



## UvA-DARE (Digital Academic Repository)

### Apps and Their Affordances for Data Investigations

Weltevrede, E.

**DOI**

[10.2307/j.ctv1qr6smr.37](https://doi.org/10.2307/j.ctv1qr6smr.37)

[10.5117/9789462989511\\_ch34](https://doi.org/10.5117/9789462989511_ch34)

[10.1017/9789048542079.035](https://doi.org/10.1017/9789048542079.035)

**Publication date**

2021

**Document Version**

Final published version

**Published in**

The Data Journalism Handbook

**License**

CC BY-NC-ND

[Link to publication](#)

**Citation for published version (APA):**

Weltevrede, E. (2021). Apps and Their Affordances for Data Investigations. In L. Bounegru, & J. Gray (Eds.), *The Data Journalism Handbook: Towards A Critical Data Practice* (pp. 246-256). (Digital Studies). Amsterdam University Press. <https://doi.org/10.2307/j.ctv1qr6smr.37>, [https://doi.org/10.5117/9789462989511\\_ch34](https://doi.org/10.5117/9789462989511_ch34), <https://doi.org/10.1017/9789048542079.035>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

## 34. Apps and Their Affordances for Data Investigations

*Esther Weltevrede*

### Abstract

Exploring app–platform relations for data investigations.

**Keywords:** apps, social media platforms, digital methods, data infrastructures, data journalism, data investigations

Recently, Netvizz, a tool to extract data from Facebook, lost access to Facebook's Page Public Content Access feature. This seems to have terminated the precarious relationship its developer, the digital methods researcher Bernhard Rieder, has maintained with the Facebook API over the past nine years.<sup>1</sup> The end of Netvizz is symptomatic of a larger shift in digital research and investigations where platforms are further restricting data collection through their application programming interfaces (APIs) and developer policies. Even though the actual effectiveness of the Cambridge Analytica methods are questioned (Lomas, 2018; Smout & Busvine, 2018), the scandal prompted a debate on privacy and data protection in social media and in turn Facebook responded by further restricting access to data from their platforms.

Since the initial announcement in March 2018,<sup>2</sup> the staggered implementation of data access restrictions by Facebook within its larger family of apps has made visible the vast network of third-party stakeholders that have come to rely on the platform for a wide variety of purposes. Apps stopped working, advertising targets have been restricted, but the party most severely hit seems to be digital researchers. This is because apps that

1 <http://thepoliticsofsystems.net/?s=netvizz>

2 <https://about.fb.com/news/2018/03/cracking-down-on-platform-abuse/>

have data collection as their primary purpose are no longer allowed. Digital researchers resisted these changes (Bruns, 2018) by arguing that they would be to the cost of research in the interest of the public good. The list of references to the Netvizz article (Rieder, 2013) comprise over 450 publications, which in reality easily exceed that amount—just consider the many student research projects making use of the tool. Similarly, an ad hoc inventory by Bechmann of studies that “could not have existed without access to API data”<sup>3</sup> comprises an impressive list of journalism, social science and other digital research publications.

Reflecting on the impact data access restrictions have on digital research, authors have contextualized these developments and periodized the past decade as “API-based research” (Venturini & Rogers, 2019) or “API-related research” (Perriam et al., 2020). These are defined as approaches to digital research based on the extraction of data made available by online platforms through their APIs. Certainly, APIs—with their data ready-made for social research—have lowered the threshold for research with social media data, not to mention that they allowed a generation of students to experiment with digital research. No technical skills are required, and for web data standards, the data is relatively clean. API-based research has also been critiqued from the onset, most notably because of APIs’ research affordances driven by convenience, affecting the researchers’ agency in developing relevant research questions (Marres, 2017).

This chapter picks up on recent calls for “post-API research” by Venturini and Rogers (2019) and the Digital Methods Initiative and focuses on the opportunities that arise in response to recent developments within social media ecosystems.<sup>4</sup> Digital research, in the sense employed in this chapter, is defined by the methodological principle of “following the medium,” responding to and interfacing methods with developments in the digital environment. In what follows I approach the recent API restrictions by arguing for the renewed need for, and potential of, creative and inventive explorations of different types of sociotechnical data that are key in shaping the current platform environments. I continue by picking up on the opportunities that have been identified by digital researchers, and adding to that by proposing a methodological perspective to study app–platform relations. In doing so I hope to offer data journalists interested in the potential of social data for storytelling (see, e.g., the chapter by Lam Thuy Vo in this volume), some

3 <https://docs.google.com/document/d/15YKeZFSUc1j03b4lW9YXxGmhYEnFx3TSy68qCrXgBEI/edit>

4 <https://wiki.digitalmethods.net/Dmi/WinterSchool2020>

starting points for approaching investigations with and about platforms and their data in the current post-API moment.

Digital methods have in common that they utilize a series of data collection and analysis techniques that optimize the use of native digital data formats. These emerge with the introduction of digital media in social life. Digital methods researchers develop tools inspired by digital media to be able to handle these data formats in methodologically innovative ways. The history of digital methods can therefore also be read as narrating a history of key data formats and data structures of the Internet; they are adaptive to changes of the media and include these in analysis. In what follows, I would like to contribute to post-API research approaches by proposing a perspective to study platforms as data infrastructures from an app–platform perspective. The impact the data access restrictions have on the larger media ecosystems attest to the fact that advanced, nuanced knowledge of platform infrastructures and their interplay with third-party apps is direly needed. It demonstrates the need for a broadened data infrastructure literacy (Gray et al., 2018), in addition to knowledge about how third-party companies and apps operate in social media environments.

## **Apps and Platforms-as-Infrastructure**

The platform data restrictions are part and parcel of developments of social media into platforms-as-infrastructure. These developments highlight the evolution of digital ecosystems' focus on corporate partnerships (Helmond et al., 2019). After a year of negative coverage following the platform's role in elections, Zuckerberg posted a note sketching out the platform's changing perspective from "connecting people" to building a "social infrastructure" (Helmond et al., 2019; Hoffmann et al., 2018).<sup>5</sup> The notion of social infrastructure both highlights social activities as the platform's core product to connect and create value for the multiple sides of the market, as well as the company's shift from a social network into a data infrastructure, extending the platform to include their websites and the larger family of 70 apps (Nieborg & Helmond, 2019).<sup>6</sup> This infrastructural turn marks a next step in the platform's ability to extend their data infrastructure into third-party apps, platforms and websites, as well as facilitating inwards integrations.

5 <https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/>

6 <https://www.appannie.com/en/>

Even though platforms-as-infrastructure receive increasing attention (Plantin et al., 2018), as do individual apps, how apps operate on and between data infrastructures is understudied and often unaccounted for. Yet apps continually transform and valorize everyday practices within platform environments. I use a relational definition of apps by focusing on third-party apps, defined as applications built on a platform by external developers, not owned or operated by the platform. When an app connects to a platform, access is granted to platform functions and data, depending on the permissions. Apps also enable their stakeholders—for example, app stores, advertisers or users—to integrate and valorize them in multiple, simultaneous ways. In other words, apps have built-in tendencies to be related to, and relate themselves within different operative data infrastructures. This specific position of third-party apps makes them particularly appropriate for studies into our platform-as-infrastructure environments.

Social media platforms pose methodological challenges, because, as mentioned, access to user-generated data is increasingly limited, which challenges researchers to consider what “social data” is anew and open up alternative perspectives. Contrary to how social media platforms offer access to user-generated data for digital research, structured via APIs, app data sources are increasingly characterized by their closed source or proprietary nature. Even though obfuscation is a widely used technique in software engineering (Matviyenko et al., 2015), efforts that render code and data illegible or inaccessible have a significant impact on digital research. These increased challenges posed by platform and app environments to circumvent or sidetrack empirical research are what colleagues and I have termed “infrastructural resistance” (Dieter et al., 2019). Instead, the data formats available for digital research today are characterized by heterogeneous data formats ranging from device-based data (e.g., GPS), software libraries (e.g., software development kits, SDKs) and network-connections (e.g., ad networks). Apps can collect user-generated data, but mostly do not offer access via open APIs, hence there is an absence of ready-made data for data investigations.

In what follows I present three different bottom-up data explorations through which digital researchers and journalists can actively invoke different “research affordances” (Wetevrede, 2016) and use these to advance or initiate an inquiry. Research affordances attune to the action possibilities within software from the perspective of, and aligned with, the interests of the researcher. This approach allows the development of inventive digital methods (Lury & Wakeford, 2012; Rogers, 2013). These

require the rethinking of the technical forms and formats of app–platform relationships by exploring their analytical opportunities. The explorations draw on recent research colleagues and I undertook, taking inspiration from but also noting challenges and making suggestions towards the type of inquiries these data sources afford to increase our understanding of the platform-as-infrastructure environment.

## Fake Social Infrastructures

The first exploration considers fake followers and their relation to Facebook's social infrastructure. Increasing attention is being paid to the extent of fake followers in social media environments from both platforms and digital research. From the perspective of the platforms, the fake follower market is often excluded in discussions of platforms as multisided markets; the fake follower market is not considered a "side" and certainly not part of the "family." Fake followers establish an unofficial infrastructure of relations, recognized by the platforms as undesirable misuses. They are unintended by the platforms, but work in tandem with and by virtue of platform mechanisms. Moreover, these practices decrease the value of the key product, namely social activity.

Colleagues and I investigated the run-up to the Brexit referendum on Twitter by focusing on the most frequently used apps in that data set (Gerlitz & Weltevrede, 2019) (see Figure 34.1). A systematic analysis of these apps and their functionalities provides insight into the mechanisms of automated and fake engagements within the platform's governance structure. In an ongoing project with Johan Lindquist, we are exploring a set of over 1,200 reselling platforms that enable the buying and selling of fake engagements on an extensive range of platforms. These initial explorations show how fake followers technically relate to platforms, both official third-party apps connecting through the API, as well as through an infrastructure of platforms unofficially connecting to social media platforms. What these initial explorations have shown is that research will have to accommodate a variety of data of automated and fake origin. Automated and fake accounts cannot (only) be treated as type or actor but as *practice*, that is situated and emerging in relation to the affordances of the medium. As shown in the case of Twitter, an account does not necessarily represent a human user, as it is accomplished in distributed and situated ways, just as a tweet is not a tweet, commonly understood as a uniquely typed post (Gerlitz & Weltevrede, 2019).

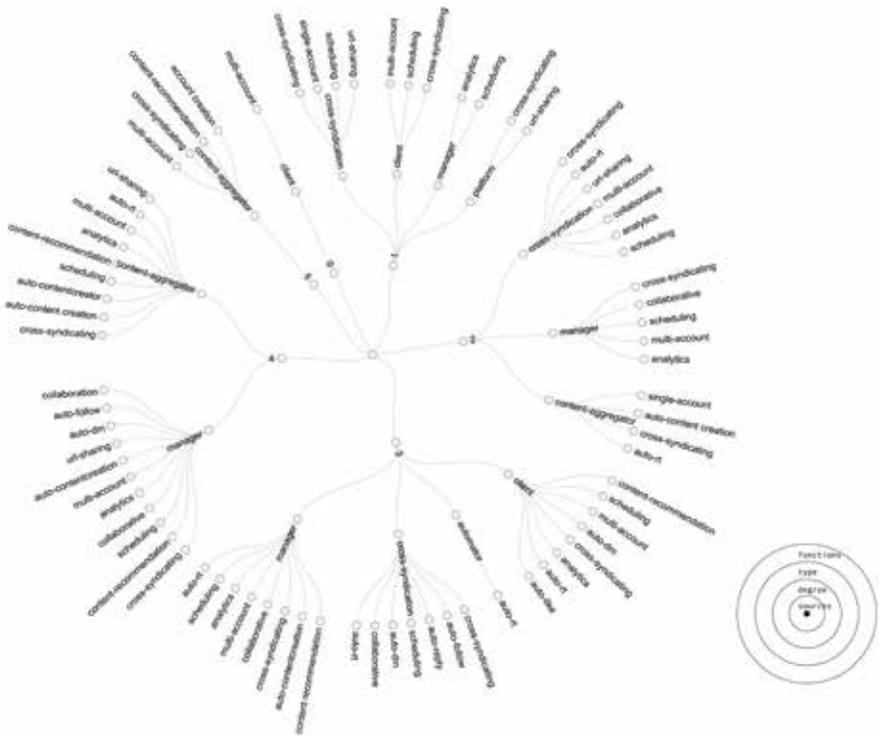


Figure 34.1. Automation functions. The dendrogram visualises the hierarchy of sources, degrees of automation, types of sources and their functions in the Brexit data set, 17–23 June 2016. Source: Gerlitz, C., & Weltevrede, E. (2019). What happens to ANT, and its emphasis on the socio-material grounding of the social, in digital sociology? In A. Blok, I. Farias, & C. Roberts (Eds.), *Companion to Actor-Network Theory*. Routledge. <https://doi.org/10.4324/9781315111667-38>

### App–Platform Relations

The second exploration considers app stores as data infrastructures for apps. Today, the main entry point to apps—for developers and users—is via app stores, where users can search for individual apps or demarcate collections or genres of apps. Building on methods from algorithm studies (Rogers, 2013; Sandvig et al., 2014), one can engage with the technicity of “ranking cultures” (Rieder et al., 2018), for example, in Google Play and the App Store. Such an undertaking concerns both algorithmic and economic power as well as their societal consequences. It can be used to gain knowledge about their ranking mechanisms and an understanding of why this matters for the circulation of cultural content.

The app stores can also be used to demarcate collections or genres of apps to study app–platform relations from the perspective of apps. In

**Table 34.1: Detected App–Platform Relations per Source Set**

Relation	Facebook	Instagram	Snapchat	Twitter
Brand (mentions)	1,449 (34.96%)	2,945 (80.03%)	614 (12.17%)	1,107 (21.80%)
Legal (mentions)	302 (7.29%)	318 (8.65%)	268 (5.31%)	305 (6.01%)
Technical (mentions)	61 (1.47%)	62 (1.68%)	70 (1.39%)	114 (2.24%)
Technical (libraries, SDK)*	83 (33.20%)	0 (0%)	0 (0%)	40 (16.13%)
Technical (HTTP requests)	156 (62.40%)	89 (35.60%)	12 (4.80%)	102 (41.13%)

\*Only for Google Play search results (N = 998)

Note. The Technical (libraries, SDKs) and Technical (HTTP requests) categories indicate apps that engage in an “official” relationship with the social medium through their APIs. From Gerlitz et al. (2019).

“Regramming the Platform” (Gerlitz et al., 2019), colleagues and I investigated over 18,500 apps and the different ways in which apps relate themselves to platform features and functionalities. One of the key findings of this study is that app developers find creative solutions to navigate around the official platform APIs, thereby also navigating around the official governance systems of platforms (Table 34.1). The app-centric approach to platforms-as-infrastructure provides insights into the third-party apps developed on the peripheries of social media platforms, the practices and features supported and extended by those apps, and the messy and contingent relations between apps and social media platforms (Gerlitz et al., 2019).

### Third-Party Data Connections

The third exploration considers app software and how it relates to data infrastructures of external stakeholders. With this type of exploration it is possible to map out how the app as a software object embeds external data infrastructures as well as the dynamic data flows in and out of apps (Weltevrede & Jansen, 2019). Apps appear to us as discrete and bounded objects, whereas they are by definition data infrastructural objects, relating themselves to platforms to extend and integrate within the data infrastructure.

In order to activate and explore the inbound and outbound data flows, we used a variation on the “walkthrough method” (Light et al., 2016). Focusing on data connections, the resulting visualization shows which data is channelled into apps from social media platforms and the mobile platform (Figure 34.2). In a second step, we mapped the advertising networks, cloud services, analytics and other third-party networks the apps connect to in order to monetize app data, improve functionality or distribute hosting to external

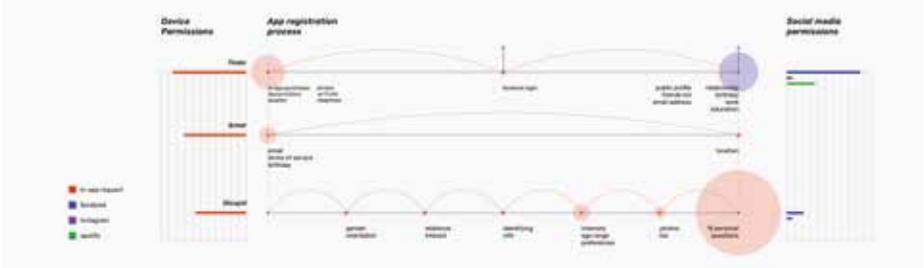


Figure 34.2. Interface walkthrough of data flows during the registration process. Source: Infrastructures of intimate data: Mapping the inbound and outbound data flows of dating apps. Computational Culture, 7. <http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-and-outbound-data-flows-of-dating-apps/>

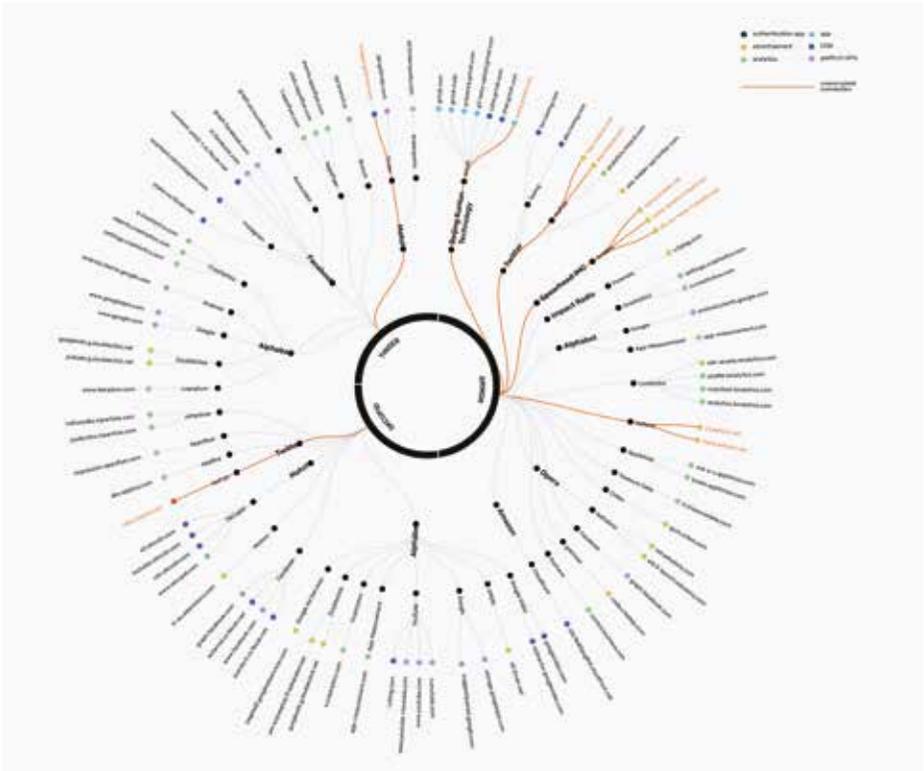


Figure 34.3. Network connections established between dating apps Tinder, Grindr and OKCupid and their third parties. Source: Infrastructures of intimate data: Mapping the inbound and outbound data flows of dating apps. Computational Culture, 7. <http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-and-outbound-data-flows-of-dating-apps/>

parties, among others (Figure 34.3). Mapping data flows in and out of apps provides critical insight into the political economy of the circulation and recombination of data: The data connections that are established, how they are triggered and which data types are being transferred to which parties.

## Conclusion

Platforms and apps are so fundamentally woven into everyday life that they often go unnoticed without any moment of reflection. This tendency to move to the background is precisely the reason why digital researchers, data journalists and activists should explore how they work and the conditions which underpin their creation and use. It is important to improve data infrastructure literacy in order to understand how they are related to different platforms and networks, how they operate between them, and how they involve a diversity of often unknown stakeholders.

In the aftermath of the Cambridge Analytica scandal, data ready-made for social investigations accessible through structured APIs is increasingly being restricted by platforms in response to public pressure. In this chapter, I have suggested that, as a response, researchers, journalists and civil society groups should be creative and inventive in exploring novel types of data in terms of their affordances for data investigations. I have explored three types of data for investigating apps. There are, moreover, multiple opportunities to further expand on this. It should be stressed that I have mainly addressed apps, yet this might offer inspiration for investigations into different data-rich environments, including smart cities and the Internet of things. A more nuanced understanding of the data infrastructures that increasingly shape the practices of everyday life remains an ongoing project.

## Works Cited

- Bruns, A. (2018, April 25). 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>
- 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*. <https://doi.org/10.1177/2056305119846486>

- Gerlitz, C., Helmond, A., Van der Vlist, F. N., & Weltevrede, E. (2019). Regramming the platform: Infrastructural relations between apps and social media. *Computational Culture*, 7. <http://computationalculture.net/regramming-the-platform/>
- Gerlitz, C., & Weltevrede, E. (2019). What happens to ANT, and its emphasis on the socio-material grounding of the social, in digital sociology? In A. Blok, I. Farias, & C. Roberts (Eds.), *Companion to actor-network theory* (pp. 345–356). Routledge. <https://doi.org/10.4324/9781315116667-38>
- Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, 5(2), 1–13. <https://doi.org/10.1177/2053951718786316>
- Helmond, A., Nieborg, D. B., & Van der Vlist, F. N. (2019). Facebook's evolution: Development of a platform-as-infrastructure. *Internet Histories*, 3(2), 123–146. <https://doi.org/10.1080/24701475.2019.1593667>
- Hoffmann, A. L., Proferes, N., & Zimmer, M. (2018). "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society*, 20(1), 199–218. <https://doi.org/10.1177/1461444816660784>
- Light, B., Burgess, J., & Duguay, S. (2016). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881–900. <https://doi.org/10.1177/1461444816675438>
- Lomas, N. (2018, April 24). Kogan: "I don't think Facebook has a developer policy that is valid." *TechCrunch*. <https://social.techcrunch.com/2018/04/24/kogan-i-dont-think-facebook-has-a-developer-policy-that-is-valid/>
- Lury, C., & Wakeford, N. (2012). *Inventive methods: The happening of the social*. Routledge.
- Marres, N. (2017). *Digital sociology: The reinvention of social research*. Polity Press.
- Matviyenko, S., Ticineto Clough, P., & Galloway, A. R. (2015). On governance, blackboxing, measure, body, affect and apps: A conversation with Patricia Ticineto Clough and Alexander R. Galloway. *The Fibreculture Journal*, 25, 10–29. <http://twentyfive.fibreculturejournal.org/fcj-179-on-governance-blackboxing-measure-body-affect-and-apps-a-conversation-with-patricia-ticineto-clough-and-alexander-r-galloway/>
- Nieborg, D. B., & Helmond, A. (2019). The political economy of Facebook's platformization in the mobile ecosystem: Facebook Messenger as a platform instance. *Media, Culture & Society*, 41(2), 196–218. <https://doi.org/10.1177/0163443718818384>
- Perriam, J., Birkbak, A., & Freeman, A. (2020). Digital methods in a post-API environment. *International Journal of Social Research Methodology*, 23(3), 277–290. <https://doi.org/10.1080/13645579.2019.1682840>
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, 20(1), 293–310. <https://doi.org/10.1177/1461444816661553>

- Rieder, B. (2013). Studying Facebook via data extraction: The Netvizz application. In *Proceedings of the 5th Annual ACM Web Science Conference* (pp. 346–355). <https://doi.org/10.1145/2464464.2464475>
- Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to “ranking cultures”: Investigating the modulation of visibility in YouTube search results. *Convergence*, 24(1), 50–68. <https://doi.org/10.1177/1354856517736982>
- Rogers, R. (2013). *Digital methods*. MIT Press.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). Auditing algorithms: Research methods for detecting discrimination on Internet platforms. International Communication Association preconference on Data and Discrimination Converting Critical Concerns into Productive Inquiry, Seattle, WA.
- Smout, A., & Busvine, D. (2018, April 24). Researcher in Facebook scandal says: My work was worthless to Cambridge Analytica. *Reuters*. <https://www.reuters.com/article/us-facebook-privacy-cambridge-analytica-idUSKBN1HV17M>
- Venturini, T., & Rogers, R. (2019). “API-based research” or how can digital sociology and journalism studies learn from the Cambridge Analytica data breach. *Digital Journalism*, 7(4), 532–540. <https://doi.org/10.1080/21670811.2019.1591927>
- Weltevrede, E. (2016). *Repurposing digital methods: The research affordances of platforms and engines* [Doctoral dissertation], University of Amsterdam.
- Weltevrede, E., & Jansen, F. (2019). Infrastructures of intimate data: Mapping the inbound and outbound data flows of dating apps. *Computational Culture*, 7. <http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-and-outbound-data-flows-of-dating-apps/>

## About the Author

**Esther Weltevrede** is an Assistant Professor of New Media and Digital Culture and coordinator of the Digital Methods Initiative at the University of Amsterdam, where she explores the various research affordances of digital media, with a specific interest in methodological innovations in the study of the infrastructures of social media platforms and third parties.