HAL
open science

Paris and Nice terrorist attacks: Exploring Twitter and web archives

Valérie Schafer, Gérôme Truc, Romain Badouard, Lucien Castex, Francesca Musiani

To cite this version:

Valérie Schafer, Gérôme Truc, Romain Badouard, Lucien Castex, Francesca Musiani. Paris and Nice terrorist attacks: Exploring Twitter and web archives. Media, War & Conflict, 2019, 12 (2), pp.153-170. 10.1177/1750635219839382 . halshs-02319683

HAL Id: halshs-02319683

https://shs.hal.science/halshs-02319683v1

Submitted on 18 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Paris and Nice Terrorist Attacks: Exploring Twitter and Web Archives

Valérie Schafer
C2DH, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Gérôme Truc
Institute for Social Sciences of Politics, CNRS, Paris, France

Romain Badouard
University of Cergy-Pontoise, Cergy-Pontoise, France

Lucien Castex
Irmeccen, Université Sorbonne Nouvelle-Paris 3, Paris, France

Francesca Musiani
ISCC, CNRS, Paris, France

Abstract

The attacks suffered by France in January and November 2015, and then in the course of 2016, especially the Nice attack, provoked intense online activity both during the events and in the months that followed. The digital traces left by this reactivity and reactions to events gave rise, from the very first days and even hours after the attacks, to a 'real-time' institutional archiving by the National Library of France (*Bibliothèque nationale de France*, BnF) and the National Audio-visual Institute (*Institut national de l'audiovisuel*, Ina). The results amount to millions of archived tweets and URLs. This article seeks to highlight some of the most significant issues raised by these relatively unedited corpora, from collection to exploitation, online stream of data to its mediation and re-composition. Indeed, Web archiving practices in times of emergency and crises are significant, almost emblematic, loci to explore the human and technical agencies, and the complex temporalities, of 'born-digital' heritage. The cases examined here emphasize the way these "emergency collections" challenge the perimeters and the very nature of Web archives as part of our digital and societal heritage, and the guiding visions of its governance and mission. Finally, the present analysis emphasizes the need for a careful contextualisation of the design process – both of original web pages or tweets and of their archived images – and of the tools deployed to collect, retrieve and analyse them.

Keywords

Terrorism, Attacks, Web, Twitter, Born-Digital Heritage, Heritagization, Web Archives

Introduction

> 'We need history because we must have rest. A pause so our consciousness can rest, so that consciousness is still a possibility – not just the seat of thought, but of practical reason, giving latitude for action.' (Boucheron, 2016, our translation)

From the moment when *'Je suis Charlie'* ('I am Charlie') appeared on the online social networks, when they vibrated intensely (Boullier, 2016) in the first days of January 2015, the impression – both individual and collective – was of an event erupting before our eyes. Within just a few days staff of the French Web legal deposit at the National Library of France (*Bibliothèque nationale de France*, BnF) and at the National Audio-visual Institute (*Institut national de l'audiovisuel*, Ina) had launched specific collections, and in November 2015 this took only hours. Other archiving followed in 2016, dedicated, for example, to the July attacks in Nice.

From the first information to appear, via research into victims, expressions of shock, reflection and polemic, to tributes and commemorations, all the different temporalities of online reaction went into the Web archives, providing a panorama of both official and popular reactions through the traces left on the Internet and on Twitter.

At the same time these reactions were entering other archival spaces, with the collection, for example, of press articles, radio and television broadcasts, administrative archiving or the gathering of tributes to victims left in public places, in the scenes of attacks and elsewhere (Bazin, 2017; Sánchez-Carretero, 2011), as carried out by the Paris Archives after the attacks of 13 November.[1] What might have appeared to be ephemeral, streaming past, became part of an archive, a heritage, as well as a source for research.

This archiving of the ephemeral, a documentation of 'the now' - in the words of the 'Documenting the Now' project initiated by a number of US universities after the death of Michael Brown and the Ferguson events,[2] is not without precedent. In 2001, following the 9-11 attacks, the 911digitalarchive[3] had been launched by US researchers. But the range and the sheer number of items collected in the wake of the 2015 attacks (12 million tweets archived by Ina in January 2015, 20 million in November 2015, and 8 million for Nice in July 2016), the reactivity of all that 'real-time' action and its repetition a few months later, represented a

departure. The massive use in 2015 of social networks that had not been part of the online landscape in 2001 offers new potential for research into the social response to attacks, alongside other resources such as the audio-visual interviews collected as part of the French 13 November research programme.[4]

The present article aims to shed light on the temporalities and implications of both archiving and research drawing on Web and Twitter archives, from the decision to collect these materials through to their first exploitation from 2016 onwards by a number of research teams.[5] This reflection seeks to open up the organisational and technical 'black boxes' of the archives and to analyse the stakes, the limitations and the methodological perspectives for researchers. This paper is organised around a three-part timeframe, from the decision to launch emergency collections through their becoming available for consultation and finally for analysis, and aims to analyse the challenge to researchers inherent in the nature, aims and content of all this documentation and data.

1. Real-Time Archiving and 'Documenting the Now'

Since the attacks of January and then of November 2015, numerous different actors from institutions, politics, the media and civil society have been mobilised to take individual or collective responsibility for the traces of terrorism's eruption in the heart of Paris. One should not forget, of course, that the French capital had previously experienced a wave of attacks in the mid-1990s and that French people had been affected by those carried out elsewhere on European soil since 11 September 2001, especially the events of Madrid in 2004 and London in 2005 (Truc, 2017). The Web had already been used at the time of these earlier attacks and its archives contain their traces, at least to a limited extent, as Jane Winters has shown (2016) in relation to the London attacks, drawing on the Web archives of the British Library and on the Internet Archive.[6] However, the 2015 attacks were followed by unprecedented emergency collection campaigns. These were launched very quickly and in some cases continue today with the inclusion of commemorative activity, for example the *#enmémoire* (in memory) collection.

1.1 Emergency collections

In the first days of January 2015, the reactivity of Web archivists in France to events was intense. Very quickly, in addition to the regular daily collections from the online press by the BnF's Web legal deposit (*Dépôt légal du Web*), and of online audio-visual content by Ina,[7] collections specifically dedicated to the attack on the offices of *Charlie Hebdo* were launched.

From the day after the attack, Ina established a collection based on Twitter, while the BnF sought to capture a broad range of reactions to events from the Web (tributes, support, analysis, critical or hostile reactions). In the following days the BnF launched an appeal for suggestions both among the network of correspondents of its own Web legal deposit, who worked on specific themes and would continue to point out content of archival interest over the following months, and on the international network of the International Internet Preservation Consortium.[8]

Figure 1: Emergency collection on *Charlie Hebdo* attacks announced by Christine Genin (BnF) on Twitter on 8 January 2015.

Between 8 and 16 January 2015, BnF collected 1,581 different Web domains on the basis of suggestions submitted. The library also decided to make a supplementary thematic collection based on one of its own selection tools which targets content from official publications, political parties, religious and protest movements. The majority of the content collected in these specific operations would otherwise have been missed in the course of regular annual collection that takes place in the autumn.

Following the attacks that happened on 13 November, the BnF and Ina again launched emergency collections in order to capture a sampling of Web reactions. The BnF added 18 new sites to its holding (such as https://www.facebook.com/PoliceNationale/ and http://www.defense.gouv.fr), as well as 43 Twitter accounts or hashtags (such as #Place_Beauvau and #attackParis), as well as including again some of the sites identified after the January attacks. This collection over five days, with a capture from Twitter four times a day and from other sites once a day, resulted in the gathering of 1.5 million URLs. Meanwhile at Ina the most significant work was with Twitter, where the autumn 2015 collection was even larger than that carried out at the start of the year. Individual researchers also carried out emergency collections whose results were added to holdings, which we discuss further below.

1.2 A long-term approach

Emergency collection campaigns are not a completely new phenomenon for archivists. At the BnF such campaigns had been launched earlier by the Web legal deposit in order to follow particular issues such as the controversy around the planned Notre-Dame-des-Landes airport in the west of France or the *Mariage pour tous* (same-sex marriage) campaign and law of 2013 (Le Follic, 2016).

Emergencies happened before the digital age. As Agnès Magnien, Director of Collections at Ina, has recalled, there had already been real-time collections of archival material on public policies such as the emergency social fund in the 1990s (emergency aid for people at risk). Another example is the emergency collections of material from ministerial offices when ministers change with every change of government. 'In short, a lesser known quality of archivists has always been their reactivity and especially, therefore, an acute awareness of current events!' (Magnien, 2015).

However, the online collections of material on the attacks have added new challenges to those already identified in the course of other exercises in archiving, notably of printed material: the streaming and volume of data, the difficulty of defining the parameters of what to capture, the need to react almost instantly, for example where Twitter is concerned, as well, sometimes, as difficulties with collection (problems with archiving Facebook posts, disappearance of content of Periscope, deletion of online messages).

These emergency collections are a challenge to archivists in terms of both time and space.

Space insofar as the repercussions of events may go beyond the usual frontiers of the legal deposit, with content circulating in different languages, countries and media. And time because, while every 'event' has a beginning, its end is harder to define. Both the BnF and Ina have therefore opted for 'long' timescales in terms of the sudden eruption of an event. Ina's collections from Twitter of hashtags like #Bataclan or *#enmémoire* (in memory) are ongoing and these archives continue constantly to grow, whenever there are hashtags recalling an earlier event when a new attack takes place (#Charlie at the time of Bataclan, or #Bataclan at the time of Nice, etc.) or the reprise of the *#porteouverte* (open door – offering hospitality) hashtag, first used in November 2015 in Paris and then in July 2016 in Nice.

2. From Collections to Corpora

The advent of online social networks clearly seems to mark a change in social reaction to terrorist attacks, which makes the archiving of their content even more important and necessary.[9] However, the real-time selection of data or the focus on Twitter clearly have implications for the thinking of researchers compiling their corpora.

2.1 Human selection and curation

Although results may look similar, not all hashtags collected have been subject to the same collection process. For example *#jenesuispascharlie* (I'm not Charlie) did not form part of Ina's first collection in 2015, but was collected retrospectively some months later when the Institute realised its interest to researchers. Some hashtags may not immediately seem relevant at a given moment, but may with time become crucial to the analysis of certain waves of opinion. Conspiracy theories or messages of support for terrorists have become a source for analysis of the counter-discourses emerging in the wake of an attack. But these circulate in spaces parallel to the main threads of discussion and with different hashtags (such as *#cheh*, meaning 'well done' in Arabic, after the *Charlie Hebdo* attack). They will not be archived unless identified as of potential research use. It may, therefore, be useful to involve researchers from the start of a collection process, in order to be aware of the broadest possible range of hashtags – Ina, in particular, has fostered such collaboration through the organisation of workshops involving both researchers and archivists.

Working with the Twitter IDs archived by the Canadian Nick Ruest[10] or by Linkfluence, Ina has carried out a 'catch-up' collection of *#jenesuispascharlie* (compiling its own archive from these IDs). Following the publication of an article by Giglietto and Lee (2015) devoted to the

#jenesuispasCharlie (I'm not Charlie) hashtag, Ina was able to complete its collection with the IDs used in this research and collected by the authors. This experience and sharing of IDs should encourage both researchers and archivists to reflect on best practice in the sharing of data and corpora. The BnF meanwhile shared its selected URLs with Archive-It (part of the Internet Archive Foundation). Archive-It has thus been able to archive some of the websites highlighted by different institutions.[11] It should be noted that in none of these cases are the actual archives shared, but only lists of the IDs and URLs held, allowing each institution to then make its own collection and thus ensure both the authenticity of every element collected and the technical homogeneity of each collection's methodology.

The case of *#enmémoire* is also an interesting one. Although it had not been separately archived at the beginning, the Ina collections already included 4,856 tweets associated with other hashtags, like #Paris, but also containing this hashtag. Examples of hashtags not previously selected can thus be located and may even be strongly represented. Ina has now retrospectively archived 13,709 tweets with this hashtag and continues to collect these.

Finally, we should note the heterogeneity of hashtags. Tags like #Paris, *#fusillade* (shooting) or *#attentat* (attack) are common and likely to have a widespread presence in archives. The hashtag *#nice* also poses problems, since it may designate the French city whose seaside promenade was the site of an attack on 14 July 2016, but equally may refer to the English adjective 'nice' and be attached to much anodyne and unrelated content. Other hashtags like *#porteouverte* or *#boycottBFM* are more specific. This last, directed at the rolling news channel BFM after the Nice attack of July 2016 (but also reappearing regularly during the 2017 presidential election campaign), also illustrates Ina's focus on the audio-visual field at the heart of its archiving perimeter.

2.2 Exhaustiveness versus representativeness

In addition to the limits of their selection criteria, collections are representative rather than exhaustive. This is notably the case with the BnF, which captures a sample from Twitter only four times a day, but also with Ina, although their Twitter collections take place more frequently. Ina, in fact, has chosen to work with the public Application Programming Interface (API) of Twitter, resulting in a vast capture limited only by Twitter's own. In practice, only 1% of the tweets posted at any moment can be collected free of charge via the API, which means that when there is a paroxysm of tweets reporting or reacting to events some are lost. According to the head of Ina's Web legal deposit,

> …At the height of the online stream in November we were only able to collect about a quarter of tweets, or even less. But the most "important" tweets tend to be retweeted, which multiplies the chances of capturing them; it's mostly the weak signals that get missed… There is currently a lot of research into the missing data, but in order to be worthwhile this should aim to be not exhaustive, but representative. For Twitter we collected limited messages, this information is archived, and notably indicates the difference between a collection by an individual and that of an archiving institution that will seek to qualify and quantify its archive. (Drugeon, 2016)

Even the 20 million tweets preserved by Ina do not constitute an exhaustive collection of everything that happened on Twitter around 13 November. On a purely statistical level, however, this limitation is not critical. The corpus is of such a size that it can be considered representative of all messages (according to random sampling methodology).

2.3 Limitations of a sociological nature

Limitations of a sociological nature can also affect representativeness. Furthermore, we do well to keep in mind Dominique Boullier's remarks underlining the extent to which these 'moments of collective effervescence have their own ecology (or echology, as Deleuze suggested)', public platforms being quite specific, 'as are the readers of a particular journal or people conversing in a particular bistro. But these traces may under particular conditions give us access to a particular process hitherto unaccounted for' (Boullier, 2016, our translation).

The choice of Twitter as principal terrain of collection partly comes down to its API, which provides the technical facilities to record all tweets and, moreover, they all belong legally in the public domain. On the sociological level, Twitter users as a group are far from representative of the population as a whole. In France, where according to the audience survey company Médiamétrie around 6 million people are active on Twitter, there is an over-representation of men (55%, with only 45% women), of young people (59% are under 34 years old), and in terms of location (one third live in the Ile-de-France region).[12] So the reactions to attacks available on Twitter are those of a limited sector of this limited population. On the other hand, Facebook is much more popular as a social network – popular in both senses of the term: many more French people are on Facebook (four times as many as on Twitter) and they are socio-demographically more diverse (although old people are under-represented). Facebook, meanwhile, only allows access to messages posted as 'public'. The possibilities in terms of archiving and constitution of a corpus from this social network are thus limited by concerns for the privacy of its users (Latzko-Toth & Pastinelli, 2013).

2.4 'Reborn-digital materials'

As historian Niels Brügger has noted (2016), the archiving of the Web implies multiple stages, in the course of which transformations may occur, to such an extent that he suggests we should be speaking of 'reborn-digital materials'. Researchers working on the born-digital heritage[13] must therefore take account of every stage in the archiving process that may modify the data being studied. Examples of this include the modifications and updates a website may undergo in the course of a collection exercise, temporal leaps when following a link between archived web pages, instances of duplication when menu buttons are not systematically reconstructed, logos or calendars, or the failure to archive certain elements (often pop-ups or advertising banners). The original form of data undergoes a number of mediations and transformations before it becomes the archive we consult as researchers (Schafer, Musiani & Borelli, 2016) and it is no less important to identify, analyse and trace back such mediations when working on an emergency collection.

We consider here some examples, the first of which is the very different forms in which tweets are archived by the BnF and by Ina. The former adopts a process, which preserves the Twitter environment and the result looks more like a screenshot, while the Ina archive highlights data and metadata, with the stream of conversation disappearing. Although it is possible to reconstruct conversations, in the Ina database they do not appear as a continuous stream, and background images disappear and are replaced by a data-tree structure. By reflecting data more than streaming, the archive here departs from the user's original browsing experience.

Another significant factor that researchers cannot ignore is the fact that retweets of a message may continue after it is archived and will not then be counted, preventing any evaluation of its popularity except at the date it is archived. We should also consider the difficulty of making sense of retweets, since communication in an emergency situation is different from the 'ordinary' usage of a network, as analysed for example in the typology proposed by danah boyd and colleagues (2010).

One final example is the fact that Ina, when it first began to share its data, chose not to include emojis until researchers pointed out the importance of retrieving the traces of these visual elements much used during events to emphasise or stand in for words. From the miniature pencil or the French flag that accompanied the *#jesuischarlie* hashtag to the joined hands that went with #prayforParis at the time of the Bataclan attacks, these elements, which

had in fact been archived, were reintegrated with content, and engineers at Ina's Web legal deposit subsequently even offered a tool for researching the emojis captured as part of a collection.

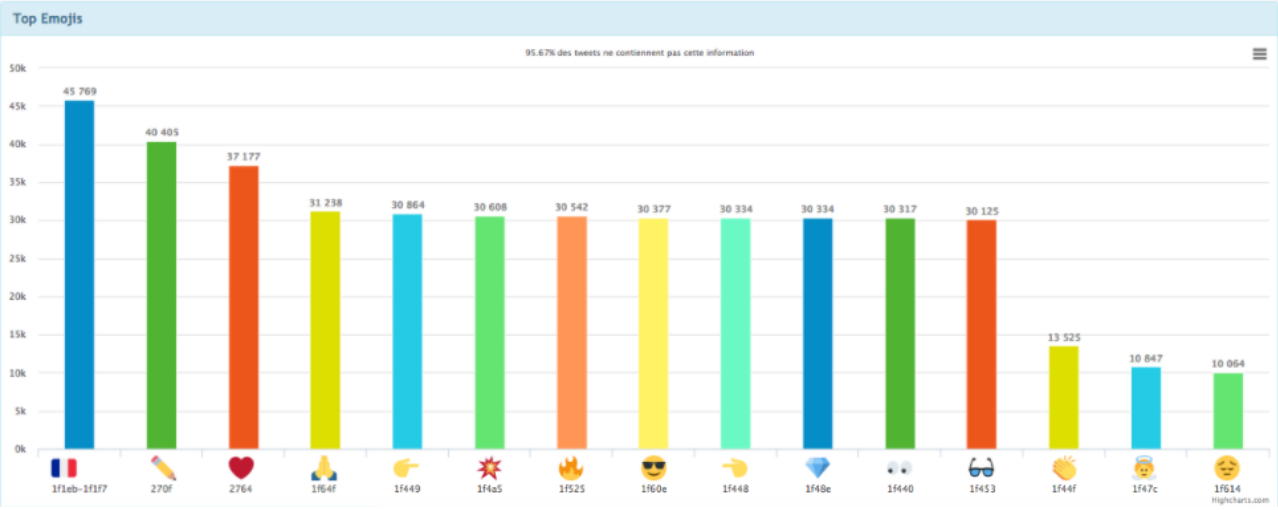

Figure 2: Search by emojis for *#jesuischarlie* © Ina

3. From Flow to Frame

Faced with these collections, whose size we have emphasised, there is an urgent need for tools of analysis. However, the tools and interfaces employed in analysis also imply mediation and the archiving institutions assume a central role, providing as they do both the data and the tools for exploiting it. Researchers must therefore exercise vigilance and make an effort to understand both the contributions these tools make to the analysis and the ways they influence it. In fact, just as there is no such thing as raw data (Gitelman, 2013), there is no such thing as a neutral tool. Methodology thus becomes a crucial issue.

3.1 Tools of analysis

Users, researchers, cannot just 'take' data from the Web legal deposit. The tools available need to meet their needs as closely as possible. This, as Thomas Drugeon notes (2016), gives rise to a two-fold concern:

> In most cases, the users who come to consult a Web legal deposit collection see it as one more source among many used in their research and will not expend an enormous amount of effort on

understanding its limitations. Some, however, will want to go further. We thus find ourselves torn between these more specific needs and those of the majority of users, for whom a tool will become incomprehensible if it is too specialised… there are nearly as many different requirements and tools as there are searches and researchers.

Ina therefore provides an infrastructure based on Elasticsearch and Kibana, allowing advanced searches of the data in its collections. Researchers can make use of an advanced research syntax in composing their requests. It is possible to use keywords, Boolean operators, meta-characters or facets, as well as to cross-reference metadata or to carry out a full-text search.

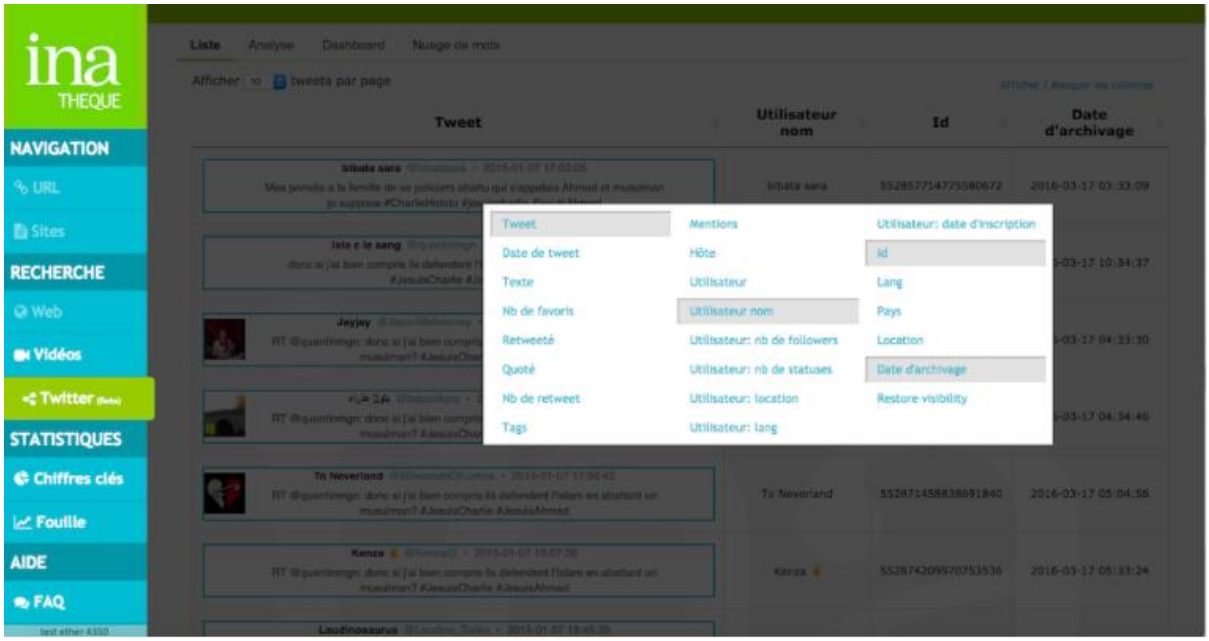

Figure 3: Search by cross-referencing metadata © Ina

Different tools of analysis are also available: timelines, word clouds, search for images (or for emojis, as discussed above), descending classifications…

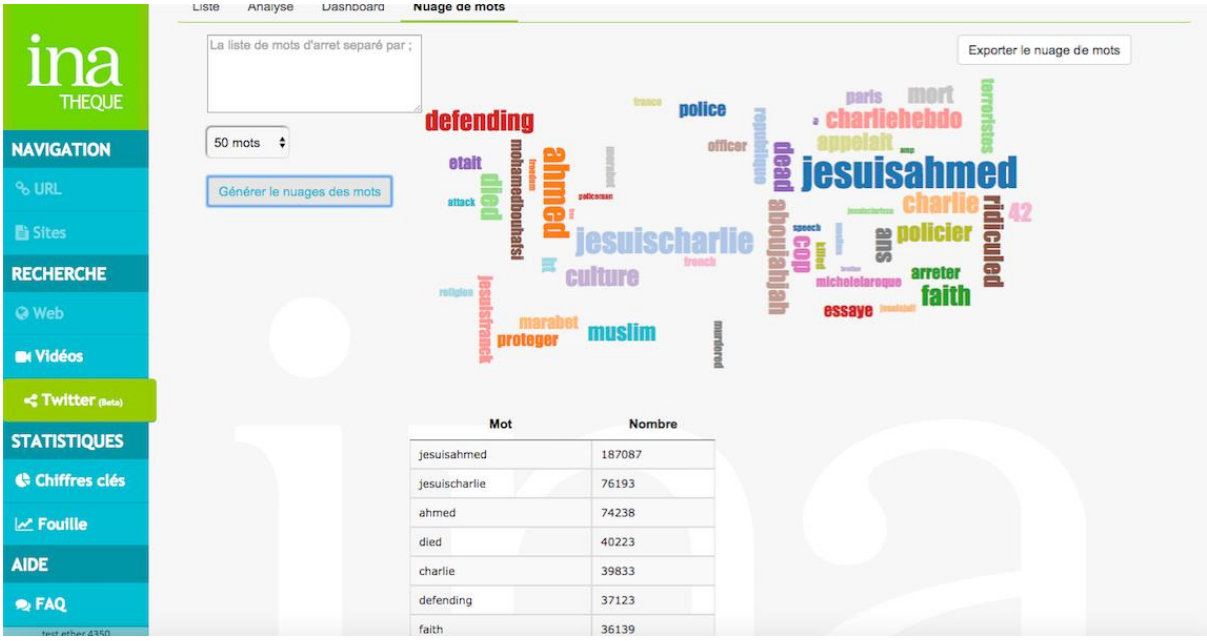

Figure 4: Tag Cloud for dedicated search for *#jesuisahmed* ('I am Ahmed') © Ina

Meanwhile the BnF – whose corpora contain both Web and Twitter archives, whereas Ina has kept the two collections separate – has also focused on facets, allowing more refined searches and the combination of selection criteria, as well as on full text.

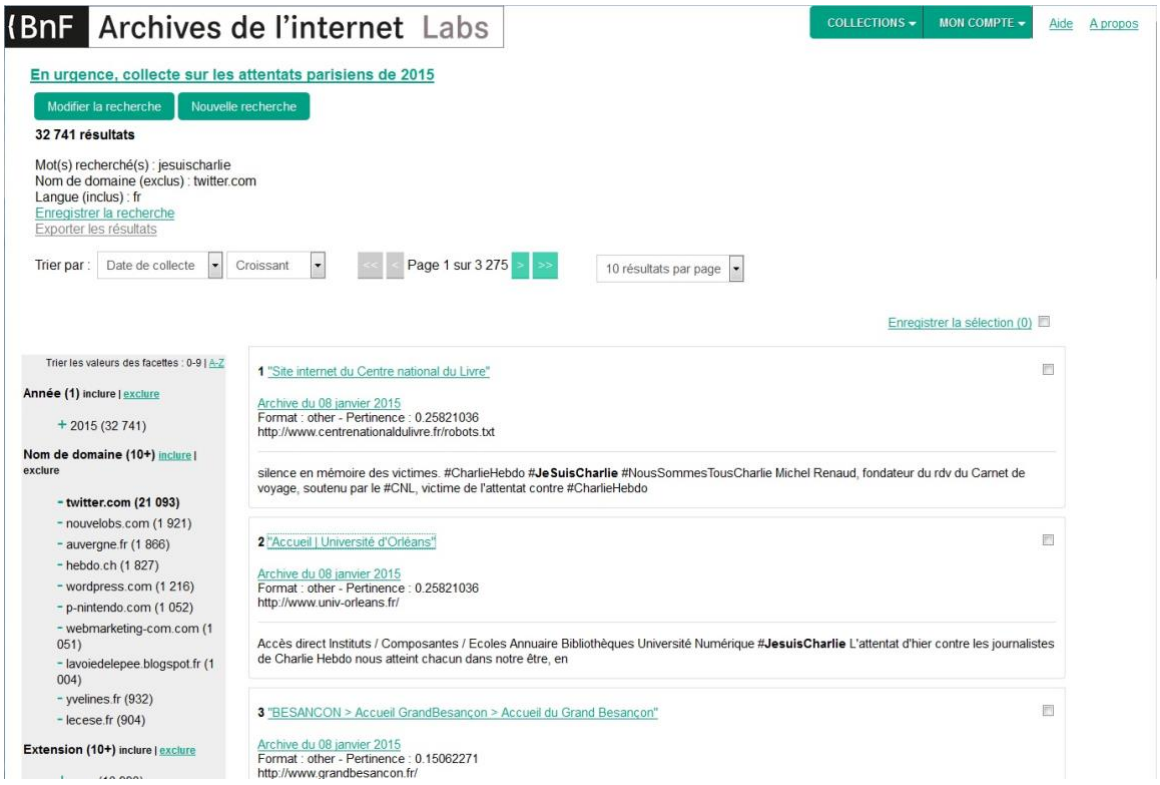

Figure 5: Full text and search by facets © BnF

Institutional archiving of this kind, moreover, does not rule out researcher-led initiatives, such as Romain Badouard's project on *#jenesuispascharlie* or the work carried out for the ANR's (*Agence nationale de la recherche* / National Research Agency) ENEID (*Éternités numériques. Identités numériques* / Digital eternities, digital identities) project[14] on the night of 13 November 2015 by Lucien Castex, who collected from the Twitter API stream in order to study the process of online commemoration. In these cases the raw data in Javascript Object Notation (JSON) format, representing more than 30 items of metadata as well as the text of the tweet (number of mentions, data, ID of tweet, ID of user, location…), was loaded into Elasticsearch in order to allow the kind of search possible at Ina. Research has been conducted with the open-source programming language R and different packages (MongoDB, Rjson, FactoMineR, ggpplot2, tm…) in order to carry out a lexical and statistical analysis of the corpus, with researchers able to draw on a whole range of tools of their choice.

3.2 Small and big data, close and distant reading

While constraints apply and choices are made – both technical and sociological – in the use of social networks as sources for the analysis of movements of opinion, the way that researchers use these sources also always implies choices. The question of adopting a mainly qualitative or quantitative approach is central, though not the only one. However, few of the analyses carried out to date have been defined by such a binary opposition. On the contrary, it is evident that 'big data' approaches allowing statistical and lexical analysis of very large corpora tend to be supplemented by other methodological approaches that can go into finer detail.

One example suffices to illustrate this point: the hashtag *#jesuiskouachi* (I am Kouachi – the Kouachi brothers carried out the *Charlie Hebdo* attack) appears to have been used more than 49,000 times on the day of the attacks, becoming a 'trending topic'. This highlights the limitations of using Twitter and its trending topics as a barometer of opinion. As demonstrated by journalist Jean-Marc Manach in a study published on the *Arrêt sur Images* (Freezeframe) website, this numerical status shows only how much and how persistently a hashtag is used, not the context in which it is used. A message denouncing a particular hashtag paradoxically helps to ensure its visibility. Jean-Marc Manach has shown how the popularity of *#jesuiskouachi* was ensured by right-wing and far-right activists intent on denouncing its use and repeatedly stating how popular the terrorists were among immigrant groups in France (Badouard, 2016).

Thus, as Niels Brügger has noted with reference to Franco Moretti's close and distant reading (2013),

> Although Moretti in the main uses the distant reading approach to the study of large amounts of digital data, I will argue that neither of these two approaches are per se inscribed or inherent in the digital material. By this I mean that simply because collections of digital material are in many cases big data, which opens up the possibility of asking and answering new types of research questions, this does not necessarily mean that they have to be approached as Big Data. (Brügger, 2015: 11)

Here, for example, an ethnography of online conversations can be very useful as a supplement to work on archived Tweets. By placing oneself in a situation of online observation and interaction, and taking account of the context of interaction, such as the technical constraints the medium places on speaking out, an online ethnography can allow us to describe in more detail the way opinions are expressed and conversations develop (Jouët & Le Caroff, 2013; Pastinelli, 2011). Meanwhile, the coding methodologies employed in analysis of online discussions (Monnoyer-Smith & Wajcik, 2012; Stromer-Galley, 2005) can allow us both to quantify the registers of expression employed by users and to qualify the messages exchanged. They thus also allow us to observe how the opinions expressed are negotiated in a discussion and can lead to the adoption of collective opinions.

3.3 Multi-temporalities, multi-spaces, multi-media

Juxtaposing data collected online with other materials documenting a posteriori off-line or non-digital reactions also allowed an enrichment of the analysis. In terms of the attacks of January and November 2015, a number of supplementary sources can be mobilised to this end. Firstly, written sources such as messages collected from the 'physical' places of the 13 November attacks by the Paris Archives (and now for the most part freely available for online consultation) may be subjected to the same types of analysis – both qualitative and quantitative – as the content collected directly from the Web. This has been done already with similar materials following the attacks in Madrid in 2004 and London in 2005 (Sánchez-Carretero, 2011; Truc, 2011, 2017).

Then there are the observations and photographs gathered at demonstrations and meetings after the attacks, from around memorial sites etc., especially in the course of the REAT

(*Recherches sur les réactions aux attentats* / Research on Reactions to the Attacks) project[15] coordinated by Gérôme Truc or, in relation to the attacks of January 2015, those collected as part of the Charlie Archive at Harvard University.[16] And finally there are individual and group interviews, such as those carried out for the Sensi-TV-T project[17] coordinated by Claire Sécail on the reception of TV images of the attacks of January and November 2015, or those conducted for the CNRS/INSERM (*Centre national de la recherche scientifique/Institut national de la santé et de la recherche médicale* / National Centre for Scientific Research/National Institute of Health and Medical Research) 13 November programme of research on memories of these attacks, which is also supported by Ina.

There is much to be gained from integrating born-digital materials into a broader empirical data ecosystem. The fact is that reactions to terrorist attacks can rarely have been as intensively documented as those to the attacks of January and November 2015. It is a challenge of real significance for researchers to come up with modes of research and of multi-media and multi-modal analysis capable of taking account of the complexity of the social process under way.

It is when juxtaposed with other empirical materials that born-digital data may be of maximum heuristic interest. Therefore it is to the collections made in real time from the Web, and especially from Twitter, at the time of these attacks that we may look for some progress towards a more finely tuned apprehension of the temporality of social reaction. We should thus be able to understand how not only the key words but also the register of reactions, the debates, polemics and so forth arising from events, circulate and pivot from one space to another. What continuities and breaks can we observe between what happens 'online' and what happens 'offline'?

On the basis of the 9-11 case, Randall Collins has already been able to demonstrate how only after 48 hours of astonishment and confusion will we see a collective response really start to organise itself around the slogans and symbols that then occupy the centre of attention in the weeks that follow, until a society slowly and gradually returns to normal – after six to nine months, if unaffected by any new attack (Collins, 2004). Today, the data at our disposal concerning reactions on Twitter to the attacks of January and November 2015 can give us a more precise grasp of what is at play in the first four hours after an attack. From this point of view the dynamic of hashtags is very instructive.

With the attacks on 9 September 2001, the slogan *'nous sommes tous Américains'* (we are all Americans) came from an editorial by Jean-Marie Colombani published the day after in *Le Monde*. On the other hand, *Libération*'s headline *'Nous sommes tous Charlie'* (we are all

Charlie) on the day after the attack on *Charlie Hebdo* was already too late – the phrase *'je suis Charlie'* (I am Charlie), which first appeared on Twitter within a few hours of the event, was adopted and will go down in history (Truc & Fléchet, 2016). First published by designer and journalist Joachim Roncin in the form of an image with *'je suis Charlie'* in white capital letters on a black background, the phrase was instantly taken up as a hashtag. It rapidly went viral and the hashtag was used on the day of the attacks nearly 3.5 million times. The phrase also quickly spread beyond the social network and by the evening of 7 January on the Place de la République, where tens of thousands had spontaneously gathered, people were chanting the slogan and waving the printed image. On TV screens and as a headline in daily newspapers, a slogan first seen on Twitter rapidly became the symbol of a national outpouring of emotion (Badouard, 2016).

The process was less clear after the attacks of 13 November. It is more difficult to say exactly which wording stood out then from the social networks: 'pray for Paris'? *'je suis Paris'* (I am Paris)?, *'nous sommes unis'* (we are united)? The director of Ina's Web legal deposit remarked (Drugeon, 2016) at the time on the wide range of hashtags compared with those that appeared after the *Charlie Hebdo* attack and rapidly came together as *#jesuischarlie* (Merzeau, 2015). The work of an archivist in such circumstances involves considerable vigilance and sensitivity to rapidly evolving trends. It is important, in particular, to be able to identify the 'carrier' of a hashtag, that is not only who comes up with it or uses it for the first time, but who gives it significance and ensures it is taken up and spread. In terms of the hashtags that appeared after 13 November 2015, the collections made by Ina allow us to distinguish the 'popular' hashtags originated by members of civil society (often personalities like singers, sportspeople or journalists), as in the cases of #prayforParis and *#jesuisParis*, and 'institutional' hashtags originated by institutions (embassies, ministries, Paris City Hall), as with *#noussommesunis* or *#fluctuatnecmergitur* (tossed but not sunk – motto long associated with the city of Paris) (Truc, forthcoming).

Using the tools of analysis developed by Ina's Web legal deposit, we can distinguish three main types of hashtags that appear in the time immediately after an attack. First, there are 'informational' hashtags arising from the need to locate and qualify an event that has just happened. On the evening of 13 November 2015, these were hashtags like #Paris, *#fusillade* and #Bataclan, which appeared with the first factual tweets. We came to know about the event as it unfolded from people who witnessed it directly. Very quickly, through exchanges on the social network between witnesses and journalists, details were revealed: the word 'shootings' gave way to 'attacks' and the hashtag #Parisattacks appeared at about 10 pm,

designating an event whose nature was now clear. At the same time, we began to see 'organisational' hashtags, by means of which a collective response to the crisis situation (help and support) was organised. These included, on the same evening, the hashtag *#porteouverte* (also used after the Nice attack), whose use peaked around midnight and then declined quickly when no longer of immediate use (we might compare this use of Twitter to the 'safety check' on Facebook). Also in this category were hashtags like *rechercheParis* (search Paris) and *#recherchePersonne* (search for person), later used to help find missing persons. Finally, once the event had been defined as a terrorist attack, 'compassionate' hashtags also started to appear and then continued to develop exponentially, rather than declining, over the following hours and days. These last enabled the rest of the Twittersphere to express support for victims. The first, that same evening, was #prayforParis, coming out of the US, which appeared in some 6.7 million tweets in the space of ten hours (source: Twitter). Its competition, from the following morning when France awoke to the shocking news, was *#jesuisParis*, inspired by *#jesuischarlie* from January 2015 (Truc, forthcoming). After 24 hours, these were the most widespread hashtags in tweets about the events. This marked the beginning of a stabilisation phase in the phrases, slogans and symbols of reaction and solidarity that dominated the scene in the following days.

When it came to calls for resistance and reconstruction, as expressed by the use of *#NousSommesUnis* (we are all united), *#TousAuBistro* (everybody to the bistro) and then *#MonPlusBeauSouvenirduBataclan* (my best Bataclan memory), these hashtags were of a different nature and came from different users. *#NousSommesUnis*, for example, took off from the moment it was used by Anne Hidalgo and the City Hall of Paris. A study of all these elements can also help to enrich our understanding of the mediatisation of death, which has also been studied in a journalistic context (Dakhlia, Quemener & Castex, 2016).

Conclusion

Terrorism's appearance at the heart of Twitter in January 2015 gave a new impetus to the institutions archiving the Web, which had already begun to consider the need for emergency collections, and a new legitimacy to their role as collectors not only fulfilling a legal deposit requirement but now also charged with preserving a collective memory. These 'critical moments' interrogate the perimeters and the very nature of the Web archives as part of our digital heritage. Moments of tension and testing, as represented by these 'emergency

collections', help to 'open the black boxes' of archiving, shedding light on how the guiding visions of our archiving institutions, of their governance and mission, and indeed of the Web itself (alive and archived), come together and can only be studied together as part of a complex but coherent ecosystem.

Beyond the importance of considering the visual and material specificities which frame the context of production and accessibility of pages or data, understanding this born-digital heritage implies an understanding of what web pages, websites, social networks, and even hyperlinks, have meant in their time.

We have also highlighted here the importance of the design process – both of an original web page or tweet and of its archived image – and of the tools deployed to collect and retrieve this. France is not the only country where these questions are being considered: Belgian and British researchers have been posing similar questions (see in particular Chambers et al., 2017).

Many questions remain to be explored and could have been further developed here, particularly with reference to the conservation of data and potential sharing, preserving and citing of the corpora collected, or to the ethical factors clearly raised. These collections also pose emphatically the question of developing digital literacy both among sociologists and among the SIC (*Sciences de l'information et de la communication* / Information and Communication Science) researchers or historians wishing to make use of this data both now and in the future. Demanding profoundly interdisciplinary approaches and a constant dialogue with archivists, they open up a fertile terrain for research at a moment when 'what we need really is *actualité* in Michel Foucault's sense: what we are becoming. This is happening now, between noise and oblivion' (Boucheron & Riboulet, cited in Calvet, 2015, our translation).

Funding

This study has been carried out within the framework of the ASAP and REAT projects, funded in 2016 by the French National Centre for Scientific Research, and of the Web90 project supported by the French National Research Agency (ANR-14-CE29-0012-01).

Notes

1 See website of Paris Archives http://archives.paris.fr/a/92/les-hommages-aux-victimes-des-attentats-sont-en-ligne/ Accessed 14 June 2017.

2 Presentation of project on 'Documenting the Now'. http://mith.umd.edu/introducing-documenting-the-now/
3 September 11 Digital Archive. Supported by several partners including Library of Congress, Smithsonian National Museum of American History, Behring Center. http://911digitalarchive.org/about
4 http://www.memoire13novembre
5 Following an appeal by the President of the CNRS, Alain Fuchs, in November 2015, after the Bataclan attacks, several dozen interdisciplinary projects on the attacks were funded, including REAT (*Recherches sur les reactions aux attentats* / Research on Reactions to the Attacks, https://reat.hypotheses.org), coordinated by Gérôme Truc, and ASAP (*Archives Sauvegarde Attentats Paris* / Back-Up Archives of the Paris Attacks, http://asap.hypotheses.org), coordinated by Valérie Schafer.
6 Founded in 1996 by Brewster Kahle, the Internet Archive Foundation has been archiving the Web since then. The archives of the Web may be consulted freely via the Wayback Machine. https://web.archive.org?
By mid-July 2017 more than 29 million web pages had been archived.
7 Since 2006 the BnF and Ina have shared responsibility for the Web legal deposit and its archiving. Ina's mission is to preserve audio-visual and associated content, while the BnF is responsible for archiving all other content created in France, amounting to several million websites, all of which are not collected with the same regularity, the online press being collected daily while other sites are only subject to annual collection.
8 IIPC members are organisations from more than 45 countries, including national, university and regional libraries and archives. http://netpreserve.org/about-us/
9 By agreement between Twitter and the Library of Congress, the latter today holds the largest archive of Twitter. Other institutions began collections on national, account or hashtag bases at the behest of Ina in 2014.
10 Nick Ruest's site contains many notes dedicated to these collections. http://ruebot.net/tag.charliehebdo
11 *Charlie Hebdo*, Archive-It. https://archive-it.org/collections/5190
12 See blogdumodérateur (moderator's blog) on Mediamétrie survey of July 2016. http://www.blogdumoderateur.com/chiffres-twitter/
13 The expression 'born-digital heritage' enables us to distinguish this from content 'converted into digital form'. See Unesco charter of 2003. http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/preservation-of-documentary-heritage/digital-heritage/concept-of-digital-heritage/
14 http://eneid.univ-paris3.fr
15 https://reat.hypotheses.org/le-projet-reat
16 Site of Charlie Archives project at Harvard. http://cahl.io/
17 On SENSI-TV-T (Sensitivity to TV treatment of terrorism) project, see https://sensitvt.wordpress.com

References

Badouard R (2016) Je ne suis pas Charlie. Pluralité des prises de parole sur le web et les réseaux sociaux (I'm not Charlie. The plurality of voices on the web and social networks). In

Lefébure P and Sécail C (eds.) *Le défi Charlie. Les médias à l'épreuve des attentats* (The challenge of Charlie. Media tested by the attacks). Paris: Lemieux Editeurs. Available at: https://hal.archives-ouvertes.fr/hal-01251253/document (accessed 5 November 2017).

Bazin M (2017) Quand la rue prend le deuil. Les mémoriaux éphémères après les attentats (When the streets go into mourning. Ephemeral memorials to the attacks). *La Vie des idées*, 26 May. Available at: http://www.laviedesidees.fr/Quand-la-rue-prend-le-deuil.html (accessed 5 November 2017).

Boullier D (2015) Charlie est un phénomène de 3ème génération (aussi) (Charlie is (also) a 3rd generation phenomenon). *SHS 3G*. Available at: http://shs3g.hypotheses.org/114 (accessed 5 November 2017).

boyd d., Golder S and Lotan G (2010) Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. *HICSS-43. IEEE:* Kauai, HI, January 6. Available at: https://www.danah.org/papers/TweetTweetRetweet.pdf (accessed 5 November 2017).

Borelli M and Schafer V (2016) Entretien avec Annick Le Follic, Chargée de collections numériques au département du dépôt légal de la BnF (Interview with Annick Le Follic, Web legal deposit at the BnF), *ASAP*. Available at: http://asap.hypotheses.org/168 (accessed 5 November 2017).

Boucheron P (2016) *Ce que peut l'histoire* (What history can do). Paris: Fayard/Collège de France.

Brügger N (2016) Digital humanities in the 21st century: Digital material as a driving force, *Digital Humanities Quarterly*, Preview, Volume 10, N° 2. Available at: http://www.digitalhumanities.org/dhq/vol/10/3/000256/000256.html (accessed 5 November 2017).

Brügger N (2015) *Humanities, digital humanities, media studies, Internet studies: An inaugural lecture.* Aarhus: CFI.

Calvet C (2015) Patrick Boucheron et Mathieu Riboulet : La question aujourd'hui est « qui est nous ? » plutôt que « qui est Charlie »? (The question today is 'Who is we ?' rather than 'Who is Charlie ?') , *Libération*, July 10. Available at: http://www.liberation.fr/societe/2015/07/10/ patrick-boucheron-et-mathieu-riboulet-la-question -aujourd-hui-est-qui-est-nous-plutot-que-qui-est-ch_1346030 (accessed 5 November 2017).

Chambers S, Mechant P, Vandepontseele S and Isbergue N (2017) Aanslagen, Attentats, Terroranschläge: Developing a special collection for the academic study of the archived web related to the Brussels terrorist attacks in March 2016. *RESAW Conference*, London. Available at: http://netpreserve.org/wac2017/abstracts/#abstract49 (accessed 5 November 2017).

Collins R (2004) Rituals of solidarity and security in the wake of terrorist attack, *Sociological Theory*, 22(1): 53-87.

Dakhlia J, Quemener N and Castex L (2016) Rythmes de mort (Rhythms of death), *Temporalités* 23. Available at: https://temporalites.revues.org/3368#quotation (accessed 5 November 2017).

Giglietto F and Lee Y (2015) To be or not to be Charlie: Twitter hashtags as a discourse and counter-discourse in the aftermath of the 2015 Charlie Hebdo shooting in France. #Microposts2015 · 5th Workshop on Making Sense of Microposts · @WWW2015. Available at: http://ceur-ws.org/Vol-1395/paper_12.pdf (accessed 5 November 2017).

Gitelman L (ed.) (2013) *Raw data is an oxymoron*. Cambridge MA: The MIT Press.

Jouët J and Le Caroff C (2013) L'observation ethnographique en ligne (Ethnographic observation online). In Barats C (ed.) *Manuel d'analyse du web en sciences humaines et sociales (*Manual of web analysis in the human and social sciences). Paris: Armand Colin, pp.147-165.

Latzko-Toth G and Pastinelli M (2013) Par-delà la dichotomie public/privé : la mise en visibilité des pratiques numériques et ses enjeux éthiques (Beyond the public/private dichotomy: Making digital practices and their ethical implications visible). *tic&société*, Vol. 7, N° 2. Available at: http://ticetsociete.revues.org/1591 (accessed 5 November 2017).

Magnien A (2016) Email exchange with V. Schafer.

Merzeau L (2015) #jesuischarlie ou le médium identité (#jesuischarlie or the identity medium), *Médium* n°43: *Charlie et les autres* (Charlie and the others), 2015/2. Available at: https://halshs.archives-ouvertes.fr/halshs-01121510 (accessed 5 November 2017).

Monnoyer-Smith L and Wojcik S (2012) Technology and the quality of public deliberation: A comparison between on and offline participation. *International Journal of Electronic Governance*, Vol. 5, N° 1: 24-49.

Moretti F (2013) *Distant Reading*. London/New York: Verso.

Pastinelli M (2011) Pour en finir avec l'ethnographie du virtuel! Des enjeux méthodologiques de l'enquête de terrain en ligne (An end to virtual ethnography! Methodological implications of researching the online field), *Anthropologie et Sociétés*, Vol. 35, No 1-2: 35-52. Available at: https://www.erudit.org/fr/revues/as/2011-v35-n1-2-as5004414/1006367ar/ (accessed 5 November 2017).

Sánchez-Carretero C, Cea A, Díaz-Mas P, Pilar M and Ortiz C (2011) On blurred borders and interdisciplinary research teams: The case of the "Archive of Mourning". *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(3), Art. 12. Available at: http://nbn-resolving.de/urn:nbn:de:0114-fqs1103124 (accessed 5 November 2017).

Schafer V, Musiani F and Borelli M (2016) Negotiating the Web of the past. Web archiving, governance and STS. *French Journal for Media Research.* Available at: http://frenchjournalformediaresearch.com/index.php?id=952 (accessed 5 November 2017).

Schafer V and Borelli M (2016) Entretien avec Thomas Drugeon, responsable du DL Web à l'Ina (Interview with Thomas Drugeon, head of Web legal deposit at Ina), on ASAP website. Available at: https://asap.hypotheses.org/tag/ina (accessed 5 November 2017).

Strommer-Galley J (2005) *Decoding deliberation online*. Paper presented at the Second Conference on Online Deliberation, Palo Alto, California.

Truc G (2011) Analyser un corpus illisible ? Le logiciel Alceste confronté à des registres de condoléances (Analysing an unreadable corpus? Alceste software faced with online condolence registers). *Langage et Société*, 135: 29-45.

Truc G (2017) *Shell shocked. The social response to terrorist attacks*. London & New York: Polity Press.

Truc G and Fléchet G (2016) Terrorism: Testing our connected solidarity. *CNRS News*. Available at: https://news.cnrs.fr/articles/terrorism-testing-our-connected-solidarity (accessed 5 November 2017).

Truc G (2017, forthcoming) Title to be decided. Ina Global.

Winters J (2016) Archiving the 7/7 bombings (7 July 2005). *ASAP Symposium*, Paris, November 29. Available at: http://asap.hypotheses.org/240 (accessed 5 November 2017).

Author biographies

Romain Badouard is Associate Professor at the University of Cergy-Pontoise (France). His researches mainly address the study of mobilizations and public opinion movements on the web and social networks.

Lucien Castex is a Research Associate at Université Sorbonne Nouvelle - Paris 3, Research Institute for Media, Culture, communication and Digital (Irmeccen). His main research focuses on Privacy, Media regulation and Internet Law with a particular interest on compliance, policy making and governance. He is secretary-general of the French Chapter of ISOC.

Francesca Musiani is tenured research fellow at the French National Centre for Scientific Research (CNRS), Institute for Communication Sciences (ISCC). Her main research interests are Internet governance (more specifically, the role of infrastructures and architectures of the Internet as tools of governance), encryption and other technologies of privacy protection 'by design', and digital heritage and heritagization.

Valérie Schafer is Professor of Contemporary European History at C2DH (Luxembourg Centre for Contemporary and Digital History) at the University of Luxembourg since February 2018. She was previously researcher at the French National Centre for Scientific Research (CNRS), where she led the ASAP project dedicated to the French Born-Digital Heritage of Terrorist Attacks. Her main research interests are the history of Internet, of the Web, of Digital Cultures and the Born-digital heritage.

Gérôme Truc is a sociologist, tenured research fellow at the CNRS - Institute for Social Sciences of Politics. Specialized in moral and political sociology, his work focuses on social responses to terrorist attacks and memorialisation process of terrorist attacks. He is the author of *Shell Shocked. The social response to terrorist attacks* (Polity Press, 2017).