THE EVOLUTION OF FACEBOOK'S GRAPH API

Marcus Burkhardt
University of Siegen

Anne Helmond
University of Amsterdam

Tatjana Seitz
University of Siegen

Fernando N. van der Vlist
Utrecht University / University of Siegen

(authors listed in alphabetical order)

Introduction

The Facebook–Cambridge Analytica data scandal continues to occupy the public debate about data privacy and protection. Central to the ongoing controversy are Facebook's data-sharing practices with third parties and their use of data for targeted advertising in political campaigns. The platform responded by changing its policies, more strictly governing third-party data access via Facebook's Graph API, and by removing particular targeting options from Facebook's Marketing API for programmatic advertisers. Facebook's application programming interfaces (APIs) are core to many ongoing concerns and enable third-party app developers to access data and functionality. In the case of Cambridge Analytica, the collection of data about millions of people via their friends via the Friends API was not a bug but a feature of how the API was designed to work. Understanding how the platform and its APIs evolve and how a leading platform such as Facebook responds to issues and concerns requires looking closely and empirically at the evolution of access points, data structures, and graph data structures (cf. Barrett and Kreiss, 2019; Helmond et al., 2019). Therefore, this paper contributes an in-depth technical and empirical perspective on the evolution of

Facebook's Graph API in its entirety from its introduction as the Facebook API (beta) in 2006 onwards, and how it evolved into one of the most significant web APIs and an integral part of contemporary advertising infrastructures and web development cultures. As we show, key changes in the Graph API evolution are characterized by phases of experimentation, standardization, commercialization, and regulation. Our study is unique in (i) its level of empirical historical granularity, utilising and reconstructing the entire collection of available materials on the Facebook Platform in the Internet Archive, (ii) in outlining an innovative research methodology for studying the evolution of a platform in terms of its APIs, and (iii) in providing a granular empirically informed history of Facebook's Graph API.

API studies

Web APIs have been core research objects for studying the politics of platforms and data sharing (e.g. Bodle, 2010), their function as data collection devices (e.g. Lomborg and Bechmann, 2014), as elements of infrastructure (e.g. Raetzsch et al., 2019; Helmond et al., 2019), and the research politics of increasingly restricted APIs (e.g. Bruns, 2019). However, few studies focus specifically on the technicity of APIs, such as on data representations, structures, or governance. Yet the technicity of APIs is crucial for understanding the politics of data sharing and how APIs represent and structure phenomena and temporarily stabilise them (cf. Pridmore, 2016). Therefore, instead of using APIs as an umbrella term for data retrieval in general, we conduct the necessary historical technical fieldwork (cf. Rieder et al., 2015) for examining the evolving architecture and interfaces of Facebook's popular web APIs. With our historical case study, we contribute to the emerging scholarship on the evolution of digital platforms, APIs, and governance (e.g. Barrett and Kreiss, 2019; Helmond & van der Vlist, 2019; Helmond et al., 2019).

Approach

Our empirical historical analysis of Facebook's Graph API is based on the entire corpus of available archived developer documentation held by the Internet Archive Wayback Machine (August 2006–February 2020). This covers the initial beta Facebook API up to the current version 6.0 of the Graph API, which was scraped from the live web. Based on this sizable corpus of archived web sources, we reconstructed the structure of the entire Graph API at different levels of granularity: from the overall structure to the individual objects or endpoints, to the associated properties, connections, and parameters or specifications of those objects.

To reconstruct and analyse the overall structure of the Graph API, we considered the networked structure of the reference documentation pages over time. We also relied on Facebook's own account of change as documented in its changelogs, in which changes to the Graph API and Marketing API are announced. These include announcements of new features, so-called breaking changes and deprecations, and information about removed or restricted data fields and edges, which can be considered important traces of governance. Looking at the evolving structure of the Graph API, we found that Facebook continues to expand with Messenger and acquisitions of Atlas, Instagram, and WhatsApp. Although these remain separated on the 'front-end', they are

increasingly integrated at the 'back-end' as reflected in the structure of the Graph API, which now is Facebook's unified foundational data infrastructure for app and business development.

We further examined how some of the most connective nodes in the Graph API – the User and the Page – evolved. These objects represent the profile pages of users and business, organisations and public figures. How are these objects conceptualised, what kind of relations can they enter into, and which operations can be performed on them? And how did the User and Page evolve? Importantly for Facebook, whose core business model is based on advertising, the user is also a targetable user. Therefore, we looked at the evolving targeting options available to target users on Facebook via the Marketing API, which has been a core component of the Graph API since 2012. These options provide another account of the user and its representation by Facebook: targetable fields do not only consist of information provided by users on their profiles but also many calculated data derivatives beyond the user's control. This approach also allows us to trace specific targeting options of interest, such as the controversial options related to race, gender, and housing exclusions (e.g. Angwin and Parris, 2016).

A scalable reading of APIs

Based on the analysis, we provide a "scalable reading" (Weitin, 2017) of the evolution of Facebook's Graph API which provides insights in how data and data flows are governed through changes in data structures and permissions. By considering the evolving structures of APIs and individual data objects, we may develop further empirically informed critiques of platforms, APIs, and their data.

References

Angwin J and Parris T (2016) Facebook lets advertisers exclude users by race. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

Barrett B and Kreiss D (2019) Platform transience: Changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review* 8(4). https://policyreview.info/articles/analysis/platform-transience-changes-facebooks-policies-procedures-and-affordances-global

Bodle R (2011) Regimes of sharing: Open APIs, interoperability, and Facebook. *Information, Communication & Society* 14(3): 320–337. DOI: 10.1080/1369118X.2010.542825.

Bruns A (2019) After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society* 22(11): 1544–1566. DOI: 10.1080/1369118X.2019.1637447.

Helmond A and van der Vlist FN (2019) Social media and platform historiography: Challenges and opportunities. *TMG – Journal for Media History* 22(1): 6–34. DOI: 10.18146/tmg.434.

Helmond A, Nieborg DB and van der Vlist FN (2019) Facebook's evolution: development of a platform-as-infrastructure. *Internet Histories* 3(2): 123–146. DOI: 10.1080/24701475.2019.1593667.

Lomborg S and Bechmann A (2014) Using APIs for data collection on social media. *The Information Society* 30(4): 256–265. DOI: 10.1080/01972243.2014.915276.

Pridmore J (2016) A social API for that: Market devices and the stabilization of digital identities. In: van der Ploeg I and Pridmore J (eds) *Digitizing Identities: Doing Identity in a Networked World*. Routledge.

Raetzsch C, Pereira G, Vestergaard LS, et al. (2019) Weaving seams with data: Conceptualizing City APIs as elements of infrastructures. *Big Data & Society* 6(1). DOI: 10.1177/2053951719827619.

Rieder B, Abdulla R, Poell T, et al. (2015) Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said". *Big Data & Society* 2(2). DOI: 10.1177/2053951715614980.

Weitin, T (2017) Scalable reading. *Zeitschrift für Literaturwissenschaft und Linguistik*. 47(1), 1–6. DOI: 10.1007/s41244-017-0048-4.