Can We Trust Open Government Data in Public Procurement?

Robin Fritzsche[a*], *Jan* Westermann[b], *Gerhard* Klassen[c], *Marc* Wyszynski[d], *Björn* Niehaves[e]

[a]AG Digital Public, University of Bremen, Bremen, Germany, robfri@uni-bremen.de, 0009-0002-8784-585X.

[b]AG Digital Public, University of Bremen, Bremen, Germany, westerma@uni-bremen.de, 0009-0009-3004-1982.

[c]AG Digital Public, University of Bremen, Bremen, Germany, gerhard.klassen@uni-bremen.de, 0000-0002-1458-6546.

[d]AG Digital Public, University of Bremen, Bremen, Germany, m.wyszynski@uni-bremen.de, 0000-0002-5292-4083.

[e]AG Digital Public, University of Bremen, Bremen, Germany, niehaves@uni-bremen.de, 0000-0002-2682-6009.

Submitted: 31 January 2025, Revised: 26 March 2025, Accepted: 21 April 2025, Published: 22 May 2025

Abstract. Evidence-based policy-making in public procurement depends on high-quality data. This case study evaluates 14 months (Nov 2023–Jan 2025) of procurement records from the EU's Tenders Electronic Daily (TED) to assess the reliability of this key data source. Our analysis uncovers significant anomalies in data accuracy and governance, raising concerns about the integrity of procurement processes and the policies built on them. We highlight the urgent need for improved data governance to support transparency, accountability, and innovation-driven procurement.

Keywords. e-procurement data, government data, data governance, public procurement.
Poster, DOI: https://doi.org/10.59490/dgo.2025.1019

1. Introduction

Public procurement plays a key role in policy-making by directing demand that can drive innovation among suppliers (Amann and Essig, 2015; Bleda and Chicot, 2020). Effective and fair procurement depends on accurate assumptions about market structures and behavior, which require systematic analysis of reliable data. Open Government Data (OGD) from e-procurement portals provides a valuable basis for such analysis, offering insights into supplier participation, market dynamics, and procurement trends. Although OGD can improve transparency and efficiency, concerns about data quality may limit its usefulness (Zuiderwijk and Janssen, 2014). This study addresses the following research question: "How good is the data quality of data provided by e-procurement portals?" We analyze public procurement data from a European OGD platform, focusing on tender awards and contract notices. The aim is to assess data quality and inform evidence-based policy-making. Our findings point to the need for more consistent data maintenance, including clear standards and controls for publishing procurement data.

2. Relevant Theory & Literature

The intersection of OGD and e-procurement has attracted significant scholarly attention (Attard et al., 2015; Zhenbin et al., 2020. OGD involves the proactive release of government-held data to promote transparency, innovation, and accountability (Charalabidis et al., 2016). While its benefits are well documented — political, economic, and operational improvements (Janssen et al., 2012) — research also highlights its risks, including privacy issues, misinterpretation, and inequalities (Zuiderwijk and Janssen, 2014).
Integrating OGD with e-procurement systems can improve efficiency and reduce corruption. Adoption, however, depends on technical, organizational, and policy factors Wang and Lo, 2016). Yet, without high data

quality, such integration may fail to deliver set goals.
Data quality remains a central concern in OGD, with dimensions such as accuracy, completeness, consistency, timeliness, and accessibility frequently cited (Batini and Scannapieco, 2016; Heinrich et al., 2017).

3. Data and Methodology

This study combines procurement data from the TED portal and firm-level data from Moody's Orbis, covering November 2023 to January 2025. Historical exchange rates from the UN Treasury were used to standardize contract values. The TED dataset includes 1.7 million contract and award notices, of which 939,165 are unique. Winner information is available in 259,487 notices, identifying 496,445 winners (209,464 unique). Orbis provides validation and geographic details for 471,303 entities. CPV codes classify 82.1% of notices, but only 27.6% disclose winner data. Value reporting is concentrated, with 93.9% of tenders falling between €4,544 and €63,379,652. Data was retrieved via the TED Search API in JSON format.

4. Results

The dataset exhibits major anomalies in total value (Fig. 1, Fig. 2). To address these, we excluded contracts with non-positive values and conducted comparisons between the full dataset and a filtered subset limited to the interquartile range (IQR).

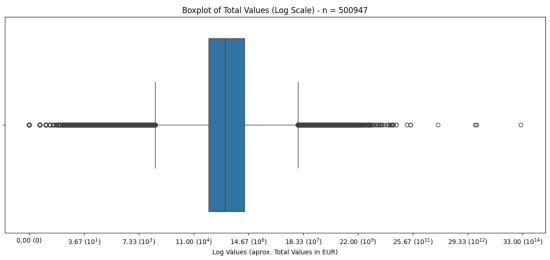

Fig. 1 – Boxplot of log scaled total values

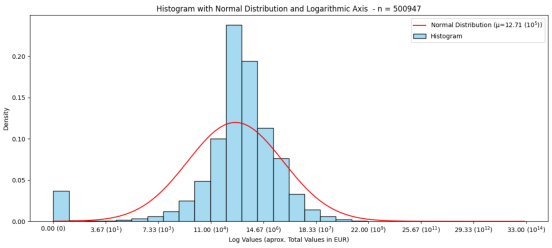

Fig. 2 – Log-transformed distribution of contract values

In the unfiltered dataset (a1, $n = 859{,}929$), Germany accounted for 19.98% of tender lots, followed by Poland (12.34%) and France (11.92%). Eastern Europe, including Poland, the Czech Republic, Romania, and Bulgaria, contributed 26.99%. After filtering out zero-value entries (a2, $n = 491{,}763$), Germany's share dropped to 9.29%, with only 45,675 of its 171,807 lots reporting a valid value. This highlights major gaps in the German data. We then compared total market volume before and after value filtering. Without restrictions (b1), the total reached €215.97 trillion, which is clearly inflated. According to the European Commission, public procurement comprises about 14% of EU GDP. Based on 2024 AMECO data (€14,122.3 billion at constant prices), this would suggest a procurement market of around €1,977.12 billion (€2,533.52 billion at current prices). Filtering to €4,544–€63,379,652 (b2) yields a revised volume of €1,256.46 billion, more aligned with expectations. Figures 1 and 2 show a right-skewed distribution with extreme outliers. Filtering changes the composition significantly. Germany's tender share drops by over 10 percentage points, and similar effects are seen in other indicators. German companies appeared as winners in 19.17% of contracts overall, but this fell to 10.92% after filtering. The average value of lots won by German firms or consortia including them was €2.51 million, dropping to €1.62 million after cleaning. These figures reflect inconsistencies due to both participation and poor data quality.

We analyzed 771,016 contracts with CPV codes across three scenarios: all classified contracts, those with positive values, and those within the filtered value range. In all cases, Division 45 (Construction Work) was dominant. The second most frequent category changed from Division 33 (Medical Equipment) in Scenario 1 to Division 71 (Engineering Services) in Scenarios 2 and 3. These shifts reflect the effect of data cleaning on sectoral distribution.

Of the 939,165 unique notices, 168,149 lacked CPV codes, indicating substantial classification gaps in the dataset.

Tab. 1 – Distribution of tender lots per CPV-Code

CPV-Division Code	Scenario 1	Scenario 2	Scenario 3
45	18.19%	7.62%	6.95%
71	10.41%	5.04%	4.77%
33	10.30%	6.76%	6.36%
90	5.95%	3.25%	3.11%
79	5.08%	3.17%	2.93%
34	4.82%	2.66%	2.54%
72	4.65%	2.96%	2.79%
50	4.43%	2.78%	2.67%
30	2.52%	1.56%	1.44%
9	2.35%	1.48%	1.36%

5. Discussion & Conclusion

A key issue is the lack of benchmarks to validate total contract values. Without a ground truth, anomalies such as zero or extremely high amounts (e.g., €1 billion) are hard to interpret, raising concerns about accuracy and intent. Our comparisons showed that outliers distort results and can lead to misleading conclusions. Missing, false, or extreme values weaken data integrity and may mask actual spending or signal fraud. Transparency requires data quality, not just quantity. TED data reveals flaws in inconsistent metadata. Better governance, oversight, and institutional capacity are needed. High-quality data is essential for accountability, policymaking, and public trust. In conclusion, TED data quality is poor. Frequent missing values and outliers limit its value for analyzing procurement and undermine the credibility of e-procurement initiatives.

Acknowledgement

- **Funding**: Research done in the frame of the research project 'Procurdat' funded by the German ministry of education and research. Funding code: 02K23A010.

References

Amann, M., & Essig, M. (2015). Public procurement of innovation: Empirical evidence from EU public authorities on barriers for the promotion of innovation. *Innovation: The European Journal of Social Science Research*, *28*(3), 282–292. DOI: https://doi.org/10.1080/13511610.2014.998641

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. DOI: https://doi.org/10.1016/j.giq.2015.07.006

Batini, C., & Scannapieco, M. (2016). *Data and information quality: Dimensions, principles, and techniques*. Springer.

Bleda, M., & Chicot, J. (2020). The role of public procurement in the formation of markets for innovation. *Journal of Business Research*, *107*, 186–196. DOI: https://doi.org/10.1016/j.jbusres.2018.11.032

Charalabidis, Y., Alexopoulos, C., & Loukis, E. (2016). A taxonomy of open government data research areas and topics. *Journal of Organizational Computing and Electronic Commerce*, *26*(1-2), 41–63. DOI: https://doi.org/10.1080/10919392.2015.1124720

Heinrich, B., Hristova, D., Klier, M., Schiller, A., & Szubartowicz, M. (2017). Requirements for Data Quality Metrics. *Journal of Data and Information Quality*, *9*(2), 1–32. DOI: https://doi.org/10.1145/3148238

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. DOI: https://doi.org/10.1080/10580530.2012.716740

Wang, H.-J., & Lo, J. (2016). Adoption of open government data among government agencies. *Government Information Quarterly*, *33*(1), 80–88. DOI: https://doi.org/10.1016/j.giq.2015.11.004

Zhenbin, Y., Kankanhalli, A., Ha, S., & Tayi, G. K. (2020). What drives public agencies to participate in open government data initiatives? an innovation resource perspective. *Information & Management*, *57*(3), 103179. DOI: https://doi.org/10.1016/j.im.2019.103179

Zuiderwijk, A., & Janssen, M. (2014). The negative effects of open government data - investigating the dark side of open data. *Proceedings of the 15th Annual International Conference on Digital Government Research*, 147–152. DOI: https://doi.org/10.1145/2612733.2612761