



Selected Papers of #AoIR2023:
The 24th Annual Conference of the
Association of Internet Researchers
Philadelphia, PA, USA / 18-21 Oct 2023

DATA QUALITY ASSESSMENT IN SMART CITIES DATAFICATION

Carl Chineme Okafor
University of Stavanger, Norway

Raul Ferrer-Conill
University of Stavanger, Norway

Sjøvaaj Helle
University of Stavanger, Norway

Abstract

Data quality facilitates data interoperability for optimal decision-making in smart cities datafication. But there are few studies on how technologists (e.g., data scientists), governance people (e.g., municipal workers), and third-party collaborators (e.g., smart city services vendors) assess data quality together in smart cities datafication. This paper offers a response to this knowledge gap, using interviews (n=10) with municipal workers, data scientists and smart city services vendors, and data structure documents (n=8) in a situated case, the Stavanger (Norway) smart city. Implicit the paper's results is that data quality is a *floating signifier* – comprising the different articulations of data scientists, municipal workers and services vendors in assessment. This generates *friction* with implications on data interoperability. This paper therefore posits that assessing data quality in smart cities datafication is ambiguous, but not empty. It fluctuates between the articulations of data scientists, municipal workers, and services vendors, with implications on data interoperability through the friction this generates.

Keywords: data quality, data interoperability, floating signifier, frictions, smart city datafication

Introduction/argumentation

Datafication is the foundation of smart cities. It provides the raw material – data – for smart cities' initiatives (Al Nuaimi *et al.*, 2015) through sensors and recognition technologies (Lomborg, Dencik and Moe, 2020), enabling data scientists, municipal workers and services vendors to understand people and social issues (Dencik *et al.*, pp 26-39, 2022) for efficient governance (Hepp, Jarke and Kramp, pp 2-3, 2022). Datafication comprises the 'data work' of collecting, managing, analysing and communicating data (Pine *et al.*, 2018, 2022), but also of data scientists, municipal

Suggested Citation (APA): Okafor, C. C., Ferrer-Conill, R, Helle, S. (2023, October). Data Quality Assessment In Smart Cities Datafication. Paper presented at AoIR2023: The 24th Annual Conference of the Association of Internet Researchers. Philadelphia, PA, USA: AoIR. Retrieved from <http://spir.aoir.org>.

workers and services vendors assessing data quality which smart cities datafication studies rarely consider.

This paper positions the data work of assessing data quality as an essential element in smart cities datafication because this enables data interoperability for optimal decision-making (Khisro, 2020); and asks, how do data scientists, municipal workers and services vendors assess data quality to enable data interoperability in smart cities datafication? The paper presents results that suggest that there is no explicit standard or model for assessing data quality amongst data scientists, municipal workers, and services vendors in Stavanger.

Deliberating on the results, the paper advances that data quality is a '*floating signifier*' comprising the different articulations of data scientists, municipal workers and services vendors who have different professional cultures but do the data work of assessing data quality in Stavanger smart city datafication. The articulations of these data worker – data scientists, municipal workers, and services vendors – in data quality assessment is ambiguous but not empty and generate '*friction*' with implications for data interoperability. This requires common grounds over the friction.

With the concepts of '*floating signifiers*' (Laclau and Mouffe, 2001; Jørgensen and Phillips, 2011) and '*frictions*' (Edwards *et al.*, 2011; Tsing, 2011), this paper articulates its arguments showing that 1) in Stavanger as a relevant case, data quality assessment in smart cities datafication materialises through the different articulations of people who engage in 'data work' (e.g., data scientists, municipal workers and services vendors), 2) friction from the different articulations of data quality assessment is often inevitable, and 3) some forms of negotiations are desirable over the friction from the different articulatory assessment of data quality.

This paper's main question of, "how do data scientists, municipal workers and services vendors assess data quality to enable data interoperability in the Stavanger smart city datafication?" is consequently answered with the support of these two sub questions: SQ1. What are the data quality assessment models that data scientists, municipal workers and services vendors engage with in the Stavanger smart city datafication? SQ2. How do the data scientists, municipal workers and services vendors negotiate their data quality assessment to enable data interoperability in the Stavanger smart city datafication?

Method and data

A single case study method (Flyvbjerg, 2011; Yin, 2018) using the Stavanger smart city is adopted in this paper. The particular focus on the smart city is because the entire operation of the smart city initiative is grounded in datafication and its integration of the social domains of technology and governance. But more importantly, we seek in this paper to find out if the different assessment of data quality has implications for data interoperability in the Stavanger smart city. In relation, this paper considers case knowledge in qualitative studies as central in how scholars understand phenomena (Flyvbjerg, 2011), which are datafication, the smart city and data quality in our case. Stavanger's smart city initiative dates back in 2016. Stavanger uses digital data and connected communication technology processes to run its services and engagements in health and welfare, education and knowledge, energy, climate, and environment, urban art, as well as governance and democracy (Stavanger City Council, 2016). As of 2023, a number of Stavanger's smart city initiatives enjoy supports from the European Union

(EU), e.g., the EU – Horizon 2020 (2015-2020) for smart cities and communities in Europe, and EU – AI4Cities, in support of artificial intelligence-based climate change mitigation efforts.

Interviews (Edwards and Holland, 2013) and documents (Bowen, 2009; Dalglis, Khalid and McMahon, 2021; Grant and Kara, 2022) are the paper's means of collecting evidence. First, we conduct interviews with 10 respondents comprising data scientists, department workers in Stavanger municipality, and third-party smart city services and solutions providers to Stavanger between May and August 2023. We also collect documents, 1) roles in system and application management/user; 2) roles in system and application management/operation supplier; 3) roles in system and application management/system manager; 4) roles in system and application management/system owner; 5) roles in system and application management/service manager; 6) roles and responsibilities in information security work; 7), data processing checklist; and 8) guidelines for information security, from Stavanger municipality in July 2023.

With no preconceived themes, we analyse our interview and document data back and forth using the thematic data analyses strategy (Braun and Clarke, 2006) to find themes that respond to the questions we pose. We refine and categorise emergent themes into 1) data quality is largely important in this smart city's datafication, 2) no explicit standard or model for assessing data quality therein, 3) silo data quality assessment is in practice, 4) inevitable friction from the silo data quality assessment in Stavanger, 5) feedbacks, metadata, documentation are potential rallying grounds over frictions. We deliberate and discuss these results through the concepts of *floating signifiers* and *frictions*.

Early results

Using the Stavanger smart city as a relevant case study, especially within the European context of smart cities datafication process, our early claim in this paper is that data quality assessment in smart cities datafication is ambiguous but it is not empty.

In there, data quality assessment fluctuates between the contextual articulations of datafication actors, e.g., data scientists, municipal workers and services vendors who do data work – a concept Bossen *et al.*, (2019) describe as a range of socio-technical practices that undergird data production and use. Data work often involves “any human activity related to creating, collecting, managing, curating, analyzing, interpreting, and communicating data” (Bossen *et al.*, 2019; Møller *et al.*, 2020), and it is key in smart cities datafication. These articulations generate friction that obstruct data interoperability, requiring common grounds on assessment to overcome friction.

Metadata, proper documentation, and feedback are the diverse forms of common grounds desirable to overcome friction from the different articulations of data quality assessment.

References

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 25.

<https://doi.org/10.1186/s13174-015-0041-5>

Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction. *Health Informatics Journal*, 25(3), 465–474.

<https://doi.org/10.1177/1460458219864730>

- Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- DalGLISH, S. L., Khalid, H., & McMahon, S. A. (2021). Document analysis in health policy research: The READ approach. *Health Policy and Planning*, 35(10), 1424–1431. <https://doi.org/10.1093/heapol/czaa064>
- Dencik, L., Hintz, A., Redden, J., & Treré, E. (2022). *Data justice*. Sage Publications Ltd.
- Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 41(5), 667–690. <https://doi.org/10.1177/0306312711413314>
- Edwards, R., & Holland, J. (2013). *What is Qualitative Interviewing?* Bloomsbury Academic. <https://books.google.no/books?id=SknDMgEACAAJ>
- Flyvbjerg, B. (2011). Case study. *The Sage Handbook of Qualitative Research*, 4, 301–316.
- Grant, A., & Kara, H. (2022). *Doing your research project with documents: A step-by-step guide to take you from start to finish*. Policy Press.
- Hepp, A., Jarke, J., & Kramp, L. (2022). New Perspectives in Critical Data Studies: The Ambivalences of Data Power—An Introduction. In A. Hepp, J. Jarke, & L. Kramp (Eds.), *New Perspectives in Critical Data Studies* (pp. 1–23). Springer International Publishing. https://doi.org/10.1007/978-3-030-96180-0_1
- Jørgensen, M., & Phillips, L. (2011). *Discourse analysis as theory and method*. Sage.
- Khisro, J. (2020). Understanding the relation between interoperability and data quality: A study of data hub development in Swedish electricity market. *International Journal of Public Information Systems*, 14(1).
- Laclau, E., & Mouffe, C. (2001). *Hegemony and socialist strategy: Towards a radical democratic politics* (2. ed). Verso.
- Lomborg, S., Dencik, L., & Moe, H. (2020). Methods for datafication, datafication of methods: Introduction to the Special Issue. *European Journal of Communication*, 35(3), 203–212. <https://doi.org/10.1177/0267323120922045>
- Møller, N. H., Bossen, C., Pine, K. H., Nielsen, T. R., & Neff, G. (2020). Who does the work of data? *Interactions*, 27(3), 52–55. <https://doi.org/10.1145/3386389>
- Pine, K., Bossen, C., Holtén Møller, N., Miceli, M., Lu, A. J., Chen, Y., Horgan, L., Su, Z., Neff, G., & Mazmanian, M. (2022). Investigating Data Work Across Domains: New Perspectives on the Work of Creating Data. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–6. <https://doi.org/10.1145/3491101.3503724>
- Pine, K. H., Bossen, C., Chen, Y., Ellingsen, G., Grisot, M., Mazmanian, M., & Møller, N. H. (2018). Data Work in Healthcare: Challenges for Patients, Clinicians and Administrators. *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 433–439. <https://doi.org/10.1145/3272973.3273017>
- Tsing, A. L. (2011). *Friction: An Ethnography of Global Connection*. Princeton University Press. <https://doi.org/10.2307/j.ctt7s1xk>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (Sixth edition). SAGE.