



Glory and Pitfalls of Digital Bricolage: How constant Technological Changes Drives Constant Changes of Humanities Methods

Luces y sombras del bricolaje digital: cómo la aceleración del cambio tecnológico reconfigura constantemente los métodos en humanidades

Luzes e sombras do bricolagem digital: como a aceleração da mudança tecnológica reconfigura constantemente os métodos nas ciências humanas

Frédéric Clavert

Universidad de Luxemburgo. Luxembourg Centre for Contemporary and Digital History
Luxemburgo

Correo electrónico: frederic.clavert@uni.lu

ORCID: <https://orcid.org/0000-0002-0237-2532>

Abstract

In 2021, Twitter announced a new version of its Application Programming Interface (API) that allowed, with constraints, the collection of Twitter data. At the same time, it introduced a new policy for researchers: those who applied for and obtained recognition as researchers by Twitter could search Twitter's whole history and harvest, theoretically, up to 10 million tweets each month. For those without enough funding to access the commercial API, this was a huge change: a real improvement, but also, for Humanities 'bricoleurs' analysing Twitter data, a major challenge. Indeed, all the tools and methodologies elaborated over a decade

Keywords

digital platform
statistical data
comparative analysis
open access

around the use of Twitter data in the Humanities, particularly in history and memory studies in our case, had to evolve. By exploring the example of the Twitter API and its evolution, this article investigates the continuous changes in our methods and ways of working, carried out under pressure from technical evolutions and their underlying business models based on data access, even when these are minor. It explores digital bricolage as an imperfect response, then the changes induced by Elon Musk's seizure of Twitter, which force a shift from bricolage to braconnage. In conclusion, we ask whether we should stop studying Twitter, though this would imply self-sabotage of our own research.

Resumen

En 2021, Twitter anunció una nueva versión de su Interfaz de Programación de Aplicaciones que le permite recopilar datos, con ciertas restricciones. Paralelamente, introdujo una nueva política: quienes soliciten y obtengan el reconocimiento como investigadores por parte de esta red social, podrán realizar búsquedas en todo el historial y recopilar (en teoría) hasta 10 millones de tuits al mes. Esta medida supuso un gran cambio para quienes no tenían acceso a la versión comercial (paga), pero también un reto para los «bricoleurs» procedentes de las humanidades que analizan los datos. Todas las herramientas y metodologías elaboradas en torno al uso de los datos de Twitter en las humanidades tuvieron que evolucionar, especialmente en los estudios de historia y memoria. A partir del ejemplo de esta Interfaz, en este artículo se investigan los cambios continuos de nuestros métodos y formas de trabajar, bajo la presión de los adelantos técnicos y sus modelos de negocio subyacentes basados en el acceso a los datos, incluso cuando son menores. Además, se explora la noción de bricolaje digital como una respuesta imperfecta a estas evoluciones y el paso al *braconnage* (caza furtiva), con los cambios inducidos tras la adquisición de Twitter por parte de Elon Musk. Con tantas modificaciones, vale preguntarse si deberíamos dejar de estudiar Twitter, aunque implicaría un autosabotaje de nuestra propia investigación.

Resumo

Em 2021, o Twitter anunciou uma nova versão de sua Interface de Programação de Aplicativos (API) que permite a coleta de dados, com certas restrições. Paralelamente, introduziu uma nova política: aqueles que solicitarem e obtiverem o reconhecimento como pesquisadores por parte dessa rede social poderão realizar pesquisas em todo o histórico e coletar (em teoria) até 10 milhões de tuítes por mês. Essa medida representou uma grande mudança para quem não tinha acesso à versão comercial (paga), mas também um desafio para os “bricoleurs” da área de ciências humanas que analisam os dados. Todas as ferramentas e metodologias desenvolvidas em torno do uso de dados do Twitter nas ciências humanas tiveram que evoluir, especialmente nos estudos de história e memória. A partir do exemplo desta interface, este artigo investiga as mudanças contínuas em nossos métodos e formas de

Palabras clave

plataforma digital
datos estadísticos
análisis comparativo
acceso abierto

Palavras-chave

plataforma digital
dados estatísticos
análise comparativa
acesso aberto

trabalho, sob a pressão dos avanços técnicos e de seus modelos de negócios subjacentes baseados no acesso aos dados, mesmo quando se trata de dados de menor porte. Além disso, explora-se a noção de “bricolagem digital” como uma resposta imperfeita a essas evoluções e a transição para a “braconnage” (caça furtiva), com as mudanças provocadas pela aquisição do Twitter por Elon Musk. Com tantas mudanças, vale a pena questionar se deveríamos deixar de estudar o Twitter, embora isso implicasse um autossabotagem de nossa própria pesquisa.

Recepción del original: 1 de septiembre de 2024.

Aceptado para publicar: 31 de marzo de 2025.



Glory and Pitfalls of Digital Bricolage: How Constant Technological Changes Drives Constant Changes of Humanities Methods

1. Introduction¹

After several months of a saga with many twists and turns, including legal ones, Elon Musk, the head of some of the most famous Silicon Valley's leading companies, bought Twitter Inc. on 27 October 2022. The epic story of this takeover then continued: layoffs of more than half of the company's employees, end of subcontracts with third-party companies, deteriorated working conditions, etc. These events have made a potential bankruptcy of Twitter credible, although not inevitable. This prospect of Twitter's demise, whether sudden or slow, has rekindled among historians—including the author of this article—² the fear of losing a priceless archive about our present.

For while Twitter had an agreement with the Library of Congress (LoC),³ the consultation of the LoC's Twitter archive is yet impossible on the one hand, the archiving is integral until 2018 and then by sampling decided on the basis of decisions linked to the political and social life of the United States on the other hand.⁴ Thus fears of a *digital dark age* are reactivated, including among those—Google for example—⁵ whose platforms are inherently unarchivable.

For the historian, losing all of Twitter—or keeping an unusable archive— would mean great losses: from oppressions, sometimes genocidal, around the world to the political tribulations of the United States since 2016, from the French Yellow Vests (*Gilets jaunes*) movement to the most recent Brazilian political events at the time of writing, tweets document the most important events of our time in a different way than newspapers or more traditional state archives.

This fear of a *digital dark age*, however, is not only linked to the risk that an unstable billionaire may pose to a web platform: the tree that is Twitter, whose data was until the end of June 2023 relatively accessible to researchers, hides the forest of

¹ I would like to thank Fred Pailler for his careful reading of a previous version of this article. Though most of its content remains valuable, this article was written in 2022 and 2023 and some of the socio-technical elements described in it might be out of date.

² Clavert, F. (2022, December 17). «Préserveons notre patrimoine numérique grâce au contre-archivage décentralisé et collaboratif». *Le Monde.fr*. https://www.lemonde.fr/idees/article/2022/12/17/preservons-notre-patrimoine-numerique-grace-au-contre-archivage-decentralise-et-collaboratif_6154820_3232.html

³ Raymond, M. (2010, April 14). How Tweet It Is!: Library Acquires Entire Twitter Archive. *Library of Congress Blog*. <https://blogs.loc.gov/loc/2010/04/how-tweet-it-is-library-acquires-entire-twitter-archive/>

⁴ It is difficult to know what has really been archived by the LoC since 2018. See: Osterberg, G. (2013, January 4). Update on the Twitter Archive at the Library of Congress. *Library of Congress Blog*. <https://blogs.loc.gov/loc/2013/01/update-on-the-twitter-archive-at-the-library-of-congress/>

⁵ Ghosh, P. (2015, February 13). Google's Vint Cerf warns of 'digital Dark Age'. *BBC News*. <http://www.bbc.com/news/science-environment-31450389>

platforms that jealously guard their data for the purposes of behavioural marketing studies and block many possible research projects. Indeed, while Twitter is a social media that is exemplary of the platforms we know, it was also a form of exception, with its rather open application programming interface (API)⁶ for researchers.⁷

The case of Facebook is much more complex: neither archived by an institution such as the *Library of Congress*, nor really accessible to researchers, researchers often have to find means whose compliance with the platforms' terms of use, legality and sometimes ethics are questionable, such as *scrapping*, i.e. the implementation of small pieces of software pretending to be human users but actually collecting data.

In this article we will focus on Twitter (now X). Twitter was much studied, as it provided researchers with tools to study it, notably *via* a socio-technical device called *Application programming interface*. Based on our own experience of research with and about Twitter, we will show how some of our methods have evolved as the ways in which the Twitter API is accessed have changed. We will then attempt to use the notion of *bricolage* as a way of adapting to these changes. Finally, we will try to understand what Elon Musk's takeover might mean for research, particularly in terms of making *bricolage* look like *braconnage*.

2. Researching with and about Twitter: from #ww1 to #covid19fr

Collecting data on social media while respecting ethical and legal rules (respecting users' privacy, following terms of use in particular) is a headache for researchers. The French project *Algopol*,⁸ for example, which explored the particular social relationships that develop on Facebook, would not be possible today, as Facebook changed its API in 2016. Hopefully, the *Algopol* project had collected enough data to produce serious results.⁹ This is a reminder of the dependence of our research, when it makes use of artefacts produced by the big web platforms, on these BigTech giants, which are often unconcerned with the needs of research. However, the example Facebook's API is extreme: this dependence is often more subtle, introducing a certain instability. For instance, by studying archived versions of Twitter documentation, Marta Severo and Timothée Giraud (2019) were able to prove that the meaning of some of the geographic metadata attached to the tweets we collect has changed over time. For those of us who use such geographic data in the context of long-term data harvesting, these documented but often poorly advertised changes can be misleading, or at least introduce inaccuracies that are not compatible with the requirements of research.

In this section, we will discuss a much better documented and less subtle change, that of the transition of the Twitter API from version 1.1 to version 2, which has also

⁶ A programming interface is a piece of software device that allows two pieces of software to share functionalities or information. In this case, software installed on our server connects to the Twitter API to obtain data (tweets) along with their metadata.

⁷ As soon as 2019, based on an analysis, notably, of the *Cambridge Analytica* scandal, Axel Bruns (2019) spoke of an *APIcalypse*, the closing down of APIs, including Facebook's in 2016 that prevented many researchers to continue their projects. In 2023, with the end of free access for researchers to X's (Twitter's) and Reddit's APIs, this APIcalypse went even further.

⁸ Consult <http://algopol.huma-num.fr/appresuits/le-projet/index.html>

⁹ For example: <http://algopol.huma-num.fr/appresuits/le-web-vu-de-facebook/index.html>

introduced a new policy towards research with the introduction of the *academic research track product*.¹⁰ We will look at two research projects we have conducted: #ww1 and #covid19fr.

2.1 The Centenary of the Great War on Twitter: *bricolage* and anticipation

In 2014, we launched a research project around the Twitter echoes of the Centenary of the Great War. From April 2014, a few weeks before the Centenary of Franz-Ferdinand's assassination, to the end of November 2019, a year after the end of the official Centenary, we collected tweets containing keywords or hashtags (keywords, acronyms or strings of characters preceded by a hash) related to the Centenary, mainly in English and French. With more than 9 million tweets (including retweets) involving various interactions between approximately 1.5 million users, the database thus created opened up numerous possibilities for analysis, notably by calling upon methodologies linked to what is now called *distant reading* (Moretti, 2007) such as forms of topic modelling or social network analysis.¹¹

But the devil is in the details, and more particularly in the details of the tweets harvesting methods. We started off very small—strictly speaking, *bricolage*—on a server installed on our private ADSL connection, we professionalized it when we changed jobs, and had later to change the pieces of software used to collect the tweets when, in 2017, Twitter doubled the size of the tweets, from 140 to 280 characters. We therefore juggle with three distinct databases: from 2014 to 2015, a first database contained the tweets collected at home using a server software called 140dev;¹² a second database was built from 2015 to 2017, using the same software; finally, a third database, from 2017 to 2019, built thanks to DMI-TCAT,¹³ was set up at the University of Luxembourg.

The changes in Twitter's data accessibility offers led us to change the collection scripts, in this case to a better software: better in terms of functionality, but also in terms of metadata harvesting. Indeed, 140dev limited the number of metadata collected with the tweets, for example by removing links to the profile pictures of the Twitter accounts whose information is collected.

More importantly, the harvesting method common to both software used in this project is the *streaming* command set of the Twitter API v1.1. At that time, it was possible to collect tweets for free, for any purpose compatible with Twitter's terms of use, within limited constraints. We could either search the history of tweets, but within a limit of 3200 tweets per request and over a period of no more than one week in the past (the so-called *search* API), or collect the stream of tweets ("firehose") at the time of their publication (the so-called *streaming* API). We chose the latter: although it did not allow us to collect freely more than 1 % of the stream of tweets (i.e., about 5 million tweets

¹⁰ <https://web.archive.org/web/20230520035516/https://developer.twitter.com/en/products/twitter-api/academic-research/product-details#academic-track>. This access to the Twitter's/X's API has been closed in June 2023.

¹¹ It is not the purpose of this paper to set out the results of this research in detail. Readers wishing to do so may consult Clavert (2018, 2021).

¹² 140dev. This set of scripts has not been developed since 2014, but was able to function properly until 2017.

¹³ *Twitter Capture and Analysis Tool* is developed by Digital Methods Initiative. For more information see Erik Borra & Bernhard Rieder (2014).

daily), it seemed sufficient for our research. We indeed significantly exceeded 1 % of the firehose, calculated over a quarter of an hour, only once, on 11 November 2018.

If the streaming API has thus provided us with a sufficient harvesting framework for our research, it comes with a condition that is much more important than the famous 1 %: anticipation. Started in 2014 as an improvisation, this collection quickly became part of a longer-term research project, covering more than 5 years. In 2014, it was difficult to foresee that the hashtag pushed by the organisers of the commemoration of the Centenary of the Battle of the Somme (1 July 2016) would be #somme100, even if this hashtag, of course, is most logical. Some hashtags have “popped up” from the database, such as #1j1p, a “rallying” hashtag of Twitter accounts who participated to the indexing of the French database of the *Morts pour la France* (Dead for France).¹⁴ Thus, analysing and regularly consulting the collected data has made it possible, in part, to anticipate or make up for omissions in the collected hashtags.

Anticipation was not the only issue with the streaming API. As it was then impossible to ‘go back in time’ beyond a week, it was difficult to compensate for some of the biases associated with the hashtag or keyword-based collection mode (D’heer et al., 2017). Among these biases, the conversational aspect is underestimated by this way of harvesting tweets. For instance, if a reply to a harvested tweet did not contain one of the collected keywords, it would not be stored in the database. But other biases of this collection method are even more troublesome. What if an event is not linked to a specific hashtag? For example, the first battle of the Marne (September 1914) is virtually absent from the corpus. What should be done when some of the tweets to be collected use a hashtag that is too generic either, for example, because it is strictly geographical (#marne, #somme or #verdun, which are widely used for present and local purposes), or because it is used for a large number of commemorations outside the centenary (#lestweforget, inspired by a poem by British author Rudyard Kipling)? What to do when a hashtag is used to evoke a prominent personality linked to the First World War, but commemorated for the whole of their life, as was the case for Jean Jaurès in 2014 or for the “année Georges Clémenceau” in 2017? It is rare that a technical solution exists.

2.2 #covid19fr: a country under lockdown on Twitter

The #ww1 project was stopped at the very beginning of December 2019. By mid-March 2020, the server we had then configured was still usable. Facing a feeling of emergency (Clavert, 2020)¹⁵ when it became clear that most Western European countries, following the Italian example, were going to decree lockdowns of their populations, including Luxembourg and France, we decided to collect data on the health crisis. Limited

¹⁴ #1j1p means “Un jour, un poilu” (“One day, one fallen soldier”). It is a (private) initiative encouraging people to go to the public database of the Dead for France and to index it, i.e., to transform images of administrative acts into text seen as text by the computer. This initiative had two aims: to pay tribute to the Poilus who died for France on the one hand, and to allow a more efficient use of this database by future historians on the other. See the dedicated website: <https://www.1jour1poilu.com/> as well as an interview with Jean-Michel Gilot, who initiated this ‘challenge’: (Gilot et al., 2018).

¹⁵ This feeling has been shared: the number of research projects, in all disciplines, to have been triggered as of March 2020 is very high, probably unprecedented for this type of crisis. See a non-exhaustive list on the World Pandemic Research Network website: <https://wprn.org/>

to French-speaking hashtags because of the 1 % constraint of the Twitter's streaming API, this project, called *#covid19fr, a country under lockdown on Twitter*, has at its heart a collection of more than 60 million tweets today. Based on the same software package as the previous project, we expected to face the same limitations, constraints and biases.

However, although tweets harvesting based on the streaming API in version 1.1 resulted in a database containing more than 63 million tweets until the 15th of March 2023, this database has not been used for the moment, for two reasons. The first is that the project was extended to Italy through joint work with Deborah Paci, historian, then at the University of Venice. The second reason is the release, in early 2021, of the Twitter API v2, but, more importantly, the creation of a new 'product', the *academic research product track*.¹⁶ Subject to a questionnaire to be filled in and a (human) evaluation by a Twitter team, a researcher recognised as such by Twitter could then collect 10 million tweets per month from the entire Twitter history. For the *#ww1* project, this would have meant, for example, being able to collect in a few days the equivalent of the entire corpus we harvested from 2014 to 2019. The *academic research product track* was closed down in June 2023, a few months after the takeover of Twitter by Elon Musk and the announcements of changes to the terms of access to the Twitter API in early February 2023.

These two reasons combined meant that the chapter and the article co-authored with Deborah Paci (Clavert & Paci, 2023) did not use the database built since 15 March 2020 through the streaming API v1.1, but rather *ad hoc* corpora, constituted and analysed for the needs of these writings thanks to the access to the entire Twitter history that was now possible. Thanks to the *twarc* tool¹⁷ developed by *Documenting the Now*,¹⁸ the collections could be precise and fast and the distant reading analyses carried out at the same time. Thus, some of the biases, limitations and constraints linked to the previous version of the API were resolved: we could go back in time to correct a corpus whose harvesting had been badly designed, for example. Anticipation was no longer an issue.

But other biases are induced by this *a posteriori* corpus building mode. Indeed, and in accordance with a certain number of legitimate ethical and legal principles related to the use of personal data, a deleted tweet, the tweets of a deleted account or of an account that has been switched to "private" mode cannot be collected *a posteriori*, whereas the streaming API allows for collection as soon as the tweets are published and before they are deleted or switched to private mode. One of the most notable effects of this limitation is the likely underestimation of controversies. In the case of the health crisis, this underestimation is possible but not very important, since our research and analysis show that, on the French side, the debates on Twitter about the French State's health policy are riddled with controversies, the most famous of which is around the personality of Professor of Medicine Didier Raoult and the use of hydroxychloroquine

¹⁶ Ternes, A. & Trujillo, L. (2021, January 26). Enabling the future of academic research with the twitter api. *Twitter Developer Platform*. https://blog.x.com/es_la/topics/product/2021/haciendo-posible-futura-investigacion-academica-twitter-api

¹⁷ Consult <https://twarc-project.readthedocs.io/en/latest/>.

¹⁸ Consult <https://www.docnow.io/>.

that he claimed. If these controversies have been underestimated, they still remain extremely present in the corpus we have collected.¹⁹

However, we wonder about the extent of the changes that our corpus on the Centenary of the Great War would have undergone if we had proceeded with a tweet harvesting in 2021 through the academic research product track. Indeed, one of the elements we think we have shown is that, at least on the French side, any commemoration is accompanied by its contestation, for various reasons. But this contestation is often low-key, quiet, with few exceptions like the Franco-German commemoration of the Centenary of the Battle of Verdun in May 2016, which was the subject of two notable controversies. Would the API v2 have led us to underestimate these forms of "memory protest"? Would the peaks of controversy in November 2014 during the inauguration of a new memorial by the then President of the Republic, François Hollande, and in May 2016 around Verdun have been as visible as they are in our current database? We cannot say, insofar as, if the conditions of the collection of tweets from 2014 to 2019 have been documented as we go along, reproducing them rigorously seems difficult and the comparison would remain hazardous.

Thus, the use of the socio-technical device that is an API to collect data related to the past and its memory on Twitter is likely to have a very strong influence on a research project and its results. It is challenging²⁰ to assess the extent of this influence, at least for the historian that we are. Nevertheless, it is certain that an answer must be given: these limits and constraints cannot be a reason for not conducting research. In the end, our response to this challenge is quite classic: it is a matter of *bricolage*.

3. Digital *bricolage* as a survival mode for humanities and social sciences research in the digital age

If we look at the history of the historical sciences in the 20th century, we can read it as an almost constant pressure exerted on historians and their method. In France, for instance, the debates between the sociologist François Simiand and the historian Charles Seignobos at the beginning of the century concerning the scientific status of history showed how history reacted to the rise of sociology (Revel, 2007). This debate, which continued in other forms and with other actors in the inter-war period with the creation by Lucien Febvre and Marc Bloch of the *Annales*, has generated strong articles (and concepts) in the historical sciences up to the present day, such as the notion of *longue durée* (Braudel, 1958), brought back into the front scene by digital humanities with the publication of the *History Manifesto* (Guldi & Armitage, 2014) along with its critical appraisal (Annales. HSS., 2015).

Our hypothesis is that, today, historical sciences are under pressure from computer sciences. A number of important digital projects claim to renew history through computing and the methods it provides. For example, the Seshat database website (François et al., 2016) claims that:

¹⁹ The results of this research are exposed in Clavert & Deborah Paci, 2023.

²⁰ Though it is not impossible: see Thibault Grison et al., 2023.

Seshat: Global History Databank was founded in 2011 to bring together the most current and comprehensive body of knowledge about human history in one place. The huge potential of this knowledge for testing theories about political and economic development has been largely untapped.²¹

The significant media coverage of this type of project²² have led to heated and very tensed debates on social media, partly based on mutual disciplinary misunderstandings (Sikk, 2020). More institutional was the exchange between Erez Lieberman Aiden and Jean-Baptiste Michel²³ on the one hand, Anthony Grafton (2011), then president of the American Historical Association on the other hand, concerning the article by a team gathered around Jean-Baptiste Michel (Michel et al., 2010) exploring the *Google Books Ngrams* tool which is based on the massive digitisation of books by Google and allows analyses based on ngrams, i.e., a sequence of *n* words. The ambition is to trace the evolution of language but also the major cultural trends through a corpus from Google Books. Here too, numerous disciplinary disagreements were heard: a historian's concern for sources cannot be satisfied by the constitution of a corpus without concern for representativeness; the criticisms addressed in the 1970s by microhistory to the quantitative historians of the 1960s and 1970s are ignored (in the sense that they are not known) by the initiators of this article; the authors, for their part, question the statistical and computer skills of the historians with whom they tried to communicate during the writing of their article.

There is a conflict of methods here. While the *big data* on which culturomics is based shares inductive reasoning with history, the historian's concern and attention to sources cannot be satisfied with the limits of reasoning linked to *big data*, which were pointed out very early on, particularly in sociology (Boyd & Crawford, 2012). The idea that the mass of data would be sufficient to ensure a form of representativeness cannot be reconciled with historical methodology.

Historians interested in Big Data and the artefacts of the past that it contains are then caught between their scientific, rational thinking, based on methods that have been tried and tested since Ranke, and a form of *pensée sauvage* that they have to call upon when they use computer tools based on a science that they know little about, by lack of training.

The term *bricolage* (or *pensée sauvage*) was defined in *La pensée sauvage* by the anthropologist Claude Lévi-Strauss (1962). There have been many uses of this concept since then as well as it was questioned and criticised as soon as the book was published (Lévi-Strauss & Ricoeur, 1963).²⁴ Here we will twist this metaphor a little, adding an adjective, *digital*, and defining *digital bricolage* as a mode of adaptation to our datafied

²¹ Consult <http://seshatdatabank.info/>.

²² Spinney, L. (2019, November 12). History as a giant data set: How analysing the past could help save the future. *The Guardian*. <https://www.theguardian.com/technology/2019/nov/12/history-as-a-giant-data-set-how-analysing-the-past-could-help-save-the-future>

²³ Lieberman Aiden, E. & Michel, J.-B. (2011, March). Thoughts/clarifications on Grafton's 'Loneliness and freedom'. *Internet Archive*. <https://web.archive.org/web/20120622082108/http://www.historians.org/Perspectives/issues/2011/1103/1103pre1.cfm>

²⁴ For a more systematic discussion of the criticisms see Anne Mélice (2009).

and networked world. This notion of *digital bricolage* has already been used in the past (Rüling & Duymedjian, 2014) and is of particular usefulness for understanding social adaptation to the changes of the digital society.

We are interested here more specifically in *digital bricolage* as practiced in the historical sciences. *Digital bricolage* is to be understood here as an academic and historian's response to technological 'disruptions', in a more general context of the neoliberalisation and globalisation of higher education and research which clearly maintains a fascination for digital innovation.

Digital bricolage is a mode of survival. How can one take an interest in collective memory artefacts on Twitter without learning late in the game how to collect data, analyse it and present the results, while at the same time understanding the methodological and epistemological limits of these approaches? Most often untrained—though it might evolve in the future—historians must then venture into the algorithmic jungle of computer tools available. This can be solved by interdisciplinary collaboration, which is not always possible. Moreover, the tools often used must be adapted for historical use. In the end, *bricolage* is a concept that can be applied to tools as well as primary sources.

We believe that our research on the Great War and the mnemonics of the health crisis fall within the scope of *digital bricolage*. We had to acquire a solid understanding of what the Twitter API is, even though it was not designed for historians (which is normal) and the vocabulary used is not that of the historian.²⁵ The very vocabulary of historians being a double-edged sword: even the term 'history', highly polysemic, is used in computing in a very different sense. Tools had to be found and diverted to collect data. While DMI-TCAT is a tool designed for research, this was not the case with *140dev*, which was designed for marketing: our use of *140dev* aroused the sceptical surprise of its developer in various email exchanges. Finally, the analysis tools are often developed by researchers, sometimes from the human and social sciences. But this has consequences: time in research is limited and documentation is often lacking. It is therefore necessary to proceed by trial and error, which is slow, sometimes uncertain, and, above all, sometimes leads to the acquisition of reflexes that could be put to better use. Thus, using software such as IRaMuTeQ (Ratinaud & Dejean, 2009) to analyse the text (an approach close to *topic modelling*) is clearly more efficient if one uses the version under development, which requires the understanding of how the *git* command (a versioning software highly used in computing development) functions on the one hand, and a constant discussion with its developers to compensate for a lack of documentation (for understandable reasons) on the other.

Digital bricolage implies a constant adaptation to the evolution of the tools used. The end of *140dev* development forced us to switch to DMI-TCAT. The evolution of the Twitter API to version 2 encouraged us to use other harvesting methods and another software, *twarc*. The mass of data collected forced us to have our analysis know-how evolved. For instance, we had sometimes to use *topic modelling* as implemented in the MALLET software (McCallum, 2002) rather than IRaMuTeQ—which is not performing

²⁵ Éric Dagiral and Fred Pailler (2018) have published a paper about using "second-hand" corpora in social sciences that enlightens some of the challenges of using tools and corpus that are not made for Humanities and Social Sciences.

topic modelling per se though its use can be the same—, which we have mastered for many years and which is part of our methodological “comfort zone”. With *digital bricolage*, there can be a huge methodological instability —though I have used IRaMuTeQ for years, it will soon cease—, which is in sharp contrast to the more traditional habits of historians which, if historians have managed to renew them, have never evolved on a such high frequency mode.

Other authors, who have not necessarily used the notion of *digital bricolage*, have noted these developments in their disciplines. The sociologist and linguist Dominique Boullier,²⁶ for example, speaks of a third generation of social sciences, focused much more on replications and propagations, i.e., the dissemination of information at high frequency than on large statistical series or opinion or field surveys. And this is another aspect of *digital bricolage*: it is not just a question of adapting to tools, but also to new concepts. For example, we studied the controversies surrounding the commemoration of the Centenary of the Battle of Verdun from the point of view of these replications, considering those controversies as ‘vibrations’, i.e. brief, dense moments of high-frequency information circulation.

Today we are wondering about how *digital bricolage* in history or memory studies will evolve. The more computer science advances, the more the machine —and also the functioning of the software— moves away from our understanding. *Digital bricolage* as we briefly defined it, allows an ephemeral but effective understanding of what historians do with computers. Will this understanding be sufficient with the emergence of *Large Language Models* (LLM) based on *transformers*, those types of neural networks that are behind the incredible rise of LLMs?²⁷ These systems, such as ChatGPT,²⁸ which has shaken up the world of higher education and research since November 2022,²⁹ are black boxes in the sense that, although their theoretical operation is understandable in broad terms, their outputs are difficult to explain, since they are based on a complex system of probabilities. Furthermore, probability is used for prediction. What can prediction mean when working on the past, if we roughly see the past as a set of events that actually happened?

The increasing use of artificial intelligence, especially machine learning, is not, however, the only limitation of *digital bricolage*. We are dependent on the willingness of the big web platforms to provide us with data. What if these platforms close down data access or simply disappear? Should we move from *digital bricolage* to *digital braconnage*?

²⁶ Boullier, D. (2016, July 18). Big data challenges for the social sciences: from society and opinion to replications. arXiv. <https://doi.org/10.48550/arXiv.1607.05034>

²⁷ Uszkoreit, J. (2017, August 31). Transformer: A Novel Neural Network Architecture for Language Understanding. *Google Research*. <https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html>

²⁸ Consult <https://chat.openai.com/>

²⁹ D’Agostino, S. (2023, January). ChatGPT advice academics can use now. *Inside Higher Ed*. <https://www.insidehighered.com/news/faculty-issues/teaching/2023/01/11/academic-experts-offer-advice-chatgpt>

4. From bricolage to braconnage?

Elon Musk's takeover of Twitter and subsequent events, including the change in access to Twitter's API, is a game changer for any research based in whole or in part on Twitter data. The drastic downsizing of the company's number of employees has reminded us that large web platforms, or parts of them, can disappear, sometimes abruptly. What is left of *myspace*? Who remembers *Orkut*, the Google social network that was closed in 2014 and which was, for example, so important in Brazil that it was managed there from 2008 to 2014?³⁰ These disappearances, whether gradual or brutal, have consequences for present-day as well as future historians in the sense that important primary sources for the understanding of our world are disappearing, sometimes partially, sometimes fully.

It has happened that archives of these arbitrarily closed services have been "rescued". This is the mission of the *Internet Archive* and its *Wayback Machine*. Unfortunately, it is understandably not possible to find in the *Wayback Machine* all pages of all websites or social media platforms beyond some static pages and some well-known users' pages. One case of rescue is emblematic but hides the forest of losses: *GeoCities*. This personal page service, typical of the web in the late 1990s and early 2000s, was bought by Yahoo. It declined to the point where it was used almost exclusively in Japan. Yahoo decided to close it down in 2009. But this closure almost resulted in the loss of the entire platform's data. It took the intervention of the *Archive Team* (Archiveteam, 2009) to recover an archive at the last moment. Subsequently exploited by Ian Milligan (2017), for example, the GeoCities helps us understand, in particular, how links were structured on the web in the second half of the 1990s. The web is in fact a factory of obsolescence (Gomez-Mejia, 2020). The history of the GeoCities archive shows us that there can be a shift from *digital bricolage* to *digital braconnage*: it required active and preventive intervention to preserve it. It was no longer a question of adapting, but of actively intervening, even against the will—or at least against the initial intent—of the data's owner.

Let's have a look at the platform that has occupied us since 2014, *Twitter*. Entering a period of instability in 2022 with Elon Musk's desire for a private takeover, this company has, since the South African billionaire took effective control in October 2022, been profoundly transformed. In February 2023 a change in the rules of access to the Twitter API was announced, via the Twitter Developers account.³¹ This announcement was implemented several months later (April but for researchers, June for researchers). It profoundly alters the way in which researchers, including in history and more broadly in the humanities and social sciences, are likely to study phenomena such as they have developed on Twitter, on the subjects that can be studied and on the results of the research itself.

³⁰ Orkut. (2023, February 22). *Wikipedia*. <https://en.wikipedia.org/w/index.php?title=Orkut&oldid=1140890642>

³¹ Twitter Developers. (2023, February 2). *Starting February 9, we will no longer support free access to the Twitter API, both v2 and v1.1. A paid basic tier will be available instead* [Tweet]. <https://twitter.com/TwitterDev/status/1621026986784337922>

The first consequence of a change in the conditions of access to the Twitter API is a probable collapse around Twitter, at least when it relies on large corpora of tweets: the price researchers, among others, have now to pay makes the access to the Twitter API unaffordable.³² Yet such research has been fundamental to understanding social media, since, albeit with significant variation, Twitter has always maintained limited but free access to its data. Of course, social media research will not stop, but it will either have to be strictly qualitative—which is fine, but we believe in the complementarity of qualitative and quantitative approaches—, have to depend on the goodwill of Twitter, or be *braconnage*.³³

This change in Twitter's policy towards researchers reinforces an aspect that is already deeply rooted in humanities and social sciences research: the division between *haves* and *have-nots*. Access to the traces of Big Data for research purposes is profoundly unequal, and research in the humanities and social sciences then becomes, at least for the fraction of history and memory researchers interested in *Big Data*, conditioned by the arbitrary policies of private companies in the *BigTech* field. This division between *haves* and *have-nots* will be stronger as it is already and will be even more accentuated by the inequalities suffered by the *global south*.

Preserving sources for the future will also be a problem. If the *Library of Congress* was archiving Twitter, in its entirety until 2018, through sampling from a North American perspective since then, we do not know what the future of this archiving will be. The difficulties linked to the archiving of Twitter closes off possibilities: studying the turmoil of US politics since 2016, analysing the beginnings of repression, preserving the traces of different activist movements in societies of the North and the South. With the closure of the Twitter API, it is the preservation of the present that becomes more difficult.

As the group *Documenting the Now* (Documenting the Now, n.d.) points out, the conditions of access to the Twitter API for researchers before 2023 also allowed them to conduct research in an ethically sound manner. Social media research is a thin path between ethical use of personal data and surveillance (Berendt et al., 2015). This path will become increasingly narrower: because while there will remain technical means of building up corpora—for example with *scrapping*— these means are only rarely designed to respect the people we study. Not only will the ridge between ethical research and surveillance become even narrower, but the risk of falling into illegality will be present at all stages of research.

Digital braconnage is understood here as a detour, not necessarily illegal—the CNIL (the French data protection commission) has authorised research projects based on *scrapping*—, of the rules imposed by *big tech* in order to continue doing research. In a sense, this concept can mobilise another, that of counter-archiving used by Anat Ben-David (2020) in relation to *Facebook*. *Digital braconnage* can even be motivated by an ethical dimension: keeping traces, archiving them, for future generations. It then becomes collective, contributory, allowing for the collaborative capacities of what has been pompously called Web 2.0.

³² Stokel-Walker, C. (2023, March 10). Twitter's \$42,000-per-Month API Prices Out Nearly Everyone. *Wired*. <https://www.wired.com/story/twitter-data-api-prices-out-nearly-everyone/>

³³ The concept of *braconnage* (poaching) is not entirely new and has been used in the context of the pandemics. See for instance Charles Parisot-Sillon, 2020.

The most recent developments in Twitter's evolution —its transformation into X— brings Twitter/X into the ranks of *big tech*, which is used to closing off access to the traces of big data. It should encourage us to modify our *bricolage* practices. It is no longer a question of answering questions posed by other disciplines, but rather of responding to the frictions of the confrontation between big tech and research.

5. Conclusion: from bricolage to braconnage... to sabotage?

Bricolage, braconnage. Sabotage? The buying of Twitter by Elon Musk is not only the saga of a businessman trying to adapt the world to his own view of freedom of speech. The long-term ambitions of Musk are apparently an X-app, an "everything app" inspired by the Chinese WeChat app.³⁴ But in the short and middle terms, this buying is first and foremost a change of nature. Over the years, Twitter inc. spent time trying to deal with racism, machism and other form of ideologies that turn into and legitimize online bullying. The end of all bans on bullying accounts, including the former US President Donald Trump's, the probable modification of Twitter ordering algorithm to favour Musk's tweets, the conversations of Musk with many US alt-right's accounts, are all signs of this change in nature.

Considering this ethical dimension —Twitter's change of nature and its possible transformation into a white alt-right social media— should we, beyond research on the alt-right, still use Twitter as a basis for research on many other topics? Will it still be possible, in this framework, to study feminist appropriation of commemorations (Smyth & Echavarria, 2021) for instance? *In fine*, Twitter's recent evolution might force us to move from *bricolage* to *sabotage*. Sabotage would mean putting an end to the fact that we have accepted the balance of power imposed by the major web platforms. It is a self-sabotage in the sense that it would force us to rethink all our methods and our ways to research those platforms.

Bibliographic references

1. Annales. HSS. (2015). La longue durée en débat. *Annales*, 70(2), 285–287. <http://www.cairn.info/revue-Annales-2015-2-page-285.htm>
2. Archiveteam. (2009). *GeoCities*. <http://www.archiveteam.org/index.php?title=GeoCities>
3. Ben-David, A. (2020). Counter-archiving Facebook. *European Journal of Communication*, 35(3), 249–264. <https://doi.org/10.1177/0267323120922069>
4. Berendt, B., Büchler, M. & Rockwell, G. (2015). Is it Research or is it Spying? Thinking-Through Ethics in Big Data AI and Other Knowledge Sciences. *KI -*

³⁴ Huang, K. (2022, October 6). What Does X Mean to Elon Musk? *The New York Times*. <https://www.nytimes.com/2022/10/06/technology/elon-musk-x.html>

- Künstliche Intelligenz*, 29(2), 223–232. <https://doi.org/10.1007/s13218-015-0355-2>
5. Borra, E. & Rieder, B. (2014). Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib*, 66(3), 262–278. <https://doi.org/10.1108/AJIM-09-2013-0094>
 6. Boyd, D. & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>
 7. Braudel, F. (1958). Histoire et Sciences sociales: La longue durée. *Annales*, 13(4), 725–753. <https://doi.org/10.3406/ahess.1958.2781>
 8. Bruns, A. (2019). After the 'APicalypse': social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544-1566. <https://doi.org/10.1080/1369118X.2019.1637447>
 9. Clavert, F. (2018). Face au passé: La Grande Guerre sur Twitter. *Le Temps Des Médias. Revue d'histoire*, (31), 173–186. https://shs.cairn.info/article/TDM_031_0173/pdf?lang=fr
 10. Clavert, F. (2020, March 31). Quel rôle pour les historiens et historiennes pendant une pandémie? *L'histoire contemporaine à l'ère numérique*. <https://doi.org/10.58079/plwp>
 11. Clavert, F. (2021). History in the era of massive data. *Geschichte Und Gesellschaft*, 47(1), 175–194. <https://doi.org/10.13109/gege.2021.47.1.175>
 12. Clavert, F. & Paci, D. (2023). 'Le goût d'un jour de fête'? Commemorating the end of the Second World War on Twitter during the lockdown: a comparison between France and Italy. In O. Fridman & S. Gensburger (Eds.), *The COVID-19 Pandemic and Memory. Remembrance, commemoration, and archiving in crisis* (pp. 221-244). Palgrave McMillan. https://doi.org/10.1007/978-3-031-34597-5_11
 13. D'heer, E., Vandersmissen, B., Neve, W. D., Verdegem, P. & Van de Walle, R. (2017). What are we missing? An empirical exploration in the structural biases of hashtag-based sampling on Twitter. *First Monday*, 22(2). <https://doi.org/10.5210/fm.v22i2.6353>
 14. Dagiral, É. & Pailler, F. (2018). Des chercheur-e-s et des tweets. Enquêter sous contraintes. In S. Lecossais & N., Quemeneur, N. (Dirs.), *En quête d'archives: bricolages méthodologiques en terrains médiatiques* (pp. 113-121). Institut National de l'Audiovisuel. <https://doi.org/10.3917/ina.lecos.2018.01.0114>

15. Documenting the Now. (n.d.). *DocNow*. <https://www.docnow.io/>
16. François, P., Manning, J. G., Whitehouse, H., Brennan, R., Currie, T., Feeney, K. & Turchin, P. (2016). A Macroscope for Global History: Seshat Global History Databank, a methodological overview. *Digital Humanities Quarterly*, 10(4). <http://www.digitalhumanities.org/dhq/vol/10/4/000272/000272.html>
17. Gilot, J.-M., Grandjean, M. & Clavert, F. (2018). 1914-1918: Quand la commémoration devient participative. *Le Temps des médias*, 2(31), 219–229. <https://www.cairn.info/revue-le-temps-des-medias-2018-2-page-219.htm>
18. Gomez-Mejia, G. (2020). La fabrique de la désuétude. Regards diachroniques sur Geocities et Myspace. In V. Schafer (Ed.), *Temps et temporalités du Web* (pp. 77–97). Presses universitaires de Paris Nanterre. <http://books.openedition.org/pupo/6088>
19. Grafton, A. (2011). Loneliness and Freedom. *Perspectives on History*, 43(3). <https://www.historians.org/publications-and-directories/perspectives-on-history/march-2011/loneliness-and-freedom>
20. Grison, T., Julliard V., Alié F. & Ecrement V. (2023). La modération abusive sur Twitter. Étude de cas sur l’invisibilisation des contenus LGBT et TDS en ligne, *Réseaux*, 1(237) 119-149. <https://www.cairn.info/revue-reseaux-2023-1-page-119.htm>
21. Guldi, J. & Armitage, D. (2014). *The History Manifesto*. Cambridge University Press.
22. Lévi-Strauss, C. (1962). *La pensée sauvage*. Plon.
23. Lévi-Strauss, C. & Ricoeur, P. (1963). *Autour de la Pensée sauvage. Réponses à quelques questions. Entretien du «groupe philosophique» d’Esprit avec Claude Lévi-Strauss* [Personal communication].
24. McCallum, A. K. (2002). *MALLET: A Machine Learning for Language Toolkit*. MALLET. <http://mallet.cs.umass.edu>
25. Mélice, A. (2009). Un concept lévi-straussien déconstruit: Le «bricolage». *Les Temps Modernes*, 5(656), 83-98. <https://doi.org/10.3917/lm.656.0083>
26. Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A. & Lieberman Aiden, E. (2010). Quantitative Analysis of Culture Using Millions of Digitized Books. *Science*, 331(6014), 176–182. <https://doi.org/10.1126/science.1199644>

27. Milligan, I. (2017). Welcome to the web: The online community of GeoCities during the early years of the World Wide Web. In N. Brügger & R. Schroeder (Eds.), *The Web as History: Using Web Archives to Understand the Past and the Present* (pp. 137-158). University College London Press. <https://doi.org/10.2307/j.ctt1mtz55k.13>
28. Moretti, F. (2007). *Graphs, Maps, Trees: Abstract Models for Literary History*. Verso.
29. Parisot-Sillon, C. (2020). Pandémie et goût de l'archive. L'ère du braconnage ? In F. Clavert C. & Muller, C. (Eds.), *Le goût de l'archive à l'ère numérique* (s.p.). Éditions de la Sorbonne. <https://gout-numerique.net/table-of-contents/archives-numerisees-la-salle-de-lecture-virtuelle-et-physique/pandemie-et-gout-de-larchive-lere-du-braconnage>
30. Ratinaud, P. & Dejean, S. (2009). *IRaMuTeQ: implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre* [Conference session]. Modélisation Appliquée Aux Sciences Humaines et Sociales (MASHS2009), Toulouse, France.
31. Revel, J. (2007). Histoire et sciences sociales. Lectures d'un débat français autour de 1900. *Mil neuf cent. Revue d'histoire intellectuelle*, 1(25), 101–126. <https://doi.org/10.3917/mnc.025.0101>
32. Rüling, C.-C. & Duymedjian, R. (2014). Digital bricolage: Resources and coordination in the production of digital visual effects. *Technological Forecasting and Social Change*, 83, 98–110. <https://doi.org/10.1016/j.techfore.2013.05.003>
33. Severo, M. & Giraud, T. (2019). La fabrique de la donnée géolocalisée. *Questions de communication*, 36(2), 43–61. <https://doi.org/10.4000/questionsdecommunication.20879>
34. Sikk, K. (2020, April 7). Cyberpunk perspectives on the big data paradigm and the discipline of history: Part I. The mindsets. *Digital History & Hermeneutics*. <https://dhh.uni.lu/2020/04/07/cyberpunk-perspectives-on-the-big-data-paradigm-and-the-discipline-of-history-part-i-the-mindsets/>
35. Smyth, H. & Echavarria, D. R. (2021). Twitter and feminist commemoration of the 1916 Easter Rising. *Journal of Digital History*, 1(1). <https://doi.org/10.1515/JDH-2021-1006?locatt=label:JDHFULL>