

Social Media Analytics: Boom and Bust?

Axel Bruns

Introduction: Big Social Data and the Computational Turn

More than any other innovation in public communication, the emergence of modern social media platforms, and the availability of data on the communication processes unfolding on such platforms, have contributed to the ‘computational turn’ (Berry, 2011) in media and communication studies. Often available from social media platforms in a clearly structured and readily analysable format through their Application Programming Interfaces (APIs), data from social media platforms provide an unprecedented and inobtrusive perspective on who engaged with whom about what, often at large scale and in close to real time. This has created new opportunities for research: beyond small-scale studies conducted through laborious participant observation (where the researchers would embed themselves in online communities and keep field notes about their experience) or other forms of direct engagement with social media users (through interviews, focus groups, and surveys), social media APIs seemed to offer not only ‘big data’ on the population-wide uptake of such new communicative tools, but ‘big *social* data’ on the public’s current interests, opinions, and sentiment, and on their social interactions with each other as they formed and severed connections, expressed and argued over personal, professional, and political views, and developed and engaged in a wide range of other social media activities and practices not foreseen by the platforms’ developers. This chapter retraces the growth of the field of research that emerged following the computational turn, and outlines the challenges for such research that have resulted from more recent developments.

Social media APIs were never explicitly designed for research purposes, of course: predominantly, they sought to enable the integration of and interactions between the different systems that operated the social media platform itself, as well as the exchange of data between these systems and the third-party applications designed to enhance those capabilities – for instance, specialist apps for professional and power users or tools providing other enhancements not anticipated by the platform itself. In the case of *Twitter*, for example, even its search function was once such a third-party addition (Burgess & Baym, 2020). But these APIs, and the data they provided, were also quickly identified as valuable new sources of insight by researchers from media and communication studies to computer science and beyond, and tools created specifically to enable researchers to access, extract, and analyse these rich new datasets from social media APIs soon multiplied.

This is true broadly for all of the wave of modern social media platforms that emerged in the first decade of the twenty-first century, but especially so for *Twitter*, for a number of reasons: its API was arguably the most accessible and most powerful (providing a live stream of public tweets matching selected keywords or hashtags); its highly public approach to communication posed fewer ethical concerns (users made a simple choice between keeping their accounts private or making them globally public, and an overwhelming majority chose to be public); and its simple and flat structure positioned it as the medium for major communicative events (as evidenced especially by hashtags from #eqnz to #blacklivesmatter; cf. Bruns & Burgess, 2015). Historically, this has led to *Twitter* being overrepresented in the scholarly literature, relative to the size of its userbase (Burgess & Bruns, 2015); even today, *Twitter* studies – and especially studies that focus predominantly on the best-known and most easily traceable feature in *Twitter*-based communication, its hashtags – continue to account for a

disproportionately large subset of the overall social media research literature (Ozkula, Reilly, & Hayes, 2020).

However, as is likely to be the case for any rapidly growing field of research, the quality and insight provided by studies drawing on such newly available ‘big social data’ was – and is – highly variable. The new-found observability of communicative patterns on these platforms at large scale tempted many social media researchers to overgeneralise: not only were the userbases of *Twitter* or *Facebook* not representative of the broader national or global digital society, and instead affected by diverging levels of interest in, literacy of, and access to these new technologies, but similarly the observations made from datasets gathered through the APIs of social media platforms were also inherently shaped by the features and functionality of those APIs themselves. Datasets of tweets containing selected hashtags do not accurately reflect the lived experiences of actual *Twitter* users who did not follow these hashtags, and might have encountered only some of these tweets as they were shared by others in their networks, for instance; datasets from *Facebook*’s API, taking into account the privacy settings that users applied to their posts and profiles, might have captured activity only from explicitly public groups and pages, but not the (arguably more important) interactions in the less visible, more private ‘personal publics’ (Schmidt, 2014) or ‘privately public’ spaces (Papacharissi, 2010: 132) surrounding users’ profile pages on the platform. Beyond the initial excitement of gaining access to such powerful sources of observational data about large-scale, real-time public communication, then, it has taken the social media research community some time to identify such limitations and articulate their impact on the results of its studies – and each new generation of Internet researchers excited about the possibilities of the computational turn in their field must necessarily undergo this learning process for itself.

In particular, it has been important to realise that the turn *towards* the computational analysis of big social data as a new source of insights into real-world communicative activities does not imply an equal and opposite turn *away from* other forms of research and critique, and most centrally that it does not suggest a preference for quantitative over qualitative methods. Indeed, the most successful and perceptive social media research tends to incorporate and blend both qualitative *and* quantitative approaches, using one to inform the other and engaging in a truly mixed-methods approach to research. For instance, research methods such as issue mapping (introduced by Rogers & Marres, 2000; updated for contemporary social media contexts by Burgess & Matamoros-Fernández, 2016) draw on a quantitative analysis of large-scale social media datasets in order to identify specific communicative acts that warrant qualitative attention through close reading and interpretation, and in doing so generate deeper insights than either of these methods would be able to produce on their own.

Other studies take an even more speculative, explicitly data-driven yet still far from purely quantitative approach: they gather data about public communication activities on social media platforms at large scale, and then computationally explore these datasets for the presence of distinct and possibly unexpected patterns. For these, they formulate explanatory hypotheses (through close reading and other qualitative, interpretive methods), and then continue with further targeted data gathering and analysis designed to prove or disprove such explanations. Common especially in the physical sciences, but thus far less so in the humanities and social sciences, this iterative approach is neither entirely computational nor manual, neither purely quantitative nor qualitative, neither simply inductive nor deductive, but more properly understood as an *abductive* research design – a ‘method by which hypotheses [are] created or discovered’

(Dixon, 2012) – and especially valuable for the identification and interpretation of previously unknown phenomena.

Following this abductive, data-driven approach, for instance, Bruns & Moon (2019) analyse the data from one full day in a national Twittersphere and not only observe those topical hashtags that are already well known, but also identify a widespread but severely underrecognised practice of ‘phatic sharing’, designed to maintain interpersonal ties, which accounts for a substantial proportion of the day’s tweeting activity. This important yet – because of the intentional absence of any major hashtags or keywords that would mark it out – inconspicuous everyday use of *Twitter* could be found only by such an explorative use of large-scale social media data, eliminating known and easily observable user practices from the dataset until only the previously unknown and unobserved remained. This and similar findings demonstrate one of the core benefits of research that utilises the large-scale datasets on public communication that can be drawn from social media platforms: they enable a particularly unobtrusive observation of communicative activity *in situ*, without affecting the object of study or requiring direct engagement with participants (in the form of interviews, surveys, or experiments): ordinary social media users largely remain unaware of being observed.

However, while such abductive, explorative, and open-ended research can thus lead to the discovery of new and unexpected user practices and communicative phenomena, these new observational data are also in danger of being accidentally or deliberately misinterpreted and misused. Famously, for example, an early and widely reported study of public communication on *Twitter* by the commercial research firm Pear Analytics characterised some 41% of all tweets as ‘pointless babble’ (Pear Analytics, 2009), in ignorance of the fact that such phatic communication often serves an important role in

maintaining social ties between users even if it is unintelligible to outsiders observing it; such communication is ‘pointless’ only if the researcher lacks the sociocultural understanding to get the point, and its blithe dismissal reflects more badly on the observer than on those being observed. Indeed, as Burgess & Baym (2020) point out, the diverging characterisation of such speech acts as ‘important’ or ‘pointless’ also reflects the researcher’s fundamental stance on whether *Twitter* – and, by extension, other social media platforms – should serve predominantly as social networks or information networks.

As noted, especially during the early stages of the development of social media analytics as a field of research in its own right it was common to see studies overgeneralise their findings beyond the practices observed. This is true not only for projections of social media findings to the digital society as a whole, but also within the confines of the platform being studied: for example, *Facebook* research that analyses engagement on public pages and groups, or *Twitter* research that captures only tweets posted to selected hashtags, cannot assume that the patterns observed will also hold for more private conversations in closed groups, around personal profiles, or through direct messages, or for everyday interactions outside of self-selecting and highly visible hashtag communities; rather, the specific data-gathering features and affordances endorsed by the platforms’ APIs have directly shaped the resultant datasets in important ways, and the limitations arising from this shaping must be recognised at all steps in the research process. More recent social media research often – but not always – does so, and many leading social media researchers actively continue to develop the means of recognising, addressing, and (where necessary) counteracting these limitations.

Finally, the ready availability of large-scale, real-time, and often very personal data on user activities on these platforms also creates the possibility for research that proactively seeks to intervene in such public communication. Many such projects are well-intentioned, yet also raise significant ethical and moral questions as well as presenting substantial potential for abuse. For instance, out of a concern that social media users may exist in monocultural ‘filter bubbles’ (Pariser, 2011) that prevent them from encountering a sufficient diversity of views, Graham & Ackland (2017) envisage a new class of ‘socialbots’ that pop such bubbles by putting users with diverging views in touch with each other. This may be beneficial in case of mild disagreements, but could have severely negative consequences in other contexts: for instance, while a socialbot of this kind might well put users who hold homophobic views in touch with more inclusive perspectives, conversely it might also confront vulnerable members of the LGBTIQ+ community with hate speech directed towards them. Other projects in this vein envisage bots that detect social media posts indicating risk-taking behaviours, depression, or suicidal ideation, and respond automatically with messages designed to express care and provide help – yet in such cases, too, it is easily possible to imagine that such messages could lead some of their addressees to respond with further self-destructive behaviour instead. Socialbots of this form would require a well-developed moral compass, therefore – but in addition to the question of *whose* morals they should adopt, such morality would also be exceptionally difficult to implement in computer code.

Beyond such thought experiments, problematic intervention at a larger scale did occur in practice in a widely criticised study of contagious emotional responses to the *Facebook* newsfeed (Kramer et al., 2014), for example. Conducted by *Facebook* researchers in collaboration with academic scholars, this study adjusted the selection

criteria of the algorithms that determine which posts from their network are highlighted to *Facebook* users, and presented one group in a large sample of nearly 700,000 users with more positive, and another with more negative content; it then measured whether users' responses reflected more positive or negative emotions as a result. Undertaken without informing or seeking consent from the users affected, the study documented significant effects from this manipulation, raising concerns especially about its impact on unwitting participants with existing vulnerabilities. Arguably, as Flick writes in a critical assessment of this case, this study 'violated the normative expectations of the very users it was studying' (2016: 26), even though it remained within the rules of *Facebook's* own Data Use Policy – this presents questions for further collaborations between in-house researchers and external scholars, and points to the important role that academic research should play in providing critical and independent scrutiny of platform providers' policies and activities.

Such critiques do not seek to diminish the overall utility of the big social data available from social media APIs, and of the research they enable, but they serve as a reminder to always consider the context, provenance, ethics, implications, and possible misuse of such datasets. Crucially, the rise of 'big data' within the digital humanities, but also well beyond this area of research, has in fact been accompanied also by the emergence of critical data studies: a field that queries the political, economic, ethical, moral, and other related aspects of working with large-scale datasets in scholarly research and other settings (for an introduction, see e.g. boyd & Crawford, 2012; Iliadis & Russo, 2016). It would be inappropriate to attempt to summarise in this chapter the detailed and sophisticated critical discussion of social media data, and of the research approaches they enable, that has already been produced as a result of such efforts – but

readers are thoroughly encouraged to engage with this important component of contemporary scholarship in the digital humanities.

From Social Media Analytics to Cambridge Analytica

As noted before, however, the APIs of most major social media platforms were never designed predominantly with research uses in mind; most principally serve internal purposes (enabling the different components of a platform's own infrastructure to connect to each other) or exist to provide data to external, commercial partners. Platforms such as *Twitter* and – to a considerably lesser extent – *Facebook* actively encouraged the development of a diverse ecosystem of third-party data users and application developers around their core products during the early years of their existence, in fact; this API-enabled ecosystem was regarded both as an indication of the overall health and potential of these platforms (and thus important in attracting further venture capital investment), and as a valuable source of new insights into their key uses and of innovative ideas for their further development. Indeed, several technologies created by third-party developers – such as the high-end *Twitter* client *TweetDeck*, in 2011, or the *Facebook* analytics suite *CrowdTangle*, in 2016 – were eventually acquired by the platform operators themselves.

Unsurprisingly – especially given that, as we have seen, social media platforms faced accusations of providing no more than spaces for ‘pointless babble’ during the early years of their existence – the development of more powerful methods and tools for analysing the volume, dynamics, and content of social media activities was a particularly important priority during these early years, and arguably remains so today. It is in this context that scholarly research also contributed most directly to the platforms’ development and public acceptance: academic research highlighted, for

instance, the critical role of social media platforms as sources of up-to-date information in crisis situations (Palen et al., 2010; Shaw et al., 2013); their important contribution to social and political debate especially in undemocratic regimes with state-controlled media systems (Papacharissi & de Fatima Oliveira, 2012); and their growing use by professional groups and communities of interest (Hermida, 2010). This work contributed to what Rogers (2014) has called the ‘debanalisation’ of social media in public perception; it was actively encouraged and instrumentalised by the platform providers themselves as they sought to position their platforms as serious communication media and attract audiences well beyond their existing in-groups of early adopters.

Perennially lagging behind *Facebook* in the size of its userbase, in its monetisation, and in the public perception of its utility as a platform, *Twitter* appeared to embrace its third-party ecosystem most openly during these early years; even today, its API remains considerably more open and powerful than those of its competitors. This accessibility encouraged the development of a number of open-source data gathering tools such as *Twapperkeeper* (subsequently *yourTwapperkeeper*) and the *Twitter Capture and Analysis Toolkit* (Borra & Rieder, 2014), which rapidly established themselves as standard mechanisms for gathering large-scale, real-time data from the platform; for *Facebook*, similar roles have been played by tools such as *Netvizz* (Rieder, 2013) or *Facepager* (Jünger & Keyling, 2017), although the private or at least semi-private nature of much *Facebook* content has always imposed greater limitations on API-based data access here than on *Twitter*, where fewer than 5% of all accounts are posting privately.

Such tools never enjoyed unqualified support from the platforms they help to study, however, even if the research they have enabled has often highlighted important and valuable uses of such platforms. Access to *Twitter* data through the standard, freely available API, for instance, is both rate-limited (allowing API clients to make only a set number of data requests within a single time window) and volume-limited (returning a live stream of posts matching the selected tracking terms that represents no more than one per cent of the total volume of all global tweeting activity at any one point in time), and rate limits have been adjusted several times over the existence of the platform, usually to reduce the number of requests per time window. Many commercial and scholarly developers initially sought to circumvent such restrictions by spreading their data requests across multiple *Twitter* accounts and API keys; *Twitter*, in turn, attempted to frustrate such practices by making it more difficult to register additional accounts and API keys. As one developer put it, the message implicit in such increasingly restrictive data access policies appeared to be: ‘thanks for getting so many people interested in Twitter. Now get lost’ (qtd. in Bucher, 2013).

Commercial and scholarly users with data requirements greater than could be satisfied through the standard platform APIs were thus faced with a choice of reducing the scale of their studies, of finding more devious ways to circumvent the rules of acceptable API use imposed by the platforms, or of buying datasets from one of a handful of commercial data resellers such as *GNIP*, *DataSift*, or *CrowdTangle*. Quite apart from the often prohibitive costs involved (which usually priced these data well outside of the reach of small-scale, publicly funded research teams, even if they may have remained affordable for larger-scale commercial analytics services), such options were themselves much reduced when *Twitter* bought *GNIP* in 2014, terminated its commercial agreement with *DataSift* in 2015, and eventually absorbed *GNIP* into its

in-house, commercial ‘enterprise API’ service in 2017, and when *Facebook* similarly acquired *CrowdTangle* in 2016. This assertion of control over data access services also signalled the platforms’ growing realisation of the substantial commercial value of social media data and social media analytics, of course.

Although perhaps no more than a side effect of such changes, this commercialisation of larger-scale social media data access had the implicit effect of excluding researchers and research teams that lacked the financial resources necessary to acquire the datasets they needed for their work; in other words, such changes privileged fully commercial or commercially funded research with immediate industry applications (for example in advertising, marketing, and public relations) over public-interest, grant-supported research that sought to examine the role of social media in the digital society (in journalism, media, communication, and other social sciences). Such concerns were raised at the time (e.g. Bruns & Burgess, 2016), and have arguably only become more critical in the meantime; the leading social media platforms have at times responded to them by offering special data access for researchers working on topics of particular public interest, but such opportunities have been fleeting at best. *Twitter*, for instance, called for submissions to a ‘Data Grants’ scheme in 2014, but awarded access to no more than six out of 1,300 applications received (Kirkorian, 2014); a more limited call for proposals to measure the health of conversations on the platform, in 2018, eventually supported only two out of more than 230 submissions (Gadde & Gasca, 2018).

Additionally, as early as 2010 *Twitter* had also promised to gift its entire tweet archive to the US Library of Congress as a scholarly resource (Stone, 2010), yet that promise gradually evaporated as both the Library’s lack of resources and expertise in

handling this large and continuously growing dataset, and disagreements between the Library and *Twitter* about appropriate processes for archiving and granting research access to the dataset, complicated matters. As a result, the Library announced in 2017 that it would shift to archiving only the tweets of a small selection of prominent users (Alaimo, 2017). This represents a significant blow not only to present-day social media researchers, but also to future historians of this critical phase in our transition to a thoroughly digital society: in their immediacy, social media serve as an invaluable ‘first draft of the present’ (Bruns & Weller, 2016) that presents critical new opportunities for historiography well beyond the study of contemporary news media.

Further, while lamenting the Library’s withdrawal of support for the *Twitter* archive, we should also remember that such plans to archive our digital present were never even on the cards for *Facebook* or other major social media platforms, arguably resulting in an even greater loss of historical data. Just as it would be almost impossible for scholars to write their histories of the first and second halves of the previous century without access to major radio and television archives, respectively, it will be exceedingly difficult to understand the sociopolitical dynamics of the first decades of the twenty-first century without such social media data. The quality, accessibility, and intelligibility of our digital legacies thus decays with each passing day unless we make more serious efforts to archive and preserve them.

If scholarly access to large-scale social media datasets had already gradually declined for several years since the heady days of permissive API rules during the establishment phases of the major social media platforms, then the 2018 scandal around social media analytics company Cambridge Analytica and its involvement with problematic political propagandists represented a rapid turn for the worse; I have

described this elsewhere with the deliberately hyperbolic term ‘APIcalypse’ (Bruns, 2019). Ironically, and despite the opportunist attempts in some of *Facebook*’s official statements on the matter to present the scandal at least in part also as a case of academic misconduct (e.g. Grewal, 2018), Cambridge Analytica’s industrial-scale extraction of personal data from unwitting *Facebook* users – who believed they were engaging with a harmless personality quiz app on the platform – results chiefly from the platform’s continued permissive attitude towards *commercial* data users even in spite of the sustained concerns raised by critical voices from the scholarly community.

Indeed, *Facebook* had to admit that Cambridge Analytica developers ‘gained access to this information in a legitimate way and through the proper channels that governed all developers on Facebook at that time’ (Grewal, 2018) – which allowed the developers not only to gather personal data from the users engaging directly with Cambridge Analytica’s apps, but also from those users’ wider networks of *Facebook* friends and contacts. Even the profiling of users’ personal and political interests on the basis of this dataset is far from an uncommon practice in social media marketing; what sets Cambridge Analytica apart from generic advertising and even political campaigns, then, is most centrally its alleged work for a number of populist and illiberal political clients. Put another way, rather than being particularly special in any way, the Cambridge Analytica case simply exposed many of the deeper issues with the commercial and political exploitation of large-scale, real-time access to big social data in a digital and datafied society that scholars in critical data studies had warned about for some time already.

Facebook reacted to the negative global publicity caused by the case by suspending some 200 other third-party apps similar to that used by Cambridge Analytica

(Pasternack, 2018). Shortly afterwards, and without prior warning to developers, it also substantially reduced the functionality of the *Facebook* API, and shut down the API of its subsidiary platform *Instagram* altogether; the latter had originally been planned for a point later in the year (Constine, 2018). Although not directly implicated in the Cambridge Analytica scandal, *Twitter*, too, soon implemented a range of administrative changes that tightened access restrictions for its API; it now requires developers ‘to provide [more] detailed information about how they use or intend to use Twitter’s APIs so that we can better ensure compliance with our policies’ (Roth & Johnson, 2018).

Such measures may be genuinely well-intentioned, but there is a considerable chance that they are misdirected. Large, well-resourced commercial and political marketing operations will almost certainly still have an opportunity to work with the major social media analytics companies that have entered into commercial partnerships with *Facebook*, *Twitter*, and other leading social media platforms, and whose business interests remain largely unaffected by these API changes; indeed, selling access to user data to the major analytics companies is part of the core business model of these social media platforms, which are therefore highly unlikely to implement any changes that would throttle such lucrative income streams. One such company, for instance, Crimson Hexagon, was only briefly suspended from accessing *Facebook* and *Instagram* data in 2018 when it was suspected of the ‘misuse of data for surveillance purposes’, but its access was swiftly restored (Pasternack, 2018). Smaller, independent teams of analysts and developers (not least also in academic rather than industry contexts), by contrast, lack such special relationships, and were forced by the new API changes to choose between stepping through a time-consuming and uncertain process of reaccrediting their data gathering tools, or walking away from their work altogether. University of Amsterdam scholar Bernhard Rieder, for instance – developer of the popular *Facebook*

analytics tool *Netvizz* – declared that he would ‘take a step back from *Netvizz*’ rather than continue to work in the ‘structurally opaque and unhelpful environment’ of the new *Facebook* app approvals process (Rieder, 2018).

It must also be noted that these new restrictions on third-party data access to social media platforms were implemented at a time when independent scrutiny of communication patterns on such platforms would have been of utmost importance. The twin political shocks of 2016 – the Brexit referendum in the UK, in June, and the election of Donald Trump as US President, in November – and the perceived role of aggressive social media campaigning in influencing public opinion relating to these votes had drawn public and scholarly attention to the dissemination of political mis- and disinformation on social media, with the help of both organic activism and inauthentic amplification, and the detection and mitigation of such attempts to interfere in democratic processes has since become a key new field of scholarly research. *Twitter*’s and, especially, *Facebook*’s introduction of new limitations on data access severely hampered the independent, critical, scholarly scrutiny of these platforms’ roles in such processes.

It is unlikely that this impact on critical social media research is entirely coincidental: ‘research by academic institutions is clearly perceived as a liability post-Cambridge Analytica. ... While there’s clearly a huge societal benefit to this research, it’s not necessarily research that benefits social media companies directly. It’s easier to say no than to figure out how to handle it properly’ (Littman qtd. in Alaimo, 2018). *Facebook*’s leadership, for instance, has a well-documented history of stonewalling and denigrating research that raises critical questions; this issue has been raised repeatedly by insiders such as former Chief Security Officer Alex Stamos (qtd. in Mac & Warzel,

2018) and co-founder Chris Hughes, who further suggests that CEO Mark Zuckerberg ‘has surrounded himself with a team that reinforces his beliefs instead of challenging them’ (2019). Zuckerberg and his inner circle appear more concerned with averting bad press for the company than with actually addressing the issues that critical scholarship has raised; to prevent such criticism by starving it of the data it requires is a logical measure from this perspective.

Critical Research in a Precarious Environment

In the wake of Cambridge Analytica, *Facebook*’s most explicit attempt to restructure its engagement with scholarly research sought to institute a new data gatekeeper: the data clearinghouse initiative Social Science One, supported by a coalition of US-based philanthropic foundations and promised access to a tightly circumscribed dataset of platform activity from *Facebook*. Guided by an ambition to establish what the organisation described grandly as a ‘new paradigm for industry-academic partnerships’ (Social Science One, 2018b), the initiative clearly placed the social media industry’s interests first, however: by its own description, it sought to enable social media platforms ‘to enlist the scientific community to help them produce social good, while protecting their competitive positions’ (Social Science One, 2018a).

For scholars, the condition for being thus enlisted was to step away from more critical questions, as Social Science One founders Gary King and Nate Persily explained in a position paper on this partnership model:

the optimal way forward ... is to find research questions that are of intellectual interest to the scientific community and either provide valuable knowledge to inform product, programmatic, and policy decisions, or are orthogonal to company interests. (King & Persily, 2018: 12)

In line with this ‘orthogonal’ approach, Social Science One’s first call for project proposals explicitly required researchers to refrain from ‘investigating internal corporate policies, decisions, or product development’ and conduct research only ‘for scientific purposes, not for purposes of advocacy, journalism, investigation of individuals, or research for competitors’ (Social Science One, 2018b). Such unambiguous encouragement to abandon more critical perspectives appears especially inappropriate at a time when – in the study of abuse, hate speech, mis- and disinformation, human moderation, algorithmic recommendations and filtering, and other key issues – the links between problematic communicative practices on social media platforms and the company policies that enable them are becoming ever more obvious.

Arguably, in fact, it is inherently impossible to study communication patterns on *Facebook* or *Twitter* without also taking into account the development of specific platform features and affordances, and the policy decisions that have guided such development (cf. the ‘platform biography’ approach pursued for the case of *Twitter* by Burgess & Baym, 2020); nor is it appropriate, and indeed ethically and morally defensible, to allow scholars to observe potentially problematic and hurtful communication practices on these platforms while explicitly preventing them from advocating for change. The purely observational stance that Social Science One’s call for proposals demands of researchers is explicitly incompatible with critical data studies and the more ‘data-activist research agenda’ (Kazansky et al., 2019: 244) that has developed amongst social scientists especially in response to the growing evidence of malignant uses of social media platforms in recent years.

Yet even in spite of this already highly restrictive approach to defining allowable research, *Facebook* has largely failed to provide Social Science One with the data it had promised: as of September 2019, the company had delivered only seven gigabytes of data, ‘a tiny fraction of the 1,000,000 gigabytes of data that had been initially promised’ (Pasternack, 2019). *Facebook*’s stated reason for this failure to honour its commitments to the initiative, and to the various research teams who had successfully responded to Social Science One’s call for proposals and were now eagerly awaiting the data on which their projects depended, was its inability to sufficiently protect the privacy of the users whose posting activities might be reflected in the dataset – yet this newfound concern for user privacy has also been suspected to be no more than a stalling tactic. Unfortunately, *Facebook* has developed a substantial track record of generating positive media coverage by announcing initiatives that emphasise its corporate social responsibility, without following through on those promises.

As a result of this lack of progress, then, Social Science One’s supporting organisations began to withdraw in late 2019. Additionally, a public statement on the initiative’s Website expressed the sentiments of some of its scientific advisors:

as members of the European Advisory Committee of Social Science One we – along with the co-chairs – are frustrated. On the one hand, we were genuinely interested in helping to build a model to support academic research, and we appreciate the efforts ... the specific data science teams within Facebook have made to this end. On the other hand, the eternal delays and barriers from both within and beyond the company lead us to doubt whether substantial progress can be made, at least under the current model.

The current situation is untenable. Heated public and political discussions are waged over the role and responsibilities of platforms in today’s societies, and yet researchers cannot make fully informed contributions to these discussions.

We are mostly left in the dark, lacking appropriate data to assess potential risks and benefits. This is not an acceptable situation for scientific knowledge. It is not an acceptable situation for our societies. (Social Science One, 2019)

Meanwhile, however, another division within the *Facebook* structure, the data analytics platform *CrowdTangle* (a previously independent company that was acquired by *Facebook* in 2016), has engaged increasingly proactively with scholarly researchers. Once largely working with corporate clients especially in marketing, brand management, and journalism, the *CrowdTangle* team announced in early 2019 that it would open up ‘access to CrowdTangle to academics and the research community ... in the next few months’ (Facebook for Media, n.d.); a public application form enabling ‘university-based researchers and academics’ to request access to *CrowdTangle* was eventually launched in July 2020 (Shiffman & Silverman, 2020), and a growing number of scholarly research teams from around the world (including the present author) have been granted access to *CrowdTangle* and its datasets.

CrowdTangle does not represent a direct substitute for the datasets Social Science One promised: it provides data only on the activities of public pages, public groups, and public verified profiles on *Facebook*, as well as covering public activities on the *Facebook*-owned *Instagram* and on independent platforms *Twitter* and *Reddit*, while Social Science One had promised to provide some insights into broader patterns of content sharing across the entire *Facebook* platform. But this does not make it any less valuable to researchers; in particular, in sharp contrast to the restrictive, ‘orthogonal’ philosophy guiding Social Science One, as of late 2020 the *CrowdTangle* application process explicitly prioritised researchers who seek to address critical issues such as the spread of misinformation, election-related activities, COVID-19 information, racial justice, and wellbeing (Shiffman, 2020); similarly, *CrowdTangle* news updates to its

community of academic users highlight new research outputs even if they are critical of *Facebook's* role in problematic communicative events, and the platform actively supports and even encourages scholarly collaboration and data sharing across institutions. Put simply, if Social Science One was a highly public initiative that promised much and delivered little, *CrowdTangle* has quietly delivered considerably greater actual benefits to researchers.

Perhaps this apparent contradiction in *Facebook's* approach to engaging with scholarly researchers can be explained by the fact that, even in spite of the cult of personality that has developed around founder Mark Zuckerberg, a large corporate organisation like *Facebook* is not a monolithic unit controlled by an all-seeing CEO, but consists instead of a multitude of internal teams and units with their own working cultures and attitudes. Where the Social Science One initiative appears to have developed from the personal relationship between fellow Harvardians Zuckerberg and King (Pasternack, 2019), only to falter once the various technical and legal departments confronted the challenge of preparing and providing actual datasets to accredited researchers, *CrowdTangle's* established practices as a once-independent platform might have better enabled it to extend its existing data provision model from commercial to scholarly users. But this also illustrates the continuing precarity of social media research that relies on access to large-scale data from the platforms' APIs: amidst the internal politics of a company like *Facebook*, there is no guarantee that *CrowdTangle's* more open and permissive approach to data provision for research purposes will not be replaced by far more restrictive rules the next time *Facebook* is embroiled in a political debate about user privacy. Indeed, in March 2021 *Facebook* announced the launch of its 'Facebook Open Research & Transparency Analytics API' (FORT), which appears to provide data that are roughly equivalent to *CrowdTangle's*

data on the activities of public *Facebook* pages, but unlike *CrowdTangle* does not cover public *Facebook* groups (Lohman & Jagadeesh, 2021). Explicitly associated with the Social Science One initiative, it remains unclear whether FORT is intended to eventually replace *CrowdTangle* access for researchers (thereby reducing the breadth of data available to scholars, and reintroducing stricter data access controls), or whether the two services will continue to coexist in the long term.

Boom, Bust, and Beyond

The precarious new equilibrium that has emerged following the early boom and post-Cambridge Analytica bust of social media analytics research presents scholars with an unenviable set of choices (also cf. Bruns, 2019). Some researchers have chosen simply to seek out new and different opportunities, away from a field that offers only uncertain returns on their investment of time and effort; *Netvizz* creator Bernhard Rieder, for instance, stated that he would ‘take a break’ from his work: ‘the increasing hostility toward independent research is creating demands that can only be tackled through more sustainable institutional forms. ... At least in my immediate environment, I do not see such forms emerge’ (Rieder, 2018). As others follow suit, this represents a considerable loss of experience and expertise for the field – especially as those scholars who have invested the most energy in developing methodological innovations quit the field in frustration over the series of increasingly insurmountable obstacles that social media platforms have chosen to place in their way.

For others, their frustration has expressed itself in diametrically opposite ways: some scholars have come to take the view that the social and societal importance of their research supersedes the rules for data access stipulated by the platforms they study – especially, of course, when those rules appear designed deliberately to prevent research

that may reflect badly on those platforms' corporate policies and practices. In particular, as the functionality of platform APIs is restricted or (as in the case of *Instagram*) removed altogether, such scholars advocate a return to Web scraping and other data acquisition practices that are not officially sanctioned – or indeed explicitly forbidden – by platform providers (e.g. Rogers, 2018). Scraping-based data gathering approaches may also require the establishment of one or more 'fake' user accounts on a platform, in order to gain access to its content in a simulated Web browser session (a practice that is itself often explicitly identified as a violation of platform rules). Alternatively, they may rely on enrolling a number of ordinary platform users who are prepared to share their browser with a scraper tool that captures aspects of their platform experience.

The latter approach was pursued for instance by investigative journalism outlet *ProPublica*'s study of targeted political advertising on *Facebook*: it launched a browser plugin that captured the ads ordinary *Facebook* users encountered during their time on the platform. This method was initially successful and captured valuable information about current advertising campaigns that was not available from *Facebook*'s own, widely announced yet severely limited Ad Library tool – but it ultimately led to an arms' race between *Facebook* and *ProPublica* developers that saw the platform attempt repeatedly to obfuscate the HTML structure of its Web pages in order to make it more difficult for the browser plugin to accurately detect the advertising content embedded therein (Merrill & Tobin, 2019). The significantly asymmetrical distribution of developer resources between the opposing sides in such conflicts, which pit small-scale journalism start-ups against multibillion-dollar transnational corporations, necessarily means that such data gathering methods remain highly precarious, then; much in the same way that their API changes and restrictions have frustrated researchers and developers, the platforms can just as easily undermine non-API approaches by making

sudden and unannounced changes to the structure of their content as it is served to their users.

Meanwhile, and possibly in addition to such non-API methods, it remains important to utilise the remaining APIs to the full extent of their limited capabilities, and to share methodological insights and approaches even more actively than before. An unintended consequence of restrictive environments tends to be that they encourage innovative workarounds to flourish, and many researchers and teams working in the field of social media analytics have responded to the precarity of their environment by testing API functionality to its limits; further, the experience of opposition and even animosity from the major platform providers, combined with a firm belief in the societal relevance and importance of the research conducted, have encouraged a greater spirit of solidarity and cooperation between the diverse groups of scholars involved in this work. This should also include further initiatives to share datasets in a secure and responsible fashion, as Weller & Kinder-Kurlanda argue (2016), but – even if services like *CrowdTangle* are beginning to allow such sharing more explicitly – current platform rules continue to complicate such forms of cross-institutional collaboration, while institutional policies as well as national and transnational legal frameworks such as the European Union’s General Data Protection Regulation (GDPR) create additional hurdles that scholars have to negotiate.

In light of this increase in scholarly solidarity, finally, it is no accident that the ‘APIcalypse’ of 2018 has also produced an unprecedented number of joint statements and open letters from the social media research community and its philanthropic supporters (e.g. Bruns, 2018; Knight First Amendment Institute, 2018; Mozilla, 2019a, 2019b). New and improved methods for API-based and alternative data gathering

approaches remain inherently precarious if platforms are able to adjust their data structures without due consideration for the impact of such changes on current research projects, and many such open letters therefore call in the first place for new formal support mechanisms for critical, independent, public-interest scholarly research; if platforms prove unwilling to support such research out of a genuine sense of their own corporate social responsibility, then the implementation of such mechanisms also requires considerable societal, institutional, and especially also political pressure. Such pressure has increased in recent years, not least also as a result of the observable growth in social media content that contains abuse, hate speech, political extremism, and mis- and disinformation about critical issues (such as the COVID-19 pandemic) – but it will be of the utmost importance that this pressure does not dissipate when platforms react to it with initiatives such as Social Science One that generate substantial positive media coverage but deliver few tangible results.

In making the public argument for better scholarly access to social media data it will also be critical not to allow platforms like *Facebook* to use the Cambridge Analytica case or any future scandals to allege widespread improper data management practices within the scholarly community, as it clearly attempted to do in its initial statements on the matter (Grewal, 2018). In many jurisdictions, scholarly research is carefully overseen by university ethics committees and similar institutional review boards, and thus subject to a far greater level of critical oversight than exists in many of *Facebook's* and other platforms' commercial partner organisations; such oversight can certainly be strengthened and systematised further – and brought in line with standard field-wide perspectives on appropriate research ethics approaches, such as those expressed in the Association of Internet Researchers' Ethical Guidelines (franzke et al., 2019) – but *Facebook's* implicit (and sometimes explicit) argument that researchers cannot be

provided with social media data because they cannot be trusted to protect its users' privacy was always disingenuous and self-serving.

Much current social media research, indeed, seeks to protect the users of *Facebook* and other platforms, and our evolving digital society as such, from these companies' failure to provide an environment where they are safe not only from the misuse of their data, but also from other major contemporary problems such as abuse, hate speech, and disinformation. The current policy and technology settings that govern the ways in which the major providers engage with the scholars who conduct critical, independent, public-interest research on their platforms continue to render this research less insightful and effective than it has the potential to be. There is therefore now a pressing need to redress the balance between commercial, scholarly, personal, and societal interests, and thereby to enable researchers to act more fully as advocates for the rights of users.

References

Alaimo, K. (2017, December 27). All Tweets Belong in the Library of Congress.

CNN. <https://www.cnn.com/2017/12/27/opinions/library-of-congress-should-keep-all-tweets-opinion-alaimo/index.html>

Alaimo, K. (2018, October 17). Twitter's Misguided Barriers for Researchers.

Bloomberg. <https://www.bloomberg.com/opinion/articles/2018-10-16/twitter-s-barriers-for-academic-researchers-are-misguided>

- Berry, D.M. (2011). The Computational Turn: Thinking about the Digital Humanities. *Culture Machine*, 12.
<http://www.culturemachine.net/index.php/cm/article/view/440>
- Borra, E., & Rieder, B. (2014). Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets. *Aslib Journal of Information Management*, 66(3), 262–278. <https://doi.org/10.1108/AJIM-09-2013-0094>
- boyd, danah, & Crawford, K. (2012). Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society*, 15(5), 662–679.
<https://doi.org/10.1080/1369118X.2012.678878>
- 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/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786>
- Bruns, A. (2019). After the ‘APIcalypse’: Social Media Platforms and Their Fight against Critical Scholarly Research. *Information, Communication & Society*, 22(11), 1544–1566. <https://doi.org/10.1080/1369118X.2019.1637447>
- Bruns, A., & Burgess, J. (2016). Methodological Innovation in Precarious Spaces: The Case of Twitter. In H. Snee, C. Hine, Y. Morey, S. Roberts, and H. Watson (Eds.), *Digital Methods for Social Science* (pp. 17-33). Palgrave Macmillan.
- Bruns, A., & Burgess, J. (2015). Twitter Hashtags from Ad Hoc to Calculated Publics. In N. Rambukkana (Ed.), *Hashtag Publics: The Power and Politics of Discursive Networks* (pp. 13–28). Peter Lang.

- Bruns, A., & Moon, B. (2019). One Day in the Life of a National Twittersphere. *Nordicom Review*, 40(s1), 11–30. <https://doi.org/10.2478/nor-2019-0011>
- Bruns, A., & Weller, K. (2016). Twitter as a First Draft of the Present – and the Challenges of Preserving It for the Future. In W. Nejdl, W. Hall, P. Parigi, & S. Staab (Eds.), *Proceedings of the 8th ACM Conference on Web Science* (pp. 183–189). ACM Press. <https://doi.org/10.1145/2908131.2908174>
- Burgess, J., & Baym, N.K. (2020). *Twitter: A Biography*. New York University Press.
- Burgess, J., & Bruns, A. (2015). Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn. In G. Langlois, J. Redden, & G. Elmer (Eds.), *Compromised Data: From Social Media to Big Data* (pp. 93–111). Bloomsbury Academic.
- Burgess, J., & Matamoros-Fernández, A. (2016). Mapping Sociocultural Controversies across Digital Media Platforms: One Week of #gamergate on Twitter, YouTube, and Tumblr. *Communication Research and Practice*, 2(1), 79–96. <https://doi.org/10.1080/22041451.2016.1155338>
- Bucher, T. (2013). Objects of Intense Feeling: The Case of the Twitter API. *Computational Culture: A Journal of Software Studies*, 3. <http://computationalculture.net/article/objects-of-intense-feeling-the-case-of-the-twitter-api>
- Constine, J. (2018, April 4). Facebook Restricts APIs, Axes Old Instagram Platform amidst Scandals. *TechCrunch*. <http://social.techcrunch.com/2018/04/04/facebook-instagram-api-shut-down/>

- Dixon, D. (2012). Analysis Tool or Research Methodology? Is There an Epistemology for Patterns? In D.M. Berry (Ed.), *Understanding Digital Humanities* (pp. 191–209). Palgrave Macmillan.
- Facebook for Media. (n.d.). CrowdTangle for Academics and Researchers. <https://www.facebook.com/facebookmedia/blog/crowdtangle-for-academics-and-researchers>
- Flick, C. (2016). Informed Consent and the Facebook Emotional Manipulation Study. *Research Ethics*, 12(1), 14–28. <https://doi.org/10.1177/1747016115599568>
- franzke, a.s., Bechmann, A., Zimmer, M., Ess, C., & Association of Internet Researchers. (2020). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>
- Gadde, V., & Gasca, D. (2018, July 30). Measuring Healthy Conversation. *Twitter Blog*. https://blog.twitter.com/official/en_us/topics/company/2018/measuring_healthy_conversation.html
- Graham, T., & Ackland, R. (2017). Do Socialbots Dream of Popping the Filter Bubble? The Role of Socialbots in Promoting Deliberative Democracy in Social Media. In R.W. Gehl & M. Bakardjieva (Eds.), *Socialbots and Their Friends: Digital Media and the Automation of Sociality* (pp. 187–206). Routledge.
- Grewal, P. (2018, March 16). *Suspending Cambridge Analytica and SCL Group from Facebook*. Facebook Newsroom. <https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/>

- Hermida, A. (2010). Twittering the News: The Emergence of Ambient Journalism. *Journalism Practice*, 4(3), 297–308. <https://doi.org/10.1080/17512781003640703>
- Hughes, C. (2019, May 9). It's Time to Break Up Facebook. *New York Times*. <https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>
- Iliadis, A., & Russo, F. (2016). Critical Data Studies: An Introduction. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716674238>
- Jünger, J., & Keyling, T. (2017). Facepager: An Application for Generic Data Retrieval through APIs. <https://github.com/strohne/Facepager>
- Kazansky, B., Torres, G., van der Velden, L., Wissenbach, K., & Milan, S. (2019). Data for the Social Good: Toward a Data-Activist Research Agenda. In A. Daly, S.K. Devitt, & M. Mann (Eds.), *Good Data* (pp. 244–259). Institute of Network Cultures. http://networkcultures.org/wp-content/uploads/2019/01/Good_Data.pdf
- King, G., & Persily, N. (2018). A New Model for Industry-Academic Partnerships. <https://gking.harvard.edu/partnerships>
- Kirkorian, R. (2014, April 17). Twitter #DataGrants Selections. *Twitter Blog*. https://blog.twitter.com/engineering/en_us/a/2014/twitter-datagrants-selections.html
- Knight First Amendment Institute at Columbia University. (2018, August 6). Knight Institute Calls on Facebook to Lift Restrictions on Digital Journalism and Research. https://knightcolumbia.org/sites/default/files/content/Facebook_Letter.pdf

Kramer, A.D.I., Guillory, J.E., & Hancock, J.T. (2014). Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

<https://doi.org/10.1073/pnas.1320040111>

Lohman, T., & Jagadeesh, K. (2021, March 24). New Analytics API for Researchers Studying Facebook Page Data. *Facebook Research Blog*.

<https://research.fb.com/blog/2021/03/new-analytics-api-for-researchers-studying-facebook-page-data/>

Mac, R., & Warzel, C. (2018, July 24). Top Facebook Security Officer’s Departing Memo: ‘We Need to Be Willing to Pick Sides’. *BuzzFeed News*.

<https://www.buzzfeednews.com/article/ryanmac/facebook-alex-stamos-memo-cambridge-analytica-pick-sides>

Merrill, J.B., & Tobin, A. (2019, January 28). Facebook Moves to Block Ad Transparency Tools—Including Ours. *ProPublica*.

<https://www.propublica.org/article/facebook-blocks-ad-transparency-tools>

Mozilla. (2019a, February 11). Open Letter: Facebook, Do Your Part Against Disinformation. *The Mozilla Blog*. [https://blog.mozilla.org/blog/2019/02/11/open-](https://blog.mozilla.org/blog/2019/02/11/open-letter-facebook-do-your-part-against-disinformation)

[letter-facebook-do-your-part-against-disinformation](https://blog.mozilla.org/blog/2019/02/11/open-letter-facebook-do-your-part-against-disinformation)

Mozilla. (2019b, March 27). *Facebook and Google: This Is What an Effective Ad Archive API Looks Like*. The Mozilla Blog.

<https://blog.mozilla.org/blog/2019/03/27/facebook-and-google-this-is-what-an-effective-ad-archive-api-looks-like>

- Ozkula, S.M., Reilly, P., & Hayes, J. (2020). Easy Data, Usual Suspects, Same Old Places? A Systematic Review of Methodological Approaches in Digital Activism Research between 1995-2019. *AoIR Selected Papers of Internet Research*, 2020. <https://doi.org/10.5210/spir.v2020i0.11298>
- Palen, L., Starbird, K., Vieweg, S. and Hughes, A. (2010). Twitter-Based Information Distribution during the 2009 Red River Valley Flood Threat. *Bulletin of the American Society for Information Science and Technology*, 36: 13–17. <https://doi.org/10.1002/bult.2010.1720360505>
- Papacharissi, Z. (2010). *A Private Sphere: Democracy in a Digital Age*. Polity.
- Papacharissi, Z., & de Fatima Oliveira, M. (2012). Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt. *Journal of Communication*, 62, 266–282. <https://doi.org/10.1111/j.1460-2466.2012.01630.x>
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. Penguin.
- Pasternack, A. (2018, August 22). Facebook Reinstates Data Firm It Suspended for Alleged Misuse, but Surveillance Questions Linger. *Fast Company*. <https://www.fastcompany.com/90219826/why-did-facebook-re-friend-a-data-firm-that-raised-spying-concerns>
- Pasternack, A. (2019, October 28). Frustrated Funders Exit Facebook’s Election Transparency Project. *Fast Company*. <https://www.fastcompany.com/90412518/facebooks-plan-for-radical-transparency-was-too-radical>

- Pear Analytics. (2009). Twitter Study – August 2009. <http://pearanalytics.com/wp-content/uploads/2012/12/Twitter-Study-August-2009.pdf>
- Rieder, B. (2013). Studying Facebook via Data Extraction: The Netvizz Application. *Proceedings of the 5th Annual ACM Web Science Conference*, 346–355. <https://doi.org/10.1145/2464464.2464475>
- Rieder, B. (2018, August 11). Facebook’s App Review and How Independent Research Just Got a Lot Harder. *The Politics of Systems: Thoughts on Software, Power, and Digital Method*. <http://thepoliticsofsystems.net/2018/08/facebooks-app-review-and-how-independent-research-just-got-a-lot-harder/>
- Rogers, R. (2014). Foreword: Debanalising Twitter: The Transformation of an Object of Study. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and Society* (pp. ix–xxvi). Peter Lang.
- Rogers, R. (2018). Social Media Research after the Fake News Debacle. *Partecipazione e Conflitto: The Open Journal of Sociopolitical Studies*, 11(2), 557–570. <https://doi.org/10.1285/i20356609v11i2p557>
- Rogers, R., & Marres, N. (2000). Landscaping Climate Change: A Mapping Technique for Understanding Science and Technology Debates on the World Wide Web. *Public Understanding of Science*, 9(2), 141–163. <https://doi.org/10.1088/0963-6625/9/2/304>
- Roth, Y., & Johnson, R. (2018, July 24). New Developer Requirements to Protect Our Platform. *Twitter Blog*. https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform.html

- Schmidt, J.-H. (2014). Twitter and the Rise of Personal Publics. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and Society* (pp. 3–14). Peter Lang.
- Shaw, F., Burgess, J., Crawford, K., & Bruns, A. (2013). Sharing News, Making Sense, Saying Thanks: Patterns of Talk on Twitter during the Queensland Floods. *Australian Journal of Communication*, 40(1), 23-40.
- Shiffman, N. (2020, November). CrowdTangle for Academics and Researchers. *CrowdTangle*. <http://help.crowdtangle.com/en/articles/4302208-crowdtangle-for-academics-and-researchers>
- Shiffman, N., & Silverman, B. (2020, July 31). CrowdTangle Opens Public Application for Academics. *Facebook Research*. <https://research.fb.com/blog/2020/07/crowdtangle-opens-public-application-for-academics/>
- Social Science One. (2018a). <https://socialscience.one/>
- Social Science One. (2018b). Requests for Proposals: Social Media and Democracy Research Grants. <https://socialscience.one/rfps>
- Social Science One. (2019, December 11). Public Statement from the Co-Chairs and European Advisory Committee of Social Science One. <https://socialscience.one/blog/public-statement-european-advisory-committee-social-science-one>
- Stone, B. (2010, April 14). Tweet Preservation. *Twitter Blog*. https://blog.twitter.com/official/en_us/a/2010/tweet-preservation.html

Weller, K., & Kinder-Kurlanda, K.E. (2016). A Manifesto for Data Sharing in Social Media Research. *Proceedings of the 8th ACM Conference on Web Science*, 166–172. <https://doi.org/10.1145/2908131.2908172>