THE COUNCIL FOR BIG DATA,
ETHICS, AND SOCIETY

October 2016

The Right to be Forgotten or the Duty to be Remembered?

Twitter data reuse and implications for user privacy

Helena Uršič, Leiden University

This case study draws on the Politwoops case and brings up the question whether politicians' statements on a publicly accessible platform should be treated their private conversations and permanently deleted if the author requires so. This situation is compared with ineffective deletion of non-politicians' personal data. The key question is who and why should have the right to have data deleted from mixed private/public platforms.

From a chatroom to a livejournal

Twitter, Inc. provides online social networking and microblogging service. The Company offers users the ability to follow other users' activity, read, and post tweets. Twitter serves customers worldwide.[1] Created in March 2006 by Jack Dorsey, Evan Williams, Biz Stone and Noah Glass, the service rapidly gained worldwide popularity. In 2013, Twitter was one of the ten most-visited websites and has been described as "the SMS of the Internet". As of May 2015, Twitter has more than 500 million users, out of which more than 332 million are active.[2] In the course of its existence, Twitter's nature as a social media network has been

[1] <http://www.bloomberg.com/quote/TWTR:US>.
[2] <https://en.wikipedia.org/wiki/Twitter>.

Data&Society

This case study was first written for the Council for Big Data, Ethics, and Society. Funding for this Council was provided by the National Science Foundation (#IIS-1413864). For more information on the Council, see: http://bdes.datasociety.net/.

changing. Most noticeably, Twitter has grown from a social network of people sharing information about their own activities and related events, into a global news media, providing real-time information generated by their users on a large scale. Throughout the years, the platform created as "*a chatroom*" has transformed into a global "*livejournal*". This change is also reflected in Twitter's welcome note, which no longer reads "*What are you doing*" but has been changed into "*What is happening*?"[3]

Twitter differs from other social media providers in the fact that it is open and public by default, unlike, for example, Facebook and Google's social networking platforms. The private nature of the latter two is also reflected in their stricter privacy policies.[4] Twitter's public nature enables easy access to some additional functions that are highly useful for social and computer scientists, such as sharing a large amount of user generated data via their APIs for further research and analysis. A vast amount of freely available public data makes Twitter highly relevant and attractive for research, marketing and analytics.

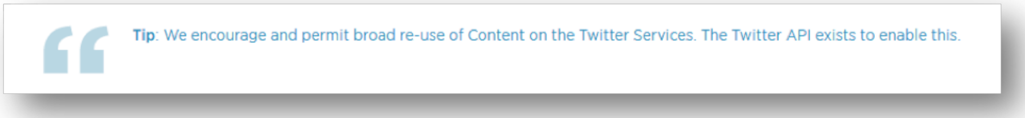
Broad data re-use permitted and encouraged

As legal scholar Jonathan Zittrain notes, today's era marks the end of "open internet" and the move to more closed ecosystems, such as Google, Apple and Facebook, whose business models rely primarily on advertising and corporate partnerships and, crucially for this case study, on reselling and sharing the data produced collectively by the platform's millions of users.[5]

An individual or an entity that seeks to exploit Twitter data (or the Content[6]) has, roughly speaking, two options: it can either use Twitter's streaming APIs to conduct a real time data analyses, or it can deploy REST APIs, which enable more sophisticated historical data analytics.[7]

In addition to provisions set by Twitter's general terms and conditions ("the Terms"), the relation between the social media provider and developers who use APIs to develop and implement services is defined by Twitter's developer agreement and policy. These documents illustrates how ambivalent Twitter's behavior is in relation to data reuse. On the one hand, Twitter encourages broad data reuse. In the Terms, they state clearly:

[3] Weller, K., Bruns, A., Burgess, J. , Mahrt, M. & Puschmann, C. (eds.), Twitter & Society, Peter Lang Publishing Inc., New York 2014, pp. 176-177.
[4] Ibid., p. 170.
[5] Zittrain, J., The future of the internet and how to stop it, Yale University Press, New haven & London, 2012.
[6] As referred to in Twitter's terms and conditions, section I.A "Definitions" <https://twitter.com/tos?lang=en>.
[7] <https://dev.twitter.com/rest/public>.



Tip: We encourage and permit broad re-use of Content on the Twitter Services. The Twitter API exists to enable this.

On the other hand, a more detailed analysis of the terms reveals a less clear picture. Twitter recently centralized the data analytics on its own platform, which leaves developers less room for maneuver.[8] In addition, developers are limited with Twitter's strict terms and policies.[9] Nevertheless, broad data reuse is still a distinctive advantage, which helps Twitter attract a considerably high number of developers.

Twitter's ephemeral nature increases users privacy expectations

All social media platforms contend with the challenge of being a mix of publicly accessible, privately controlled platforms on which users have a wide range of privacy expectations—they are simultaneously public soap boxes and gated communities. Each platform balances these expectations in its own way, with unique consequences. Twitter supports a several ways of communication, which vary in the degree of privacy protection. Tweets are public postings that everyone can see. Direct messages enable users to have private conversations with other Twitter users.[10] In addition, Twitter distinguishes regular and verified Twitter accounts. The latter are marked with a blue badge and used to establish authenticity of identities of publicly recognizable individuals and brands on Twitter.[11]

Twitter's basic proposition is that all the content generated by users (apart from the content shared via direct messages) is public by default. However, the open nature of the platform is not incompatible with the ideas of privacy rights neither it means that Twitter is no longer bound by legal requirements. Under the EU law, Twitter is considered a data controller, which translates into the duty to comply with data protection rules.[12]

Interestingly, the recent research has shown that despite the public nature of its platform,

[8] In particular the termination of the contract with Datasift received a lot of media attention. See Nick Halstead, Twitter Ends its Partnership with DataSift – Firehose Access Expires on August 13, DataSift Blog, April 11, 2015 <http://blog.datasift.com/2015/04/11/twitter-ends-its-partnership-with-datasift-firehose-access-expires-on-august-13-2015/>.

[9] Puschmann, C. & Burgess, J., The politics of Twitter Data, HIIG Discussion Paper Series, Discussion Paper No. 2013-01 2016, p. 7.

[10] <https://support.twitter.com/articles/14606#>.

[11] <https://support.twitter.com/articles/119135#>.

[12] Controller is the one that determines the purposes and the means of the processing of personal data. Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L 281.

Twitter users still have considerably high privacy expectations. Zimmer & Proeferes explain that this is because of the ephemeral nature of Twitter as social media.[13] Instant messaging on Twitter is carried out differently than perceived by an average user. The momentary relevance of tweets they post often contrasts the fact that in reality messages remain accessible far longer than expected.[14]

The need for privacy protection is acknowledged in Twitter's Privacy Policy[15] and, even more directly, through the Terms and Developer Policy. The latter urges APIs developers to respect users privacy and control over the Content posted on Twitter.[16]

Politwoops and celebrities' rights to their own data

As explained above, the data generated on Twitter is not meant to lie dormant. Twitter tries to make use of it by sharing datasets with third parties—mostly advertising and commerce companies—developing its own data analytics, enabling API streaming and managing their service. The examples below show two ways of how personal data reuse can be carried out through collaboration with API developers. The first example relates to politicians' data, while the second one concerns an average user's data.

In relation to politicians' data, the issue of data reuse was widely discussed in the media after Twitter had terminated *Politwoops* access to its application program interface (API).[17] *Politwoops,* which advertised themselves as "*the only comprehensive collection of deleted twits by politicians that offers a window into what they hoped you didn't see*", used APIs to identify politicians' deleted tweets and (re)publish them on their own platform even after the politician had chosen to make them no longer visible on the Twitter platform itself.[18]

Twitter's standard terms and conditions urge APIs users to cease processing of the tweets that have been deleted by original posters from Twitter as soon as Twitter flags such twits

[13] Weller, K., Bruns, A., Burgess, J. , Mahrt, M. & Puschmann, C. (eds.), Twitter & Society, Peter Lang Publishing Inc., New York 2014, pp. 170.

[14] Similarly, using @ sign in a tweet have far-reaching consequences, which a regular user hardly predict. Namely, those twits are shared not only with a single person mentioned in the tweet but distributed widely to all his/her followers. *Supra*, note 3, p.171.

[15] <https://twitter.com/privacy?lang=en>.

[16] Section 3 <https://dev.twitter.com/overview/terms/policy>.

[17] However, the platform was then reset and today it operates normally. Hern, A., Twitter blocks access to political transparency organisation Politwoops, The Guardian, August 24, 2015 <http://www.theguardian.com/technology/2015/aug/24/twitter-blocks-access-political-transparency-organisation-politwoops> .

[18] <http://politwoops.eu/>.

and alerts.[19] However, *Politwoops* used its access in the opposite manner of how Twitter intended, by using the alerts as a way to keep track of and share deleted tweets.[20]

Stating that "honouring the expectation of user privacy for all accounts is a priority for us, whether the user is anonymous or a member of Congress" the company moved to block *Politwoops'* access to Twitter data.[21] This was driven not only by the violation of the terms, but also—and maybe predominantly—by privacy concerns at a time when tech giants were regularly coming under fire from European regulators for alleged violations of the relatively strict European laws.[22]

Politwoops was a hard case since it required from Twitter to strike the right balance between two opposing values: protection of politicians' right to control their personal data and public demand for more transparency in political discussions. It is not surprising that it received massive media attention and triggered heated discussions.[23] At its heart, this is a question of whether politicians' statements on a publicly accessible platform should be treated as their own personal data or as a form of public speech similar to quotes in newspapers or statements made at public events.

The right of non-celebrities to control their own data

Consider whether the outcome would be any different if a twit, processed by APIs developers as part of their data streams, was deleted by an average user and not a politician. In this case public transparency would only play a minor role (if any) as an ethical justification for doing so.

An average Twitter user would often delete a tweet that she has posted recklessly or in hurry, and would expect that once a posting disappears from her timeline, it is not publicly available anymore. However, if such a tweet has been shared with third parties, the situation gets more complicated. Consider a statement from an executive of a small Italian

[19] "[…] your Service should execute the unfavorite and delete actions by removing all relevant messaging and Twitter Content, not by publicly displaying to other end users that the Tweet was unfavorited or deleted." <https://dev.twitter.com/overview/terms/agreement-and-policy>.

[20] Noriega, M., Delete your twits, rewrite history? The Politwoops controversy, explained, Vox.com, August 26, 2015 <http://www.vox.com/explainers/2015/8/26/9211117/politwoops-delete-twits>.

[21] Hern, A,, Twitter blocks access to political transparency organisation Politwoops, August 24, 2015 <https://www.theguardian.com/technology/2015/aug/24/twitter-blocks-access-political-transparency-organisation-politwoops August 24, 2015>.

[22] Stupp, C., Twitter shuts down transparency tracker, EurActiv.com, August 25, 2015 <http://www.euractiv.com/sections/infosociety/twitter-shuts-down-mp-transparency-tracker-317027>.

[23] See for example Bump, P.: Twitter's terrible decision to block Politwoops https://www.washingtonpost.com/news/the-fix/wp/2015/06/03/twitters-terrible-decision-to-block-politwoops/

enterprise that uses Twitter data as part of its business model:[24]

> "When a user deletes a tweet on his (or her) account, Twitter wants me to delete the tweet I gathered from streaming API the very same moment. I've been confirmed that nobody totally fits Terms and conditions, so we are in good company, and that in Italy, we use the uttermost attention about this matter."

Article 12, para. (b) of the EU Data protection directive obliges data controllers, i.e. an individual or a legal person who determines the means and purposes of data processing, to stop processing of data, which is inaccurate or incomplete, if a data subject requires so. As APIs users have the ability to decide on the means and the purposes for which personal data streamed from Twitter will be used, they fall under the definition of controller and hence they need to comply with the directive's requirement.[25] When a user deletes a statement from her Twitter timeline, this is a clear indication of her desire to have the data deleted. According to Twitter's terms of access, following such an action a compliant API user should remove the critical data from the database they have built via Twitter's API.

> Article 12
> Right of access
> Member States shall guarantee every data subject the right to obtain from the controller:
> (a) […]
> (b) as appropriate the rectification, erasure or blocking of data the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data;
> (c) […].

Twitter's Developer Policy reflects the above requirement as it urges API users to stop processing the deleted content. The requirement is elaborated in section 3 (a) of the policy:

> Take all reasonable efforts to do the following, provided that when requested by Twitter, you must promptly take such actions:
>
> i. Delete Content that Twitter reports as deleted or expired;

[24] Arrigo, A., Data reuse – can you really do it? LinkedIn Pulse, Jun 23, 2015 <https://www.linkedin.com/pulse/data-reuse-can-you-really-do-alessandro-arrigo?trk=prof-post>.

[25] Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data, [1995] OJ L 281.

ii. Change treatment of Content that Twitter reports is subject to changed sharing options (e.g., become protected); and
iii. Modify Content that Twitter reports has been modified.

However, effective deletion of tweets from SME's databases is unlikely to happen. The reasons are twofold: [26] first, national data protection authorities lack resources and knowledge to perform the physical raids necessary to check a developer's databases for compliance; second, Twitter does not check APIs developers' compliance. This is not surprising. As Twitter has no direct commercial interest in enforcing the deletion rule, why would they even bother? As long as their own terms and conditions are in line with the legal requirements and disclose their disagreement such conduct,[27] Twitter can easily shake off the burden.

Discussion questions

Contrary to the public outrage following the *Politwoops* case, no such reaction is noticed when personal data of millions of users is unjustly kept by third parties. This state-of-affairs raises the following dilemmas:

- What are risks and concerns related to business models that disable effective deletion of someone's public posts?
- Should there be an ethical and/or legal distinction between public figures using social media and regular users with regards to their right to delete their data? What should determine whether any given user is a public figure or a regular user?
- Do you consider Twitter a public forum or a private platform? What about Facebook? Instagram? Other social media platforms? Explain why.
- Should Twitter be required to act as a gatekeeper and actively monitor its API developers' compliance?
- Is the burden that the the right to deletion imposes on social media companies and third parties too heavy?

[26] The interviews were conducted as part of Eudeco project, a European Commission founded project <http://data-reuse.eu/>.
[27] Twitter terms – developer agreement & policy – see section I.2 and I.3. "Maintain the Integrity of Twitter's products" and "Respect Users' Control and Privacy".

- In countries where data protection laws are less strict or do not exists at all, should reuse of deleted twits be tolerated by Twitter?

- Can you think of any technical enablers that would help small and medium size enterprises comply with the requirements and fulfill Twitter users expectations in a less intrusive way?