


# Between control and empowerment: Data quality in border and migration management

**Working Paper****Author(s):**

Leese, Matthias 

**Publication date:**

2024-11

**Permanent link:**

<https://doi.org/10.3929/ethz-b-000705001>

**Rights / license:**

In Copyright - Non-Commercial Use Permitted

**Originally published in:**

CURATE Working Paper 4

**Funding acknowledgement:**

MB22.00035 - CURATE: The data practices of European law enforcement and border control (SBFI)

# Between control and empowerment: Data quality in border and migration management

CURATE Project Working Paper No. 4 – November 2024

Matthias Leese  
[mleese@ethz.ch](mailto:mleese@ethz.ch)

This Working Paper conceptually explores the role of data quality in European migration and border management. Drawing on the notion of legibility as well as the dual nature of identification, it examines how forms of datafied representation – and notably their reliability – impact both state power and migrant experiences. The analysis highlights that data quality should be understood as a tool for both empowerment and control, with implications for both regulation and oversight.

“Data quality is fast becoming a hot topic, as demand for high-quality data continues to grow” (European Union, 2021d: 7). This is how the European Union, in its recent “Data Quality Guidelines,” frames the importance of data quality vis-à-vis growing efforts towards administrative transparency and evidence-based policymaking. As aggregated databases and open data initiatives proliferate across Europe, the question of data quality has indeed been taken out of the technical-bureaucratic confines of the back office and put on the political agenda – and migration and border control make no exception in this regard.

As the Council of the EU emphasizes, border management and migration in Europe “are information driven and today information is

the most valuable asset; [and] their efficiency is totally dependent on digital technologies, infrastructure and large-scale IT systems” (Council of the European Union, 2020a: 2). Against this backdrop, the diagnosed quality of available data does, however, not live up to expectations. As the Council’s Information Management Strategy (2016: 6) has highlighted, databases for border and migration management in the EU are riddled with “inappropriate use of data fields, data inconsistencies, use of incorrect data formats, insertion of records with missing data and insertion of poor quality biometric samples.”

Such shortcomings are, almost needless to say, politically hardly tolerable in a domain where reliable data are required “to avoid false positive matches and thus to avoid further action being taken on the basis of false information” (Council of the European Union,

2020b: 2). Subsequently, several initiatives to improve the quality of data used for the governance of migration and borders in Europe have been launched over the past years, tackling both the making of data at the national level and the pooling and redistribution of data at the European level (Council of the European Union, 2020a; eu-LISA, 2020; European Union, 2020, 2021a, 2021b, 2021c).

Despite its political prioritization, academic engagement with data quality has so far has largely been confined to technical and managerial perspectives and the concept remains understudied in (critical) border and migration scholarship (for notable exceptions, see Forti, 2022; van Rossem, 2021). To understand data quality and its implications for the governance of mobility, this Working Paper builds on Scott’s notion of legibility, as well as on Caplan and Torpey’s concept of the ambiguity of identification. Data quality, in this context, provides a unique window into how states make up the mobile populations they intend to govern – and more specifically into the challenges, frictions, and resistances

### **Data quality and government**

Data quality is a concept that is not easy to pin down. From a philosophical perspective, it can be understood as the question how empirical phenomena are represented in digital ways (Coopmans et al., 2014). This involves asking whether and how something can be represented in the first place, as well as how well a particular representation lives up to the actual phenomenon. From a technical perspective, data quality is about characteristics such as the completeness, timeliness, or semantic and syntactic integrity of a dataset (Batini and Scannapieca, 2006). Both these perspectives open up the concept for complex debates about measurements, accuracy, and informational value. Such

of identifying and datafying people on the move.

The Working Paper is organized as follows. It first provides a definition of the concept of data quality, opting for a relational understanding that defines the quality of data in accordance with their use case contexts. In doing so, it also engages the specific accuracy requirements for data in governmental contexts. It then discusses the ambiguity of data-based identity and identification in the relationship between the state and the individual, explicating how identification can be both an instrument of empowerment and of repression – and how this duality is especially important in border and migration contexts where state interventions occur on a continuum between threat assessments and the protection of vulnerable people. The Working Paper concludes by spelling out how data quality in this context provides a unique window not only into state rationales, but also into the practical challenges and strategies of producing reliable information about mobile populations.

debates do, however, tend to have limited value for the study of data quality in practical contexts.

A more common conceptual approach, often found in applied literature that engages business and management, is thus to define data quality in relation to particular use cases (Lee et al., 2006; Loshin, 2001; Wang et al., 2002). Such an understanding starts with the task at hand and then asks whether the available data are suitable for the completion of this task. If so, they are considered to be of good quality. Importantly, a relational understanding means that data that are considered adequate in one context might be considered inadequate in another. Some use cases, such as targeted advertising, might

tolerate lower data quality due to logics of scale and probability, while other domains such as nuclear plant operations demand high reliability due to the potential for catastrophic outcomes.

In governmental contexts, there is a usually an expectation that the data that the state possesses about its citizens are accurate and complete (Leese, 2022a). In fact, as Scott (1998) has argued in his seminal book “Seeing Like a State”, the systematic collection and administration of data can be understood as the hallmark of modern forms of government. As Scott shows from a historical perspective, the datafication of humans and nature is in fact what made the world “legible” from a bureaucratic point of view, thus allowing states to understand and manage what was going on within the confines of their sovereign territory. Legibility, in this context, refers to the process by which states gather information about society in a way that renders it visible and manipulable for administrative and political purposes.

As Scott highlights, early forms of datafication have for the first time allowed governments to systematically “see” and understand the populations, lands, and economies they were set to administer and control. A pertinent example for the creation of legibility through datafication are land surveys and cadastral mapping. Whereas previously, land tenure systems were often based on local customs which could vary from region to region and even village to village, datafication schemes imposed standardized systems of property ownership that made land easier to tax and regulate. Such simplification (e.g., dividing land into easily measurable, uniform plots) allowed the state to control and extract revenue from agricultural production more efficiently.

In regard to the relations between states and humans, Scott shows how censuses and national registries have constituted key steps towards systematic population management.

By means of classifying people in standardized categories (e.g., by age, gender, occupation, ethnicity, etc.), these measures historically marked a key step towards the creation of individual administrative identities and the enrolment of these identities into public databases. This allowed for systematic knowledge production about citizens and residents, accumulating information about property, military service, tax payments, and so on.

Notably, scholars have more recently explored how the making and management of populations is contingent on the digital tools and databases that are used in the process. As Scheel (2020) shows in the Estonian national context, the data-related practices of public statistics offices define who/what is included or excluded, thus playing a key role in the production of official registries and bureaucratic records. Importantly, in doing so he shows how populations should not be understood as fixed, pre-existing entities, but rather as “enacted” (Law and Urry, 2004; Mol, 2002) through datafication in a performative way in the first place.

The idea of enactment is particularly relevant in border and migration contexts. States usually have existing records about their citizens and residents due to their enrolment in official population registries and other public administration databases. This is not the case when it comes to people on the move. Foreign nationals who seek to enter the territory of the state must thus, from a state perspective, be rendered knowable through the creation of a digital record that is based, for instance, on the data from machine readable, globally standardized travel documents (Salter, 2003), on the previous application for a visa (Salter, 2006), or on information from one of the many official travel pre-registration schemes such as the US ESTA, the Canadian eTA, or the new European ETIAS (Gäckle, 2020).

Matters of legibility become, however, substantially more complicated in less regulated forms of border crossings. When people enter a state’s territory via unregulated sea or land borders, in some cases without carrying passports or other documents, this puts state authorities in a challenging position where they need to find ways to identify individuals and assess their credibility in the context of the asylum system, as well as their harmlessness in security contexts. Against the backdrop of global migration, the struggle for legibility of foreigners can be understood as a major driving force towards data sharing and biometric forms of identification (Leese, 2022b), as well as the erection of facilities for the identification and bureaucratic enrolment of asylum seekers and refugees at states’ borders (Tazzioli and Garelli, 2018).

Against this backdrop, scholars have more recently studied the ways in which migration comes to be enacted and rendered governable through datafication. Scheel et al. (2019) have, for instance, demonstrated how migration is rendered knowable and governable through the production of statistics, the registration of migrants, and the creation of biometric databases. These practices, so they argue, should be understood as inherently political,

as they influence how migration is perceived, managed, and controlled. For example, the collection of migration statistics or the use of biometric data in border control systems not only generate knowledge about migration but also affect the lives and movements of migrants by shaping the policies that regulate them.

Similarly, Pelizza (2020) has introduced the concept of “alterity processing” to show how migrant registration and identification procedures simultaneously enact both individual migrants and emergent bureaucratic and political orders. In doing so, she argues that through the data infrastructures and bureaucratic practices involved in migration management, individuals who are unknown to European authorities are transformed into “European-legible” identities, such as migrants, refugees, or criminals. These processes do not simply reflect existing realities but actively construct both individual and collective identities. Pelizza thus emphasizes that different registration and identification procedures compete to legitimize different chains of actors, data, and metadata – with some actors being included in certain processes while others being excluded.

### **Between control and empowerment**

As we have seen, the development of official records and the state’s capacity to identify individuals have been pivotal in the emergence of modern public administration. However, notions of identification and legibility tend to shift meaning in relation to the political contexts within which they are applied. As Caplan and Torpey (2001) have argued, while identification enables states to manage and govern populations more effectively, it can also heighten state power and control, often to the detriment of individual freedoms.

Historically, as Caplan and Torpey show, identification systems have facilitated extensive surveillance schemes, particularly in authoritarian regimes like Nazi Germany, where population registers were used to track, target, and repress entire groups. Similarly, in Rwanda, identification documents that classified people by ethnicity were a direct tool in the orchestration of the 1994 genocide. These examples illustrate the dangers of codifying identity in ways that can exclude, harm, or even annihilate certain populations. When individuals are reduced to mere entries in a bureaucratic system, state administration

indeed has the power to strip them of their personal autonomy and leave them vulnerable to extreme repression. The bureaucratic overreach that accompanies such surveillance capacities underscores how easily identification systems, when misused, can become instruments of oppression.

However, identification systems do not solely serve repressive functions, but they can also be powerful tools for empowerment. Caplan and Torpey in this context emphasize that identification is equally a mechanism for providing individuals with rights and recognition. Identification documents are essential for registering births, enabling access to education, participating in voting processes, and receiving social services. In this way, being counted and registered in state systems grants individuals legal standing and the opportunity to engage in civic life. In societies where legal recognition is linked to rights and participation, identification thus becomes a gateway to empowerment rather than subjugation.

The dual nature of identification is particularly significant in the context of borders and migration, where the balance between state control and individual rights is a delicate one. Borders represent spaces where the state's regulatory power meets

individuals' need for recognition and protection. Here, the ambiguity of identification becomes acutely relevant. Migrants and refugees, often lacking formal documentation, are vulnerable to being classified into rigid categories such as "legal" or "illegal." Without proper identification, they can be denied entry, asylum, or basic human rights, which reinforces state control and limits their autonomy. At the same time, identification can offer migrants a path toward inclusion, granting them legal status and enabling access to essential services such as health care or legal counseling.

This tension between exclusion and inclusion highlights the need for vigilant oversight of identification systems. This is particularly the case because migrants and asylum seekers, as Tsagarousianou (2024) argues, are prone to experiencing datafication in embodied ways, when material objects like wristbands and bureaucratic papers serve as markers of identity that must be carried for identity purposes within legal and administrative frameworks, thus determining access to rights such as asylum and prospects for future inclusion. As Tsagarousianou claims, in this way, migrants become especially exposed to the ambiguity inherent in identification systems that operate along lines of both care and control, serving to protect as much as to restrict.

## Conclusions

As we have seen, the ability to make foreigners legible is central in governing borders and migration. Data quality, in this context, concerns the question how accurately the state can "see" and understand mobile populations such as migrants and refugees. These individuals are often undocumented or have inconsistent records, making them difficult to categorize within the state's bureaucratic systems. This issue is

particularly acute in border and migration contexts, where states must continuously render these populations legible through a combination of data production, verification, and integration into official databases. The need to account for people who may have crossed borders irregularly or lack documentation adds layers of complexity, as state authorities rely on biometric data, travel histories, and global information-sharing systems to assess individuals' identities and legal statuses.

Studying data quality in these contexts provides a valuable window into how states attempt to resolve these challenges. By examining the practical difficulties that can arise – such as missing information, errors in biometric matching, or issues with data sharing between national and international systems – we can gain insights into the operational mechanisms that underpin state governance. These processes reveal the ways in which governments manage migration, balance security concerns, and ensure that individuals are correctly identified for asylum, refusal of entry, or other migration-related decisions. Such an analytical angle can also expose the frictions and resistances within these systems, as technological and human shortcomings, political decisions, and infrastructural limitations all shape how effectively states can manage borders and the people who cross them.

By focusing on the quality of data used in border and migration governance, we are also able to uncover the underlying rationales and logics of state authority in these domains. The emphasis on data-driven governance, especially in the digital age, highlights the

increasing reliance on technologies like biometrics and large-scale databases to ensure effective population management. These tools are, however, not neutral. Rather, they reflect broader state interests, such as national security, economic management, or humanitarian obligations, and can both empower and disenfranchise individuals.

While this chapter has used the European Union context to illustrate the importance of data quality in border and migration governance, these questions have broader implications. In fact, states around the world grapple with the need for accurate data to manage migration, enforce border security, and uphold international obligations such as the protection of refugees. In many of these contexts, the challenges of poor data quality are magnified by resource constraints, political instability, or inadequate infrastructure. And yet, the need for legibility remains just as critical. Therefore, while the EU may serve as a prominent example of these challenges, the need for high-quality data in border and migration governance is a global issue, shaping the way states interact with increasingly mobile populations worldwide.

## References

- Batini C and Scannapieca M (2006) *Data Quality: Concepts, Methodologies and Techniques*. Berlin/Heidelberg: Springer.
- Caplan J and Torpey J (2001) Introduction. In Caplan J & Torpey J (eds.) *Documenting Individual Identity: The Development of State Practices in the Modern World*. Princeton: Princeton University Press, 1-12.
- Coopmans C, Vertesi J, Lynch M and Woolgar S (2014) Introduction: Representation in Scientific Practice Revisited. In Coopmans C, Vertesi J, Lynch M & Woolgar S (eds.) *Representation in Scientific Practice Revisited*. Cambridge: MIT Press, 1-12.
- Council of the European Union (2016) 13258/16: Renewed Information Management Strategy (IMS) - 5th Action List - State of Play, 14 October. Brussels.
- Council of the European Union (2020a) 11824/20: Roadmap for Standardisation for Data Quality Purposes, 11 November. Brussels.
- Council of the European Union (2020b) 11825/20: European Data Quality Day - Summary of the Replies, 11 November. Brussels.
- eu-LISA (2020) Data Quality and Interoperability: Addressing the Capability Gaps Through

- Standardisation: eu-LISA 12th Industry Roundtable, 3-5 November. Tallinn.
- European Union (2020) Commission Implementing Decision (EU) 2020/2165 of 9 December 2020 on laying down rules for the application of Regulation (EU) 2018/1861 of the European Parliament and of the Council as regards the minimum data quality standards and technical specifications for entering photographs and dactyloscopic data in the Schengen Information System (SIS) in the field of border checks and return (notified under document C(2020) 8599). Brussels: Official Journal of the European Union.
- European Union (2021a) Commission Implementing Decision (EU) 2021/31 of 13 January 2021 on laying down rules for the application of Regulation (EU) 2018/1862 of the European Parliament and of the Council as regards the minimum data quality standards and technical specifications for entering photographs, DNA profiles and dactyloscopic data in the Schengen Information System (SIS) in the field of police cooperation and judicial cooperation in criminal matters and repealing Commission Implementing Decision (EU) 2016/1345 (notified under document C(2020) 9228). Brussels: Official Journal of the European Union.
- European Union (2021b) Commission Implementing Regulation (EU) 2021/2224 of 16 November 2021 laying down the details of the automated data quality control mechanisms and procedures, the common data quality indicators and the minimum quality standards for storage of data, pursuant to Article 37(4) of Regulation (EU) 2019/818 of the European Parliament and of the Council. Brussels: Official Journal of the European Union.
- European Union (2021c) Commission Implementing Regulation (EU) 2021/2225 of 16 November 2021 laying down the details of the automated data quality control mechanisms and procedures, the common data quality indicators and the minimum quality standards for storage of data, pursuant to Article 37(4) of Regulation (EU) 2019/817 of the European Parliament and of the Council. Brussels: Official Journal of the European Union.
- European Union (2021d) Data.europa.eu: Data Quality Guidelines. Luxembourg: Publications Office of the European Union.
- Forti M (2022) Data Quality Issues in EU Digital Procedures of Border Control. *Diritti Fondamentali* (3): 230-244.
- Gäckle N (2020) Taming Future Mobilities: Biopolitics and Data Behaviourism in the European Travel Information and Authorisation System (ETIAS). *Mobilities* 15(2): 257-272.
- Law J and Urry J (2004) Enacting the Social. *Economy and Society* 33(3): 390-410.
- Lee Y W, Pipino L L, Funk J D and Wang R Y (2006) *Journey to Data Quality*. Cambridge/London: MIT Press.
- Leese M (2022a) Data Quality in Governance: A Definition and a Research Agenda. CURATE Working Paper No. 1, November 2022. doi:10.3929/ethz-b-000581434.
- Leese M (2022b) Fixing State Vision: Interoperability, Biometrics, and Identity Management in the EU. *Geopolitics* 27(1): 113-133.
- Loshin D (2001) *Enterprise Knowledge Management: The Data Quality Approach*. Burlington: Morgan Kaufmann.
- Mol A (2002) *The Body Multiple: Ontology in Medical Practice*. Durham: Duke University Press.
- Pelizza A (2020) Processing Alterity, Enacting Europe: Migrant Registration and Identification as Co-construction of Individuals and Politics. *Science, Technology, & Human Values* 45(2): 262-288.



- Salter M B (2003) *Rights of Passage: The Passport in International Relations*. Boulder: Lynne Rienner Publishers.
- Salter M B (2006) The Global Visa Regime and the Political Technologies of the International Self: Borders, Bodies, Biopolitics. *Alternatives: Global, Local, Political* 31(2): 167-189.
- Scheel S (2020) Biopolitical Bordering: Enacting Populations as Intelligible Objects of Government. *European Journal of Social Theory* 23(4): 571-590.
- Scheel S, Ruppert E and Ustek-Spilda F (2019) Enacting Migration Through Data Practices. *Environment and Planning D: Society and Space* 37(4): 579-588.
- Scott J C (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven/London: Yale University Press.
- Tazzioli M and Garelli G (2018) Containment Beyond Detention: The Hotspot System and Disrupted Migration Movements Across Europe. *Environment and Planning D: Society and Space* 38(6): 1009-1027.
- Tsagarousianou R (2024) The Datafication of Migrant Bodies and the Enactment of Migrant Subjectivities: Biometric Data, Power and Resistance at the Borders of Europe. *Media, Culture & Society* 46(4): 688-705.
- van Rossem W (2021) Data Quality in Data Infrastructures of Migration and Border Control. Paper presented at the 8th STS Italia Conference, Trento, Italy.
- Wang R Y, Ziad M and Lee Y W (2002) *Data Quality*. New York/Boston/Dordrecht/London/Moscow: Kluwer Academic Publishers.

Suggested citation: Leese M (2024) Between control and empowerment: Data quality in border and migration management. CURATE Working Paper No. 4, November 2024. doi: 10.3929/ethz-b-000705001