only shown three examples out of a multitude of GLAM-Wiki activities taking place around the world. These often use the same tools that we have highlighted, but also different and customised tools as well. Even within the same project, different countries can use different generations of tools to accomplish the same activities. These tools are constantly evolving and collaboratively improved to fit local situations and optimise a global workflow.

Additionally, the way that museums operate in this context is changing. There are challenges that can be understood as broader issues for heritage preservation institutions in a network society. They impose the effort of an institution permanently updating its workflows to international standards and best practices. At the same time, there is a need to adapt and create tools and routines appropriate to local contexts, in line with an institution's collections, budget, and work conditions.

It is important to emphasise that the cases reported here are affected by a shared problem in Global South cultural activities: the absence of public policies aimed at improving teams and technical tools in cultural institutions. In other words, for cultural institutions like the ones discussed here the effort required to keep updated is enormous and depends on engagement that requires creativity and advocacy in favour of free and democratic dissemination of culture and knowledge.

At this point, we return to Castells' statements on the necessary local embeddedness of global projects supported by the digital culture. Changes in the emergence of a web community are relatively positive, as we could see in the cases we have treated in this paper, since they contribute to a relatively bottom-up and participatory social engagement with heritage consciousness and preservation. Yet, it is worth stressing the unequal economic development that characterises the consumerist capitalist economy and its impact on the digital infrastructure distribution. We face the risk of becoming dependent on the technological products that big for-profit corporations offer, for which information–and also culture–is a commodity. To overcome this risk, public funding is necessary for a more egalitarian distribution of technical infrastructure for digital culture initiatives.

In this paper, we show that Levy's understanding of collective intelligence must be framed within the local realities of the Global South, particularly Brazil. The collaborative system of social relationships to solve technical problems and develop practices for digital preservation and dissemination is structured by the prevailing lack of institutional support and resources, thus what is conceived as social intelligence is more importantly creativity to work around technical challenges.

Funding and Acknowledgements

João Alexandre Peschanski's research activity is supported by the Research, Innovation and Dissemination Center for Neuromathematics (FAPESP 2013/07699– 0) and the Faculdade Cásper Líbero Interdisciplinary Research Center. The parts of the research activity by Solange Ferraz de Lima and João Alexandre Peschanski that are described in this paper are supported by the FAPESP research project 2021/06767-8. We thank Lodewijk Gelauff for useful conversations about Wiki Loves Monuments.

Authors' contribution

All authors contributed equally to this research and article.

References

- Alves, É. P. F., Burley, P. R., & Peschanski, J. A. (2021). Structuring bibliographic references: Taking the journal Anais do Museu Paulista to Wikidata. In L. M. Bridges, R. Pun, & R. A. Arteaga (Eds.), Wikipedia and academic libraries: A global project (pp. 260-276). Michigan Publishing. https://doi.org/10.3998/mpub.11778416.ch17.en
- Alves, É. P. F., Lih, A., Knipel, R., & Ånäs, S. (2021, May 26). Wikidata Lab XXIX: Processo de roundtripping [Workshop]. Wikidata Lab, 29th, online. https://pt.wikipedia.org/w/index.php?title=Wikipédia:Edit-a-thon/Atividades_em_portugu ês/Wikidata_Lab_XXIX&oldid=61248109
- Araújo, G. V. F. de, & Knipel, R. (2018, November 3). What are the good models for GLAM pages on-wiki? [Presentation]. GLAMWiki Conference 2018, Tel Aviv, Israel. https://commons.wikimedia.org/wiki/File:GLAM_WIKI_2018_-_What_are_the_good_models_for_GLAM_pages_on-wiki_-_Giovanna_Fontenelle_and_Richard_Knipel.webm
- Azevedo, P. O. de. (1987). Por um inventário do patrimônio cultural brasileiro. Revista do Patrimônio Histórico e Artístico Nacional, 22, 82–85.
- Azevedo, P. O. de, & Lima, V. L. R. C. (Eds.). (1975). Inventário de Proteção do Acervo Cultural da Bahia (Vol. 1). Secretaria da Indústria e Comércio.
- Azzellini, É. C., & Peschanski, J. A. (2020, September 18). Opening up the collection of a closed museum in Brazil. Open GLAM. https://medium.com/open-glam/opening-up-thecollection-of-a-closed-museum-in-brazil-f58f05bef5b2
- Brefe, A. C. F. (2003). História nacional em São Paulo: O Museu Paulista em 1922. Anais do Museu Paulista (10–11), 79–103. https://doi.org/10.1590/S0101-47142003000100006
- Carvalho, J. C. de. (1888). Meteorito de Bendegó: Relatório apresentado ao ministerio da agricultura, commercio e obras publicas e a sociedade de geographia do Rio de Janeiro sobre a remoção do meteorito de Bendengó do sertão da provincia da Bahia para o Museu Nacional. Museu Nacional. Imprensa Nacional. https://w.wiki/4JM4

- Carvalho, V. C. de, Marins, P. C. G., & Lima, S. F. de. (2021). Curadoria em museus de história. Anais do Museu Paulista: História e Cultura Material, 29, 1-24. https://doi.org/10.1590/1982-02672021v29e40
- Castells, M. (2001). A Galáxia Internet: Reflexões sobre a Internet, negócios e a sociedade. Jorge Zahar Editor.
- Castells, M. (2013). Redes de indignação e esperança: Movimentos sociais na era da internet (2a edição). Zahar.
- Dearborn, C., & Meister, S. (2017). Failure as process: Interrogating disaster, loss, and recovery in digital preservation. Alexandria: The Journal of National and International Library and Information Issues, 27(2), 83–93. https://doi.org/10.1177/0955749017722076
- Evenstein, S. (2019, March 5). OERs, Emerging Technologies and AI in Education: The Case Study of Wikidata [Presentation]. Artificial Intelligence and Frontier Technologies for OER.

https://commons.wikimedia.org/wiki/File:Wikidata_Presentation_at_UNESCO_by_Shani _Evenstein.pdf

- Guedes, M. T. (1987). Inventário nacional dos bens imóveis tombados. Revista do Patrimônio Histórico e Artístico Nacional, 22, 86–89.
- Idiegbeyan-Ose, J., Ifijeh, G., Ilogho, J., Iwu-james, J., & Izuagbe, R. (2018). Disaster in libraries in Developing Countries: The need for digital preservation of information resources. https://doi.org/10.4018/978-1-5225-6936-7.ch017
- International Council on Monuments and Sites. (1964). The Venice Charter: International Charter for the Conservation and Restoration of Monuments and Sites. https://www.icomos.org/charters/venice_e.pdf
- Kaffee, L.-A., Piscopo, A., Vougiouklis, P., Simperl, E., Carr, L., & Pintscher, L. (2017). A Glimpse into Babel: An Analysis of Multilinguality in Wikidata. Proceedings of the 13th International Symposium on Open Collaboration, 1–5. https://doi.org/10.1145/3125433.3125465
- Knauss, P. (2018). Cartografia dos Museus de História no Brasil—Uma escrita em movimento. Anais 200 anos de museus no Brasil: desafios e perspectivas. https://www.museus.gov.br/wp-content/uploads/2019/12/Anais-200anosMuseusBrasil_FINAL.pdf
- Kumar, P. (2019). Learning from the Past and Preparing for the Future: Cases and Tools for Cultural Heritage during Disasters [PHD Thesis, IMT School for Advanced Studies Lucca]. http://e-theses.imtlucca.it/292/1/Kumar_phdthesis.pdf
- Kumar, P. (2020). Crowdsourcing to rescue cultural heritage during disasters: A case study of the 1966 Florence Flood. International Journal of Disaster Risk Reduction, 43, 101371. https://doi.org/10.1016/j.ijdrr.2019.101371
- Lévy, P. (2007). A inteligência coletiva: Por uma antropologia do ciberespaço (10a edição). Edições Loyola.
- Lima, S. F. de. (2017). Aprovação de proposta de parceria. https://files.numec.prp.usp.br/public/3e421f

- Magrini, S. (2018). WebGIS, Cultural Heritage, Preservation, Tourism and the Wiki world: A case study from Emilia Romagna (Italy). JLIS, 9. https://doi.org/10.4403/jlis.it-12474
- Martins, D. L. (2021, June 24). Wikidata Lab XXX: Roundtripping process for Tainacan [Workshop]. Wikidata Lab, 30th, online. https://pt.wikipedia.org/w/index.php? title=Wikipédia:Edit-a-thon/Atividades_em_português/ Wikidata_Lab_XXX&oldid=61411744
- Meneses, U. T. B. de, Arantes Neto, A. A., Carvalho, E. de A., Magnani, J. G. C., & Azevedo, P. O. D. de. (2006). A cidade como bem cultural: Áreas envoltórias e outros dilemas, equívocos e alcance da preservação do patrimônio ambiental urbano. [Debate]. Patrimônio : atualizando o debate. https://repositorio.usp.br/item/001636193
- Motta, F., & Silva, R. R. da. (2020). A adoção de tecnologias digitais na reconstrução do Patrimônio: Relato da experiência do Museu Nacional, Brasil. Informação & Sociedade: Estudos, 30. https://doi.org/10.22478/ufpb.1809-4783.2020v30n2.52260
- Nora, P. (Ed.). (1992). Les lieux de mémoire (3rd ed., Vol. 3). Gallimard.
- Nora, P. (1993). Entre memória e história: A problemática dos lugares (Y. A. Khoury, Trans.). Revista Projeto História, 10, 7–28.
- Peschanski, J. A. (2018, September 10). After a catastrophic fire at the National Museum of Brazil, a drive to preserve what knowledge remains. Wikimedia Foundation. https://wikimediafoundation.org/news/2018/09/10/national-museum-brazil-fire/
- Peschanski, J. A. (2021). Variedades de processos de difusão digital colaborativa: Descrição e análise de iniciativas GLAM-Wiki no Brasil. Resgate: Revista Interdisciplinar de Cultura, 29, e021006–e021006. https://doi.org/10.20396/resgate.v29i1.8659966
- Peschanski, J. A., Araújo, G. V. F. de, & Alves, É. P. F. (2018). Developing tGLAM: A landing-page generator for GLAM initiatives. This Month in GLAM, VIII(IX). https://outreach.wikimedia.org/wiki/GLAM/Newsletter/September_2018/ Single#Brazil_report
- Pires, D. de O. (Ed.). (2017). 200 anos do Museu Nacional. Associação Amigos do Museu Nacional.

https://www.museunacional.ufrj.br/200_anos/doc/200_anos_do_Museu_Nacional.pdf

- Posada, E. J. R., Berdasco, Á. G., Canduela, J. A. S., Sanz, S. N., & Saorín, T. (2012). Wiki Loves Monuments 2011: The experience in Spain and reflections regarding the diffusion of cultural heritage. Digithum, 0(14). https://doi.org/10.7238/d.v0i14.1472
- Rachman, Y. B., & Afidhan, S. (2018). Digital Disaster Preparedness of Indonesian Special Libraries. Preservation, Digital Technology & Culture, 47(2), 54–59. https://doi.org/10.1515/pdtc-2018-0009
- Rinehart, A. K., Prud'homme, P.-A., & Huot, A. R. (2014). Overwhelmed to action: Digital preservation challenges at the under-resourced institution. OCLC Systems & Services, 30(1), 28–42. https://doi.org/10.1108/OCLC-06-2013-0019
- Sá, D. M. de, Sá, M. R., & Lima, N. T. (2018). O Museu Nacional e seu papel na história das ciências e da saúde no Brasil. Cadernos de Saúde Pública, 34. https://doi.org/10.1590/0102-311X00192818

- Schwarcz, L. M. (1989). O nascimento dos museus brasileiros. In S. Miceli, F. Massi, Instituto de Estudos Econômicos, Sociais e Políticos de São Paulo, Financiadora de Estudos e Projetos (Brazil) (Eds.), História das ciências sociais no Brasil. Vértice/IDESP.
- Thota, P., & Elmasri, R. (2021). Web Scraping of COVID-19 News Stories to Create Datasets for Sentiment and Emotion Analysis. The 14th PErvasive Technologies Related to Assistive Environments Conference, 306–314. https://doi.org/10.1145/3453892.3461333
- Tosun, E., & Bostan, S. (2021). Disaster and emergency preparedness in Istanbul museums. <u>Museum Management and Curatorship, 1–20.</u> <u>https://doi.org/10.1080/09647775.2021.1969678</u>
- Vieira, G. S. (2012). IPAC Inventário de Proteção do Acervo Cultural: Os modelos da Bahia e Pernambuco nas décadas de 1970 e 1980. Revista Tempo Histórico, 4(1), 1–14.
- Wiki Movimento Brasil. (2020, August 20). Guia para a elaboração da estratégia de difusão digital do Museu Paulista. wmnobrasil.org. https://wmnobrasil.org/museu-paulista-guia/index.html
- Wiki Movimento Brasil. (2021, May 19). O Museu Paulista na Cultura Digital. Wikiversidade. https://pt.wikiversity.org/wiki/O_Museu_Paulista_na_Cultura_Digital