

# Anne Helmond

## on Researching the History of the Web

**KEYWORDS:** web historiography; web archives; web ecosystems; source code analysis; software studies

Anne Helmond, assistant professor of New Media and Digital Culture at the University of Amsterdam and member of its Digital Methods Initiative, is primarily known among her peers for her work on the ‘platformization’ of the internet. Helmond herself defines ‘platformization’ – a notion she first coined in her PhD dissertation – as ‘the rise of the platform as the dominant infrastructural and economic model of the social web.’ The process of platformization is characterized by the ‘dual logic of social media platforms’ expansion into the rest of the web and, simultaneously, their drive to make external web and app data platform ready’.<sup>1</sup> In recent years, she has further pursued her interest in the web’s past, which she first developed in her dissertation, in a project entitled ‘App Ecosystems: A Critical History of Apps’, for which she received a grant from the Netherlands Organization for Scientific Research (NWO, lasting until 2020).<sup>2</sup>

In this interview, we set out to discuss some of the intricacies of doing web history. First, we talk about the current state of web historiography (and its various pitfalls) and the need for a so-called ‘website ecology’, which redirects researchers’ perspectives from the single website and its content to its various contexts on the larger web, as well as its entanglements with other web objects and actors. In this context, we also touch upon the role of web archives and the various obstacles they present for scholarly research. Next, we move on to some concrete examples, taken from Helmond’s recent and ongoing research, of what such an approach can deliver.

– Eef Masson and Karin van Es



*Anne Helmond.*

*EM/KvE*: In a recent article, you have discussed the current state of web historiography.<sup>3</sup> Can you talk a little, first, about the primary interests within this tradition so far, and the hurdles today's scholars need to take in researching the web's past?

*AH*: The research area of web historiography is concerned with the study of the methods that web historians employ to write histories of the web.<sup>4</sup> In particular, it aims to put web history on the agenda of internet researchers, and considers which methods are available to us – or should be newly developed – to write histories of the web.<sup>5</sup> As one of the key figures in the field of web history, Niels Brügger, has argued: 'a better understanding of the web of the past is an important condition for gaining a more complete understanding of the web of today.'<sup>6</sup>

Important tools (and corpora) for writing histories of the web are national and international web archives (e.g. that of the National Library of the Netherlands, or the Internet Archive). These web archives come with specific 'research affordances' that allow for particular histories to be written.<sup>7</sup> For example, most web archives work with so-called 'seed lists' that contain the starting points for web crawlers to crawl and archive websites. In a recent talk, Mark Graham, the technical director of the Internet Archive's Wayback Machine, explained that this archive's seed list contains over 7000 crawls from a wide range of sources (Wide crawls, Twitter, Archive Team, Survey Crawls, TLD Crawls, Alexa Internet, Domain Crawls, 300+ Wikipedia Sites, 3rd parties, Save Page Now, WordPress, Amber, Perma.cc, Custom Crawls, YouTube, LinkArchiver, 600+ Archive-It, Top News, Top Sites and News Grabber). Together, these 7000 starting points crawl 1.5 billion URLs per week.<sup>8</sup> This means that researchers, unless they run their own archival web crawlers or make use of professional web crawling services, have to rely on what has been crawled *for* them.

Furthermore, all the dominant web crawlers focus on crawling single websites, and not their larger context – thereby privileging the single website as the main unit of analysis.<sup>9</sup> As a consequence, many web historical studies focus on the changing content of a single website or a set of websites, or on their interlinking patterns.<sup>10</sup>

*EM/KvE*: In your own work, you want to shift attention to the history of website *ecosystems*, taking a perspective you term 'website ecology'. Can you explain what you mean by this – and where the inspiration for this notion came from? Specifically, can you explain how such a perspective problematizes the idea of the website as a bounded object?

*AH*: In response to the dominant modes of using web archives to examine the evolution of a website, a set of websites, or their interlinking patterns, I aim to investigate what other research affordances archived websites may have. In my work, I aim to move beyond the *content* of archived websites, focusing rather on their archived *context*. One of my core arguments here is that we should consider the value of the archived source code of websites, since these provide valuable insights into the larger context that websites operate in.<sup>11</sup> By going into the source code of archived websites, we can see which technologies have been used to build them, but also which third-party elements are embedded in these sites – thus showing us which third-party actors are technically and commercially connected to websites. This enables us to study the

infrastructural and commercial evolution of the web at large, but from the perspective of a single website.<sup>12</sup>

In employing the biological metaphor of the ‘ecosystem’, I try to highlight that websites should not be seen as static and self-contained objects, but rather, as ecosystems that are inhibited and shaped by third parties through various interactions between the object (the website) and its larger context (the web). Carlos Scolari and Marianne van den Boomen have argued that (biological) metaphors, while they come with their limitations, play an important role in conceptualizing new media technologies and in developing media studies theory at large.<sup>13</sup> For me, the use of the ecosystem metaphor, and website ecology as the study of a website’s ecosystem, enables me to consider the complex socio-technical relations between websites and third parties connected to them. In doing so, I see the website as a porous object setting up channels for data flows with third parties.<sup>14</sup>

*EM/KvE:* Which kinds of methods does a website ecology involve? And which role do (existing) web archives play in this type of scholarship?

*AH:* Historical website ecology draws on such existing approaches as source code analysis, which considers the source code as ‘a social and technical artefact’ that displays the contextual development of the coded software object – like a piece of software or a website – thus enabling ‘an “archaeology” of software processes.’<sup>15</sup> That is, the source code bears the encoded signs of software development which may be used to uncover the evolution of software and websites. It further draws from approaches developed in the area of software studies, in an effort to turn to the historical material traces of software and to consider ‘the culture of software in constructing histories of the web’, thereby bringing a ‘software studies lens to web historiography’.<sup>16</sup>

Web archives are important in this approach as they provide the source material to view and analyse the ‘back-ends’ of websites, that give access to the larger contexts that websites are embedded in as well as their production contexts. For example, in a study with colleague Esther Weltevrede, I analysed the source code of a set of archived Dutch blogs to understand the evolution of the Dutch blogosphere not only in terms of how blogs interlink, but also in terms of how they are technically built and which functions they contain.<sup>17</sup> Another approach, developed with David Nieborg and Fernando van der Vlist, aims to move from the single website to the larger techno-commercial context that websites are embedded in. From this perspective, we examined how social media platforms engage in official partnerships with data providers and marketing agencies.<sup>18</sup> We used the Internet Archive’s Wayback Machine to investigate Facebook’s changing platform-industry partnerships over time, and gained insights into who Facebook partners with and what the role of these partners is in relation to the platform. Such studies can offer a much-needed historical perspective onto contemporary controversies around, for instance, Cambridge Analytica, the emergence and role of third-party data providers such as Acxiom, and the role of third-party apps that have been built on top of Facebook.

*EM/KvE:* One set of connections that you have explored so far is those involving trackers. Which insights into historical website ecologies have you obtained in researching those?

AH: In my case study on historical tracking ecosystems, I focused on the data connections created by trackers on websites over time, to study changes in the techno-commercial environment of the web at large.<sup>19</sup> One thing I found was a decline in trackers on websites over time, which initially was a counter-intuitive finding. However, another related large-scale study confirmed my hypothesis that this decline in trackers is actually a sign of media concentration with larger tracking companies such as Google buying up smaller players in the industry.<sup>20</sup> For me, historical source code analysis opens up new ways to investigate the political economy of the web and the techno-commercial evolution of the web.

EM/KvE: How are you planning to expand on your study of web history in the next few years?

AH: In a forthcoming contribution, I focus on one of the core elements of the web, the hyperlink, to trace its evolution and uptake in various technical systems, including proto-hypertext systems, hypertext systems, search engines and social media platforms on the world wide web as well as in the mobile app ecosystem. The purpose of such an approach is to use this central web object 'as a way to understand social, technical, and commercial transformations of the web' and in the mobile app space.<sup>21</sup> In my current projects I am further exploring the research affordances of web archives beyond studies of the content or link structure of a single website. In one project, with Nieborg and Van der Vlist, I am further developing methods to examine the techno-commercial evolution of the web and platform-industry partnerships.

In my most recent, NWO-funded project, I am developing methods to use web archives to study the evolution of mobile apps and the app ecosystem. In doing so, I wish to draw attention to how contemporary and popular media objects of use and study, such as mobile apps, could be studied in the future. As fleeting objects that resist archiving, how can apps as software objects be archived for future historians? What kinds of tools and methods do we need in order to understand how Facebook evolved from a single website into a family of 72 apps?<sup>22</sup>

## Notes

The work discussed in this interview is part of the research programme Innovational Research Incentives Scheme Veni, with project number 275-45-009, which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

1. Anne Helmond, "The Platformization of the Web: Making Web Data Platform Ready," *Social Media + Society* 1, no. 2 (July–December 2015): 1–11, especially 1 and 8 (quotes), <https://doi.org/10.1177/2056305115603080>.
2. Netherlands Organization for Scientific Research (NWO), "Veni Awards 2017," NWO website, <https://www.nwo.nl/en/research-and-results/programmes/Talent+Scheme/awards/veni+awards/veni+awards+2017> (accessed 22 August 2018).
3. Anne Helmond, "Historical Website Ecology: Analyzing Past States of the Web Using Archived Source Code," in *Web 25: Histories from the First 25 Years of the World Wide Web*, ed. Niels Brügger (New York: Peter Lang Publishing, 2017), 139–55.
4. Meghan Dougherty and Steven M. Schneider, "Web Historiography and the Emergence of New Archival Forms," in *The Long History of New Media: Technology, Historiography, and Newness in Context*, ed. David W. Park, Nicholas W. Jankowski, and Steve Jones (New York: Peter Lang Publishing, 2011): 253–262; Kirsten Foot and Steven M. Schneider, "Object Oriented Web Historiography," in *Web History*, ed. Niels Brügger (New York: Peter Lang Publishing, 2010), 61–79; Niels Brügger, "Web Historiography and Internet Studies: Challenges and Perspectives," *New Media & Society* 15, no. 5 (May 2013): 752–764; Helmond, "Historical Website Ecology".
5. For more on the programmatic role of web historiography, see Brügger, "Web Historiography and Internet Studies".

6. Ibid., 752.
7. Esther J.T. Weltevrede, "Repurposing Digital Methods: The Research Affordances of Platforms and Engines" (unpublished PhD thesis, University of Amsterdam, 2016).
8. Anne Helmond, "Verslag van de internationale web archiving conferentie in Jerusalem," Anne Helmond home page (blog post), 6 May 2018, <http://www.annahelmond.nl/2018/05/06/verslag-van-de-internationale-web-archiving-conferentie-in-jerusalem/> (accessed 22 August 2018).
9. See Brügger, "Web Historiography and Internet Studies"; Richard Rogers, "Doing Web History with the Internet Archive: Screencast Documentaries," *Internet Histories* 1, no. 2 (2017): 160–72, <https://doi.org/10.1080/24701475.2017.1307542>.
10. Brügger, "Web Historiography and Internet Studies"; Rogers, "Doing Web History with the Internet Archive"; Helmond, "Historical Website Ecology".
11. Helmond, "Historical Website Ecology".
12. Ibid.
13. Carlos A. Scolari, "Media Ecology: Exploring the Metaphor to Expand the Theory," *Communication Theory* 22, no. 2 (May 2012): 204–22; Marianne van den Boomen, *Transcoding the Digital: How Metaphors Matter in New Media (Theory on Demand series, no. 14)* (Amsterdam: Institute of Network Cultures, 2014), <http://networkcultures.org/blog/publication/no-14-transcoding-the-digital/>.
14. Helmond, "The Platformization of the Web".
15. Cleidson de Souza, Jon Froehlich, and Paul Dourish, "Seeking the Source: Software Source Code as a Social and Technical Artifact," in *Proceedings of the 2005 International ACM SIGGROUP Conference on Supporting Group Work (Sanibel Island, Florida, 6–9 November 2005)* (New York: ACM, 2005), 197–206, <http://dl.acm.org/citation.cfm?id=1099239>, specifically p. 206.
16. Quotes from Megan Sapnar Ankerson, "Historicizing Web Design: Software, Style, and the Look of the Web," in *Convergence, Media, History*, ed. Janet Staiger and Sabine Hake (New York, NY: Routledge, 2009), 192–203, specifically 195. For more on software studies, see Matthew Fuller, *Software Studies: A Lexicon* (Cambridge, MA: MIT Press, 2008), and Matthew G. Kirschenbaum, "Virtuality and Vtml: Software Studies after Manovich," *Electronic Book Review*, 29 August 2003, <http://www.electronicbookreview.com/thread/technocapitalism/morememory> (accessed 22 August 2018).
17. Esther Weltevrede and Anne Helmond, "Where Do Bloggers Blog? Platform Transitions within the Historical Dutch Blogosphere," *First Monday* 17, no. 2 (February 2012), <https://doi.org/10.5210/fm.v17i2.3775>.
18. Anne Helmond, David B. Nieborg, and Fernando N. van der Vlist, "The Political Economy of Social Data: A Historical Analysis of Platform-Industry Partnerships," in *8th International Conference on Social Media & Society: Social Media for Good or Evil (Toronto, Canada, 28–30 July 2017)* (New York: ACM, 2017), 1–5, <https://doi.org/10.1145/3097286.3097324>.
19. Helmond, "Historical Website Ecology".
20. See Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner, "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016," in *Proceedings of the 25th USENIX Security Symposium (Austin, Texas, 10–12 August 2016)* (Berkeley, CA: USENIX, 2016), 997–1013, [https://www.usenix.org/system/files/conference/usenixsecurity16/sec16\\_paper\\_lerner.pdf](https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_lerner.pdf).
21. Anne Helmond, "A Historiography of the Hyperlink: Periodizing the Web Through the Changing Role of the Hyperlink," in *The SAGE Handbook of Web History*, ed. Niels Brügger and Ian Milligan (London: SAGE Publications Ltd., forthcoming).
22. David B. Nieborg and Anne Helmond, "The Political Economy of Facebook's Platformization in the Mobile Ecosystem: Facebook Messenger as a Platform Instance," *Media, Culture & Society* (2018, forthcoming).

## Biography

**Anne Helmond** is assistant professor of New Media and Digital Culture at the University of Amsterdam. She is a member of the Digital Methods Initiative research collective, where she focuses her research on the infrastructure of social media platforms and apps. Her research interests include digital methods, software studies, platform studies, app studies, infrastructure studies and web history. She currently holds a Veni grant from the Netherlands Organization for Scientific Research (NWO) for the project "App Ecosystems: A Critical History of Apps" (2017–2020).