



Selected Papers of #AoIR2025:
The 26th Annual Conference of the
Association of Internet Researchers
Niterói, Brazil / 15 – 18 Oct 2025

ACCESS IS NOT ENOUGH! RECONCEPTUALIZING DATA QUALITY AS A PUBLIC VALUE IN TIMES OF PLATFORMIZATION AND EUROPEAN REGULATION

Yannik Peters
GESIS – Leibniz Institute for the Social Sciences

Katrin Weller
GESIS – Leibniz Institute for the Social Sciences

While the initial phase of social science engagement with platform data was referred to as "wild west" (Puschmann, 2019), questions regarding data quality have gained significant attention in recent years. This has also led to new approaches for structuring the dimensions of data quality. For example, Daikeler et al. (2024) have proposed the distinction between intrinsic and extrinsic data quality. While intrinsic data quality refers to the accuracy of data (and viewing quality-criteria from "within" the data), extrinsic data quality places a stronger focus on its usability (and takes on the perspective of data users in different contexts). Different approaches also exist that reflect data quality dimensions via "error frameworks", a format adopted from survey methodology research (Groves and Lyberg 2010, Sen et al. 2021). Several frameworks and data quality tools also distinguish between measurement and representation errors, such as the KODAQs Toolbox. Measurement errors rather refer to construct validity (how well topics/phenomena are measured), its operationalization and measurement. Representation errors, on the other hand, are more connected to the target population (who do we study) and related biases.

With this contribution we link recent work from the area of data quality research with critical reflections on platformization (Poell et al. 2019, van Dijck et al. 2018) and argue that the quality of online platform data cannot solely be understood as a methodological, internal research construct, but also as a contested concept in which various actors from politics, the tech industry and academia negotiate and represent specific interests. Digital platforms, for example, have severely restricted access to their data in recent years, such as the Meta Group (Bruns 2019, Freelon 2018, Mimizuka 2025). Previously, business models were established in which researchers often worked with free API

Suggested Citation (APA): Peters, Y., Weller, K. (2025, October). *Access is not enough! Reconceptualizing Data Quality as a Public Value in Times of Platformization and European Regulation*. Paper presented at AoIR2025: The 26th Annual Conference of the Association of Internet Researchers. Niterói, Brazil: AoIR. Retrieved from <http://spir.aoir.org>.

versions with limited access to data. For example, Twitter's free Streaming API allowed only to collect a sample of 1% of all Tweets, with the Rest API it was only possible to collect data of the past seven days. Higher data quality could only be obtained through monetized API access (Salvatore et al. 2020). In this respect, digital data quality became an economic good as data quality indicators like *completeness*, *timeliness*, and *accuracy* were very much determined by digital platforms. After a brief phase of free data access for researchers on Twitter via the Research API, the commoditization of data quality was further advanced after the Twitter takeover by Elon Musk and the rebranding of the platform to X.

This strategy of digital platforms has now increasingly become the subject of political efforts of regulation, especially in the European Union (EU). The EU's special position with regard to the regulation of digital platforms is essentially due to the fact that the EU does not possess its own platform ecosystem but is primarily dependent on the American (e.g. Meta, Microsoft, Amazon, Google, X, etc.) and, to a lesser extent, the Chinese platforms (e.g. Baidu, Alibaba, Tencent, ByteDance). Therefore, the European path lies above all in the regulation of digital platforms against the backdrop of its (European) public values (van Dijck 2021). The EU's recent major regulatory packages of digital platforms are the Digital Services Act (DSA), the Digital Markets Act (DMA), and the AI Act. The DSA includes provisions for access to platform data by researchers. Article 40 specifically governs researchers' access to platform data, with Article 40.12 addressing access to public platform data and Article 40.4 pertaining to non-public platform data. In recital 97 of the DSA very large online platforms are supposed to give "access to data from [...] vetted researchers affiliated to a research organization" while the research itself is required to address "systemic risks" (Art. 40.4 and 40.12) in the EU or "risk mitigation measures" (Art. 40.4). As a main reason for this, the DSA names "supporting their [research organizations] public interest mission". In this sense, data access can be considered a public value or at least serving public values. It remains to be seen how the changing legal situation leads to novel data access options and increased data quality (Giglietto/Terenzi 2024, Brown et al. 2024). Even though data access is an important and fundamental part of external data quality (with accessibility also being represented by the letter A in the FAIR principles (Wilkinson et al. 2016) that advocate better research data management structures), other intrinsic and extrinsic components of data quality like *accuracy* or *completeness* are still underrepresented in the DSA or the finalized *Delegated Regulation on data access provided for in the Digital Services Act*¹. Also, platforms are still providing APIs containing incorrect data, e.g. the Tik Tok API (Darius 2024). With a special focus on data donation studies, Hase et al. have most recently criticized the missing transparency and completeness of data provided by digital platforms. Very large online platforms (VLOPs) are also engaging in parts of the current public consultation processes, e.g. by providing responses to the EU's draft of the *Delegated Regulation*. However, a noteworthy development in this

1

https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13817-Delegated-Regulation-on-data-access-provided-for-in-the-Digital-Services-Act_en

regard is the inclusion of data quality in Recital 13 of the Delegated Regulation following the second feedback round. Specifically, “evidence of poor quality or unreliability of such data” – that is, data also available from other sources – is recognized as a legitimate reason or justification for researchers to request or obtain this data through the Digital Services Coordinators from the platforms. Conversely, this implies the establishment of a minimum standard of data quality for the data that platforms must provide under the DSA. This is also due to the immense effort of researchers and NGOs in the public consultation processes.

With this contribution, we link the current scientific debates on data quality to discourses of platformization and platform power. In this respect, data quality is proving to be a contested construct whose meaning is currently constituted somewhere between research standard, economic good and public value. Furthermore, we problematize that the DSA should also focus, besides data access, on data quality. The remaining risk is that even if platforms grant access to data, the actual purpose of conducting reliable and socially relevant research on systemic risks can be undermined by a lack of data quality. The question that arises is what possible solutions exist for this problem. Overall, we identify the need for new, additional actors to provide long-term accountability and supportive infrastructure for data quality assessment. Such new actors could be, for example, networks of academic institutions, academic associations, or research infrastructure institutions. They could serve extrinsic data quality by providing long term archival structures for data access, they could foster communication about intrinsic data quality by developing data documentation standards, they could create and provide benchmark datasets and data quality indicators or create auditing mechanisms for APIs or other data access modes. In addition, they could mediate between research communities and platform providers and coordinate data needs in the community. As next steps, a broader picture of the relation between platform providers, researchers and legal frameworks needs to be drawn and transferred into sustainable research infrastructure.

References

- Brown, M. A., Gruen, A., Maldoff, G., Messing, S., Sanderson, Z., Zimmer, M. (2024). Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations. *arXiv*. Computers and Society. <https://doi.org/10.48550/arXiv.2410.23432>
- Bruns, A. (2019). After the ‘APIcalypse’: social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, 22(11), 1544–1566. <https://doi.org/10.1080/1369118X.2019.1637447>
- Daikeler, J., Fröhling, L., Sen, I., Birkenmaier, L., Gummer, T., Schwalbach, J., Silber, H., Weiß, B., Weller, K., & Lechner, C. (2024). Assessing Data Quality in the Age of Digital Social Research: A Systematic Review. *Social Science Computer Review*, 0(0). <https://doi.org/10.1177/08944393241245395>

Darius, P. (2024). Researcher Data Access Under the DSA: Lessons from TikTok's API Issues During the 2024 European Elections. *Tech Policy Press*.
<https://www.techpolicy.press/-researcher-data-access-under-the-dsa-lessons-from-tiktok-s-api-issues-during-the-2024-european-elections/>

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 35(4), 665–668. <https://doi.org/10.1080/10584609.2018.1477506>

Giglietto, F., Terenzi, M. (2024). The State of Social Media Research APIs & Tools in the Digital Service Act Era. White Paper. *Prompt*.
<https://disinfo-prompt.eu/posts/5TVAc369E7KfgFi4Pcots6>

Groves, R. M., Lyberg, L. 2010. Total Survey Error: Past, Present, and Future. *Public Opinion Quarterly*, Volume 74, Issue 5, 849–879, <https://doi.org/10.1093/poq/nfq065>

Hase, V., Ausloos, J., Boeschoten, L., Pfiffner, N., Janssen, H., Araujo, T., Carrière, T., de Vreese, C., Haßler, J., Loecherbach, F., Kmetty, Z., Möller, J., Ohme, J., Schmidbauer, E., Struminskaya, B., Trilling, D., Welbers, K., Haim, M. (2024). Fulfilling data access obligations: How could (and should) platforms facilitate data donation studies? *Internet Policy Review*, 13(3),
<https://policyreview.info/articles/analysis/fulfilling-data-access-obligations>

Mimizuka, K., Brown, M. A., Yang, K.-C., & Lukito, J. (2025). Post-Post-API Age: Studying Digital Platforms in Scant Data Access Times. *arXiv*.
<https://doi.org/10.48550/arXiv.2505.09877>

Poell, T. & Nieborg, D. & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1425>

Puschmann, C. (2019). An end to the wild west of social media research: a response to Axel Bruns. *Information, Communication & Society*, 22(11), 1582–1589.
<https://doi.org/10.1080/1369118X.2019.1646300>

Salvatore, C., Biffignandi, S., Bianchi, A. (2021). Social Media and Twitter Data Quality for New Social Indicators. *Social Indicator Research*, 156, 601–630.
<https://doi.org/10.1007/s11205-020-02296-w>

Sen, I., Flöck, F., Weller, K., Weiß, B., Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *Public Opinion Quarterly*, 85(S1) 399–422. <https://doi.org/10.1093/poq/nfab018>

van Dijck, J. (2021). Governing trust in European platform societies: Introduction to the special issue. *European Journal of Communication*, 36(4), 323-333.
<https://doi.org/10.1177/02673231211028378>

van Dijck, J., Poell, T., de Waal, M. (2018). *The Platform Society*. New York: Oxford Academic. <https://doi.org/10.1093/oso/9780190889760.001.0001>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3.
<https://doi.org/10.1038/sdata.2016.18>