

Life Online: The Digital Footprint Gap*

Moritz Büchi

University of Zurich

Christoph Lutz

BI Norwegian Business School

Marina Micheli

University of Zurich

Position Paper for the Partnership for Progress on the Digital Divide 2017
International Conference, 24–26 May 2017, San Diego, California, USA.

Gaps in the Understanding of Digital Divides and their Impacts

The digital inequality literature has focused on the antecedents and effects of differentiated Internet use. The scientific and public policy relevance of the digital divide is based on the assumption that those who are able to effectively use the Internet might increase their social, economic, cultural, and human capital, and this could exacerbate social inequalities (DiMaggio et al., 2004). But individuals' digital footprints—the sum of their digitally traceable behavior and online presence—can also lead to beneficial and adverse outcomes, short-term or long-term, individual or societal. What users *do* online matters; however, what *is* online about them also has consequences. Digital footprints entail results both in the sense of immediate gratifications (e.g., enjoyment) and of outcomes (e.g., finding a job). Digital divide scholarship should consider how online traces and their consequences vary according to socio-demographic variables and traditional markers of inequalities. From this premise, we develop the concept of “digital footprint gap,” suggest viable methods to investigate it, and reflect on action plans that could connect research, policy, and practice.

Digital footprints are not only the product of active participation through content production and sharing, but also of “passive participation” (Lutz & Hoffmann, 2017). This includes low-involvement forms of participation, algorithmically generated data, and information made available by other users (Marwick & boyd, 2014). Social media platforms afford many simple user actions, such as liking, favoriting, following, or commenting, which are not necessarily considered active participation, but nonetheless

* All authors contributed equally to this work.

contribute to a digital footprint. Platforms and online services even generate data without active user involvement. Just turning on an Internet-enabled device means that an IP address is revealed; clicking a like button is not only a social signal to another user but also informs the platform's advertising profiles; and publishing a blog post is not only a service to its readers but also influences search engine indexing. Finally, Internet users can "be participated" without their knowledge or against their will (Casemajor et al., 2015). Examples include tagging, endorsements, ratings, and comments on the visible end of the spectrum, searches (e.g., googling someone) and various automated data analyses on the less visible end. With the increasing pervasiveness of networked technology (e.g., Internet of Things, social robotics) we expect such passive participation to increase.

The term "digital footprint gap" was first employed by Robinson et al. (2015, p. 571) in the context of digital inequalities over the life course. "[T]he day-to-day lives of babies and infants are broadcast to other parties and sometimes the world at large, often without the knowledge and consent of the child" (p. 571). The term describes differences in the amount of online traces between individuals or population groups. We extend this concept, considering not only the quantity, but also the quality and, most importantly, the implications of online traces. How are users with different socio-economic backgrounds aware of and able to manage their footprints? Do the outcomes of quantitatively and qualitatively varying digital footprints exacerbate inequalities?

Despite the prominence of privacy as a research topic (Baruh et al., 2017), little scholarship uses a digital inequalities perspective to study digital footprints. The literature on inequalities in online content creation has shown that age is a decisive factor, with young users creating more online content than older adults (Blank; 2013; Correa, 2010; Hargittai & Walejko, 2008; Hoffmann et al., 2015; Schradie, 2011). Socio-economic status and gender effects are less clear for online participation and it is important to differentiate types of online content produced (Blank, 2013). However, such findings do not necessarily apply to the digital footprint gap. For example, can we assume that because young users create more content online than older users they are also better represented in big data? Analysis of digital footprint gaps needs to consider not only active online participation, but also low-involvement forms, algorithmically generated data, and data posted by other users.

Research and Policy Recommendations

Future research investigating digital footprint gaps should combine different methodological approaches. While big data analyses offer benefits compared with self-re-

ported data (Shiffman et al., 2008), disadvantages of digital traces include overrepresentation of certain sources due to data accessibility (e.g., Twitter API), noise (e.g., through bots), engagement invisible to machines (e.g., subtweeting), and neglect of human self-awareness as well as the ecological context (Tufekci, 2014).

When broader representativeness is important, most digital trace data is inadequate (Blank, 2016; Hargittai, 2015) and traditional social science methods are better suited. Representative surveys could measure digital skills related to platform algorithms and privacy settings (Büchi et al., 2016). Qualitative interviews combined with social media profile analysis could also be a valuable method (Dubois & Ford, 2015). During interviews, respondents could discuss content they have posted, but also what has been posted by others (comments, likes, retweets), as well as by the platform itself (a news item in the feed). Moreover, interviews could be enriched by search engine use so that respondents could look for their digital traces and discuss the results with the interviewer. Finally, interviews with social groups particularly affected by digital traces could investigate how digital footprint gaps are perceived and enacted. Young users and micro-celebrities (e.g., Instagram influencers) would be groups to scrutinize (Micheli, 2016; Abidin, 2015). Such actor-focused methods could inform “social analytics,” i.e., how users make sense of their own digital footprints or how they “reflect upon, and adjust, their online presence and the actions that feed into it, through the use of analytics” (Couldry et al., 2016, p. 119). Beyond this, media content analyses of negative passive participation, for example in the form of doxing and online harassment, could help case study selection. Finally, digital methods and software studies could offer useful insights to understand how platforms generate data and the implications for digital inequalities (Light et al., 2016; Rogers, 2009).

Although the concept of the digital footprint gap has not been thoroughly investigated, the right to be forgotten currently presents a way for individuals to control (access to) their footprints. We propose additional policy directions. First, more funding should be allocated to understanding how inequalities are (re)produced through digital footprints. This includes the critical study of algorithms, for example how algorithmic discrimination occurs (Sandvig et al., 2014) or how digital footprints influence reality construction and social order (Just & Latzer, 2016). In particular, policymakers should allow researchers more access to the ways platforms manage and curate digital footprints, i.e., they should enhance transparency through platform auditing. Second, practices favoring digitally active individuals or placing particular burden on disadvantaged communities (e.g., through unfair exclusion from economic opportunities and unjustified targeting via predictive analytics) should be banned (Madden et al., 2017). Third, literacy and skills initiatives might be embedded into institutions like schools

and companies to enhance general awareness and mitigate adverse outcomes from digital footprints.

Collaborative Partnerships and Action Plans

The topic of digital footprint gaps is complex as it involves various actors, from individual users and their network, to platform providers, data brokers, civil rights groups, and government institutions. Consequently, solutions to reduce systematic inequalities need inclusive collaboration between different stakeholders. In particular, the voices of the users most affected by digital footprints—or the lack thereof—should be heard. Offline and online roundtable discussions and open town hall meetings could be a first step in bringing the actors together and letting them voice their concerns. Moreover, journalists could collect interesting stories, put them together in approachable formats (videos, infographics, gamified storytelling), and thus put the issue on the agenda.

References

- Abidin, C. (2015). Communicative intimacies: influencers and perceived interconnectedness. *Ada: A Journal of Gender, New Media, and Technology*, 8. Retrieved from <http://adanewmedia.org/2015/11/issue8-abidin/>
- Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: a meta-analytical review. *Journal of Communication*. Advance online publication. <https://doi.org/10.1111/jcom.12276>
- Blank, G. (2013). Who creates content? *Information, Communication & Society*, 16(4), 590–612. <https://doi.org/10.1080/1369118X.2013.777758>
- Blank, G. (2016). The digital divide among Twitter users and its implications for social research. *Social Science Computer Review*. Advance online publication. <https://doi.org/10.1177/0894439316671698>
- Büchi, M., Just, N., & Latzer, M. (2016). Caring is not enough: the importance of Internet skills for online privacy protection. *Information, Communication & Society*. Advance online publication. <https://doi.org/10.1080/1369118X.2016.1229001>
- Casemajor, N., Couture, S., Delfin, M., Goerzen, M., & Delfanti, A. (2015). Non-participation in digital media: toward a framework of mediated political action. *Media, Culture & Society*, 37(6), 850–866. <https://doi.org/10.1177/0163443715584098>
- Correa, T. (2010). The participation divide among “online experts”. *Journal of Computer-Mediated Communication*, 16(1), 71–92. <https://doi.org/10.1111/j.1083-6101.2010.01532.x>
- Couldry, N., Fotopoulou, A., & Dickens, L. (2016). Real social analytics: a contribution towards a phenomenology of a digital world. *The British Journal of Sociology*, 67(1), 118–137. <https://doi.org/10.1111/1468-4446.12183>

- DiMaggio, P., Hargittai, E., Celeste, C., & Shafer, S. (2004). From unequal access to differentiated use: a literature review and agenda for research on digital inequality. In K. Neckerman (Ed.), *Social inequality* (pp. 355–400). New York, NY: Russell Sage Foundation.
- Dubois, E., & Ford, H. (2015). Trace interviews: An actor-centered approach. *International Journal of Communication*, 9, 2067–2091. Retrieved from <http://ijoc.org/index.php/ijoc/article/view/3378>
- Hargittai, E. (2015). Is bigger always better? Potential biases of big data derived from social network sites. *The ANNALS of the American Academy of Political and Social Science*, 659(1), 63–76. <https://doi.org/10.1177/0002716215570866>
- Hargittai, E., & Walejko, G. (2008). The participation divide: content creation and sharing in the digital age. *Information, Communication & Society*, 11(2), 239–256. <https://doi.org/10.1080/13691180801946150>
- Hoffmann, C. P., Lutz, C., & Meckel, M. (2015). Content creation on the Internet: a social cognitive perspective on the participation divide. *Information, Communication & Society*, 18(6), 696–716. <https://doi.org/10.1080/1369118X.2014.991343>
- Just, N., & Latzer, M. (2016). Governance by algorithms: reality construction by algorithmic selection on the Internet. *Media, Culture & Society*, 39(2), 238–258. <https://doi.org/10.1177/0163443716643157>
- Light, B., Burgess, J., & Duguay, S. (2016). The walkthrough method: an approach to the study of apps. *New Media & Society*. Advance online publication. <https://doi.org/1461444816675438>
- Lutz, C., & Hoffmann, C. P. (2017). The dark side of online participation: exploring non-, passive and negative participation. *Information, Communication & Society*, 20(6), 876–897. <https://doi.org/10.1080/1369118X.2017.1293129>
- Madden, M., Gilman, M. E., Levy, K. E., & Marwick, A. E. (2017). Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans. *SSRN Electronic Journal*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2930247
- Marwick, A. E., & boyd, d. (2014). Networked privacy: how teenagers negotiate context in social media. *New Media & Society*, 16(7), 1051–1067. <https://doi.org/10.1177/1461444814543995>
- Micheli, M. (2016). Social networking sites and low-income teenagers: between opportunity and inequality. *Information, Communication & Society*, 19(5), 565–581. <https://doi.org/10.1080/1369118X.2016.1139614>
- Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. J., & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, 18(5), 569–582. <https://doi.org/10.1080/1369118X.2015.1012532>
- Rogers, R. (2009). *Digital methods*. Cambridge, MA: MIT Press.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: research methods for detecting discrimination on internet platforms. Paper presented at the *Data and discrimination: converting critical concerns into productive inquiry* preconference of the

- 64th Annual Meeting of the International Communication Association. May 22, 2014, Seattle, WA, USA. Retrieved from <https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf>
- Schradie, J. (2011). The digital production gap: the digital divide and Web 2.0 collide. *Poetics*, 39(2), 145–168. <https://doi.org/10.1016/j.poetic.2011.02.003>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32. <https://doi.org/10.1146/annurev.clinpsy.3.022806.091415>
- Tufekci, Z. (2014). Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *ICWSM '14: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (pp. 505–514). Palo Alto, CA: AAAI. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8062/8151>