

## Repurposing Digital Methods in a Post-API Research Environment: Methodological and Ethical Implications

*Alessandro Caliendo*

### **How to cite**

Caliandro, A. (2021). Repurposing Digital Methods in a Post-API Research Environment: Methodological and Ethical Implications. [Italian Sociological Review, 11 (4S), 225-242]  
Retrieved from [<http://dx.doi.org/10.13136/isr.v11i4S.433>]

[DOI: 10.13136/isr.v11i4S.433]

### **1. Author information**

*Alessandro Caliendo*

Department of Political and Social Sciences, University of Pavia, Pavia,  
Italy

### **2. Author e-mail address**

*Alessandro Caliendo*

E-mail: [alessandro.caliandro@unipv.it](mailto:alessandro.caliandro@unipv.it)

### **3. Article accepted for publication**

Date: January 2021

Additional information about  
**Italian Sociological Review**  
can be found at:

**About ISR-Editorial Board-Manuscript submission**



# *Repurposing Digital Methods in a Post-API Research Environment: Methodological and Ethical Implications*

Alessandro Caliandro\*

Corresponding author:  
Alessandro Caliandro  
E-mail: [alessandro.caliandro@unipv.it](mailto:alessandro.caliandro@unipv.it)

## **Abstract**

Especially after the social media curtailing brought about the Cambridge Analytica scandal, it has become increasingly difficult to do social media research by using digital methods as well as *following the medium*. This condition brings in new methodological and ethical challenges. This article proposes some methodological strategies to ‘repurpose’ digital methods in a post-API research environment. The discussion draws on three case studies: 1) studying Instagram stories: the scraping dilemma; 2) studying smartphone in everyday contexts: researching digital environments not connected to APIs; 3) studying fake news on Twitter: dealing with increasingly useless APIs data. For each case study methodological and ethical implications are examined. In conclusion, the article suggests that a possible viable strategy to repurpose digital methods in a post-API era is to *follow the natives* (along with the *medium*), that is, to take advantage of the natively digital methods through which social media users manage their own data as well as emic conception of what is ethical (or at least acceptable) regarding the handling of their own data.

Keywords: digital methods, post-API, follow the natives.

## **1. Introduction**

The term digital methods describes “the deployment of online tools and data for the purposes of social and medium research. More specifically, they derive from online methods, or methods of the medium, which are reimagined and repurposed for research” (Rogers, 2017: 75). The digital methods paradigm is premised on the idea of conceiving the Internet more as a source of methods

---

\* Department of Political and Social Sciences, University of Pavia, Pavia, Italy.

rather than an object of study. From this derive the famous principle of *follow the medium* (Rogers, 2013), meaning researchers should learn from and take advantage of the *natively digital methods* that digital environments, such as search engines or social media platforms, employ to gather, order, organise, rank and rate digital data – as with APIs, algorithms, tags, likes, hashtags, etc. (Caliandro, 2018). The main scope of digital methods is to repurpose all these natively digital methods for social research scopes.

Such methodological framework was developed in a period of general openness and accessibility of the Internet and Internet data. In the early days, digital methods research focused on the so-called *politics of the web* (Rogers, 2010); studying, for example, Google as an epistemological machine (that is, as a nonhuman actor playing a key role in the production, circulation and ordering of knowledge within the contemporary society) (Rogers, 2018), or the practices of links exchanging between websites (along with the consequent emerging of (new) social hierarchies) (Rogers, 2009).

An important step forward in digital method research and practice happened around the beginning of 2010s, with the dramatic diffusion of social media platforms (like Facebook, Twitter, and Instagram) and their successive ‘colonization’ of the web (Helmond, 2015). The real gamechangers here were the public APIs released by the principal social media platforms on the market, which gave researchers access to a rich pool of digital data – rich from both a quantitative and qualitative point of view (Russell, 2013). (A sheer example of this is the Instagram API, which before 2015 allowed developers to get data from hashtag and/or users’ profiles with no limitation in terms of quantity and time frame – along with a plentiful set of metadata, such as: post id, comment count, like count, location, post link, hashtags, mentions caption, type, url image, user id author, username author, date post). The access to social media APIs started a little revolution in the digital methods field, insofar social media platforms permitted researchers to explore, not only the socio-technical structures shaping online communication (e.g., the logic of Google’s PageRank), but also cultural processes emerging from users’ everyday digital practices (Giglietto, Rossi, Bennato, 2012; Marres, Gerlitz, 2015) – (consider, for example, studies on digital social formations (Arvidsson et al., 2016; Marres, Gerlitz, 2019) or the semantic structure of online public discourses (Marres, 2015; Geboers, Van De Wiele, 2020). It is not a case that in the last decade digital methods scholars produced a conspicuous stream of research that casted a light upon the most pivotal and compelling socio-cultural phenomena characterizing the contemporary digital era, like echo-chambers (Colleoni, Rozza, Arvidsson, 2014), political bots (Bessi, Ferrara, 2016), algorithmic culture (Airoldi, Beraldo, Gandini, 2016), digital surveillance (Van der Vlist, 2017), fake news (Gray, Bounegru, Venturini, 2020).

Things started changing from 2018 on, when the infamous Cambridge Analytica scandal was brought at the attention of the public thanks to the investigative work of The Guardian’s journalist Carole Cadwalldr<sup>1</sup>. Cadwalldr inquiry revealed that “Cambridge Analytica’s apparent data analytics and psychographic voter targeting capabilities, building especially on social media data obtained from Facebook with the help of a ‘personality test’ app that gathered information from up to 87 million users” (Bruns, 2019: 1547). In response to the scandal, and in order to better protect the privacy of their own users, Facebook (along with other platforms) began a policy of closure and restriction of its formerly open APIs. Axel Bruns (2019) contends that the Cambridge Analytica scandal represented a convenient pretext for social media companies like Facebook and Twitter to make their data, progressively, inaccessible. A move that does nothing but increasing the commercial value of those data (since the business model of social media platforms consists precisely in selling users’ data to third parties (Srnicek, 2017), rather than increasing the users’ privacy. In fact, those users’ data that are no more retrievable via public APIs, are still accessible, under the payment of a fee, to private companies – which use them for business and marketing purposes. Thus, it is not a case that academics have been amongst those most affected by the social media APIs curtailing – (we all remember with grief the sad shut down of Netvizz (Rieder, 2018)). Again, Axel Bruns (2018) provocatively argued that this initiative had the consequence (more or less explicit) of silencing social researchers, who, through their empirical studies, more than once pointed to the various ‘information disorders’ populating social media environments (Bounegru et al. 2018). Therefore, social researchers find themselves in the paradoxically situation of living in a *platformised* and *datafied society* (Van Dijck et al. 2018), where access to platforms and their data is becoming more and more difficult. This condition urges digital methods to ‘repurpose themselves’, that is, to review their protocol of research and devise new methodological strategies to explore digital environments.

## **2. Doing digital methods in non-API environments: follow the natives!**

Starting from these premises, the objective of this article is to propose some new methodological strategies to conduct digital research in a post-API era. In order to meet this objective, I will present and discuss three different research ‘scenarios’ pertaining to the application of digital methods in non-API environments:

---

<sup>1</sup> See <https://www.theguardian.com/news/series/cambridge-analytica-files>

- 1) Studying Instagram stories: the scraping dilemma.
- 2) Studying smartphone in everyday contexts: researching digital environments not connected to APIs.
- 3) Studying fake news on Twitter: dealing with increasingly useless APIs data.

To describe each scenario, I will draw on three different research projects, which I recently developed and published (with the collaboration of some colleagues from University of Milano and University of Milano-Bicocca). I selected these three case studies for different reasons: i) they allow discussing the practice of digital research in different non-API environments, since they are based on researches conducted respectively on Instagram, smartphones and Twitter; ii) for the same reason, they allow highlighting how the methodological strategy I propose can be adapted to different digital environments; iii) they bring to light a variety of ethical issues related to the diverse range of techniques and tools that can be used to collect data from social media and mobile devices; iv) last, but not least, they all deal with phenomena deemed particularly relevant in recent years (that is, the diffusion of ephemeral social media content (Abidin, 2021); ageing and the use of digital technologies (Rosales, Fernández-Ardèvol, 2020); COVID-19 epidemic and misinformation (Bruns, Harrington, Hurcombe, 2020)). For each case study, I will discuss the related methodological and ethical implications. Based on the discussion, I propose the methodological strategy of *following the native* (Latour, 2005), that is, – beyond observing how digital devices structure online communication and interaction (*follow the medium*) – paying attention to how users use digital devices as well as how such use impacts on the structure of online communication and interaction (Caliandro, Gandini, 2017).

### **3. Studying Instagram stories: the scraping dilemma**

This case study refers to a research published in the article ‘From archive cultures to ephemeral content, and back: studying Instagram Stories with digital methods’ (Bainotti, Caliandro, Gandini, 2020), and discusses possible methodological strategies to collect and analyse Instagram stories. Instagram stories are slippery pieces of data, since they disappear from the platform by 24 hours after their publication. Plus, they are specifically not retrievable through the Instagram API. The curtailing of social media APIs, which followed the Cambridge Analytica scandal (Bruns, 2019), has resulted in the impossibility of obtaining some kinds of social media data, such as Instagram Stories (Bruns, 2018). Specifically, as far as Instagram Stories are concerned, the walls put up may be also linked to their content – usually they display private and personal

information – as well as to their significant business value (Puschmann, 2019) – most of the interactions among influencers and consumers unfold through Instagram Stories (Warren, 2019). Therefore, this is a case in which one cannot *follow the medium*. Nevertheless, another viable methodological strategy to study Instagram stories consists in *following the natives* (Latour, 2005), that is, taking advantage of the “natively digital methods by which Internet users capture and archive Instagram Stories themselves” (Bainotti, Caliandro, Gandini, 2020: 4). Specifically, to collect and analyse digital stories it is possible to take advantage of one of the many apps or online free tools to save Instagram stories.

For the research discussed in this paragraph, we took advantage of StorySaver<sup>2</sup>, a freely available tool to scrape Instagram Stories. The functioning of this tool replicates, in some way, the ephemerality that is typical of Instagram Stories: in fact, it allows to visualise Stories posted by users with a public profile and to download them (as .jpg or .mp4 files), within the 24 hours during which the Stories are available on the Instagrammer’s profile. Through this tool we collected a corpus of 292 Instagram stories (that were randomly selected from Instagram by following the generic hashtag #happy), that was organised in an ad hoc excel spreadsheet. We then performed a qualitative visual analysis (Altheide, 1987; Rose, 2016) on the data collected (see Table 1).

TABLE 1. Instagram Stories classification based on qualitative visual analysis.

Categories	Occurrences (n.)	Occurrences (%)
Portrait	89	30,6
Composition	64	22,0
Materiality	38	13,0
Setting	41	14,0
Celebration	30	10,3
Food & Beverage	23	8,0
Others	6	2,1
<b>Total</b>	<b>292</b>	<b>100</b>

The problem with this kind of methodological strategy is that one has to employ scraping (i.e. IT techniques that allow researchers to obtain digital entities from the HTML code of the webpage in which they are located (Weltevrede, 2016)) to collect data, and these techniques have a controversial (ethical) status in social research methods (Landers, Brusso, Cavanaugh, 2016). In fact, although scraping is not illegal per se (Waterman, 2020), nevertheless it is a practice opposed by social media platforms for a variety of reasons. First, this is because scraping carries the risk of crashing the website under study. In fact, in order to obtain the whole content of a webpage, a script has to make a large number of ‘calls’ to the server hosting the webpage. If a large number of

<sup>2</sup> <https://www.storysaver.net/>.

calls is made multiple times, by more than one user and on many pages, then the risk of crashing the server is quite high. Second, this is due to the fact that scraping allows to collect data from users with private profiles – thus, to obtain information that users are not necessarily happy to share. Last, a similar issue concerns the willingness of social media companies to share data with developers and researchers (this is strictly linked to the significant economic value of data, for platforms (Zuboff, 2019)): specifically, through scraping it is possible to access also those data that social media companies do not intend to share. Yet, as Venturini and Rogers (2019) argue, some forms of scraping need to be considered a ‘necessary evil’ for social research, if performed conscientiously. In order to manage the collection and management of Instagram Stories ‘conscientiously’ we devised a set of (ethical) practices. First, we did not programme our own scraping script (instead, as mentioned, we used StorySaver). In this way, we avoided bypassing the platform’s restrictions or blocks (Chellapilla, Larson, Simard, 2005), disguising the non-human identity of the collector of data (Von Ahn, Blum, Hopper, 2003), or accessing content protected by privacy settings or passwords (Franzke, Bechmann, Zimmer, 2019). By doing so, we complied with Instagram Terms of Service (Fiesler, Lampe, Bruckman, 2016). Also, once the data collection was completed, we managed the data ensuring the respect of users’ privacy and that no harm was caused to them. Specifically: i) we privileged the collection of Stories not dealing with sensitive content (e.g., political views, or sex); ii) we did not show in the published article any material (e.g., screenshots, usernames,) that might allow to identify the user’s profile or to obtain user’s personal information; iii) we analysed data in an aggregated and clustered form and presented them through coding categories; iv) we did not share data with third parties. By adopting this approach, we did not ‘[break] the law, or [put] a burden on a site’s servers, or potential(ly) harm [...] users’ (Fiesler, Beard, Keegan, 2020: 10), thus we avoided incurring in the ethical problems that are normally ascribed to scraping techniques.

Moreover, I would like to put forth a last point. By briefly searching the Web, it is possible to see a plentitude of tools for saving Instagram stories, as well as YouTube channels that aggregate stories. If one looks at the statistics about apps’ downloading provided by Google Play, several interesting data emerge. For example, one sees that the app *Story Saver per Instagram*<sup>3</sup> has been downloaded 1M+ times, *Story Saver Downloader & Repost Video and Post*<sup>4</sup> 5M+

---

<sup>3</sup><https://play.google.com/store/apps/details?id=com.story.saver.instagram.save.download.video&hl=it&gl=US>.

<sup>4</sup><https://play.google.com/store/apps/details?id=com.magic.story.saver.instagram.video.downloader&hl=it&gl=US>.



times, and *Story Saver for Instagram – Assistive Story*<sup>5</sup> 5M+ times. These statistics further confirm that users customarily employ online scrapers to save and archive Instagram stories as well as consider this practice, somehow, legit.

Lastly, it is worth stressing that while conducting our research we consciously decided to follow the call by prominent digital scholars to actively engage in identifying new, alternative, effective methods and tools to keep social media research alive in an era of APIs curtailing (Bruns, 2018; Kazansky, Torres, van der Velden, 2019).

#### **4. Studying smartphone in everyday contexts: researching digital environments not connected to APIs**

This case study refers to a research published in the article ‘Older People and Smartphone Use in Everyday Life: An Inquire on Digital Sociality of Italian Older Heavy Users’ (Caliandro et al., 2021), which (among the other things) discusses the possible methodological strategies to collect and analyse data from users’ smartphones.

Smartphones have a stunning global diffusion and uptake. According to the *Digital In 2020* report (WeAreSocial, 2020), there are currently 5.1 billion of unique mobile users worldwide (67% of the global population), while the smartphone users amount to 3.5 billion (circa), that is 45% of the world population. On average people spend 6 hours and 42 minutes on Internet per day; nearly half of this time (48%) is spent on smartphones (WeAreSocial, 2020). As a matter of fact, smartphones have a profound impact on everyday social relations, since they mediate a large part of our distance as well as face-to-face interactions (Campbell, 2019). Notwithstanding, their capillary diffusion, there is still a scarcity of empirical social research on smartphones – if compared to social research on social media. For example, the Digital Methods Initiative has no specific programmes dedicated to smartphones – (if we exclude the App Studies Initiative, which, by the way, is not focused on smartphones per se, but rather it is more concerned with ‘appification’ of the web and society (Dieter et al., 2019). In my opinion this is due (also) to the practical and technical difficulties in retrieving data from users’ smartphones. A way to overcome these difficulties is to take advantage of tracking devices. In our case, we used a commercial App (Rescue Time), which we installed on the smartphones of 30 volunteers aged 62-76 – (recruited thanks to the collaboration with the association Auser Monza-Brianza, aimed at promoting

---

<sup>5</sup><https://play.google.com/store/apps/details?id=com.popularapp.storysaver&hl=it&gl=US>.

active ageing). The app remained installed on the participants' smartphone from the 24<sup>th</sup> of January 2019 to the 24<sup>th</sup> of February 2019 and gathered 75,089 data. Rescue Time tracks exclusively log data, such as the number users' accesses to WhatsApp during the day, but not the content of the conversations taking place on the platform. Specifically, Rescue Time registers which is the application accessed, the number of times the application is accessed, as well as the duration, in seconds, of each access. Rescue Time automatically codes the collected log data into 9 categories: Utilities, Communication & Scheduling, Entertainment, Social Networking, Reference & Learning, News & Opinion, Shopping, Business, Design & Composition.

Rescue Time allows following users' daily smartphone activities for long periods of time, and with a degree of granularity that is almost impossible to reach when applying more traditional techniques of data collection (for example, participant observation) (Ørmen, Thorhauge, 2015). Moreover, the collection of log data allows researchers obtaining precise measurements on smartphone activities, avoiding the risks of systematic misestimation (Boase, Ling, 2013), typically linked to other research tools, like questionnaires or self-tracking sheets.

Beyond collecting log data, to get a more comprehensive understanding of the socio-cultural everyday practices that participants developed around their smartphones, we took advantage of two qualitative methods: collective interviews and face-to-face semi-structured interviews. Differently from a focus group, the objective of the collective interviews was not to elicit a collective debate on a given topic, but rather comment and interpret together the outputs of Rescue Time. In fact, the sense of many data released by Rescue Time is very hard to decipher without the participants' cooperation (Aipperspach et al., 2006). For example, we noticed a dramatic drop of smartphones activities during the weekends. This odd result was immediately clarified by participants, who explained that during the weekends they are usually more engaged in 'real-life activities' (e.g., relaxing in front of the TV or hanging out with friends/relatives) and so, physiologically, they have less time to spend on their smartphone. Also, collective interviews were strategical to construct the interview schedule for the (subsequent) individual interviews, since helped us to identify 'hot topics' to be touched during the face-to-face encounters with participants (e.g., the importance of 'media ideologies' (Fernández-Ardèvol et al., 2020), which has been a pivotal topic in both face-to-face interviews and the published article).

Lastly, it is important to stress that we did not simply use Rescue Time for our empirical study, rather we *re-purposed* it for social research scopes. In fact, Rescue Time is an app originally designed to support users in digital detox

exercises as well as in managing their work productivity<sup>6</sup> – whereas we took advantage of the app to explore forms of sociality. In addition to this, in order to better answer our research questions<sup>7</sup>, we had to perform some operations of data re-coding and polishing. For example, we re-grouped WhatsApp, Facebook, and YouTube under the same category of ‘Social media’, since, customary, Rescue Time assigns those platforms to distinct categories, respectively: ‘Communication and Scheduling’, ‘Social Networking’ and ‘Entertainment’. Plus, qualitative methods were necessary to make sense of most of the log data.

Considering the private nature of the data that Rescue Time permits to extract from the participants’ smartphones, some considerations related to privacy are due. First, it is important to remark that Rescue Time, by default, can only track log data and not the specific content of the website a user logs in. Therefore, in this way, the privacy of participants is protected, at least partially. Anyway, all the participants signed an informed consent, which, among the other things, granted them the possibility to uninstall the app at any moment and without notice or explanation. Surprisingly, at the moment of installation, participants were not so much concerned about privacy issues, rather about technical ones. They wanted to be reassured that the app would have not consumed their data and/or band – (something that Rescue Time does not do). When later, during the collective interviews, we asked participants if they were unconformable with a ‘spying app’ in their phones, they replied they were not. As a matter of fact, most of them claimed of having soon forgotten to have an app running in the background of their phones – (something that they normally do with all the apps already installed on their mobile devices). Therefore, differently from what initially expected, we accounted for a limited *Hanthonne effect*. Something that can be, in turn, ascribed to the so called big ‘brother effect’, that is, the diffused social awareness (and somehow sense of impotence) about the systems of data surveillance operating on the Internet, personal devices and public venues (Mani, Chouk, 2019). Nevertheless, participants seemed quite annoyed about the fact that Rescue Time made their batteries consuming faster – (something we forgot to mention at the beginning of the project). This taught us that, for further research employing tracking apps, besides all the standard assurances on privacy protection, it is important to precisely inform participants about all the possible technical consequences

---

<sup>6</sup> <https://www.rescuetime.com/>

<sup>7</sup> The research questions were the following: *How older people use smartphone in their everyday life? How is the smartphone used by older people in the construction of social relations in everyday life? Which are the advantages and disadvantages of such use perceived by older people? Which kind of norms and meanings older people assign to the construction of social relations via smartphone?*

of those apps on the correct functioning of their mobile devices. Therefore, along the ethical guidelines of social research related to privacy protection, digital researchers must also consider questions of ‘natively digital ethics’, which have to do with concerns and issues intrinsically embedded in the digital technologies employed for empirical research, as well as participants’ perceptions of the functioning of such technologies.

### **5. Studying fake news on Twitter: dealing with increasingly useless APIs data**

This case study refers to a research published in the article ‘Fake news, COVID-19 e Infodemia: un esempio di ricerca sociale in real-time su Twitter’ (Caliandro, Anselmi, Sturiale, 2020), which discusses the methodological implications of studying fake news on Twitter.

In this research, in order to monitor the spread of an infodemic in Italy – that is the epidemic spread of fake news related to COVID-19 that poses serious problems for public health (Zarocostas, 2020) – we analyzed the processes of circulation of fake news within the Italian Twittersphere during the health emergency. Drawing upon the analysis of 7,237,581 tweets, we concluded that it is not possible to observe a real infodemic within the Italian Twittersphere. In fact, only 0,001% tweets contained links from fake news websites and 1,44% commented fake news. Plus, all these tweets tended to circulate only in very specific periods and within closed communities. Furthermore, we found that the 61% of fake news detected dealt with the topic of immigration. More specifically, they consisted in false information regarding the responsibility of migrant people for the diffusion of coronavirus. These kinds of fake news do not represent, per se, a threat to public health; on the contrary, they can be clearly identified as part of the strategies adopted by Italian right-wing populist movements to get online visibility in order to put forth their political positions.

In my opinion, these results are not totally independent from the increasing interventions of digital platforms aimed at policing the creation of content on social media. Also, we must acknowledge that maybe Twitter is not the best place in which monitoring the circulation of fake news. The progressive diffusion of debunking initiatives has probably made the producers of fake news more cautious: they tend not to publish fake news within public social media environments and migrate them to more private one (like WhatsApp or Telegram group). Moreover, the recent censorship initiatives enacted by social media platforms might have exacerbated this trend. In fact, precisely during the pandemic, Twitter publicly announced that the platform would have taken action in flagging and removing fake content circulating within its social

network (Heilweil, 2020). For how noble this initiative can be, it has the effect of making Twitter less transparent: how does Twitter identify fake news? How does Twitter define them? Which kinds of algorithms does it use to discriminate the false positives? It is very unlikely that a private company like Twitter will be willing to answer publicly these kinds of questions. Moreover, the Twitter censorship makes very difficult for social researchers to study fake news, that is, to understand their mechanisms of circulation and main rhetoric – without necessary impeding them to circulating in private digital venues (Rogers, 2020).

Finally, it is important to highlight that some of the problems occurring in researching fake news on Twitter derive also from the procedures through which they are operationalized in digital methods research. The main procedure for identifying fake news on social media consists in following the medium, that is, in detecting URLs coming from fake news outlets (e.g., Sputniknews or Imolaoggi). Nevertheless, for the issues discussed above, this strategy is not always viable. In our case, we tried to overcome this problem by *following the natives*, that is, by detecting the presence of keywords related to fake news within the text of the tweets posted by users. This strategy allowed us to study the impact of fake news within the public discourse, by observing how fake news' rhetoric impacts on the micro-narrative of 'regular' users. As a matter of fact, this line of thought is similar to Rauchfleisch e Kaiser' reflections (2020) on bots; basically, they argue that it is not sufficient to count the number of bots within a stream of tweets, rather it is important to understand their impact on users' activities. Therefore, as already stressed in the previous paragraphs, a viable strategy to repurpose digital methods in a post-API era is to follow the natives, besides the medium.

## 6. Conclusion

In this short essay, I discussed some methodological and ethical strategies to repurposing digital methods in a post-API era. Specifically, I proposed to *follow the natives* (along with the medium), that is, to take advantage of the natively digital methods through which Internet users manage their own data as well as emic conception of what is ethical (or at least acceptable) regarding the handling of their own data.

As far as the method is concerned, I proposed to *follow the natives*, when following the medium is not possible or somehow limited by APIs restrictions. In the case studies presented, following the natives has two declinations: a) taking advantage of the natively digital methods through which users collect and organize their own digital data (i.e., Instagram Stories, Rescue Time); b) observing the strategies through which users use those digital devices that

structure streams of online communication (i.e., fake news and URLs). Moreover, I showed how following the natives translate also into the employment of qualitative methods – (which formally do not belong to the digital methods paradigm and which the paradigm considers as ‘virtual methods’ (i.e., interviews, content analysis, text analysis). In this sense, I would advocate a more systematic integration between digital ethnography and digital methods (Caliandro, 2018). In fact, digital ethnography permits “researchers to observe online dynamics through the same interfaces as the actors they study” (Venturini, Rogers, 2019: 536-537), as well as access digital environments not accessible through APIs (Semenzin, Bainotti, 2020). The ‘follow the natives’ principle applies also to the ethical aspects discussed in the article. Specifically, we saw the general attitude Instagram users have towards the archiving of Instagram Stories, as well as the natively digital ethical principles (related to privacy) proposed by smartphone users. Anyhow, regarding ethical implications of digital methods research a last reflection it is due. When it comes to digital social research, it is not possible to establish univocal guidelines. As Floridi and Taddeo (2016) maintain, digital ethical guidelines cannot be universal as well as set once and for all. Instead, they must be aimed at being contextual, that is, elaborated and adjusted according to the specific digital platforms under study, the kinds of data collected, the kinds of tools used to collect data, the objectives of research, the research question. Drawing on what has been said throughout the article, I would add that in a post-API era, researchers should find a (very complicated) balance between a data activist stance and the concern for protecting participants from harm and guaranteeing them a proper balance of benefit and burden deriving from their participation in the research. In this regard, probably, when it comes to digital research, the point is not simply safeguarding the privacy of the single individuals from which we extract data, but also trying to redistribute to participants the value we extract from their data (Floridi, 2017) – at least partially. In the case of social research, this may translate in sharing with participants, as much as possible, the knowledge we extract from data. Nevertheless, as the last case study shows (‘fake news on Twitter’), this practice seems, again, more and more hampered by social media APIs restrictions.

Admittedly, the reflections put forth in this essay have some limitations. First, the points discussed in the article stem from only three case studies, which, furthermore, were conducted by the author himself. Anyhow, the scope of this article was to kick start a fruitful conversation on the destiny of social media research in a post-API era, rather than propose generalizable results or universal theoretical concepts. Second, the article discusses digital methods drawing quite exclusively on Rogers’ theorizations and works, overlooking other prominent approaches such as *interface methods* (Marres, Gerlitz, 2015) or

*trace methods* (Salganik, 2018; Veltri, 2021). Nevertheless, this choice had a precise rationale. In this contribution I wanted to highlight the methodological issues concerning the possibility to ‘follow the medium’ within post-API environments – and, therefore, I focused on Rogers, since ‘*follow the medium*’ is the core epistemological principle at the base of his conception of digital methods (Rogers, 2009). Finally, it is worth mentioning, that, in the period I was finalising the present article, a big change has occurred in the world of social media APIs. Both Facebook and Twitter made their APIs open to academics. Facebook through CrowdTangle, a “Facebook-owned tool that tracks interactions on public content from Facebook pages and groups, verified profiles, Instagram accounts, and subreddits” (Shiffman, 2021a); Twitter by “opening up its full tweet archive to academic researchers for free” (Statt, 2021). This is undoubtedly a great achievement, that due to timing reasons I could not discuss in this article. Anyhow, this positive step forward needs to be thoroughly scrutinized by the academic community. In fact, regarding CrowdTangle for example, the access to Facebook is conditioned to the submission of an application in which the academic must demonstrate that her/his research focuses on “Misinformation, Elections, COVID-19, Racial justice, Well-being” (Shiffman, 2021b) – which are, for sure, crucial topics but do not cover the entire spectrum of social research issues.

## References

- Abidin, C. (2021), From “networked publics” to “refracted publics”: A companion framework for researching “below the radar” studies, *Social Media+ Society*, 7(1), 1-13.
- Airoldi, M., Beraldo, D., Gandini, A. (2016), Follow the algorithm: An exploratory investigation of music on YouTube, *Poetics*, 57, 1-13.
- Altheide, D.L. (1987), Reflections: Ethnographic content analysis, *Qualitative Sociology*, 10(1), 65–77.
- Aipperspach, R., Rattenbury, T.L., Woodruff, A., Anderson, K., Canny, J.F., Aoki P. (2006), *Ethno-mining: integrating numbers and words from the ground up*, Electrical Engineering and Computer Sciences University of California at Berkeley Tech Report UCB/EECS-2006, 125, <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-124.html>.
- Arvidsson, A., Caliendo, A., Airoldi, M., Barina, S. (2016), Crowds and value: Italian directioners on Twitter, *Information, Communication & Society*, 19(7), 921–939.

- Bainotti, L., Caliendo, A., Gandini, A. (2020), From archive cultures to ephemeral content, and back: studying Instagram Stories with digital methods, *New Media & Society*, <https://doi.org/10.1177/1461444820960071>.
- Bessi, A., Ferrara, E. (2016), Social bots distort the 2016 US Presidential election online discussion, *First Monday*, 21(11/7), <https://firstmonday.org/article/view/7090/5653>
- Boase, J., Ling, R. (2013), Measuring mobile phone use: Self-report versus log data, *Journal of Computer-Mediated Communication*, 18(4), 508-519.
- Bounegru, L., Gray, J., Venturini, T., Mauri, M. (2018), *A Field Guide to 'Fake News' and Other Information Disorders*, Amsterdam, Public Data Lab, <https://zenodo.org/badge/DOI/10.5281/zenodo.1136272.svg>
- Bruns, A. (2018), *Facebook shuts the gate after the horse has bolted, and hurts real research in the process*, Internet Policy Review, <https://policyreview.info/>.
- Bruns, A. (2019), After the 'APIcalypse': social media platforms and their fight against critical scholarly research, *Information, Communication & Society*, 22(11), 1544–1566.
- Bruns, A., Harrington, S., Hurcombe, E. (2020), '<? covid19?> 'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook, *Media International Australia*, 177(1), 12-29.
- Caliandro, A., Gandini, A. (2017), *Qualitative Research in Digital Environments: A Research Toolkit*, London, Routledge.
- Caliandro, A. (2018), Digital Methods for Ethnography: Analytical Concepts for Ethnographers Exploring Social Media Environments, *Journal of Contemporary Ethnography*, 47(5), 551-578.
- Caliandro, A., Anselmi, G., Sturiale, V. (2020), *Fake news, Covid-19 e Infodemia: un esempio di ricerca sociale in real-time su Twitter*, Mediascapes Journal, <https://ojs.uniroma1.it/index.php/mediascapes/article/view/17105>.
- Caliandro, A., Garavaglia, E., Sturiale, V., Di Leva, A. (2021), Older People and Smartphone Use in Everyday Life: An Inquire on Digital Sociality of Italian Older Heavy Users, *The Communication Review*, 1-32, <http://dx.doi.org/10.1080/10714421.2021.1904771>.
- Campbell, S.W. (2019), From frontier to field: Old and new theoretical directions in mobile communication studies, *Communication Theory*, 29(1), 46-65.
- Chellapilla, K., Larson, K., Simard, P.Y. (2005), Computers beat humans at single character recognition in reading based Human Interaction Proofs (HIPs), *CEAS 2005 – Conference on Email and Anti-Spam, July 21-22, 2005, Stanford University, California*, <https://www.semanticscholar.org/paper/Computers-beat-Humans-at->



- Single-Character-in-basedChellapillaLarson/e405e0f1bca55b7d0e17656533f63fd025c2dd58.
- Colleoni, E., Rozza, A., Arvidsson, A. (2014), Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using Big Data, *Journal of communication*, 64(2), 317-332.
- Dieter, M., Gerlitz, C., Helmond, A., Tkacz, N., van Der Vlist F.N., Weltevrede, E. (2019), Multi-situated app studies: Methods and propositions, *Social Media+ Society*, 5(2), 1-15.
- Fiesler, C., Lampe, C. Bruckman, A.S. (2016), Reality and perception of copyright terms of service for online content creation, *CSCW'16: Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing*, 27 February–2 March, 2016, San Francisco, California (pp. 1450–1461), New York: ACM.
- Fernández-Ardèvol, M., Belotti, F., Ieracitano, F., Mulargia, S., Rosales, A., Comunello, F. (2020), “I do it my way”: Idioms of practice and digital media ideologies of adolescents and older adults, *New Media & Society*, <https://doi.org/10.1177/1461444820959298>.
- Fiesler, C., Beard, N., Keegan, B.C. (2020), *No robots, spiders, or scrapers: legal and ethical regulation of data collection methods in social media terms of service*, cmci.colorado.edu., [https://cmci.colorado.edu/~cafi5706/ICWSM2020\\_datascraping.pdf](https://cmci.colorado.edu/~cafi5706/ICWSM2020_datascraping.pdf).
- Floridi, L. (2017), Group privacy: A defence and an interpretation, In Taylor, L., Floridi, L., van der Sloot, B. (eds.), *Group Privacy: New Challenges of Data Technologies*, pp. 83-100, Cham, Springer.
- Floridi, L., Taddeo, M. (2016), *What is data ethics?*, *Philosophical Transactions of the Royal Society A*, 374: 20160360. <http://dx.doi.org/10.1098/rsta.2016.0360>.
- Franzke, A.S., Bechmann, A., Zimmer, M. (2019), *Internet research: ethical guidelines 3.0: association of Internet researchers*, Association of Internet Researchers, <https://aoir.org/reports/ethics3.pdf>.
- Geboers, M.A., Van De Wiele, C.T. (2020), Machine Vision and Social Media Images: Why Hashtags Matter, *Social Media+ Society*, 6(2), <https://doi.org/10.1177/2056305120928485>.
- Giglietto, F., Rossi, L., Bennato, D. (2012), The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source, *Journal of technology in human services*, 30(3/4), 145-159.
- Gray, J., Bounegru, L., Venturini, T. (2020), ‘Fake news’ as infrastructural uncanny, *New Media & Society*, 22(2), 317-341.
- Helmond, A. (2015), The platformization of the web: Making web data platform ready, *Social Media+ Society*, 1(2), <https://doi.org/10.1177/2056305115603080>.

- Heilweil, R. (2020), *Twitter now labels misleading coronavirus tweets with a misleading label*, Vox, [https://www.vox.com/recode/2020/5/11/21254889/twitter-coronavirus-covid-misinformation-warnings-labels?\\_c=1](https://www.vox.com/recode/2020/5/11/21254889/twitter-coronavirus-covid-misinformation-warnings-labels?_c=1).
- Kazansky, B., Torres, G., van der Velden, L. (2019), Data for the social good: toward a dataactivist research agenda, In Daly, A., Devitt, S.K., Mann, M, (eds.), *Good Data*, pp. 244–259, Amsterdam, Institute of Network Cultures.
- Landers, R.N., Brusso, R.C., Cavanaugh, K.J. (2016), A primer on theory-driven web scraping: automatic extraction of Big Data from the Internet for use in psychological research, *Psychological Methods*, 21(4), 475–492.
- Latour, B. (2005), *Reassembling the Social: An Introduction to Actor-Network-Theory*, Oxford, Oxford University Press.
- Mani, Z., Chouk, I. (2019), Impact of privacy concerns on resistance to smart services: does the ‘Big Brother effect’ matter?, *Journal of Marketing Management*, 35(15/16), 1460-1479.
- Marres, N. (2015), Why map issues? On controversy analysis as a digital method, *Science, Technology & Human Values*, 40(5), 655–686.
- Marres, N., Gerlitz, C. (2015), Interface methods: Renegotiating relations between digital social research, STS and sociology, *The Sociological Review*, 64(1), 21–46.
- Marres, N., Gerlitz, C. (2019), Social media as experiments in sociality, In Marres, N., Guggenheim, M., Wilkie, A. (eds.), *Inventing the Social*, pp. 253-287, London, Mattering Press.
- Ørmen, J., Thorhauge, A.M. (2015), Smartphone log data in a qualitative perspective, *Mobile Media & Communication*, 3(3), 335-350.
- Puschmann, C. (2019), An end to the wild west of social media research: a response to Axel Bruns, *Information, Communication & Society*, 22(11), 1582–1589.
- Rauchfleisch, A., Kaiser, J. (2020), *The False positive problem of automatic bot detection in social science research*, Berkman Klein Center Research Publication, <https://cyber.harvard.edu/news/2020-03/false-positive-problem-automaticbot-detection-social-science-research>.
- Rieder, B. (2018, August 11), *Facebook’s app review and how independent research just got a lot harder*, The Politics of Systems, <http://thepoliticsofsystems.net/2018/08/facebooks-app-review-and-how-independent-research-just-got-a-lot-harder/>.
- Rogers, R. (2009), *The End of the Virtual*, Amsterdam, Amsterdam University Press.
- Rogers, R. (2010), Internet research: The question of method: A keynote address from the YouTube and the 2008 election cycle in the United States conference. *Journal of Information Technology & Politics*, 7, 241–260.
- Rogers, R. (2013), *Digital Methods*, Cambridge, MIT Press.

- Rogers, R. (2017), Foundations of Digital Methods: Query Design, In Schäfer, M.T., van Es, K. (eds.), *The Datafied Society: Studying Culture through Data*, pp. 75-94, Amsterdam, Amsterdam University Press.
- Rogers, R. (2018), Aestheticizing Google critique: A 20-year retrospective, *Big Data & Society*, 5(1), <https://doi.org/10.1177/2053951718768626>.
- Rogers, R. (2020), Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media, *European Journal of Communication*, 35(3), 213-229.
- Rosales, A., Fernández-Ardèvol, M. (2020), Ageism in the era of digital platforms, *Convergence*, 26(5/6), 1074-1087.
- Rose, G. (2016), *Visual Methodologies: An Introduction to Researching with Visual Materials*, London, Sage.
- Russell, M. (2013), *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More*, Sebastopol, O'Reilly Media.
- Salganik, M. J. (2018), *Bit by bit. Social research in a digital age*, Princeton, Princeton University Press.
- Semenzin, S., Bainotti, L. (2020), *The use of Telegram for the non-consensual dissemination of intimate images: gendered affordances and the construction of masculinities*, SocArXiv Papers, <https://osf.io/preprints/socarxiv/v4f63/>.
- Shiffman, N. (2021a), *Citing CrowdTangle Data*, Help.Crowdtangle, <https://help.crowdtangle.com/en/articles/3192685-citing-crowdtangle-data>.
- Shiffman, N. (2021b), *CrowdTangle for Academics and Researchers*, Help.Crowdtangle, <https://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers>.
- Srnicek, N. (2017), *Platform capitalism*, Cambridge, Polity Press.
- Statt, N. (2021). *Twitter is opening up its full tweet archive to academic researchers for free: to help researchers better study online discourse and platform trends*, The Verge, <https://www.theverge.com/2021/1/26/22250203/twitter-academic-research-public-tweet-archive-free-access?fbclid=IwAR0zeojPZR06wk-TGmKdrXTM5oU3rvijNJ4H2WU0Rwoeo6QJi6YUY5Z11J8>.
- Van Dijck, J., Poell, T., De Waal, M. (2018), *The platform society: Public values in a connective world*, Oxford, Oxford University Press.
- Van der Vlist, F.N. (2017), Counter-Mapping Surveillance: A Critical Cartography of Mass Surveillance Technology After Snowden, *Surveillance & Society*, 15(1), 137-157.
- Veltri, G.A. (2021), *La ricerca sociale digitale*, Milano, Mondadori.
- Venturini, T., Rogers, R. (2019), 'API-based research' or how can digital sociology and journalism studies learn from the Facebook and Cambridge analytica data breach, *Digital Journalism*, 7(4), 532-540.

- Von Ahn, L., Blum, M., Hopper, N.J. (2003), CAPTCHA: using hard AI problems for security, *EUROCRYPT*, 2003, 297–311.
- Warren, J. (2019, October 24), *10 strategies to drive traffic and sales from Instagram*, Later, <https://later.com/blog/drive-traffic-from-instagram/>.
- Waterman, T. (2020, January 29), *Web scraping is now legal*, Medium, <https://towardsdatascience.com/web-scraping-is-now-legal-6bf0e5730a78>.
- WeAreSocial, (2020), *Digital in 2020*. <https://wearesocial.com/digital-2020>.
- Weltevrede, E. (2016), *Repurposing digital methods: the research affordances of platforms and Engines*, PhD Dissertation, Amsterdam, University of Amsterdam.
- Zarocostas, J. (2020), How to fight an infodemic, *The Lancet*, 395(10225), 676.
- Zuboff, S. (2019), *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, London, Profile Books.