

ARTICLE



#Cuéntalo: the path between archival activism and the social archive(s)

Vicenç Ruiz Gómez^a and Aniol Maria Vallès^b

^aBoard member of the Society of Catalan Archivists and Records Managers, Catalonia, Spain; ^bMember of the Open Data and Transparency Working Group, Society of Catalan Archivists and Records Managers, Catalonia, Spain

ABSTRACT

The goal of this article is to explain our experience, as Society of Catalan Archivists and Records Managers (AAC) members, in the field of social web archiving. To that end, we have structured it in three main parts, the first of which is to show the importance of archival science as a political tool in the framework of the information society. The second part focuses on the path followed by the AAC from its first steps taken to preserve social web hashtags, in order to gain technical expertise, to the reflection on the theoretical background required to go beyond the mere collection of social web content that led us to the definition of a new type of archival fonds: the social fonds. Finally, the third part sets out the case study of #Cuéntalo. Thanks to our previous experiences, this hashtag, created to denounce male violence, enabled us to design a more robust project that not only included the gathering and preservation of data but also a vast auto-categorisation exercise using a natural language processing algorithm that assisted in the design of a dynamic and startling data visualisation covering the 160,000 original tweets involved.

KEYWORDS

Archival activism; big data; social archive; male violence; natural language processing

Introduction

On 26 April 2018, a Spanish Court imposed an incredibly light sentence on five men, known as ‘The Wolfpack’, who were convicted of gang-raping a young woman. The sentence aroused a wave of indignation across the country. Two days later, the journalist Cristina Fallarás published the original tweet on the hashtag #Cuéntalo (the Spanish version of #MeToo), aimed at women, giving a first-person account by a woman who had suffered sexual abuse. At that time, we were already leading an ongoing project to monitor and record socially relevant hashtags promoted by the Society of Catalan Archivists and Records Managers, from which we are members, the most significant being the one arising from Les Rambles terrorist attack in Barcelona, #NoTincPor (#IamNotAfraid), and those linked to the #CatalanReferendum, using the methodology and tools developed by ‘Documenting the Now’. But it was with #Cuéntalo that we decided to explore a model of social media archiving for the gathering, contextualisation and diffusion of such data, since we understood that this hashtag was created as a digital

community archive and used as a tool of reparation and civic empowerment in the fight against male violence.

Traditionally, archives are considered the by-product of an activity, but the digital transformation of society – as will be explained in the first section – has turned ‘recordkeeping’ itself into a civic/political tool and, even, a means of production. This *material archival turn*¹ should oblige us as professionals to choose sides clearly: to collaborate with surveillance capitalism mechanisms or to cooperate with the exercise of digital sovereignty that can only be achieved by helping people to become their own archivists. It is no longer simply a technical matter of compiling records, but also an ethical and political issue.

In fact, and this is the object of the second section, this societal transformation urges us archivists to redefine basic principles, such as provenance, and even attempt to define new types of fonds, since new creators that do not share the juridical and individual nature of the traditional ones appear. By collectively using social media platforms, they produce massive fonds, as in our case study: #Cuéntalo. As we will explain in the third part, through capturing and categorising the 160,000 original tweets included in #Cuéntalo hashtag, we have succeeded in transforming data into information, but we have not yet been able to turn this information into social evidence,² since we have been unable to convince any institution, either public or private, to fund the participatory recordkeeping platforms and practices needed to build a real social archive. In order to clarify some concepts, we must say that we do not use ‘archive’ in the Foucauldian or Derridean sense, but rather as an ‘archival institution’, which includes ‘fonds’ but also ‘archivists’ (although in this case, we should more properly speak about ‘archivers’, as proposed by Ketelaar)³ and the ‘policies’ defined for the arrangement of records and ‘archives’. Hence, the term ‘social archive’ does not refer to a ‘social media’ collection, but to an archive built by social and not just ‘professional’ consent.

Infosphere and surveillance capitalism: the archive society

The so-called fourth industrial revolution has for the first time in history made it possible to have at our disposal massive recording tools, such as social media platforms, that allow all strata of society to provide documentary evidence of their activities. Apparently, this has brought about the end of the monopoly on archives held by states, large corporations and social elites. We say ‘apparently’, because it is highly probable that in the future we will be unable to access all the records generated today by the social networks. This does not just affect each one of us individually; it is on a collective level that it has the most far-reaching consequences. The horizontal and free nature of platforms such as Twitter, Facebook and Instagram make them almost the only media (together with instant messaging apps such as Whatsapp, Telegram or Signal) through which social movements and non-institutional groups document and record their existence and activity.

As a result, what we are witnessing is the consolidation of the network society that Manuel Castells defined as ‘a society whose social structure is made of networks powered by microelectronics-based information and communication technologies’;⁴ in other words, a society in which practically everything we do can be catalogued, communicated and processed in the form of data. However, this is not something that can be done in

a planned and ordered way by each of us individually, but rather is undertaken by agents consisting mainly of governments and above all by large corporations. In order to grasp the extent of this transformation, the contributions made by Luciano Floridi and Shoshana Zuboff are especially important.

According to Floridi, we are indeed experiencing the fourth revolution, but one that is not only industrial but also epistemological; about our very conception of the world itself. Thus, the transformations wrought by Copernicus (the Earth is not the centre of the universe), Darwin (evolution equates us with all other living beings) and Freud (the unconscious shows that the human mind is not rational), are now being succeeded by the omnipresence of information technologies that by means of the infosphere has brought us to the threshold of a new period, hyperhistory (equivalent to the passage from prehistory to history).⁵

Floridi understands the infosphere as ‘the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations’.⁶ This does not only refer to cyberspace but also includes the analogue spaces of information and ‘offline’. In fact, in the broadest sense of the term, he is suggesting that the infosphere could become a synonym of reality, given that the third-order technologies on which the fourth revolution is based make us think about the world informationally and make the world we experience informational.⁷ The degree of development of the infosphere is that which indicates the transition towards the hyperhistory of society, where ‘ICTs and their data-processing capabilities are not just important but essential conditions for the maintenance and any further development of societal welfare, personal well-being, and overall flourishing’.⁸

Nevertheless, this momentous transformation is not taking place in a politically or economically neutral environment. Floridi devotes some attention to the political context, but not so much to the economic situation, about which we believe that Zuboff’s analysis is of greater assistance. Floridi provides a brief history of political power, starting with the Peace of Westphalia (1648), which ushered in the era of sovereign states, understood as the main agents of information that legislate for or at least seek to control the technological means involved in the life-cycle of information, such as education, the census, taxes, criminal records and so on.⁹ However, in the long run, with this purpose and this control, states will gradually drive us towards an information society in which other institutional agents across the world have already gained the upper hand, a state of affairs endorsed by the Bretton Woods agreements. Furthermore, we are currently witnessing the emergence of a new ‘informational order’ consisting of a multi-agent system largely dominated by big corporations who compete with states for the control of political power.¹⁰

By analysing four factors (power, geography, organisation and democracy), Floridi shows how ICTs are undermining the predominant role played by states. We wish to stress the significance of ICTs and states for the impact they have on the reformulation of important archival concepts, such as ‘producer’, and linked this to the principle of provenance itself. Where organisation is concerned, Floridi shows how ICTs ‘fluidify’ the topology of politics, given that they promote ‘through management and empowerment, the agile, temporary, and timely aggregation, disaggregation, and reaggregation of distributed groups “on demand”, around shared interests, across old, rigid boundaries, represented by social classes, political parties, ethnicity, language barriers, and so forth’.

Indeed, these groups may acquire enormous influence in a purportedly direct democracy, but one which in practice achieves a consensus through the media, especially via social networks.¹¹ Needless to say, while being fluid and temporary, these 'on-demand' groups generate a trail of their activities, and thereby become producers of documentary evidence. The fact that they may not constitute a legal entity or possess a defined organic structure does not mean that these activities lack importance (in fact, Floridi tends to maintain the opposite); neither does the apparent incoherence of their documentary procedures prevent their records from being understood as fonds.

As mentioned above, Floridi does not analyse the economic logic as exhaustively as politics or technology. Shoshana Zuboff, on the other hand, does focus her attention on that issue, on the understanding that all the transformations connected with the new technologies and Big Data arise from a new phase of capitalism. Accordingly, she proposes the concept of Surveillance Capitalism, understood as the form of capitalism that by means of massive data extraction seeks 'to predict and modify human behavior as a means to produce revenue and market control'.¹² Strictly speaking, Big Data is not a matter of a technological effect, but rather constitutes the very basis of a new logic of accumulation that converts the populace into a resource by means of an extractive process carried out 'in absence of dialogue or consent despite the fact that they signal both facts and subjectivities of individual lives'.¹³ Indeed, it is precisely these subjectivities that endow Big Data with its value, given that the main business of the large technology companies consists of auctioning them off to their real clients, who are in fact other companies that wish to sell their products to consumers whose everyday habits and patterns of consumption they require for that purpose. In other words, advances in the computerised analysis of massive data spring primarily from the need to optimise the different phases of production in which information returns to its source (people) in the form of *surveillance assets*, with the aim of generating profit and, through the associated transactions produced thereby, trigger a new cycle of extraction.¹⁴ The capacity of producing this new type of assets is that which provides the big technological companies with value and, more accurately, the new tycoons of surveillance capitalism.¹⁵

In short, whether we situate ourselves in the interpretive framework of the infosphere or that of surveillance capitalism, it seems clear that the creation, reception, maintenance, use and elimination of information (all that is referred to in our discipline as recordkeeping or simply archiving) currently go beyond the mere reflection of an activity and becomes the means of production, accumulation and control of an entirely new social logic. We find ourselves at a point in time where the position of each one of us in the social hierarchy depends on those who design and control the flow of information and those who generate knowledge and decide how it is to be reused. Thus, we are living in an archival society. But the problem is that we neither design nor control the flow of our information: it is not *our* archive. As Amelia Acker and Adam Kriesberg have pointed: 'Social media platforms support active communities where citizens access, create and engage with information, news and contemporary digital culture ... Yet most social media platforms lack long-term preservation and access strategies for citizens, journalists and research institutions that support scholarly inquiry and access to reliable evidence and authentic information'.¹⁶

The type of archival practice we prioritise will, therefore, affect not only the management of information and the documentary heritage but also may increasingly become

a tool of civic counterpower. It is precisely here where the shift from archival activism to the design of the social archive is found: it is not just a matter of technology, but also a political and ethical issue. Empowering people with archival autonomy¹⁷ is one of the primary goals that we as professionals should strive to achieve in the current context of surveillance capitalism, where every facet of our identity and our activities flow through public and private data platforms. Precisely because society itself has become an archive, it is in the interests of democracy that we should endeavour to make people their own archivists.

Obviously, this goal is currently beyond the scope of the legal framework of most Catalan (and, by extension, worldwide) archival institutions, since they are constrained by political mandates or funding limitations. It is for that very reason that three years ago the AAC, as a non-governmental professional association, established a social web archival activism project.

The AAC approach to social media archiving

In this second section, we aim to explain the particular path followed by the AAC in order to gain expertise in social web archiving, not only from a technological perspective but as a means to reflect on its civic implications. Or, more precisely, the implications for archival theory and methodology when used as a tool for social activism. First, we will summarise our first experiences, especially, around two events that had a great impact on Catalan society between August and October 2017: the terrorist attack on Les Rambles and the Catalan Referendum of Independence. In the second part of this section, we will focus on the theoretical and methodological framework, developed from the learnings and limitations of these experiences, that later allowed us to face the #Cuéntalo challenge.

Our first steps in social web archival activism

In 2014, the second author of this article presented ‘El tractament arxivístic de Twitter’ (‘The Archival Processing of Twitter’), the Final Project of his Master course at the Escola Superior d’Arxivística i Gestió de Documents (School of Archival Science and Records Management), supervised by professor Joan Soler, director of the Arxiu Històric de Terrassa (Terrassa Historical Archives).¹⁸ In this work, he analyses various cases concerning the preservation of social networks, most of which are theoretical, but some already put into practice and currently under development.

This initial experience has enabled us to learn about the limitations involved in dealing with the documentation of social networks, not only in regard to digital technology but also in the legal sphere and in issues of access. Gaining access to the original Twitter database proved to be difficult, at least for us, since the budgets we are obliged to work with in archiving in Catalonia made the task unaffordable.

However, it was very clear from that work that there was indeed a system which, if pursued, could yield some useful results, and that was through the Twitter API (Application Program Interface). This was an undertaking that given its new versions has turned out to be increasingly laborious,¹⁹ due in part to the growth in interest in this system and also because Twitter is exercising a stricter control over the information it stores.²⁰ Nevertheless, these restrictions have to some extent stabilised connections with

the social network and have protected Twitter from the frequent crashes to which it was subject in its early days.

Thanks to the initial approaches made in these early experiences, we were able to try out Twitter functions, such as downloading the personal archives of each user or taking a look at formats like JSON and CSV that were unknown to us, but which today are commonly used in open data. For example, NARA archives are recommending CSV for 'Plain-text delimited or marked-up structured data files' and 'Structured Data/Spreadsheets'.²¹ However, in spite of the progress, we were making in this area, our professional activity in the following years led us to work in other archival fields and those initial approaches were put on hold.

As a result of the terrorist attacks in Barcelona and Cambrils, in 2017 the AAC undertook a line of research involving activity devoted to exposing the need to preserve the documentary evidence generated by the public through the social networks. Thus, two days after the attacks, the first author of this article posted a communication on his professional blog in which he outlined the need to preserve and process from an archival perspective the content generated on the social networks through the hashtag *#NoTincPor* (*#IamNotAfraid*), with the message 'do not be afraid of democratically building the documentary heritage' as 'part of the struggle against terror and against all those who (one way or another) seek to profit from it', and to 'work together to preserve the memory of these (and other) attacks by means of the archives, understood as trustworthy institutions, through the social participation and contextualisation of information as a means of reparation and a "Work of Mourning"'.²²

In this same article, one of the examples that stand as a basic reference was the Documenting the Now collective²³ and its development of tools and a methodology for the capture and preservation of data. Inspired by the methodology and theoretical ideas of this collective, set up as a result of the protests in Ferguson, Missouri, USA to create a record of the *#BlackLivesMatter* movement,²⁴ and thanks to their *Twarc*²⁵ tool for archiving Twitter, we focused our efforts on the collection of content in the *#NoTincPor* hashtag between 17 and 28 August 2017, which consisted of more than 50,000 tweets. It was, thus, a first experience that allowed us to preserve the data produced in this social platform, but lacked of any deeper attempt to understand its archival nature.

Subsequently, on the occasion of the Referendum held on 1 October 2017, we tried out a new approach. Since it was possible to foresee the celebration of the referendum and the social impact it would have, on 20 September, the AAC put out an initial call for the archives in the country to participate actively in the collection of fonds generated by the public at large, in order to complete the information produced institutionally. In parallel with this, and in collaboration with Ed Summers, a professor at the Maryland Institute of Technology and a member of *DocNow*, we organised the monitoring and gathering of the main hashtags associated with the *#CatalanReferendum*. The *#ArxivemelMoment*²⁶ (*#ArchivingTheMoment*) campaign attracted a significant following (with the participation of more than forty archives) and, as a result of the activity on the networks, yielded a dataset enabling eight and a half million tweets to be recorded between 19 September and 5 October, amounting to some 56 gigabytes with JSON profiles, the most ambitious data capture we have achieved to date.

We continued our monitoring of different data preservation projects, keeping a close eye on the various ideas and suggestions that arose from different sources, whether those

coming from institutional archives (the Library of Congress or the UK National Archives) or from collectives such as those mentioned above (Documenting the Now and others), which, while receiving institutional support, also put forward projects of a more activist nature.

Our closest contacts were with Documenting the Now, because they possessed tools that were made public with an open-source licence, which enabled us to use them for our own project. This was important, since not only did it provide us with a theoretical framework and a project for the future, it also showed us that we could begin to take action and undertake live data capture. That was how, after the initial experience with *#NoTincPor* (*#IamNotAfraid*) and *#CatalanReferendum*, we finally determined to carry on the work with these tools and maintain constant contact and share our experiences.

We repeatedly conducted trials during the final months of 2017 and throughout 2018, with a somewhat hacker mentality, in order to test the tools and instal them on different platforms: the most unusual case was to try them out in Google Cloud Shell,²⁷ a virtual machine provided by Gmail to all its users, but not very widely known. However, we were also doing trials with Twitter to test the limitations with its API, analysing what kind of data we were receiving and the information that could be extracted from it. That was how we discovered the one-week limit for capturing tweets, as well as the time it took for the system to download what we had requested from the API. Thus, we eventually decided to reproduce our work environment in the most popular operating systems. The fact that the tool we chose, Twarc, was very simple, intuitive and fully operational for our purpose, enabled us to use it in the most popular operating systems, and together with the availability of Ubuntu Shell on Windows,²⁸ the possibilities of working smoothly multiplied even more.

This work with Twarc also determined the method we would use to conduct the data capture, chosen from those available through this software.²⁹ If we chose 'filter', the live capture function, it would oblige us to keep the capture window open and hope that the system would not fail at any time. This option was restrictive, because it only allowed us to harvest current metadata rather than those that were added or updated in the following days. For that reason, we adopted the 'search' function, which enabled us to plan data collection and draw up a schedule to cover the periods of interest, as well as to leave a few days margin to include and update metadata from the tweets. At the same time, we had to devise a model for describing the captures, based on the metadata that are usually required for the publication of datasets, consisting of a brief description of the content, the steps for how the datasets are to be collected and formatted, as well as a licence and the minimal descriptive data of their volume and format. Thus, we were able to say that by early 2018 we had a working method that enabled us to deal with large-volume data capture using our very own means.

Defining a social fonds

In parallel with the acquisition of technical expertise, the AAC embarked on a process of theoretical reflection with the aim of advancing archival processing beyond harvesting in a dataset of one or more hashtags. Thanks mainly to the experience gleaned with *#CatalanReferendum*, we were able to arrive at a definition – provisional, while awaiting the indispensable critical response of the archival community – of what we refer to as

social fonds and in which we have identified three processing priorities. We understand the social fonds to be the entire body of records created by a community of users through a participatory dynamic on social platforms around a shared interest or event. This new type of documentary fonds is characterised as being produced by a collective creator – recall that in traditional fonds, whether personal, familial or commercial, etc., the creator is a single juridical or physical person – and generated from a private or exclusive platform where the producers are not owners, and done so by means of a short-term formation process, but nevertheless producing a large volume of documents. The three processing priorities that derive from the nature of this type of fonds are as follows:

a) a participative model of archival processing is essential for its collaborative generation;

b) it is necessary to adopt proactive measures to ensure free and democratic access in the mid- and long term; and

c) given its massive nature, it is vital to be equipped with tools that automate the analysis (evaluation, description) and use of the information.

Bearing in mind both the definition and the main features of social documentary fonds, we realised that in order to build the most suitable framework of action, it was necessary to have recourse to the entire set of theoretical and methodological approaches generated by post-custodial archiving. For the sake of brevity, we will not attempt to summarise this archival current. Instead, we prefer to mention the main contributions that have influenced our approach.

Thus, the first characteristic to be addressed was its generation on a private platform. As outlined in the previous section, our main reference point was Documenting the Now, although we have also followed other projects and solutions that Morgan E. Currie and Britt S. Paris have called archival data activism.³⁰ Perhaps one of the best examples is the Environmental Data and Governance Initiative (EDGI), created with the aim of counter-acting Donald Trump's denial of climate change. Among the measures adopted by EDGI, which form part of the community-driven Data Refuge project, is the design of a web application (ArchiversSpace) for managing the entire life-cycle of datasets and storing them in an open public archive. Likewise, by means of the Data Together data management model, all the captured datasets can be processed in a distributive and precipitative way.³¹

However, capturing data from private platforms is just the first step in transforming surveillance assets into civic counterpower. It is necessary to take into account the second characteristic of social fonds: they consist of a collective creator made up of all the users on a social platform who have generated the documents. The material existence of a collective creator reinforces the idea of societal provenance put forward by Tom Nesmith,³² which Jeanette Bastian extended with her definition of a community of records 'as the aggregate of records in all forms generated by multiple layers of actions and interactions between and among the people and institutions within a community. Layers of records parallel the active life of the community itself'.³³ Anne Gilliland has adapted, in part, these reflections to conceptualise some types of 'networked records' produced in digital network milieus. It is the case of the *multi-provenance bureaucratic record* and the *record created by the crowd* that clearly challenge the traditional notion of provenance and pose serious problems for 'the descriptive standards community [that] has resisted building more

complexity into descriptive standards regarding provenance... Born-networked, multi-provenancial records, such as those generated by organisational or scientific research collaborations, or within large-scale social media or other Web 2.0 ... make it impossible to continue to ignore this issue, no matter how problematic it might be for archival arrangement and description practices'.³⁴

As we see, intellectual principles (social provenance, parallel provenance, community of records) in the offline world have become an empirical reality in our *onlife* society.³⁵ The logic behind the social fonds is assimilated into that of the community archives, given that these are non-institutional fonds produced by (often) subaltern collectives. Likewise, the fonds are generated not only in terms of memory but also as an act of civic and political affirmation.³⁶ It is precisely here where the archival bond endows the document generated with coherence.

As we soon shall see, #Cuéntalo could be conceived as a community archive, since most of the tweets are posted by victims of violence that, at least while the hashtag remains active, make up a community of records.³⁷ However, the difference with the traditional community archives is that they arise from the use of a private technological platform and the fluidity of the community itself. Social fonds creators can usually be more closely identified with 'on-demand' infosphere fluid groups than with the stable (though subaltern) communities of records. W. Lance Bennett and Alexandra Segerberg typified this leap from traditional collective action to the logic of connective action developed through social media platforms.³⁸ Contrary to the former, connective action does not require great organisational resources neither previous social network relationships. Instead, social media become 'flexible organizations in themselves, often enabling coordinated adjustments and rapid action aimed at often shifting political targets, even crossing geographic and temporal boundaries in the process'. In this sense, both authors pointed the need that the digital devices and content used 'may remain behind on the web to provide memory records or action repertoires that might be passed on via different mechanisms associated with more conventional collective action such as rituals or formal documentation'.³⁹

In fact, it is especially from a recordkeeping perspective that these groups can become communities: their participation throughout the entire execution of archival processing is the *sine qua non* when it comes to generating community awareness by creating their own arsenal of accountability⁴⁰ with the aim of achieving a greater social justice.⁴¹

In this sense, the theoretical and methodological framework developed by Monash University around different projects of obvious social relevance is incontestably an example to follow. In particular, the participatory recordkeeping models designed by Livia Iacovino and Greg Rolan, the latter furthermore forming part of a holistic project such as the 'Archives and Rights of the Child', based on the records continuum paradigm.⁴² But, since there is no institutional mandate to start the archival processing of social fonds and their creators may be unaware of this initial processing, the 'rights approach' must be completed with the theoretical model of radical empathy proposed by Michelle Caswell and Marika Cifor, in which archivists play the role of 'caregivers, bound to records creators, subjects, users, and communities through a web of mutual responsibility'.⁴³ Mainly for a social fonds like #Cuéntalo, where it was essential 'to remind ourselves not to erase differences between bodies, and not to reinforce hierarchies that permanently position some as caregivers and others as care recipients'.⁴⁴

Finally, the massive volume of data from the social fonds requires the use of automated analysis procedures, such as algorithms. In this regard, it is worth emphasising the pathway opened up by a group of archival, computer engineer and data scientist professionals who proposed the formal structure of Computational Archival Science, defined as a ‘transdisciplinary field concerned with the application of computational methods and resources to large-scale records/archives processing, analysis, storage, long-term preservation, and access, with the aim of improving efficiency, productivity, and precision in support of appraisal, arrangement and description, preservation, and access decisions’.⁴⁵ In other words, in the absence of automatic categorisation techniques, based on natural language processing algorithms, for example, or on data visualisation algorithms, it is almost impossible to undertake the archival processing of the social fonds. We need to be able to work with algorithms that enable us to obtain representative vectors for keywords, for example, and by learning how to use this mechanical operation (unattainable with human capacity alone) we achieve the statistical development of a body of content, thereby obtaining a comprehensible series of graphs and visualisations.⁴⁶ In this way, we could certainly capture Big Data (millions of tweets in a dataset), but it would still be difficult to transform it into information, and finally into evidence.

These three concepts – data, information and evidence – are often employed as equivalents, but if we really value our discipline we would do well to differentiate between them carefully.⁴⁷ As recently summarised by Laura Millar: ‘*data* are some combination of elements of raw content, such as numbers or letters, and *information* is contextualised data, or data infused with layers of meaning. A *record* captures information or data in a fixed medium; it is a ‘whole’ thing: an email, a report, or a text message. *Evidence* is any source of information that provides demonstrable proof. We cannot say that a Facebook post is only information, or a photograph is always evidence, or a database is just data. If the source – data element, photograph album, or membership database – can be used to provide proof of actions, transactions, or decisions, then it has evidential value’.⁴⁸

#Cuéntalo: a failed (?) attempt to build a social archive

In accordance with these relations between data, information, records and evidence, we will divide the exposition of our archival project for #Cuéntalo into two parts. The first concerns what has already been done; that is, capturing the massive data and turning it into coherent information. The second concerns what should be done in order to turn those social web records into social evidence.

From data to information

The sentence in the ‘Wolfpack Case’ was made public on 26 April 2018 and triggered a wave of indignation. The hashtag #NoesAbusoEsViolación (#ItIsNotAbuseItIsRape) emerged on Twitter, and others such as #JusticiaPatriarcal (#PatriarchalJustice) were recovered. This initial reaction was superseded by an authentic tsunami when Cristina Fallaràs (@LaFallaras a Twitter) posted the first tweet with the #Cuéntalo hashtag in which she urged women victims of abuse and rape to speak out.⁴⁹

As explained in previous sections, the AAC has extensive experience in the capture of hashtags on Twitter, and we were quickly able to see that this outcry of indignation could

provide a fresh opportunity for obtaining highly useful datasets for those engaged in work and research into massive data, otherwise known as Big Data. The *#NoesAbusoEsViolación* hashtag was the first to be captured; it was, in fact, the first to appear and was extremely active during those early days. Shortly afterwards, however, it became apparent that *#Cuéntalo* was gaining a much greater and more resounding following, clearly outstripping the earlier hashtags.

On this occasion, seeing that it was a hashtag with enormous potential, it was decided that instead of conducting a single isolated capture, a long-term capture would be much more appropriate. The plan was to harvest tweets over intervals of days, for example, on weekdays and at weekends, in order subsequently to unite all the tweets recorded in a single dataset. However, the difficulty was that this strategy had to be carried out against the clock, given the narrow margin of just a few days allowed by Twitter for the capture of tweets, not to mention the sheer volume that had to be downloaded in such a short space of time. Thus, the data captured during the first few days were extended to cover whole weekends, thereby exhausting until the last minute the 7-day window of opportunity for freely obtaining the datasets, thus avoiding their purchase at an unaffordable cost.

The capture of the *#Cuéntalo* dataset lasted for two weeks, from 27 April to 13 May 2018, with the result of 2,111,998 tweets and 12.8GB, figures which would open up a whole series of research possibilities, primarily because of the impact of the communication surge and equally for the evolution of the exchange of messages over subsequent days. It would, therefore, be possible for us to enter the field of data visualisation and contribute a proactive archival vision, with this ambition, and in the knowledge that we could now move from two dimensions, where we possessed visualisations of point-to-point conversations (from nodes to vectors), to three dimensions, where we could see the complete volume of conversations (both in depth and dimension) and their changes over time.

Nevertheless, we still needed help in order to take this next step, so we contacted the journalist and data activist Karma Peiró whom we knew for her extensive experience in data processing and visualisation. She explained that contacts already existed among journalists, but what they needed was someone like us who had managed to obtain the data, and so everything began to fall into place. It was also thanks to her that we got in touch personally with Cristina Fallarás (who originally launched the *#Cuéntalo* hashtag), who immediately provided us with the chance to broaden the scope and capacity of the project. At the same time, we realised the need for some computational muscle to extend the capture even further. Once again, it was Karma Peiró who set up contacts with data scientists and personnel at the Barcelona Supercomputing Centre (Centre Nacional de Supercomputació), especially with the researcher Fernando Cucchiatti. After an initial meeting in May 2018, both he and his team joined the project and obviously proved to be vital for designing both the analysis and the visualisation of the content, which consisted of the 160,000 original tweets collected in the dataset.

To summarise briefly, we found testimonies from women from 60 different countries, posted in only two weeks; 790,000 individual users; 160,000 original stories, 50,000 of which were first-person accounts, one out of every ten referring to murders, one out of seven to rape, three out of ten to sexual assault, one out of six to abuse, one out of three to harassment, and also one out of three referring to incidents to which women had been subjected while out walking alone. There were also 3,500 references to cases of rape and

sexual assault against minors under 18 years of age and more than 1,000 against minors under 12 years of age.⁵⁰ Many of these accounts had been made public for the first time. We were dealing with a dataset consisting of a private, personal content so staggering that it increased the complexity of its sensitive treatment.

Such sensitive treatment is indeed found in the logic behind the design of the visualisation ([Figure 1](#)). It consists of a large circle of dots in which each colour is significant; the red dots indicate tweets about physical assaults (murder, rape, sexual assault and abuse) and the pale pink dots indicate the other cases. The circle is read from right to left as if moving around a clock with a circular face, while the dots are arranged according to the time when they were written and posted in the different countries over a period of two weeks.

One hundred sixty-thousand original tweets (discarding retweets and ‘likes’) were analysed for the visualisation, which consisted of analysing the content of the written messages in the tweets as well as the interaction to which these had given rise, which would involve an even more complex analysis. These 160,000 tweets were broken down into those that referred to personal experiences, and those that recounted third-person experiences of victims who had either been afraid to speak about the incidents before, or had not previously been Twitter users, or because they had died, as well as tweets that expressed surprise and support for the movement. Tweets that contained advertising or graphic images, as well as those from small groups of trolls or those of a mocking or taunting nature were also included, even though they constituted a very small minority of the tweets analysed.

In order to form an idea of what keywords it was necessary to identify in the analysis, 16 members of Fernando Cucchietti’s team⁵¹ devoted themselves to classifying manually and categorising the content of 10,632 randomly selected tweets. Of these tweets, 31.03% were first-person testimonies, 8.91% were second-person testimonies, 40.18% tweets in support of and 3.12% against the movement, the remaining 16.69% being ‘others’. Extrapolating these percentages to the total number of tweets, they have an error ranging

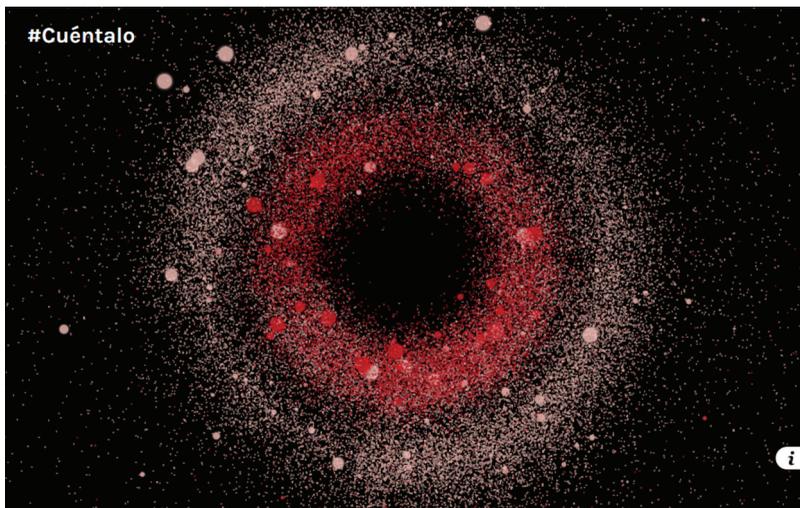


Figure 1.

from 1.5% for the tweets against to 3% for the testimonies and almost 6% for the tweets in favour. Of the first- and second-person tweets (almost 40%), 3.92% referred to murder, 5.59% to rape, 11.18% to sexual assault, 6.27% to abuse, 14.19% to harassment, 11.78% to fear and 19.23% to disgust/anger/sadness (since the tweets might refer to different cases at the same time, the percentages do not add up to 100%). The methodology enables these percentages to be extrapolated with a margin of error of 1% in cases of murder and rape, 3% for assault, abuse and harassment, and 6% for the categories of fear and indignation.⁵² The self-excluding categories range from first- and second-person tweets to those expressing support and solidarity with the movement and those opposing it, as well as some random tweets (because they were in other languages or because of spam). Since the categorisation is based on written texts, the tweets based solely on images could not be classified and are included in the 'others' category.

Once the categorisation was completed, the content of the tweets was analysed; the testimonies of these women convey a feeling of dread and insecurity, whether they concern murder or repeated torture, or every kind of physical, verbal and virtual abuse, as well as rape and harassment. It was not possible to classify them according to the age of the victims or their aggressors, nor by the types of aggressor (strangers, friends, relatives, etc.), due to technical limitations unavoidable at that time. For that reason, it was evident that the categorisation would be imperfect and subject to improvement in various aspects, such as the legal or social evaluation, or a clearer specification of the types of assaults and whether or not they constituted rape.

The conception of what form the visualisation would take varied throughout the process of development, and although at the start it seemed that the conversations might be represented by a tree structure, later it appeared that a linear narrative over time could well be a suitable formulation. This search for a form that one could follow over the passage of time lasted from 27 April to 13 May, and while the linear time form seemed the most appropriate, it restricted the possibility of adding more tweets from people who wished to do so in the future. So little by little, we opted for a circular representation that enabled all the tweets to be seen as well as each individual one. And by recovering the original idea of *#Cuéntalo*, Fernando's team met with Cristina Fallarás and between them, they decided that the circular form would be ideal, but that the messages and personal testimonies should be located in the centre of the circle, with the outer ring containing those expressing support and the remaining comments arranged around the circumference. Furthermore, the dots indicating each tweet were arranged so that they could be viewed clockwise rotating from right to left and going from day to day.

The final circle structure for the visualisation conveys a feeling of safety and protection, and at the same time, a sense of introspection. Each tweet is represented by a dot, forming altogether a cloud-like constellation, stunning for its magnitude and also because of the repercussion of the messages it contains. It constitutes a metaphor that enables us to see the individuality of each testimony in the form of a tweet, and a time scale showing how and when they were posted. The dark centre of the circle is surrounded by lighter coloured dots indicating messages about physical assaults and the red dots indicating cases of extreme violence. Most of the tweets referring to physical assaults are located in the centre and are marked in red because they concern the most harrowing accounts.⁵³

This can all be seen on the project website (www.proyectocuentalo.org).⁵⁴ A communication channel was also set up so as to enable the producers of the tweets to withdraw them if they wish to do so.

From the social web to the social evidence

As explained in the previous section, cooperation among the different professional profiles by means of the creation of a community of practice has enabled us to transform the Big Data from #Cuéntalo into a monumental resource of coherent information consisting of 160,000 documents captured from the social network, together with a total of more than two million interactions. However, in order for this to be converted into a social archive, that is, for the information obtained to have evidential value, it has to be considered socially as proof of the actions described. As Brien Brothman summarised:

It is debatable whether evidence can so simply just be kept or preserved. One could object, as we do here, that evidence does not exist in or for itself. *Someone . . .* has to come along (later) and discover and *use* records for a particular purpose and to serve an identifiable interest, to imagine and interpret what (and whose) interests and purposes they will serve. Our claim is that records are arguably created by an almighty originator. Evidence however cannot be so created. Evidence rather arises out of processes of social negotiation *after the fact*.⁵⁵

Unfortunately, despite having designed the continuation of the project in two phases for the implementation of this participative archive, we have been unable to obtain the necessary funding and no archival institution has consented to join the project. As stated above, the objective was to employ the participative archival model as the basis (from evaluation to documentary description), and also to include the automated processes for the use of massive data and the dissemination of the results. Perhaps more importantly, it was necessary to develop a user-friendly technological platform to enable the archives and other institutions to deal with the processing of massive data (at least on a basic level), even though we have been unable to count on the assistance of engineers or other qualified personnel. We summarise below the basic outlines of the project, in case it may prove helpful for other or future experiences.

The first phase (Design and Implementation) was expected to be conducted over a period of eighteen months. The first six months would be devoted to completing the design of the methodological framework and the selection of the relevant tools, while the remaining twelve months would be dedicated to the implementation of the precipitative archival processing as a tool for the analysis and visualisation of massive data. Despite consisting of a social fonds with an international scope, we believed that it was essential to try it out first at a local level. Given that most of the project members were closely linked to Barcelona, and that a large number of the original tweets (more than 4,000) were geolocalised in the city, we opted for the Catalan capital as the archival community for an initial trial run. The idea was to facilitate preliminary meetings between tweet producers, archivists and ICT engineers. It was also clear, however, that the virtual infrastructure had to be implemented in the most innovative way possible.

First of all, we wanted to test a trustworthy repository for datasets and social web content, Dataverse or B2safe being two possible options in order to guarantee their

preservation. In particular, we would have liked to focus on the design of the participatory recordkeeping platform (probably using Mukurtu), and the definition of an ethical digital framework for archiving the social web.⁵⁶ We eventually asked the BSC-CNS team how much time would be needed to design an open-source tool for analysing massive data and how much it would cost. In reply, they came up with a very detailed proposal. The tool in question had to be easy to use for professional profiles of both journalists and social researchers as well as others. It had to be capable of retrieving millions of tweets as well as reading them and summarising them mechanically, and then identifying them according to subject matter, determining the relation between them and detecting the relevant words they contained. It also had to enable the analysis of the structure and properties of the dynamic network in which these tweets appeared, their different authors and actors and also the roles they played.

The intention was that this tool would be developed by a group devoted to the analysis and visualisation of data, supported by a group from the Barcelona Super Computing Centre skilled in natural language processing. It was also necessary to define the profiles of potential users and find volunteers for the usability tests, added to which was the design of the architecture and application; the definition of the local technologies and the server; the costing of the design of functional prototypes; a graphic interface design team; data scientists; software architecture experts, together with application developers; personnel for documenting processes and designers of dissemination material, not to mention ensuring the provision of basic technical support for tackling any new bugs that might appear.

As regards the second phase (Integration and Pluralisation), a period of eighteen months was scheduled. The main objective was to obtain institutional sustainability and a political mandate for the archiving of social fonds. To that end, it was necessary to include at least one powerful administration for integrating the participatory platform into its institutional archival system so that a regulated policy of capturing datasets and contents from the social network could be undertaken. This was a vital step for addressing the extension and adoption of an initial archival structure with a global reach for the *#Cuéntalo* movement. Ideally, it was possible to build a federation of nodes with other geographical areas (especially, the rest of Spain and Latin America) for sharing the experience and methodology of the initial archive's community.

There was also another goal to be achieved. After converting the information into evidence by archival processing, it would then be a question of promoting interdisciplinary research and pluralising the theoretical and technological knowledge generated by the project. We were aware of the need to initiate academic research, together with the organisation of workshops and seminars (hackathons, conferences, etc.), in order to socialise the results from the outset, since we believed that this was the best way to generate the archival autonomy of the body politic, understood as a tool for political empowerment, which we regard as uppermost in our professional mission.

Conclusion

As may be seen, our pursuit of the social archive has so far failed to reach its goal, in spite of which we should not believe that it has led us to a dead end. We have at least managed to establish a place for *#Cuéntalo* among the different archival projects whose aim is to

tackle the challenge of preserving and managing paricipitatively the information generated collectively from the social network by the public at large. Among other important issues, also at stake is the construction of a public memory under democratic control. That is why, even in the absence of a legal mandate, the Society of Catalan Archivists and Records Managers has wished to set an example for institutional archive services to act as communities of practice, seeking solutions that can be socialised so that subsequently all the collectives who may be interested are able proactively to manage the documentation they generate in production platforms they do not administrate.

In the light of the results obtained in the first phase completed in the *#Cuéntalo* project, we have at least been able to contribute an entire series of tools to empower the victims of male violence: the preservation of the dataset, the analysis and categorisation of 160,000 original tweets and a dynamic dataviz, among others. Altogether, it has helped to generate a massive collective support and social recognition. Furthermore, we trust that it has brought us to a turning point for the professional archivists in our country. In the wake of *#Cuéntalo*, the value of records (and future archives) for the current social web material is no longer an issue that is open to doubt. It remains to be seen whether the institutional Catalan archives are prepared to prioritise the processing of these archives as part of their usual business.

But, more importantly, what would be really necessary is a structural cooperation policy between the archival community and collectives that create these social fonds, based on three basic principles: to encourage these collectives to manage and keep their archives; to allow the participation of these collectives in public appraisal committees; and to let archivists, as nodal professionals between administration and communities transform their technical expertise into a social asset for building trust. In short, to turn institutional archives from ‘memory’ institutions to ‘citizen trust’ institutions. To do so institutional archives should act as communities of practice, testing solutions that can be socialised so that later each interested group may proactively take care of its records generated in environments (private platforms) that it does not own. As professionals, we must ensure not only the democratic control of the construction of public memory but also the need to socially build our present democracy.

Notes

1. In contrast with the philosophical or other theoretical disciplines, archival turn started in 1990s due to the influence of, especially, Foucauld and Derrida works; this archival turn is material since information governance techniques are used as means of production for economic and social profits.
2. ‘Evidence is a social construct, just like financial currencies or geopolitical boundaries. There is no “right” way to create a record. People will decide what they will document, how they will document it, and why they will document it, according to their own needs, technologies, politics, cultures, and customs’. Laura A. Millar, *A Matter of Facts. The Value of Evidence in an Information Age*, ALA-SAA, Chicago, 2019, p. 39.
3. Eric Ketelaar, ‘Tacit narratives: the meanings of Archives’, *Archival Science*, vol. 1, no. 2, 2001, pp. 131–141.
4. Manuel Castells, ‘Informationalism, networks, and the network society: a theoretical blueprint’, in Manuel Castells (ed.), *The network society: a cross-cultural perspective*, Edward Elgar, Northampton, Massachusetts, 2004, available at <<https://annenberg.usc.edu/sites/>

- default/files/2015/04/28/Informationalism%2C%20Networks%20and%20the%20Network%20Society.pdf>, accessed 15 January 2020.
5. Luciano Floridi, *The 4th Revolution: How the Infosphere is reshaping human reality*, Oxford University Press, Oxford, 2014. We started writing this article before the COVID-19 lockdown in Spain. Needless to say that especially this part dedicated to the Archive Society has become even more relevant to understand the way data through private platforms determine our existence.
 6. *ibid.*, p. 41. This definition ought not be completely alien to the archival discipline. We should note the similarity with the ‘fonds’ proposed by ISAD(G): ‘Records in their entirety, regardless of form or medium, organically created and/or accumulated and used by a particular person, family, or corporate body in the course of that creator’s activities and functions’.
 7. In this broadest sense, infosphere aligns with the records continuum paradigm: Frank Upward, ‘Structuring the records continuum part one: Post-custodial principles and properties’, *Archives and Manuscripts*, vol. 24, no. 2, 1996, pp. 268–285 and ‘Structuring the records continuum part two: Structuration theory and recordkeeping’, *Archives and Manuscripts*, vol. 25, no. 1, 1997, pp. 10–35; Sue McKemmish, ‘Placing records continuum theory and practice’, *Archival Science*, vol. 1, no. 2, 2001, pp. 333–359.
 8. Floridi, p. 4.
 9. *Ibid.*, p.171. Once again, from the archival perspective, this period ought to be familiar to us, since it is none other than the ‘crucial phase in the history of archives’, according to the classic article by Henry Bautier, ‘La phase cruciale de l’histoire des archives: la constitution des dépôts d’archives et la naissance de l’archivistique (XVIè- début du XIXè siècle)’ (‘The crucial phase in the history of archives: the constitution of archival depots and the birth of archival science [XVIth – beginning of the XIXth century]’), *Archivum. Revue Internationale des Archives*, 1968, vol. XVIII, pp. 139–149.
 10. As of 2018, the five companies with the greatest market value are technological enterprises (Apple, Amazon, Alphabet – i.e., Google – Facebook and Alibaba). The volume of business they handle exceeds the GDP of most countries in the world. However, in terms of real profits, Apple it is the only Big Tech company that figures in the leading positions (information sourced from the website <www.statista.com>).
 11. Floridi, pp. 176–177.
 12. Shoshana Zuboff, ‘Big other: surveillance capitalism and the prospects of an information civilization’, *Journal of Information Technology*, vol. 20, 2015, pp. 75–89.
 13. *ibid.*, p. 79. According to Zuboff, the five most important sources of Big Data are: computer-mediated economic transactions, sensors embedded in the Internet of Things, corporate and government databases, private and public surveillance cameras, and non-market activities such as Facebook ‘likes’, Google searches, emails, location, communication patterns ... literally ‘every click’.
 14. *ibid.*, p. 80.
 15. *ibid.*, p. 81. This explains the apparent incongruence between market value and real profits, as remarked in Note 10. As Zuboff summarises: ‘This new market form has quickly developed into the default business model for most online companies and startups, where valuations routinely depend upon “eyeballs” rather than revenue as a predictor of remunerative surveillance assets’.
 16. A Acker and A Kriesberg, ‘Tweets May Be Archived: Civic Engagement, Digital Preservation and Obama White House Social Media Data’ *Proceedings of the Association for Information Science & Technology*, vol. 54, no. 1, 2017, pp. 1–9.
 17. Joanne Evans, Sue McKemmish, Elizabeth Daniels and Gavan McCarthy, ‘Self-determination and Archival Autonomy: Advocating Activism’, *Archival Science*, vol. 15, no. 4, 2015, pp. 337–68.
 18. Aniol Maria Vallès, ‘El tractament arxivístic de Twitter’, Master Dissertation [online], supervised by Joan Soler Jiménez, Universitat Autònoma de Barcelona, Escola Superior d’Arxivística i Gestió de Documents, 2016, available at <<https://ddd.uab.cat/record/167528/>>, accessed 7 April 2020.

19. Michael Sippey, 'Changes coming in Version 1.1 of the Twitter API', 16 August 2012, <https://blog.twitter.com/developer/en_us/a/2012/changes-coming-to-twitter-api.html>, accessed 7 April 2020.
20. Developer Agreement and Policy – Twitter Developers: Content redistribution <<https://developer.twitter.com/en/developer-terms/agreement-and-policy>>, accessed July 7th, 2020.
21. Preservation Action Plan: Structured Data/Plain Text National Archives and Records Administration (NARA) <https://github.com/usnationalarchives/digital-preservation/blob/master/Structured_Data_Formats/NARA_PreservationActionPlan_StructuredData_20200629.pdf> and Preservation Action Plan: Structured Data/Spreadsheets National Archives and Records Administration (NARA) <https://github.com/usnationalarchives/digital-preservation/blob/master/Structured_Data_Formats/NARA_PreservationActionPlan_Spreadsheets_20200629.pdf>, accessed 7 July 2020.
22. Vicenç Ruiz Gómez. '#NoTincPor: l'arxivística del dol' ('#IamNotAFraid: the archiving of mourning'), 19 August 2017, available at <<https://empoweringarchives.wordpress.com/2017/08/19/notincpor-larxivistica-del-dol/>>, accessed 7 April 2020.
23. Documenting the Now, <<https://www.docnow.io>>, accessed 7 April 2020.
24. Lisa Peet, 'Documenting the Now Builds Social Media Archive', *Library Journal*, 2 May 2016, available at <<http://lj.libraryjournal.com/2016/05/digital-content/documenting-the-now-builds-social-media-archive>>, accessed 7 April 2020.
25. DocNow/twarc: A command line tool (and Python library) for archiving Twitter JSON, available at <<https://github.com/DocNow/twarc>>, accessed 7 April 2020.
26. 'Arxivar l'avui per guanyar el futur' ('Archiving the present to win the futur'), Associació d'Arxivers – Gestors de Documents de Catalunya, 4 October 2017, available at <<https://arxivers.com/campanyes/arxivar-lavui-per-guardar-el-futur/>>, accessed 7 April 2020.
27. Google Cloud Shell, <<https://console.cloud.google.com/cloudshell/>>, accessed 7 April 2020.
28. Microsoft, 'Windows Subsystem for Linux Installation Guide for Windows 10'. available at <<https://docs.microsoft.com/en-us/windows/wsl/instal-win10>>, accessed 22 March 2020.
29. DocNow/twarc, 'A command line tool (and Python library) for archiving Twitter JSON', GitHub, available at <<https://github.com/DocNow/twarc#usage>>, accessed 7 April 2020.
30. Morgan E Currie and Britt S Paris, 'Back-ups for the future: archival practices for data activism', *Archives and Manuscripts*, vol. 46, no. 2, 2018, pp. 124–142.
31. *ibid.*, pp. 134–138.
32. Tom Nesmith, 'The concept of societal provenance and records of nineteenth-century Aboriginal-European relations in Western Canada: implications for archival theory and practice', *Archival Science*, vol. 6, no. 3–4, 2006, pp. 351–360; 'Still fuzzy, but more accurate. Some thoughts on the "ghosts" of archival theory', *Archivaria*, no. 47, 1999, pp. 136–150.
33. Jeanette A Bastian, *Owning memory: how a Caribbean community lost its archives and found its history*, Libraries Unlimited, London, 2003.
34. Anne J Gilliland, 'Reconceptualizing records, the archive and archival roles and requirements in a networked society', *Knygotyra*, no. 63, 2014, pp. 17–34. Specifically for Twitter records, Gilliland concludes that archivists 'should be focusing on ways in which different kinds of records can productively and effectively be extracted from these corpora' and that 'automated retrieval directly from an extensive corpus of born-networked materials may be the only way to find material' (p. 30).
35. Quite an irony for the evolution of the provenance concept itself, which started as a physical arrangement method. *Onlife experience*, as used by Luciano Floridi, refers to the fact that 'the digital online world is spilling over into the analogue-offline world and merging with it' (Floridi, p. 43).
36. Andrew Flinn, Mary Stevens and Elizabeth Shepherd, 'Whose memories, whose archives? Independent community archives, autonomy and the mainstream', *Archival Science*, vol. 9, no. 1–2, 2009, pp. 71–86; Andrew Flinn, 'The impact of independent and community archives on professional archival thinking and practice', in Jenny Hill (ed.), *The future of Archives and Recordkeeping: A reader*, Facet Publishing, London, 2011, pp. 145–170.

37. In the case of other social documentary fonds, this community will not be so homogeneous, but will express opposed visions and opinions (as, for example, in #CatalanReferendum). However, this diversity should be regarded as something positive, since it enables the capture of a multiplicity of voices capable of putting an end to the silence of secular archives that affect the current documentary heritage, consisting of traditional fonds in which a single record creator imposes his or her only documental voice through control of the genesis and the documentary circuits.
38. W Lance Bennett and Alexandra Segerberg, 'The logic of connective action', *Information, Communication & Society*, vol. 15, no. 5, 2012, pp. 739–768.
39. *ibid.*, p. 753. Within connective action, two kind of networks are identified: Organisational Enabled Networks and Self Organising Networks. This last one fits with #Cuéntalo case, according with its characteristics: 'little or no organizational coordination of action, large scale personal access to multi-layered social technologies, communication content centers on emergent inclusive personal action frames, personal expression shared over social networks'.
40. Livia Iacovino, 'Archives as Arsenals of Accountability', in Terry Eastwood and Heather MacNeil (eds.), *Currents of Archival Thinking*, Libraries Unlimited, Santa Barbara, 2010, pp. 181–212.
41. Wendy Duff, Andrew Flinn, Karen Suurtamm and David Wallace, 'Social justice impact of archives: a preliminary investigation', *Archival Science*, vol. 13, no. 4, 2013, pp. 317–348.
42. Livia Iacovino, 'Rethinking Archival, Ethical and Legal Frameworks for Records of Indigenous Australian Communities: A Participant Relationship Model of Rights and Responsibilities', *Archival Science*, vol. 10, no. 4, 2010, pp. 353–372; Greg Rolan, 'Agency in the archive: a model for participatory recordkeeping', *Archival Science*, vol. 17, no. 3, 2017, pp. 195–225; Joanne Evans, Sue McKemmish and Greg Rolan, 'Critical Approaches to Archiving and Recordkeeping in the Continuum', *Journal of Critical Library Information Studies*, vol. 1, no. 2, 2017, pp. 1–38, available at <<https://journals.litwinbooks.com/index.php/jclis/article/view/35>>, accessed 8 April 2020.
43. Michelle Caswell and Marika Cifor, 'From Human Rights to Feminist Ethics: Radical Empathy in the Archives', *Archivaria*, no. 81, 2016, pp. 23–43.
44. *ibid.*, p. 32.
45. Richard Marciano, Victoria Lemieux, Mark Hedges, Maria Esteva, William Underwood, Michael Kurtz and Mark Conrad, 'Archival Records and Training in the Age of Big Data', in J Percell, LC Sarin, PT Jaeger and JC Bertot (eds), *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education*, Advances in Librarianship, vol. 44B, Emerald Publishing, 2018, pp. 179–199.
46. Feliu Serra Burriel, '#Cuéntalo. Construint una memòria plural del present' ('#Cuéntalo. Building a plural memory of the Present'), Barcelona Supercomputing Centre – Centro Nacional de Supercomputación, XVII Congrés d'Arxivística i Gestió Documental de Catalunya, 2019, available at <<https://www.youtube.com/watch?v=0WM2wlvY-LA>> (starts at 12:00), accessed 7 April 2020.
47. Geoffrey Yeo, 'Concepts of Record (1): Evidence, Information, and Persistent Representations', *The American Archivist*, vol. 70, no. 2, 2007, pp. 315–343; Luciana Duranti, 'Whose truth? Records and archives as evidence in the era of post-truth and disinformation', in Caroline Brown (ed.), *Archival Futures*, Facet Publishing, London, 2018, p. 19–32.
48. Millar, p. 28
49. @lafallaras, <<https://twitter.com/lafallaras/status/989916242717954048?lang=en>>. Here is the English translation of this tweet: 'I have started the day by proposing #Cuéntalo because I think that ALMOST ALL OF US have suffered some kind of sexual aggression. THREAD'.
50. The analysis of some 3 million tweets reveals more than 160,000 stories of sexual assault. Project Cuéntalo is available at <<http://blog.proyectocuentalo.org/2019/01/07/lanalisi-de-casi-3-milions-de-tuits-revela-mes-de-160-mil-histories-dagressions-sexuals/>>, accessed 23 March 2020.

51. Fernando Cucchietti's team consisted of: Sol Bucalo, Luz Calvo, Carlos Carrasco, Fernando Cucchietti, Artur García Saez, Carlos García Calatrava, David García Povedano, Juan Felipe Gómez, Camilo Arcadio González, Guillermo Marín, Irene Meta, Patricio Reyes, Feliu Serra, Diana Fernanda Vélez, María Coto and Laura Gutierrez.
52. #Cuéntalo – The Visualisation, The BSC Viz Corner, available at <<http://www.bsc.es/viz/corner/?p=210&lang=es>>, accessed 23 March 2020.
53. It is no coincidence that the form of Dataviz chosen reminds very much of the topographical representation of the records continuum created by Frank Upward, nor that the four categories chosen in regard to the producers of the tweet also resemble the different degrees of agency defined in Greg Rolan's participatory recordkeeping model.
54. We were assisted in the execution of this phase of the project by a grant from the Ajuntament de Barcelona (Barcelona City Hall).
55. 'Afterglow: Conceptions of Record and Evidence in Archival Discourse', *Archival Science*, vol. 2, no. 3–4, 2002, p. 334.
56. DocNow, 'Ethical Considerations for Archiving Social Media Content Generated by Contemporary Social Movements: Challenges, Opportunities, and Recommendations', April 2018, available at <<https://www.docnow.io/docs/docnow-whitepaper-2018.pdf>>, accessed 8 April 2020.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by the Barcelona City Council [Direct grant 18S16280 1312/18].

Notes on contributors

Vicenç Ruiz Gómez, PhD, is currently an archives and records manager administrator at the Official Mail Unit of the European Parliament, with a previous broad professional experience both in the public (municipal and regional administrations) and corporate sector. He has also been a professor at the Master of Archival Science and Records Management of the Autonomous University of Barcelona from 2013 to June 2020. He has been a Board Member of the Society of Catalan Archivists and Records Managers since 2013.

Aniol Maria Vallès is an archivist specialising in digital environment. He holds a degree in History (Rovira i Virgili University) and a Master in Archival Science and Records Management (Autonomous University of Barcelona). He has worked both in the archives of the Generalitat de Catalunya and in the archives of Aigües de Barcelona (AGBAR-SUEZ), as well as in various municipal public archives. He is also part of the Working Group on Transparency and Open Data of the Society of Catalan Archivists and Records Managers (AAC-GD), in which he has participated in the development of the 'Maturity Model in Records Management for Open Government'.

ORCID

Vicenç Ruiz Gómez  <http://orcid.org/0000-0002-5843-6072>